

LinearRegresssionGenderScopus2minimalRobustLoop.R

MikeT

Wed Jun 12 06:58:35 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

```

```

#FieldNeedsExtraTime <- c(1311, 3100, 2700, 2709, 2714, 2719,2722, 2734,
2744, 2747, 3104, 3107, 1701, 1706, 3302,3303,3309, 3320, 3322)

```

```

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$YMresiduals <- resid(YearLm)
YearGenderLm <- lm(YMresiduals~FirstAuthorFemale,
data=AllScopusDataOlderFirstGendered)
print(bartlett.test(YMresiduals~FirstAuthorFemale,
data=AllScopusDataOlderFirstGendered)) #Homogeneity of Variances test - this
is the key test - should not fail by much, but might becuse older years have
fewer females and higher variance
plot(YearGenderLm, which = 1) #Residuals vs. fitted. Check that width
of residual bar does not vary too much for homoscedasticity; Look out for
cone shape
#If problems here, check the outliers in regression 1
}, error = function(e) return("failed Heteroscedasticity checks"))
tryCatch({
#####
# Test for different team sizes for male and female authors - first and
last gendered
#####
#First author gender 2018
GenderAuthors2018 <-
split(AllYearsBothGendered2018$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)
    }
  })

```

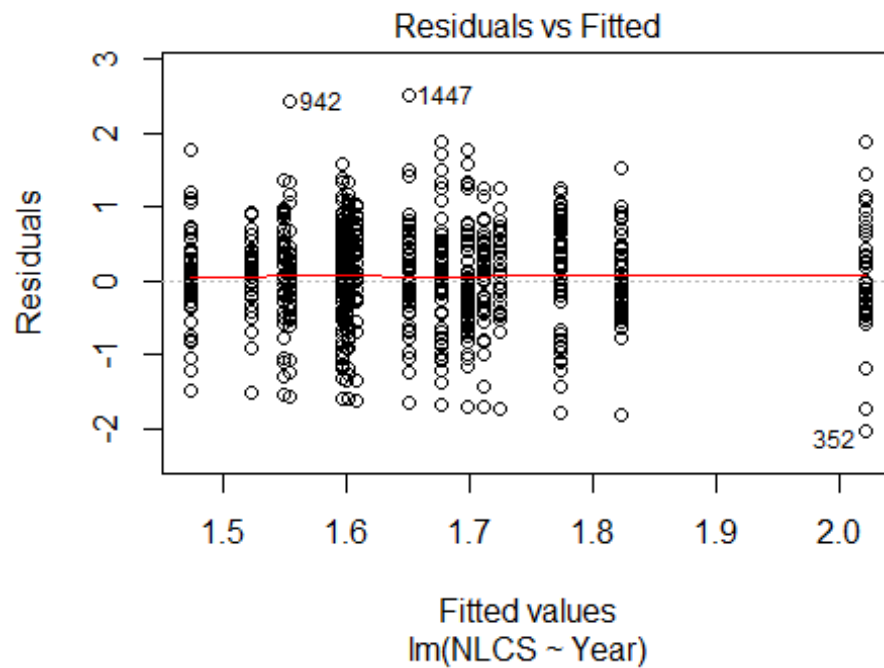
```

    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"))
  }
}

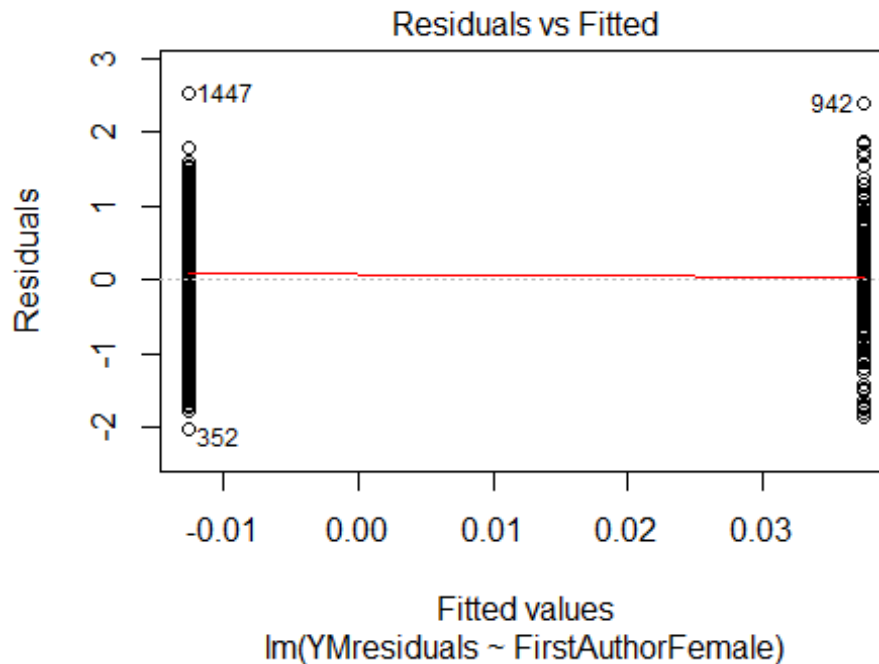
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1000"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   84   74   64   67   75   78   68   73   66   55   70   85   87   78   77
## 2011 2012
##   91   93
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   39   47   48   35   44   30   47   58   48   37   57   57   62   64   46
## 2011 2012
##   63   70
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   41   45   33   41   29   42   50   43   32   55   53   53   56   42

```

```
## 2011 2012
## 55 64
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 36, df = 16, p-value = 0.003
```

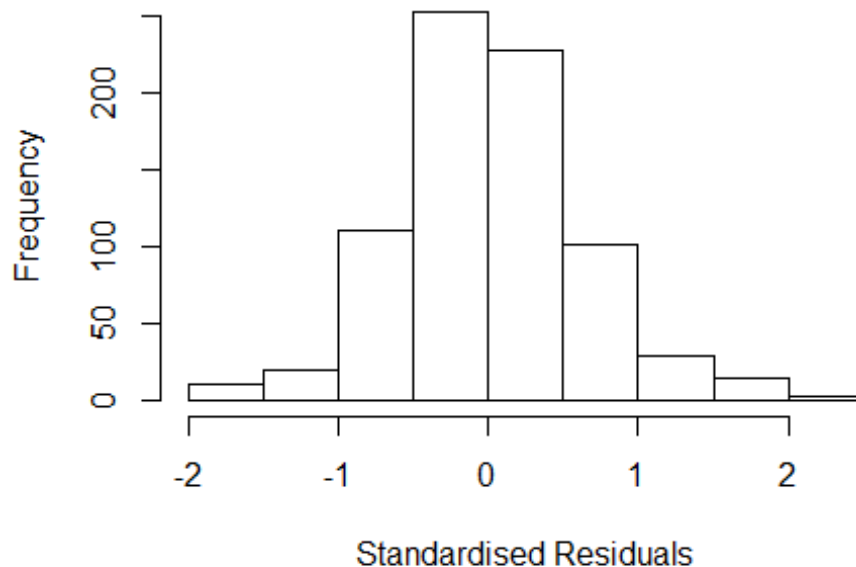


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.22, df = 1, p-value = 0.6
```

```
## [1] "Female first author team size 2018 geometric mean: 4.37065317944501"
## [1] "Male first author team size 2018 geometric mean: 4.15077961485921"
##
## 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: 3.94357305225206"
## [1] "Male last author team size 2018 geometric mean: 4.32678933334408"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5800, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "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  1.082 1      1.040
## UniqueAuthors    1.904 4      1.084
## Year             2.119 16      1.024
```

Residuals from first and last author and team size



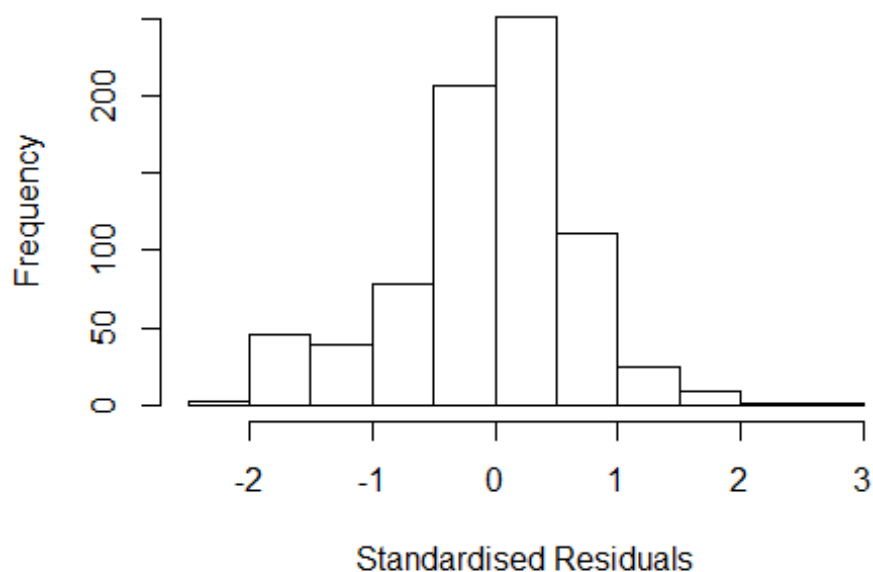
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.88126 -0.37036 -0.00733  0.35289  2.45611
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.77953    0.14309   5.45 6.9e-08 ***
## FirstAuthorFemale1 -0.05446    0.05271  -1.03  0.302
## LastAuthorFemale1 -0.06002    0.05985  -1.00  0.316
## UniqueAuthors2    0.92639    0.10798   8.58 < 2e-16 ***
## UniqueAuthors3    0.90591    0.10733   8.44 < 2e-16 ***
## UniqueAuthors4    1.10169    0.10715  10.28 < 2e-16 ***
## UniqueAuthors5    1.06971    0.09604  11.14 < 2e-16 ***
## Year1997          0.20873    0.16376   1.27  0.203
## Year1998          0.20300    0.15162   1.34  0.181
## Year1999          0.49411    0.21680   2.28  0.023 *
```

```

## Year2000      -0.05625    0.15532   -0.36    0.717
## Year2001      0.14584    0.16519    0.88    0.378
## Year2002     -0.08105    0.13043   -0.62    0.535
## Year2003      0.02779    0.15041    0.18    0.853
## Year2004     -0.16916    0.13889   -1.22    0.224
## Year2005     -0.09124    0.13284   -0.69    0.492
## Year2006     -0.10495    0.13177   -0.80    0.426
## Year2007     -0.00402    0.12716   -0.03    0.975
## Year2008     -0.07146    0.12711   -0.56    0.574
## Year2009      0.08649    0.13071    0.66    0.508
## Year2010      0.10805    0.14015    0.77    0.441
## Year2011      0.18254    0.15028    1.21    0.225
## Year2012      0.04336    0.13049    0.33    0.740
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.328, Adjusted R-squared:  0.309
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 699 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0014 0.8410 0.9490 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      1.30e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.107 1      1.052
## LastAuthorFemale  1.119 1      1.058
## Year              1.229 16      1.006

```

Residuals from first and last author



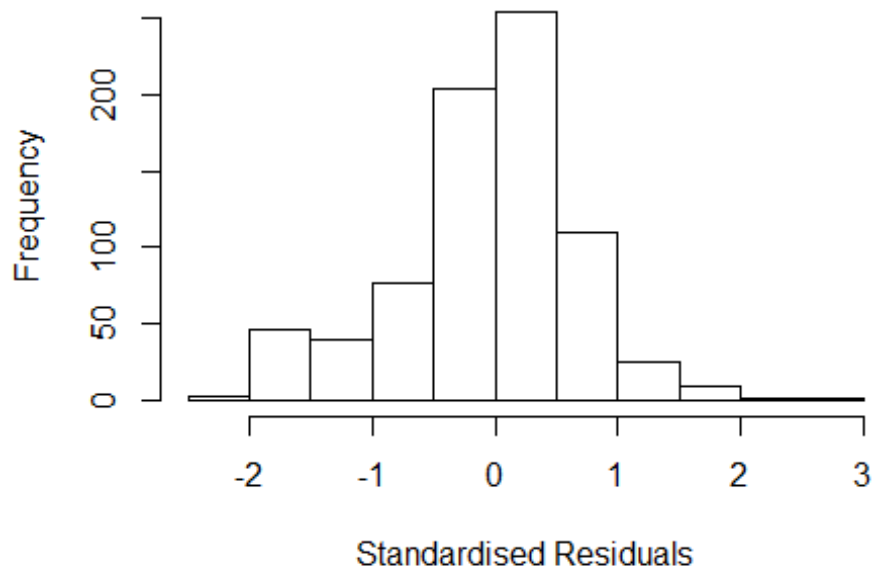
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1447 33847172327 4.158 2007    1000    2    2.501
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1386 -0.3983  0.0154  0.4038  2.5007
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.73255    0.14112   12.28  <2e-16 ***
## FirstAuthorFemale1  0.00626    0.05869    0.11    0.92
## LastAuthorFemale1 -0.05413    0.06977   -0.78    0.44
## Year1997          0.04874    0.19126    0.25    0.80
## Year1998          0.09164    0.16817    0.54    0.59
## Year1999          0.40605    0.20649    1.97    0.05 *
## Year2000         -0.10954    0.18434   -0.59    0.55
## Year2001          0.13771    0.18784    0.73    0.46
## Year2002         -0.15107    0.15336   -0.99    0.32
## Year2003         -0.06327    0.17107   -0.37    0.71
## Year2004         -0.22928    0.16076   -1.43    0.15
## Year2005         -0.15874    0.17073   -0.93    0.35
```

```

## Year2006      -0.21055    0.16351   -1.29    0.20
## Year2007      -0.07521    0.16590   -0.45    0.65
## Year2008      -0.11455    0.15909   -0.72    0.47
## Year2009      -0.08158    0.17101   -0.48    0.63
## Year2010       0.06137    0.16496    0.37    0.71
## Year2011       0.14474    0.19286    0.75    0.45
## Year2012      -0.00808    0.16210   -0.05    0.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.574
## Multiple R-squared:  0.0471, Adjusted R-squared:  0.0242
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 697 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0187 0.8230 0.9480 0.8620 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.30e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.101 1      1.049
## Year      1.101 16      1.003

```

Residuals from first author



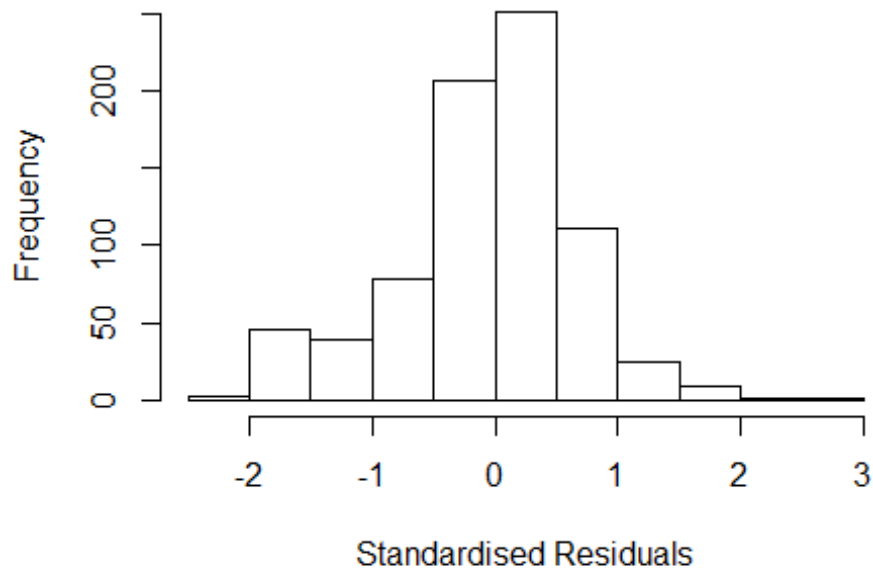
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1447 33847172327 4.158 2007    1000    2    2.501
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1299 -0.3948  0.0243  0.4018  2.5115
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.72681    0.14139   12.21  <2e-16 ***
## FirstAuthorFemale1 0.00138    0.05927    0.02    0.98
## Year1997        0.04248    0.19054    0.22    0.82
## Year1998        0.09671    0.16859    0.57    0.57
## Year1999        0.40308    0.20557    1.96    0.05 .
## Year2000       -0.11314    0.18408   -0.61    0.54
## Year2001        0.14009    0.18763    0.75    0.46
## Year2002       -0.15179    0.15353   -0.99    0.32
## Year2003       -0.06709    0.17106   -0.39    0.70
## Year2004       -0.22792    0.16095   -1.42    0.16
## Year2005       -0.15748    0.17107   -0.92    0.36
## Year2006       -0.21066    0.16379   -1.29    0.20
```

```

## Year2007          -0.08030    0.16550   -0.49    0.63
## Year2008          -0.11961    0.15910   -0.75    0.45
## Year2009          -0.08628    0.17080   -0.51    0.61
## Year2010           0.05548    0.16453    0.34    0.74
## Year2011           0.15011    0.19358    0.78    0.44
## Year2012          -0.00528    0.16261   -0.03    0.97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.574
## Multiple R-squared:  0.0468, Adjusted R-squared:  0.0252
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 696 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0161 0.8230 0.9470 0.8620 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.30e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.115 1          1.056
## Year            1.115 16          1.003

```


Residuals from last author



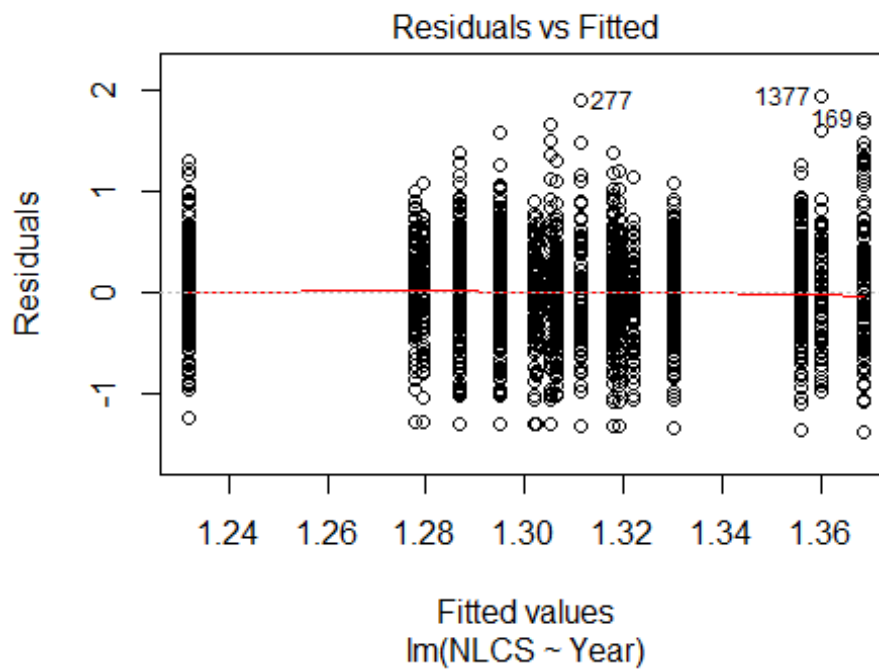
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1447 33847172327 4.158 2007    1000    2    2.501
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1404 -0.3957  0.0173  0.4019  2.4994
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.7348     0.1389   12.49  <2e-16 ***
## LastAuthorFemale1 -0.0532     0.0706   -0.75    0.45
## Year1997          0.0486     0.1911    0.25    0.80
## Year1998          0.0912     0.1680    0.54    0.59
## Year1999          0.4057     0.2065    1.96    0.05 *
## Year2000         -0.1104     0.1842   -0.60    0.55
## Year2001          0.1373     0.1878    0.73    0.46
## Year2002         -0.1515     0.1534   -0.99    0.32
## Year2003         -0.0650     0.1697   -0.38    0.70
## Year2004         -0.2309     0.1598   -1.45    0.15
## Year2005         -0.1586     0.1705   -0.93    0.35
## Year2006         -0.2110     0.1634   -1.29    0.20
```

```

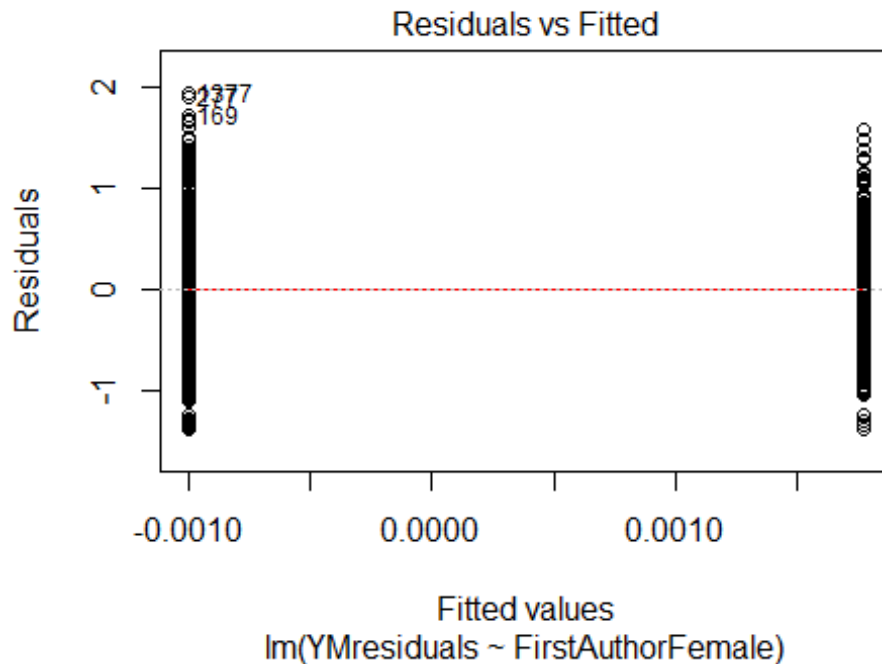
## Year2007          -0.0762      0.1657   -0.46      0.65
## Year2008          -0.1151      0.1589   -0.72      0.47
## Year2009          -0.0823      0.1709   -0.48      0.63
## Year2010           0.0601      0.1643    0.37      0.71
## Year2011           0.1451      0.1930    0.75      0.45
## Year2012          -0.0087      0.1618   -0.05      0.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.574
## Multiple R-squared:  0.0473, Adjusted R-squared:  0.0257
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 698 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0183 0.8240 0.9480 0.8620 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.30e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 768"
## [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
## 222 152 143 128 145 125 152 121 155 172 197 229 264 309 402
## 2011 2012
## 563 768
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 131 78 82 71 75 50 99 80 108 119 142 155 178 237 290
## 2011 2012

```

```
## 455 590
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 117 66 69 64 56 42 85 66 94 100 125 132 154 213 251
## 2011 2012
## 390 517
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 140, df = 16, p-value <2e-16
```

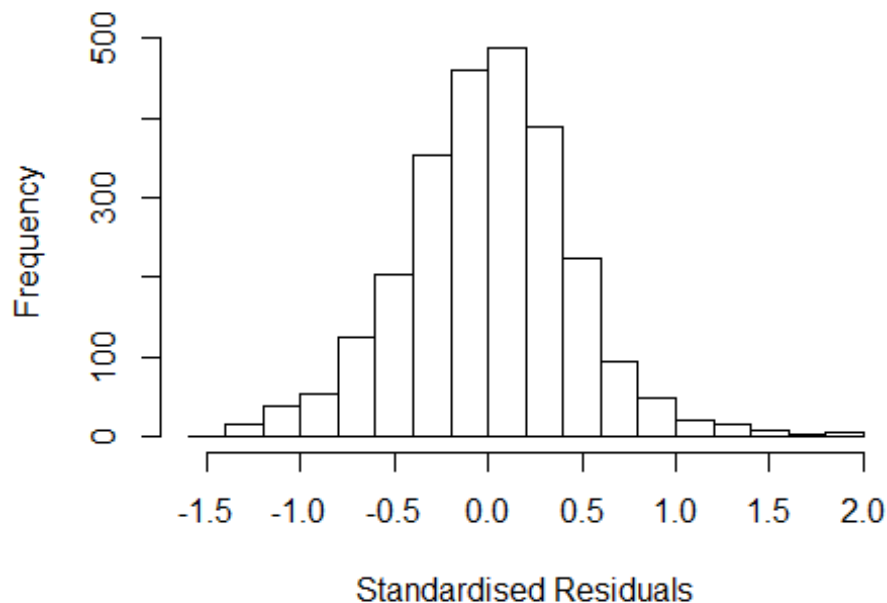


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 23, df = 1, p-value = 1e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 4.56275160319015"
## [1] "Male first author team size 2018 geometric mean: 3.9918526675761"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.36585491088979"
## [1] "Male last author team size 2018 geometric mean: 4.1741677544028"
##
## 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 1.120 1          1.059
## LastAuthorFemale  1.057 1          1.028
## UniqueAuthors    1.509 4          1.053
## Year              1.602 16         1.015
```

Residuals from first and last author and team size



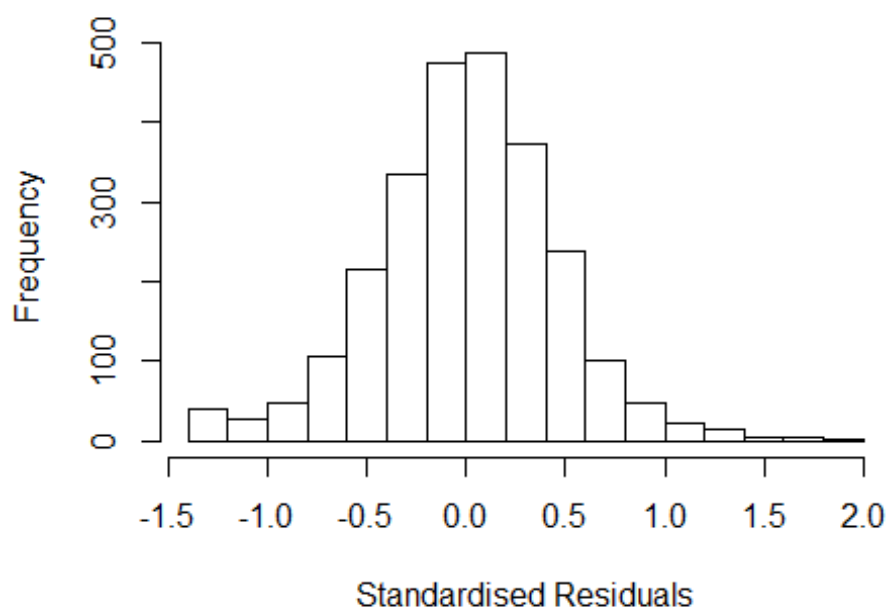
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.46076 -0.27141 0.00912 0.26938 1.91426
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20222 0.07338 16.38 < 2e-16 ***
## FirstAuthorFemale1 -0.02135 0.01844 -1.16 0.24698
## LastAuthorFemale1 0.00397 0.02205 0.18 0.85729
## UniqueAuthors2 0.16946 0.03624 4.68 3.1e-06 ***
## UniqueAuthors3 0.14288 0.03729 3.83 0.00013 ***
## UniqueAuthors4 0.21834 0.03979 5.49 4.5e-08 ***
## UniqueAuthors5 0.25854 0.03709 6.97 4.0e-12 ***
## Year1997 -0.04339 0.10402 -0.42 0.67663
## Year1998 -0.00239 0.09279 -0.03 0.97949
## Year1999 -0.00463 0.09227 -0.05 0.95996
```

```

## Year2000      0.00800      0.08977      0.09  0.92903
## Year2001     -0.02672      0.10119     -0.26  0.79173
## Year2002     -0.06294      0.08459     -0.74  0.45696
## Year2003     -0.03283      0.09063     -0.36  0.71724
## Year2004     -0.01163      0.08465     -0.14  0.89073
## Year2005     -0.00399      0.08423     -0.05  0.96220
## Year2006     -0.00118      0.07941     -0.01  0.98815
## Year2007     -0.05868      0.07876     -0.75  0.45634
## Year2008      0.02540      0.07985      0.32  0.75044
## Year2009     -0.03982      0.07606     -0.52  0.60068
## Year2010     -0.07233      0.07658     -0.94  0.34497
## Year2011     -0.09903      0.07332     -1.35  0.17695
## Year2012     -0.17015      0.07268     -2.34  0.01931 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.042, Adjusted R-squared:  0.0337
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 217 weights are ~= 1. The remaining 2324 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8640 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      3.94e-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 1.074 1      1.036
## LastAuthorFemale 1.039 1      1.019
## 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
## -1.39751 -0.27441  0.00797  0.27391  1.99109
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.31768    0.07061   18.66  <2e-16 ***
## FirstAuthorFemale1 -0.00445    0.01827   -0.24    0.81
## LastAuthorFemale1  0.00870    0.02202    0.40    0.69
## Year1997        -0.01637    0.10532   -0.16    0.88
## Year1998        -0.00656    0.09539   -0.07    0.95
## Year1999         0.01864    0.09461    0.20    0.84
## Year2000         0.04042    0.09027    0.45    0.65
## Year2001        -0.02646    0.10362   -0.26    0.80
## Year2002        -0.03675    0.08482   -0.43    0.66
## Year2003        -0.00677    0.09144   -0.07    0.94
## Year2004         0.03086    0.08527    0.36    0.72
## Year2005         0.01368    0.08599    0.16    0.87
```

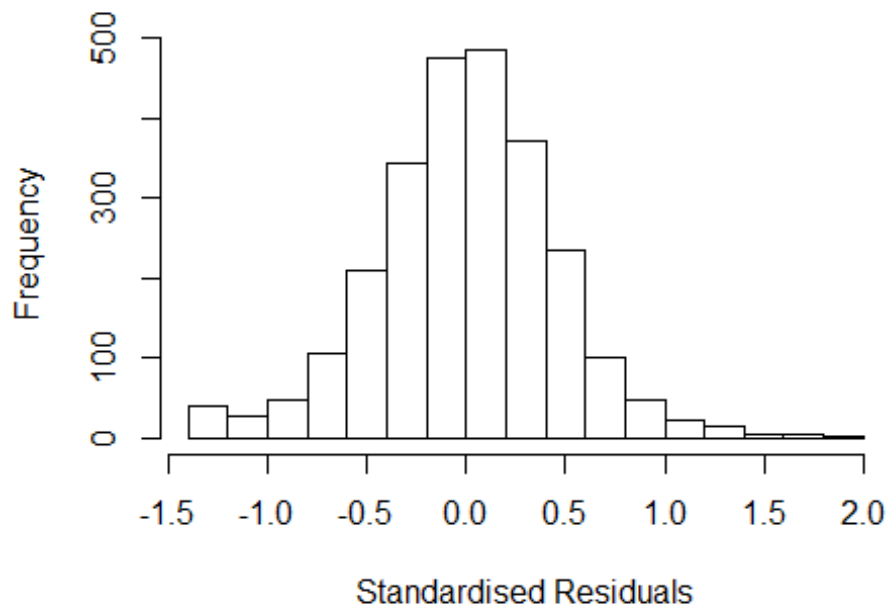


```

## Year2006          0.01025      0.07992      0.13      0.90
## Year2007          -0.02266      0.08023     -0.28      0.78
## Year2008          0.07112      0.07999      0.89      0.37
## Year2009          0.01873      0.07631      0.25      0.81
## Year2010          -0.00652      0.07639     -0.09      0.93
## Year2011          -0.03277      0.07306     -0.45      0.65
## Year2012          -0.09114      0.07238     -1.26      0.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.415
## Multiple R-squared:  0.0112, Adjusted R-squared:  0.00418
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## observation 550 is an outlier with |weight| = 0 ( < 3.9e-05);
## 219 weights are ~ = 1. The remaining 2321 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8640 0.9510 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          3.94e-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 1.065 1          1.032
## Year              1.065 16          1.002

```

Residuals from first author



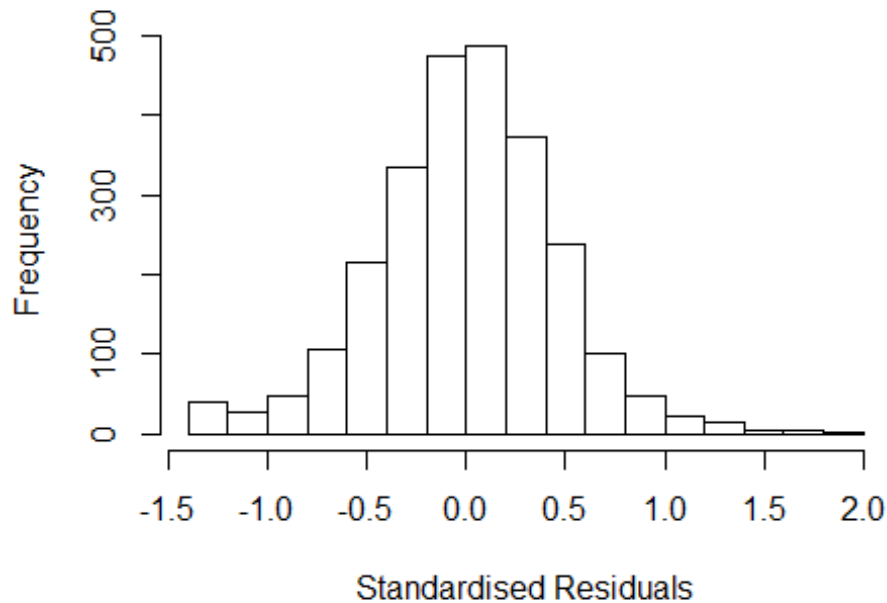
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3902 -0.2756  0.0107  0.2727  1.9894
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.31824    0.07056   18.68  <2e-16 ***
## FirstAuthorFemale1 -0.00354    0.01821   -0.19    0.85
## Year1997      -0.01529    0.10522   -0.15    0.88
## Year1998      -0.00489    0.09517   -0.05    0.96
## Year1999       0.01907    0.09458    0.20    0.84
## Year2000       0.04121    0.09029    0.46    0.65
## Year2001      -0.02612    0.10370   -0.25    0.80
## Year2002      -0.03665    0.08483   -0.43    0.67
## Year2003      -0.00561    0.09145   -0.06    0.95
## Year2004       0.03141    0.08519    0.37    0.71
## Year2005       0.01469    0.08591    0.17    0.86
## Year2006       0.01105    0.07986    0.14    0.89
```

```

## Year2007          -0.02196    0.08025   -0.27    0.78
## Year2008          0.07193    0.07993    0.90    0.37
## Year2009          0.02012    0.07614    0.26    0.79
## Year2010         -0.00542    0.07632   -0.07    0.94
## Year2011         -0.03168    0.07301   -0.43    0.66
## Year2012         -0.09013    0.07233   -1.25    0.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.415
## Multiple R-squared:  0.0112, Adjusted R-squared:  0.00453
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## observation 550 is an outlier with |weight| = 0 ( < 3.9e-05);
## 213 weights are ~= 1. The remaining 2327 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8650 0.9510 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      3.94e-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 1.031 1          1.015
## Year            1.031 16          1.001

```

Residuals from last author



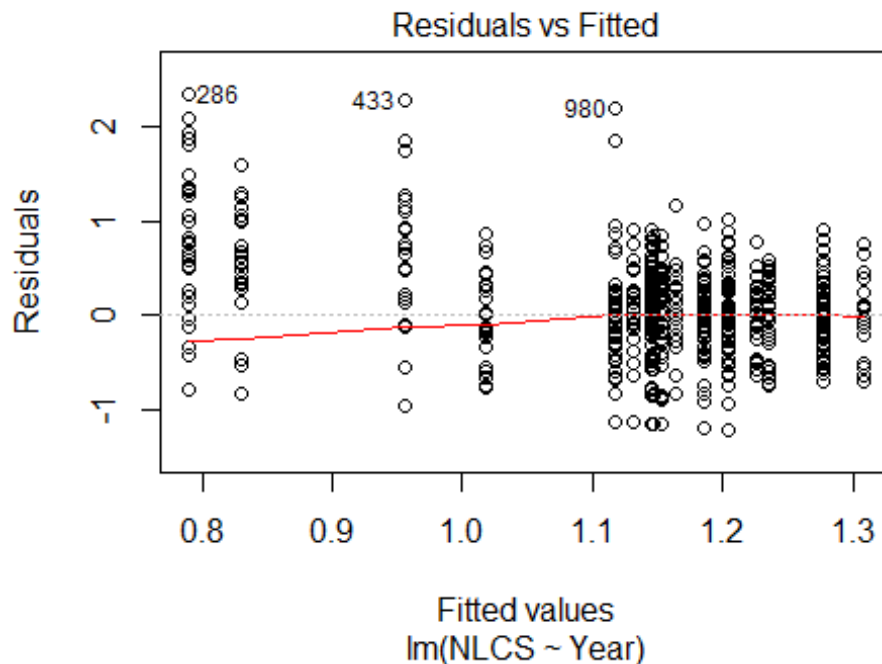
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3953 -0.2742 0.0087 0.2755 1.9926
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31668 0.07024 18.75 <2e-16 ***
## LastAuthorFemale1 0.00805 0.02196 0.37 0.71
## Year1997 -0.01586 0.10509 -0.15 0.88
## Year1998 -0.00659 0.09533 -0.07 0.94
## Year1999 0.01899 0.09440 0.20 0.84
## Year2000 0.04018 0.09020 0.45 0.66
## Year2001 -0.02689 0.10351 -0.26 0.80
## Year2002 -0.03723 0.08473 -0.44 0.66
## Year2003 -0.00723 0.09135 -0.08 0.94
## Year2004 0.03063 0.08517 0.36 0.72
## Year2005 0.01350 0.08590 0.16 0.88
## Year2006 0.00956 0.07977 0.12 0.90
```

```

## Year2007      -0.02274    0.08015   -0.28    0.78
## Year2008      0.07061    0.07986    0.88    0.38
## Year2009      0.01785    0.07622    0.23    0.81
## Year2010     -0.00709    0.07632   -0.09    0.93
## Year2011     -0.03345    0.07299   -0.46    0.65
## Year2012     -0.09225    0.07222   -1.28    0.20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.415
## Multiple R-squared:  0.0112, Adjusted R-squared:  0.00452
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 550 is an outlier with |weight| = 0 ( < 3.9e-05);
## 220 weights are ~= 1. The remaining 2320 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0007 0.8650 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      3.94e-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: 2541"
## [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
## 280 165 141 61 57 56 49 62 69 88 83 84 60 46 33
## 2011 2012
## 41 35
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 60 43 42 27 26 30 34 40 45 51 57 60 37 36 21

```

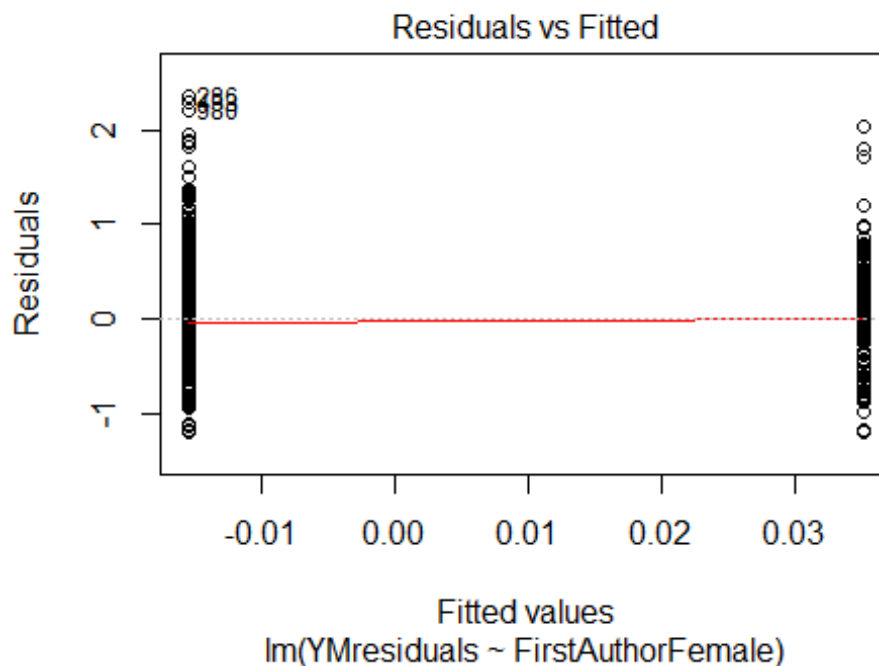
```
## 2011 2012
## 27 18
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 50 36 40 21 23 23 29 34 36 46 48 51 34 30 19
## 2011 2012
## 22 16
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 160, df = 16, p-value <2e-16
```



```
##
## 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: 2.64111740110372"
## [1] "Male first author team size 2018 geometric mean: 2.4949584045823"
## Warning in wilcox.test.default(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 = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.24351597244694"
## [1] "Male last author team size 2018 geometric mean: 2.76412108112213"

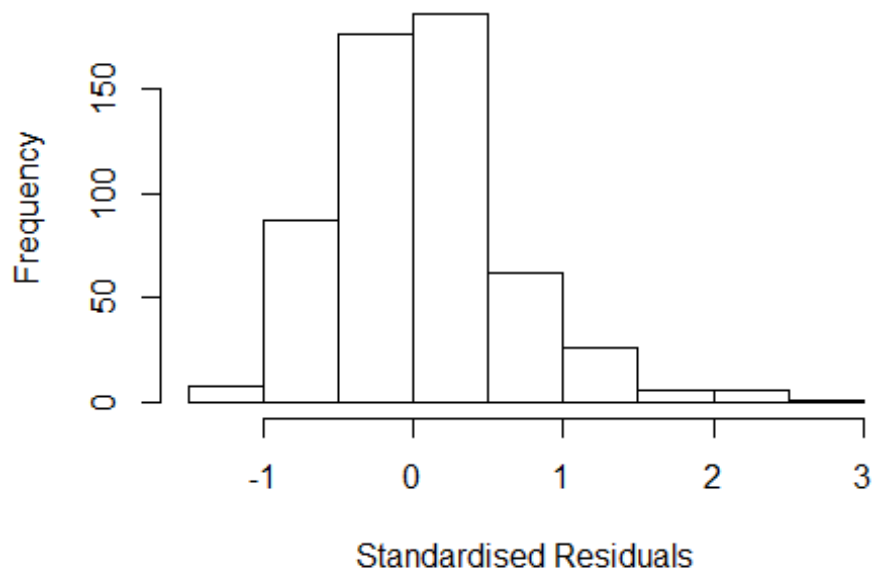
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 58, p-value = 0.4
## 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	1.373	1	1.172
LastAuthorFemale	1.396	1	1.182
UniqueAuthors	2.871	4	1.141
Year	3.752	16	1.042

Residuals from first and last author and team size



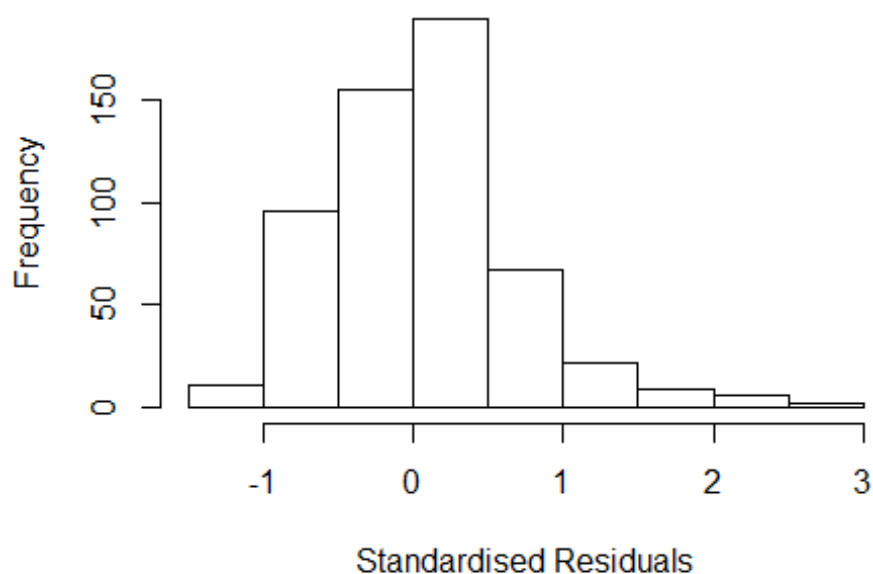
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 286 0029693941 3.125 1996    1101      3      2.868
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2378 -0.3493  0.0247  0.3665  2.8680
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.2570    0.1883   1.36 0.17298
## FirstAuthorFemale1  0.0434    0.0560   0.78 0.43826
## LastAuthorFemale1 -0.0712    0.0697  -1.02 0.30733
## UniqueAuthors2    0.3447    0.0753   4.58 5.8e-06 ***
## UniqueAuthors3    0.2296    0.0813   2.82 0.00493 **
## UniqueAuthors4    0.4142    0.0984   4.21 3.0e-05 ***
## UniqueAuthors5    0.3018    0.0951   3.17 0.00159 **
## Year1997          0.1409    0.6158   0.23 0.81904
## Year1998          0.2215    0.3017   0.73 0.46323
## Year1999          0.6361    0.2176   2.92 0.00362 **
```

```

## Year2000          0.6405      0.2083      3.08  0.00221 **
## Year2001          0.7162      0.2146      3.34  0.00091 ***
## Year2002          0.6721      0.2069      3.25  0.00123 **
## Year2003          0.5556      0.2105      2.64  0.00854 **
## Year2004          0.6701      0.2062      3.25  0.00123 **
## Year2005          0.6850      0.2005      3.42  0.00068 ***
## Year2006          0.7380      0.1946      3.79  0.00017 ***
## Year2007          0.6984      0.1976      3.54  0.00044 ***
## Year2008          0.7470      0.2042      3.66  0.00028 ***
## Year2009          0.6493      0.2220      2.92  0.00360 **
## Year2010          0.6744      0.2151      3.14  0.00181 **
## Year2011          0.7247      0.2167      3.34  0.00088 ***
## Year2012          0.7348      0.2267      3.24  0.00126 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.463
## Multiple R-squared:  0.231, Adjusted R-squared:  0.2
## Convergence in 47 IRWLS iterations
##
## Robustness weights:
## 4 observations c(40,42,73,83) are outliers with |weight| = 0 ( <
0.00018);
## 41 weights are ~= 1. The remaining 513 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0074 0.8340 0.9370 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           1.79e-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 1.293 1 1.137
## LastAuthorFemale 1.209 1 1.100
## Year 1.334 16 1.009

```

Residuals from first and last author



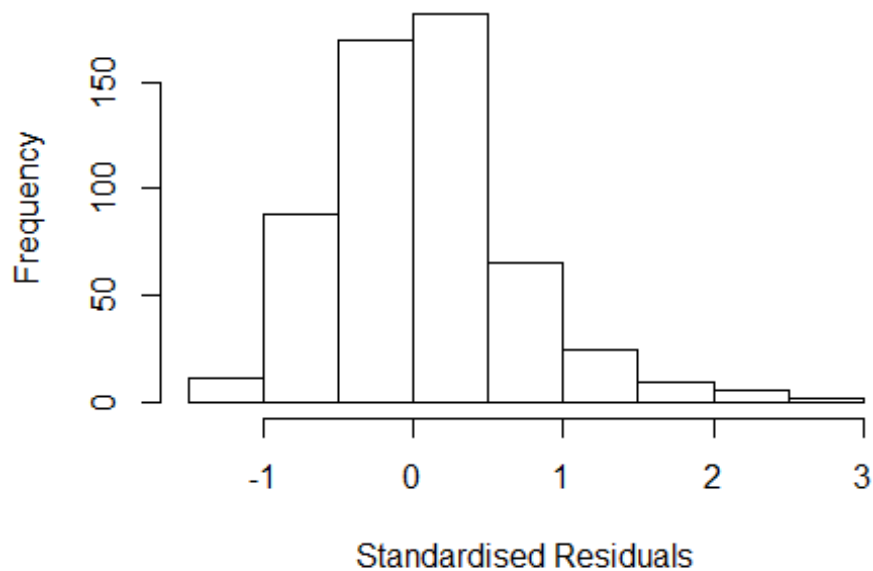
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 286 0029693941 3.125 1996    1101     3     2.717
## 433 0030827339 3.221 1997    1100     2     2.538
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2135 -0.4075  0.0321  0.3796  2.7170
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4080    0.1636   2.49  0.01294 *
## FirstAuthorFemale1 0.0947    0.0566   1.67  0.09484 .
## LastAuthorFemale1 -0.0969    0.0708  -1.37  0.17136
## Year1997         0.2748    0.5457   0.50  0.61482
## Year1998         0.3056    0.2927   1.04  0.29677
## Year1999         0.7054    0.1961   3.60  0.00035 ***
## Year2000         0.6051    0.1943   3.11  0.00194 **
## Year2001         0.8035    0.1915   4.20  3.2e-05 ***
## Year2002         0.7397    0.1904   3.88  0.00012 ***
## Year2003         0.5748    0.1879   3.06  0.00233 **
## Year2004         0.8077    0.1779   4.54  6.9e-06 ***
```

```

## Year2005          0.7940      0.1722      4.61  5.0e-06 ***
## Year2006          0.8386      0.1675      5.01  7.5e-07 ***
## Year2007          0.7967      0.1737      4.59  5.6e-06 ***
## Year2008          0.8331      0.1790      4.66  4.1e-06 ***
## Year2009          0.7182      0.2062      3.48  0.00054 ***
## Year2010          0.7496      0.1984      3.78  0.00018 ***
## Year2011          0.8089      0.1856      4.36  1.6e-05 ***
## Year2012          0.8523      0.1985      4.29  2.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.474
## Multiple R-squared:  0.183, Adjusted R-squared:  0.156
## Convergence in 49 IRWLS iterations
##
## Robustness weights:
## 5 observations c(1,40,42,73,254)
## are outliers with |weight| = 0 ( < 0.00018);
## 45 weights are ~= 1. The remaining 508 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0016 0.8210 0.9340 0.8690 0.9790 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.79e-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 1.143 1      1.069
## Year              1.143 16      1.004

```

Residuals from first author



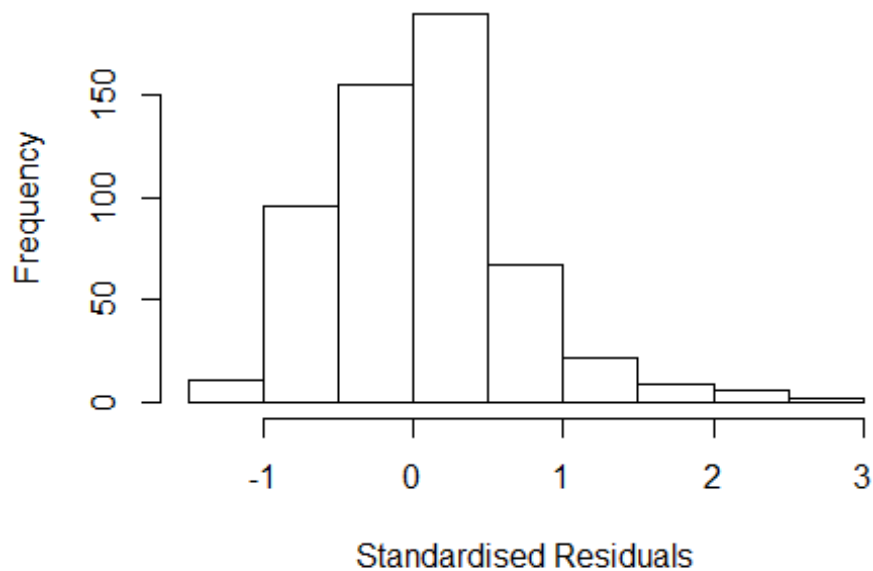
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 286 0029693941 3.125 1996    1101     3     2.717
## 433 0030827339 3.221 1997    1100     2     2.538
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.285 -0.370  0.025  0.385  2.755
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.3698    0.1642   2.25 0.02475 *
## FirstAuthorFemale1 0.0776    0.0529   1.47 0.14252
## Year1997        0.1983    0.5786   0.34 0.73196
## Year1998        0.3193    0.3206   1.00 0.31966
## Year1999        0.7436    0.1966   3.78 0.00017 ***
## Year2000        0.6375    0.1951   3.27 0.00115 **
## Year2001        0.8335    0.1927   4.33 1.8e-05 ***
## Year2002        0.7803    0.1921   4.06 5.6e-05 ***
## Year2003        0.6012    0.1907   3.15 0.00170 **
## Year2004        0.8372    0.1807   4.63 4.5e-06 ***
## Year2005        0.8148    0.1722   4.73 2.8e-06 ***
```

```

## Year2006          0.8618      0.1686      5.11  4.5e-07 ***
## Year2007          0.8164      0.1747      4.67  3.8e-06 ***
## Year2008          0.8508      0.1813      4.69  3.4e-06 ***
## Year2009          0.7563      0.2093      3.61  0.00033 ***
## Year2010          0.7747      0.1991      3.89  0.00011 ***
## Year2011          0.8361      0.1875      4.46  1.0e-05 ***
## Year2012          0.8927      0.1979      4.51  7.9e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.455
## Multiple R-squared:  0.203, Adjusted R-squared:  0.178
## Convergence in 42 IRWLS iterations
##
## Robustness weights:
## 7 observations c(1,32,40,42,73,78,254)
## are outliers with |weight| = 0 ( < 0.00018);
## 40 weights are ~= 1. The remaining 511 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0056 0.8270 0.9310 0.8630 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      1.79e-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 1.137 1      1.066
## Year      1.137 16      1.004

```

Residuals from last author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 286 0029693941 3.125 1996    1101     3     2.717
## 433 0030827339 3.221 1997    1100     2     2.538
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2464 -0.4073  0.0368  0.3791  2.7070
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4180    0.1629   2.57 0.01055 *
## LastAuthorFemale1 -0.0680    0.0679  -1.00 0.31702
## Year1997        0.1934    0.5137   0.38 0.70673
## Year1998        0.3142    0.3074   1.02 0.30717
## Year1999        0.6955    0.1956   3.56 0.00041 ***
## Year2000        0.6165    0.1978   3.12 0.00193 **
## Year2001        0.8284    0.1950   4.25 2.5e-05 ***
## Year2002        0.7718    0.1959   3.94 9.3e-05 ***
## Year2003        0.5826    0.1890   3.08 0.00216 **
## Year2004        0.8282    0.1802   4.60 5.3e-06 ***
## Year2005        0.8156    0.1743   4.68 3.6e-06 ***
```

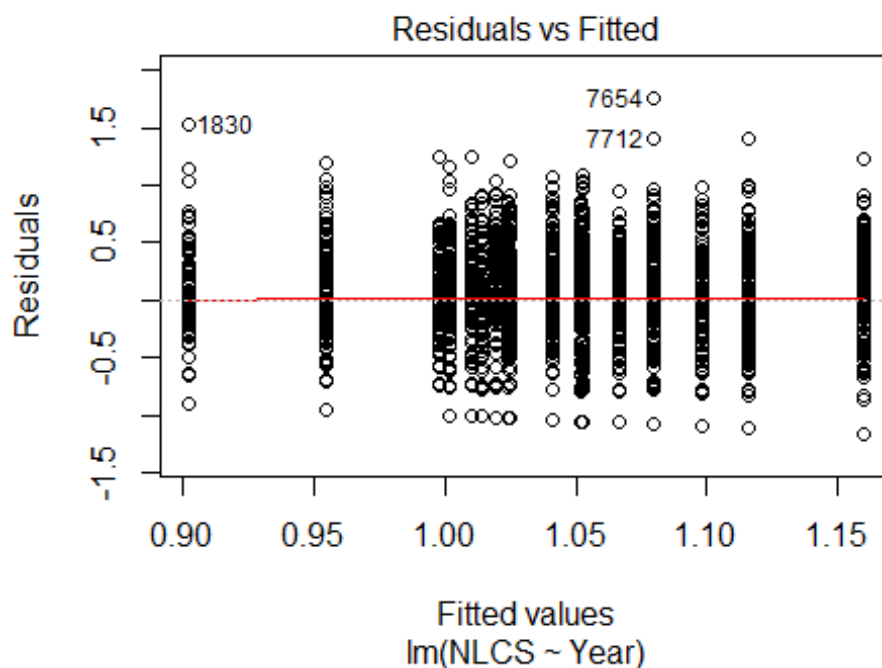
```

## Year2006          0.8599      0.1695      5.07  5.4e-07 ***
## Year2007          0.8073      0.1747      4.62  4.8e-06 ***
## Year2008          0.8364      0.1804      4.64  4.5e-06 ***
## Year2009          0.7388      0.2118      3.49  0.00053 ***
## Year2010          0.7652      0.2020      3.79  0.00017 ***
## Year2011          0.8197      0.1870      4.38  1.4e-05 ***
## Year2012          0.8702      0.1995      4.36  1.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.456
## Multiple R-squared:  0.197, Adjusted R-squared:  0.172
## Convergence in 38 IRWLS iterations
##
## Robustness weights:
## 8 observations c(1,32,40,42,73,78,83,254)
## are outliers with |weight| = 0 ( < 0.00018);
## 37 weights are ~= 1. The remaining 513 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0258 0.8330 0.9250 0.8660 0.9790 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.79e-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: 558"
## [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
## 429 406 369 352 394 408 377 362 390 383 393 394 351 454 381
## 2011 2012
## 375 394
##

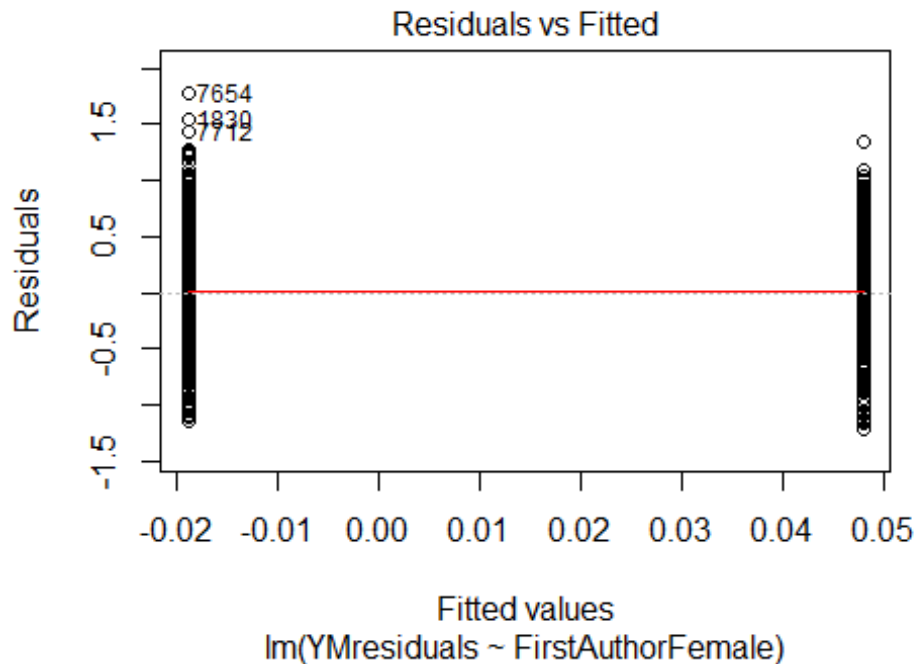
```



```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 133 111 102 130 87 102 133 103 150 139 151 145 159 200 176
## 2011 2012
## 182 203
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 106 96 84 109 71 84 110 81 124 115 126 124 132 167 148
## 2011 2012
## 151 170
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 33, df = 16, p-value = 0.008
```

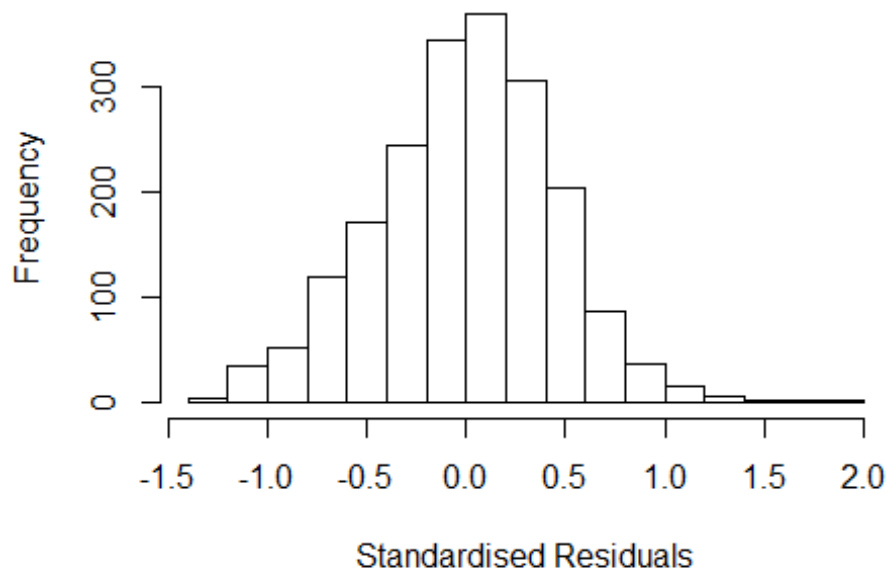


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.6, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.95085199244841"
## [1] "Male first author team size 2018 geometric mean: 3.14797263407039"
##
## 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: 2.70128895096073"
## [1] "Male last author team size 2018 geometric mean: 3.23291795430713"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, 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 1.078 1      1.038
## LastAuthorFemale  1.076 1      1.037
## UniqueAuthors    1.246 4      1.028
## Year              1.348 16     1.009
```

Residuals from first and last author and team size



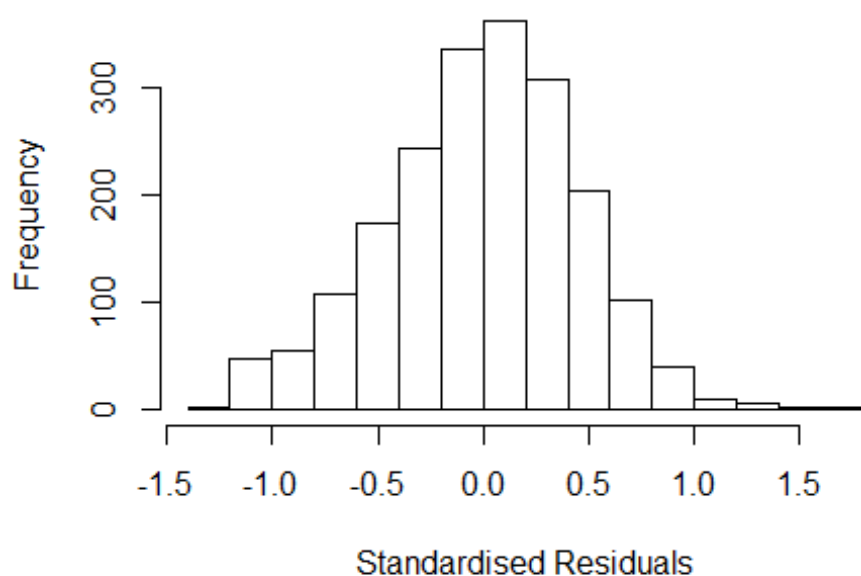
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2457 -0.2947 0.0194 0.2956 1.8435
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.89781 0.05936 15.13 < 2e-16 ***
## FirstAuthorFemale1 0.07524 0.02253 3.34 0.00085 ***
## LastAuthorFemale1 0.00170 0.02516 0.07 0.94608
## UniqueAuthors2 0.13724 0.04128 3.32 0.00090 ***
## UniqueAuthors3 0.18262 0.04034 4.53 6.3e-06 ***
## UniqueAuthors4 0.24498 0.04211 5.82 7.0e-09 ***
## UniqueAuthors5 0.28807 0.04208 6.85 1.0e-11 ***
## Year1997 -0.05595 0.06955 -0.80 0.42126
## Year1998 -0.09337 0.06538 -1.43 0.15343
## Year1999 -0.01988 0.07067 -0.28 0.77846
```

```

## Year2000      -0.16856    0.07569   -2.23  0.02606 *
## Year2001      0.00124    0.07997    0.02  0.98763
## Year2002     -0.14105    0.07034   -2.01  0.04509 *
## Year2003     -0.11882    0.07151   -1.66  0.09678 .
## Year2004     -0.10683    0.06393   -1.67  0.09488 .
## Year2005     -0.09518    0.06289   -1.51  0.13029
## Year2006     -0.10436    0.06401   -1.63  0.10321
## Year2007     -0.04188    0.06292   -0.67  0.50578
## Year2008     -0.08673    0.06738   -1.29  0.19822
## Year2009     -0.01721    0.06221   -0.28  0.78210
## Year2010      0.01274    0.06490    0.20  0.84443
## Year2011      0.02601    0.06250    0.42  0.67732
## Year2012     -0.04458    0.06469   -0.69  0.49078
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0616, Adjusted R-squared:  0.0512
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 200 weights are ~= 1. The remaining 1798 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0368 0.8640 0.9490 0.9020 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.01e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.066 1 1.033
## LastAuthorFemale 1.060 1 1.030
## Year 1.101 16 1.003

```

Residuals from first and last author



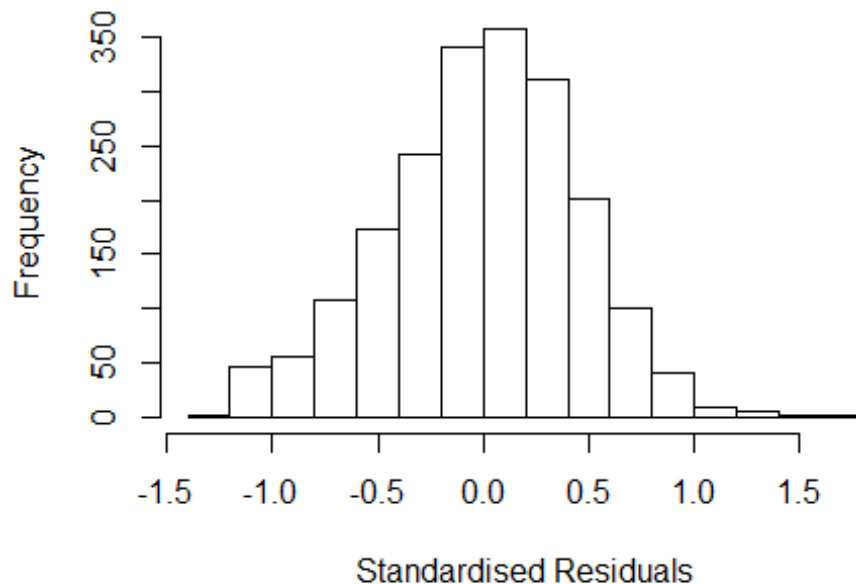
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2057 -0.3054  0.0194  0.3079  1.7781
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.04170    0.05217   19.97 < 2e-16 ***
## FirstAuthorFemale1  0.08428    0.02298    3.67  0.00025 ***
## LastAuthorFemale1 -0.00611    0.02563   -0.24  0.81162
## Year1997        -0.05409    0.07084   -0.76  0.44522
## Year1998        -0.08027    0.06851   -1.17  0.24146
## Year1999        -0.01180    0.07219   -0.16  0.87017
## Year2000        -0.15014    0.07878   -1.91  0.05684 .
## Year2001         0.03107    0.08059    0.39  0.69992
## Year2002        -0.11659    0.07220   -1.61  0.10653
## Year2003        -0.07662    0.07208   -1.06  0.28794
## Year2004        -0.07515    0.06472   -1.16  0.24567
## Year2005        -0.04527    0.06400   -0.71  0.47943
```

```

## Year2006      -0.05637    0.06394   -0.88  0.37811
## Year2007      0.00414    0.06360    0.07  0.94812
## Year2008     -0.03512    0.06807   -0.52  0.60593
## Year2009      0.03769    0.06216    0.61  0.54435
## Year2010      0.06684    0.06553    1.02  0.30788
## Year2011      0.08584    0.06388    1.34  0.17916
## Year2012      0.01425    0.06586    0.22  0.82872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.446
## Multiple R-squared:  0.026, Adjusted R-squared:  0.0171
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 179 weights are ~= 1. The remaining 1819 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0762 0.8650 0.9480 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      5.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.053 1      1.026
## Year      1.053 16      1.002

```

Residuals from first author



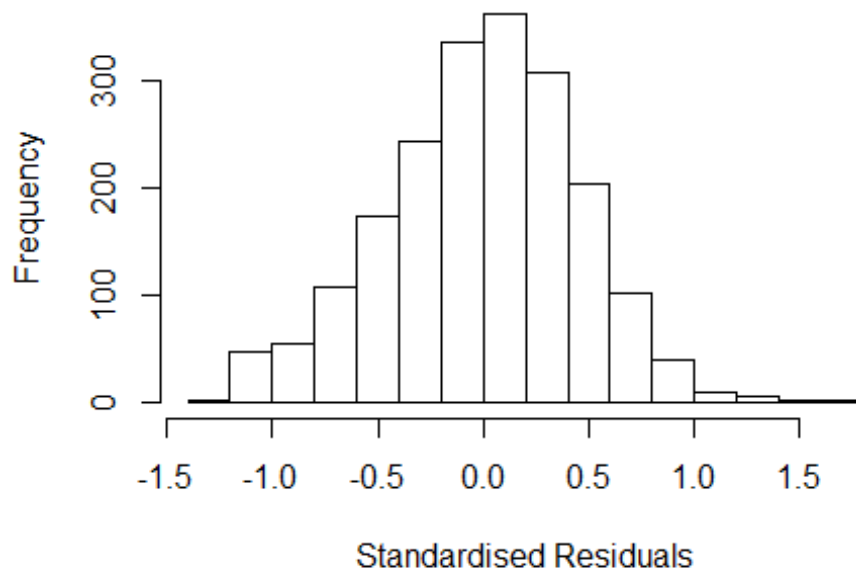
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2099 -0.3059 0.0182 0.3077 1.7794
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04046 0.05160 20.16 < 2e-16 ***
## FirstAuthorFemale1 0.08350 0.02285 3.65 0.00026 ***
## Year1997 -0.05306 0.07039 -0.75 0.45105
## Year1998 -0.08021 0.06848 -1.17 0.24163
## Year1999 -0.01154 0.07214 -0.16 0.87290
## Year2000 -0.14984 0.07868 -1.90 0.05701 .
## Year2001 0.03156 0.08042 0.39 0.69475
## Year2002 -0.11633 0.07217 -1.61 0.10717
## Year2003 -0.07654 0.07206 -1.06 0.28834
## Year2004 -0.07500 0.06467 -1.16 0.24631
## Year2005 -0.04535 0.06395 -0.71 0.47830
## Year2006 -0.05620 0.06388 -0.88 0.37910
```

```

## Year2007          0.00477    0.06341    0.08  0.94010
## Year2008         -0.03521    0.06804   -0.52  0.60493
## Year2009          0.03775    0.06213    0.61  0.54348
## Year2010          0.06673    0.06553    1.02  0.30867
## Year2011          0.08589    0.06383    1.35  0.17861
## Year2012          0.01416    0.06585    0.22  0.82971
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.446
## Multiple R-squared:  0.026, Adjusted R-squared:  0.0176
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 184 weights are ~= 1. The remaining 1814 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0751 0.8640 0.9480 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      5.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.051 1          1.025
## Year            1.051 16          1.002

```


Residuals from last author



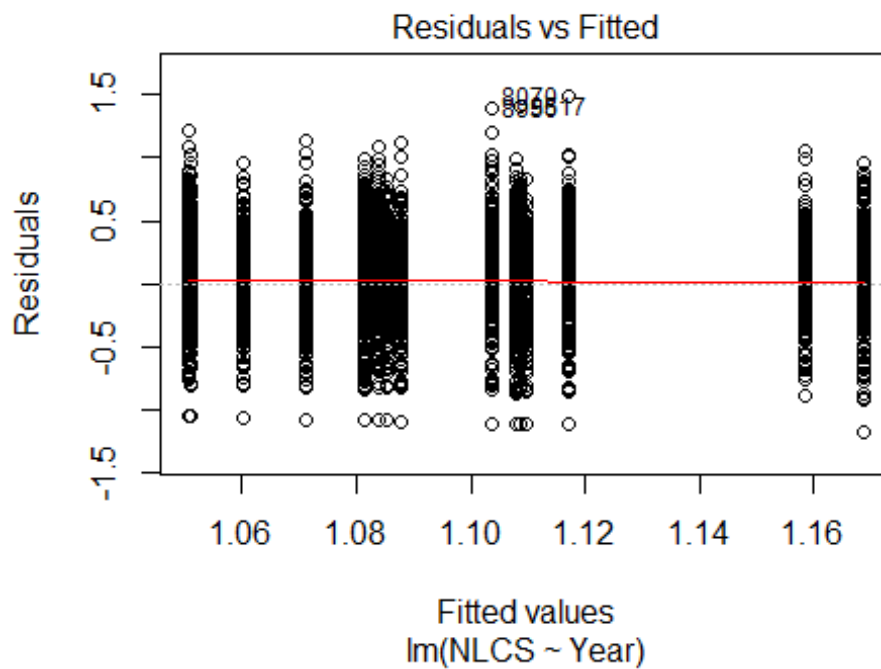
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1608 -0.2990 0.0132 0.3057 1.7482
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.049795 0.051943 20.21 <2e-16 ***
## LastAuthorFemale1 0.007048 0.025566 0.28 0.783
## Year1997 -0.047290 0.070243 -0.67 0.501
## Year1998 -0.075576 0.068182 -1.11 0.268
## Year1999 0.000444 0.072396 0.01 0.995
## Year2000 -0.140718 0.078168 -1.80 0.072 .
## Year2001 0.046820 0.080351 0.58 0.560
## Year2002 -0.103683 0.072425 -1.43 0.152
## Year2003 -0.059091 0.071538 -0.83 0.409
## Year2004 -0.065856 0.064506 -1.02 0.307
## Year2005 -0.027942 0.063556 -0.44 0.660
## Year2006 -0.038658 0.063195 -0.61 0.541
```

```

## Year2007      0.020461  0.063175  0.32  0.746
## Year2008     -0.019904  0.067773 -0.29  0.769
## Year2009      0.052769  0.061611  0.86  0.392
## Year2010      0.085910  0.064870  1.32  0.186
## Year2011      0.103979  0.062978  1.65  0.099 .
## Year2012      0.035977  0.064833  0.55  0.579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.446
## Multiple R-squared:  0.0192, Adjusted R-squared:  0.0108
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 175 weights are ~= 1. The remaining 1823 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0904 0.8660 0.9490 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      5.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1998"
## [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
## 602 558 570 497 530 487 489 492 476 436 534 530 590 604 581
## 2011 2012
## 572 622
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 248 214 255 264 189 158 252 235 247 241 279 265 323 299 309
## 2011 2012

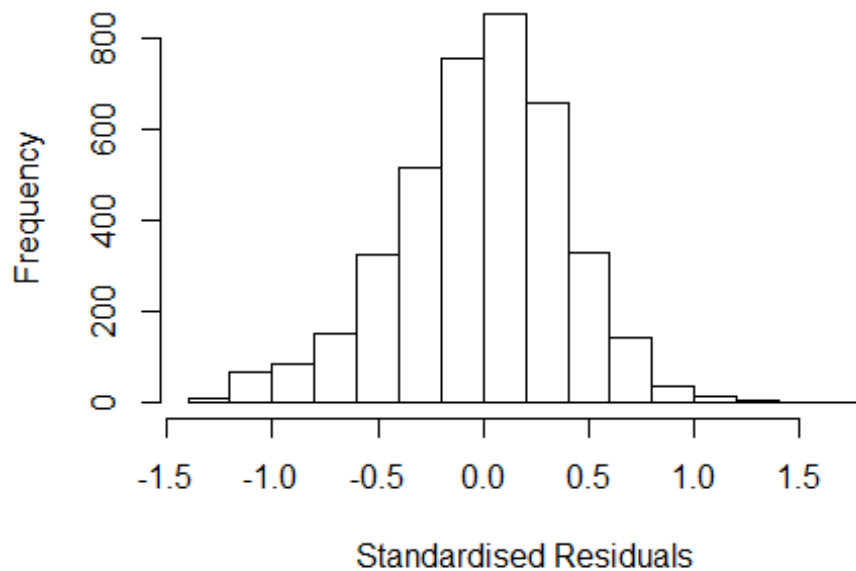
```

```
## 327 330
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 215 186 229 227 168 143 229 207 213 213 249 235 285 266 276
## 2011 2012
## 293 299
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 67, df = 16, p-value = 4e-08
```



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


Residuals from first and last author and team size



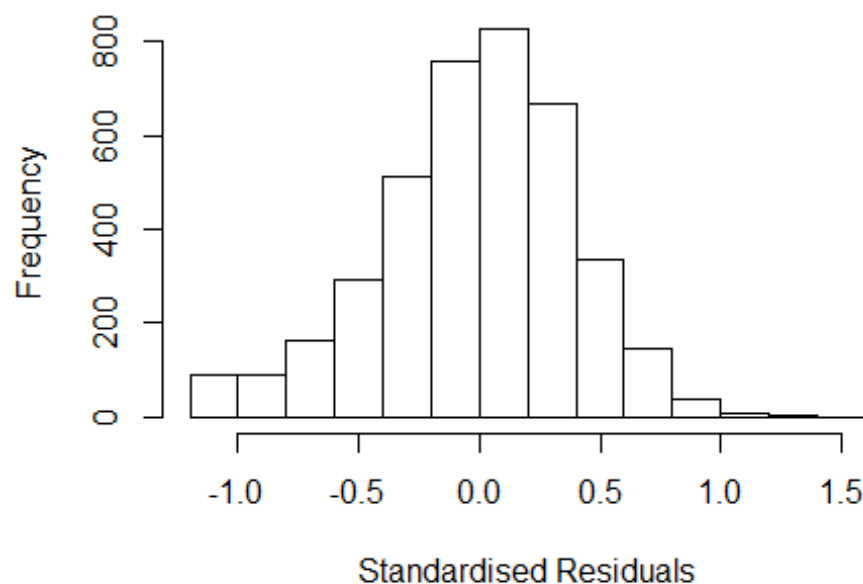
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3002 -0.2519  0.0179  0.2521  1.6049
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10080    0.03544   31.06 < 2e-16 ***
## FirstAuthorFemale1  0.00579    0.01327    0.44  0.66264
## LastAuthorFemale1 -0.01030    0.01539   -0.67  0.50340
## UniqueAuthors2     0.11462    0.02200    5.21  2.0e-07 ***
## UniqueAuthors3     0.12068    0.02343    5.15  2.7e-07 ***
## UniqueAuthors4     0.20387    0.02720    7.50  8.1e-14 ***
## UniqueAuthors5     0.18013    0.02874    6.27  4.1e-10 ***
## Year1997          -0.04782    0.04472   -1.07  0.28499
## Year1998          -0.09762    0.04156   -2.35  0.01888 *
## Year1999          -0.13996    0.04118   -3.40  0.00068 ***
```

```

## Year2000      -0.09257    0.04466   -2.07  0.03826 *
## Year2001      -0.10569    0.04890   -2.16  0.03074 *
## Year2002      -0.05518    0.03754   -1.47  0.14167
## Year2003      -0.08281    0.03849   -2.15  0.03149 *
## Year2004      -0.09397    0.04098   -2.29  0.02189 *
## Year2005      -0.10603    0.04018   -2.64  0.00835 **
## Year2006      -0.10098    0.03939   -2.56  0.01040 *
## Year2007      -0.13820    0.03928   -3.52  0.00044 ***
## Year2008      -0.10273    0.03974   -2.59  0.00977 **
## Year2009      -0.09701    0.04111   -2.36  0.01833 *
## Year2010      -0.09703    0.03952   -2.46  0.01412 *
## Year2011      -0.11608    0.03983   -2.91  0.00359 **
## Year2012      -0.16111    0.04287   -3.76  0.00017 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0279, Adjusted R-squared:  0.0225
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 335 weights are ~= 1. The remaining 3598 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0286 0.8690 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          2.54e-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 1.064 1 1.032
## LastAuthorFemale 1.041 1 1.020
## Year 1.061 16 1.002

```

Residuals from first and last author



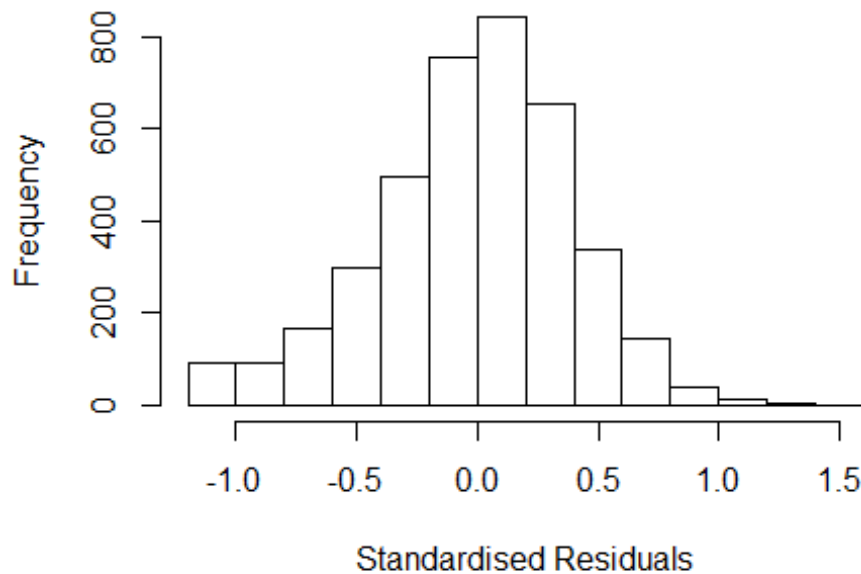
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1898 -0.2574  0.0168  0.2529  1.4867
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1898     0.0314   37.94  <2e-16 ***
## FirstAuthorFemale1  0.0130     0.0133    0.97   0.3296
## LastAuthorFemale1 -0.0151     0.0155   -0.97   0.3321
## Year1997          -0.0399     0.0451   -0.89   0.3753
## Year1998          -0.0849     0.0415   -2.05   0.0407 *
## Year1999          -0.1164     0.0412   -2.82   0.0048 **
## Year2000          -0.0783     0.0442   -1.77   0.0769 .
## Year2001          -0.0770     0.0487   -1.58   0.1144
## Year2002          -0.0324     0.0376   -0.86   0.3894
## Year2003          -0.0608     0.0387   -1.57   0.1162
## Year2004          -0.0692     0.0410   -1.69   0.0911 .
## Year2005          -0.0847     0.0403   -2.10   0.0357 *
```

```

## Year2006          -0.0783      0.0396    -1.98    0.0479 *
## Year2007          -0.1087      0.0398    -2.73    0.0064 **
## Year2008          -0.0735      0.0400    -1.84    0.0664 .
## Year2009          -0.0673      0.0412    -1.63    0.1028
## Year2010          -0.0652      0.0396    -1.65    0.0996 .
## Year2011          -0.0817      0.0399    -2.05    0.0407 *
## Year2012          -0.1163      0.0431    -2.70    0.0069 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared:  0.00558,    Adjusted R-squared:  0.00101
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 316 weights are ~= 1. The remaining 3617 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.086  0.870  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      2.54e-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 1.045 1      1.022
## Year              1.045 16      1.001

```


Residuals from first author



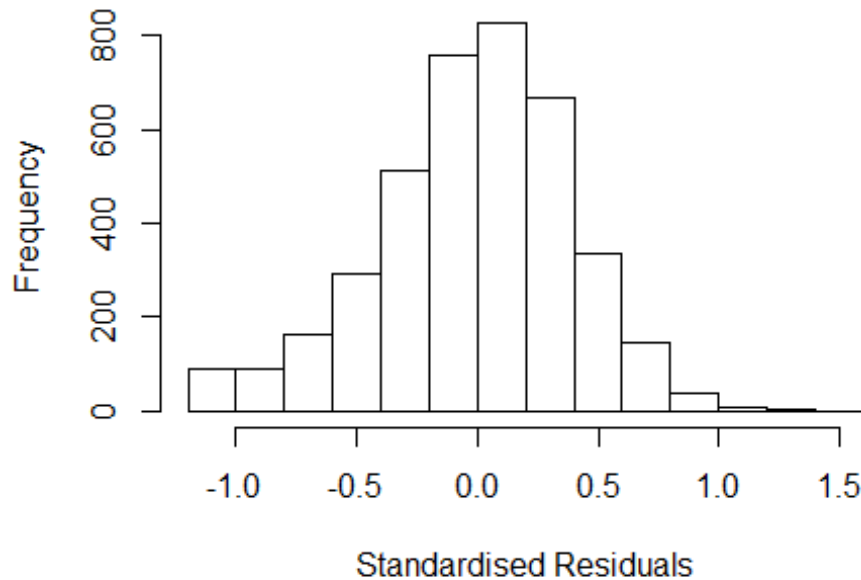
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1987 -0.2567 0.0162 0.2530 1.4892
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1881 0.0314 37.90 <2e-16 ***
## FirstAuthorFemale1 0.0106 0.0132 0.80 0.4228
## Year1997 -0.0409 0.0450 -0.91 0.3638
## Year1998 -0.0851 0.0415 -2.05 0.0402 *
## Year1999 -0.1171 0.0412 -2.84 0.0045 **
## Year2000 -0.0795 0.0442 -1.80 0.0721 .
## Year2001 -0.0783 0.0487 -1.61 0.1082
## Year2002 -0.0327 0.0376 -0.87 0.3834
## Year2003 -0.0614 0.0387 -1.59 0.1122
## Year2004 -0.0698 0.0409 -1.70 0.0884 .
## Year2005 -0.0859 0.0403 -2.13 0.0330 *
## Year2006 -0.0789 0.0396 -1.99 0.0463 *
```

```

## Year2007          -0.1090      0.0398   -2.74   0.0061 **
## Year2008          -0.0743      0.0400   -1.86   0.0632 .
## Year2009          -0.0683      0.0412   -1.66   0.0973 .
## Year2010          -0.0665      0.0395   -1.68   0.0925 .
## Year2011          -0.0834      0.0398   -2.09   0.0365 *
## Year2012          -0.1177      0.0430   -2.74   0.0062 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared:  0.00534,    Adjusted R-squared:  0.00102
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 313 weights are ~= 1. The remaining 3620 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0844 0.8700 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      2.54e-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 1.023 1      1.011
## Year              1.023 16      1.001

```

Residuals from last author



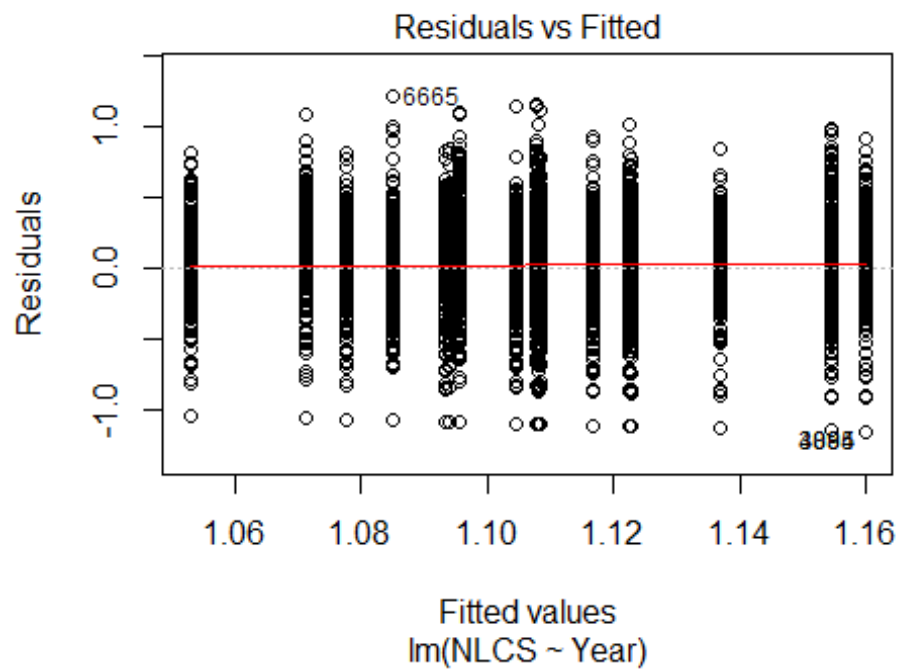
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1921 -0.2541 0.0173 0.2544 1.4814
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1921 0.0313 38.11 <2e-16 ***
## LastAuthorFemale1 -0.0124 0.0154 -0.80 0.4224
## Year1997 -0.0391 0.0450 -0.87 0.3856
## Year1998 -0.0839 0.0415 -2.02 0.0435 *
## Year1999 -0.1155 0.0412 -2.80 0.0051 **
## Year2000 -0.0765 0.0441 -1.73 0.0830 .
## Year2001 -0.0753 0.0487 -1.55 0.1220
## Year2002 -0.0299 0.0374 -0.80 0.4240
## Year2003 -0.0583 0.0386 -1.51 0.1310
## Year2004 -0.0669 0.0408 -1.64 0.1016
## Year2005 -0.0834 0.0403 -2.07 0.0383 *
## Year2006 -0.0761 0.0395 -1.93 0.0540 .
```

```

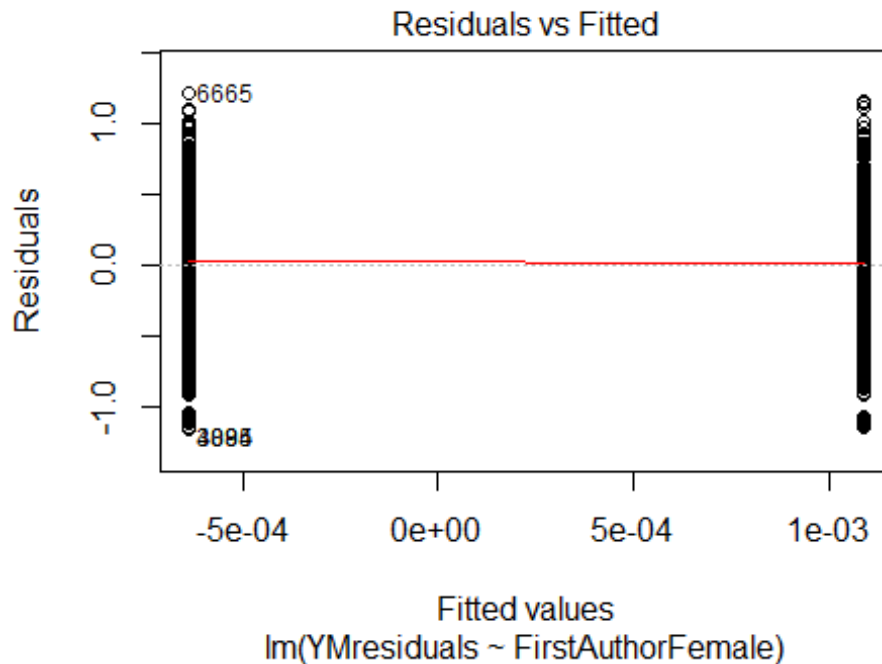
## Year2007          -0.1062      0.0397   -2.67    0.0076 **
## Year2008          -0.0705      0.0399   -1.77    0.0770 .
## Year2009          -0.0643      0.0411   -1.57    0.1175
## Year2010          -0.0621      0.0395   -1.57    0.1157
## Year2011          -0.0781      0.0396   -1.97    0.0488 *
## Year2012          -0.1134      0.0429   -2.64    0.0083 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared:  0.00534,    Adjusted R-squared:  0.00102
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 306 weights are ~= 1. The remaining 3627 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0888 0.8700 0.9510 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.54e-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: 3933"
## [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
## 593 564 500 480 498 456 419 424 396 443 463 471 585 532 509
## 2011 2012
## 545 574
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 324 287 274 258 202 145 250 253 235 293 298 314 392 357 345
## 2011 2012

```

```
## 375 402
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 291 257 247 230 178 133 228 221 206 259 259 271 350 315 300
## 2011 2012
## 338 367
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 82, df = 16, p-value = 6e-11
```

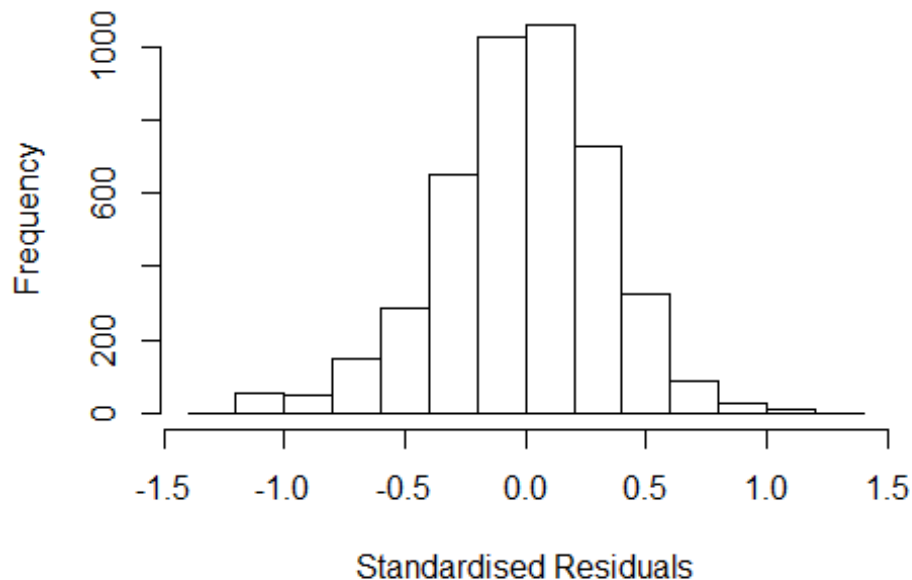


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 1e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 3.33595932989801"
## [1] "Male first author team size 2018 geometric mean: 3.57101223383029"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.32044932429154"
## [1] "Male last author team size 2018 geometric mean: 3.50587710479069"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9000, 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 1.069 1      1.034
## LastAuthorFemale  1.044 1      1.022
## UniqueAuthors     1.148 4      1.017
## Year              1.206 16      1.006
```

Residuals from first and last author and team size



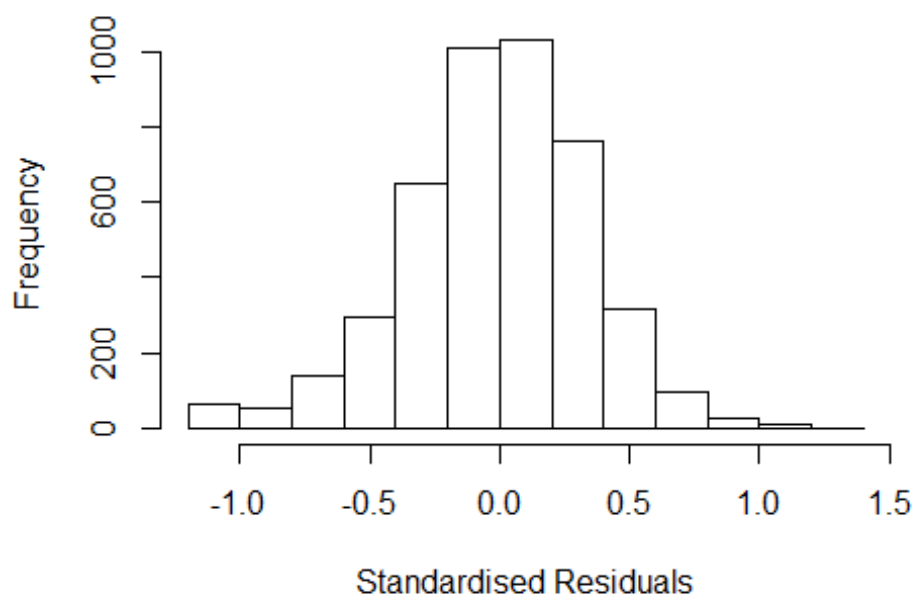
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21084 -0.21520  0.00211  0.21263  1.33846
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06475    0.02977   35.77 < 2e-16 ***
## FirstAuthorFemale1 -0.00691    0.01060   -0.65  0.51448
## LastAuthorFemale1 -0.00546    0.01299   -0.42  0.67401
## UniqueAuthors2     0.11985    0.02230    5.37 8.1e-08 ***
## UniqueAuthors3     0.10944    0.02265    4.83 1.4e-06 ***
## UniqueAuthors4     0.16832    0.02409    6.99 3.2e-12 ***
## UniqueAuthors5     0.21522    0.02473    8.70 < 2e-16 ***
## Year1997         -0.04106    0.03253   -1.26  0.20699
## Year1998         -0.01532    0.03350   -0.46  0.64754
## Year1999         -0.04073    0.03233   -1.26  0.20778
```

```

## Year2000      -0.01832    0.03303   -0.55  0.57919
## Year2001      -0.01420    0.03854   -0.37  0.71268
## Year2002       0.00471    0.03205    0.15  0.88310
## Year2003      -0.05696    0.03152   -1.81  0.07080 .
## Year2004      -0.10880    0.03271   -3.33  0.00089 ***
## Year2005      -0.07424    0.03089   -2.40  0.01628 *
## Year2006      -0.08873    0.02974   -2.98  0.00287 **
## Year2007      -0.10120    0.02971   -3.41  0.00066 ***
## Year2008      -0.08532    0.02824   -3.02  0.00253 **
## Year2009      -0.07173    0.02879   -2.49  0.01277 *
## Year2010      -0.09142    0.03022   -3.03  0.00250 **
## Year2011      -0.08984    0.03198   -2.81  0.00499 **
## Year2012      -0.10341    0.03065   -3.37  0.00075 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.323
## Multiple R-squared:  0.0339, Adjusted R-squared:  0.0291
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 351 weights are ~= 1. The remaining 4099 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0475 0.8700 0.9530 0.8950 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.25e-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 1.054 1 1.027
## LastAuthorFemale 1.040 1 1.020
## Year 1.077 16 1.002

```


Residuals from first and last author



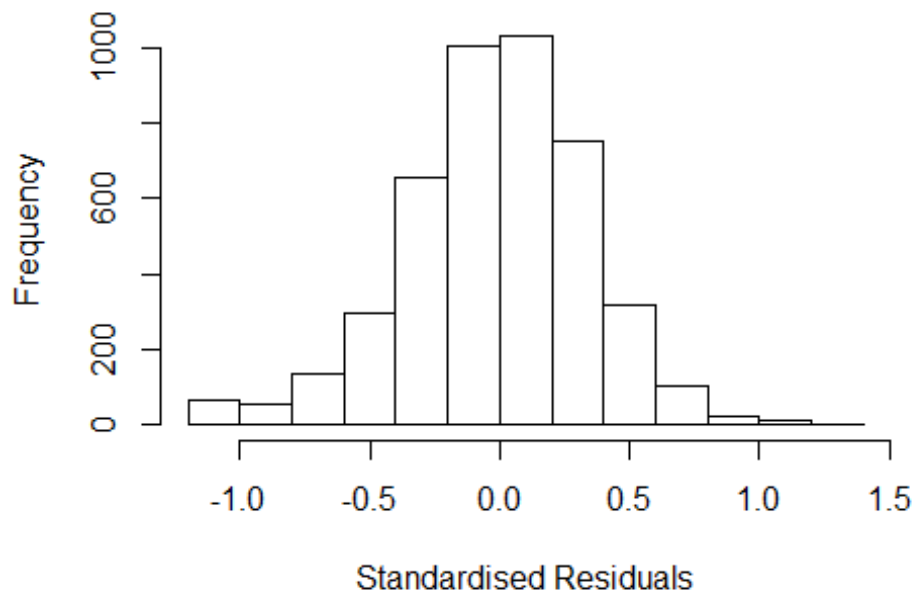
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19278 -0.22078 0.00263 0.21755 1.21589
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17101 0.02381 49.18 <2e-16 ***
## FirstAuthorFemale1 0.00237 0.01067 0.22 0.8240
## LastAuthorFemale1 -0.01066 0.01303 -0.82 0.4135
## Year1997 -0.03872 0.03290 -1.18 0.2393
## Year1998 -0.02243 0.03432 -0.65 0.5134
## Year1999 -0.04091 0.03291 -1.24 0.2140
## Year2000 -0.01026 0.03367 -0.30 0.7606
## Year2001 0.00307 0.03831 0.08 0.9361
## Year2002 0.02177 0.03241 0.67 0.5018
## Year2003 -0.03994 0.03187 -1.25 0.2101
## Year2004 -0.09168 0.03332 -2.75 0.0060 **
## Year2005 -0.05780 0.03123 -1.85 0.0643 .
```

```

## Year2006      -0.06776    0.03043   -2.23   0.0260 *
## Year2007      -0.08490    0.03034   -2.80   0.0052 **
## Year2008      -0.06291    0.02872   -2.19   0.0286 *
## Year2009      -0.05061    0.02921   -1.73   0.0832 .
## Year2010      -0.06379    0.03056   -2.09   0.0369 *
## Year2011      -0.05942    0.03265   -1.82   0.0689 .
## Year2012      -0.07860    0.03127   -2.51   0.0120 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.326
## Multiple R-squared:  0.00822,    Adjusted R-squared:  0.00419
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 356 weights are ~= 1. The remaining 4094 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.134  0.870  0.952  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      2.25e-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 1.047 1      1.023
## Year      1.047 16      1.001

```

Residuals from first author



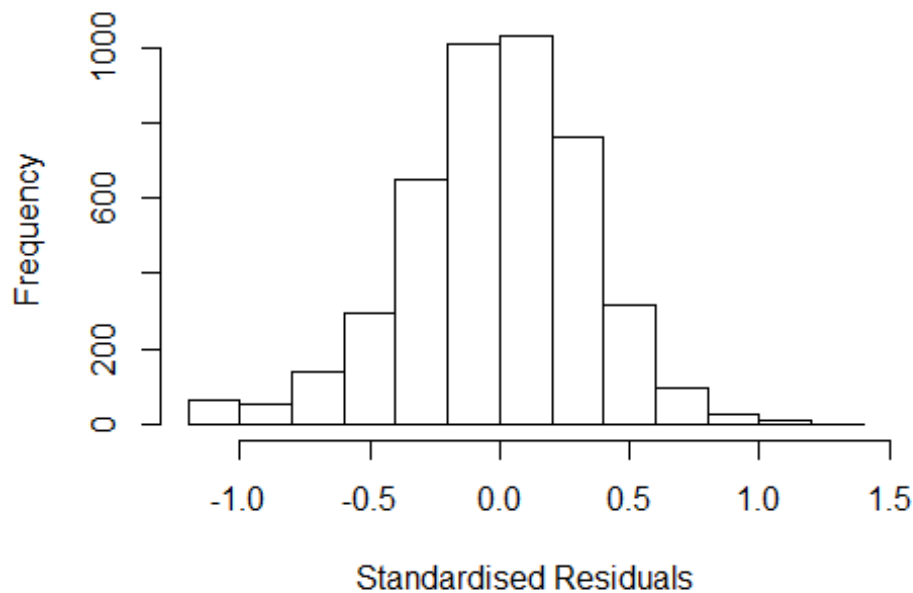
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19186 -0.22152 0.00247 0.21715 1.21745
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16979 0.02381 49.13 <2e-16 ***
## FirstAuthorFemale1 0.00128 0.01064 0.12 0.9040
## Year1997 -0.03929 0.03285 -1.20 0.2318
## Year1998 -0.02307 0.03426 -0.67 0.5008
## Year1999 -0.04060 0.03291 -1.23 0.2174
## Year2000 -0.01062 0.03364 -0.32 0.7523
## Year2001 0.00229 0.03829 0.06 0.9524
## Year2002 0.02207 0.03240 0.68 0.4957
## Year2003 -0.04001 0.03188 -1.26 0.2095
## Year2004 -0.09228 0.03329 -2.77 0.0056 **
## Year2005 -0.05804 0.03122 -1.86 0.0630 .
## Year2006 -0.06836 0.03040 -2.25 0.0246 *
```

```

## Year2007          -0.08524    0.03031   -2.81    0.0049 **
## Year2008          -0.06322    0.02871   -2.20    0.0277 *
## Year2009          -0.05168    0.02912   -1.77    0.0760 .
## Year2010          -0.06447    0.03055   -2.11    0.0349 *
## Year2011          -0.06031    0.03259   -1.85    0.0643 .
## Year2012          -0.07954    0.03121   -2.55    0.0109 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.326
## Multiple R-squared:  0.00808,    Adjusted R-squared:  0.00427
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 361 weights are ~= 1. The remaining 4089 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.133  0.870  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      2.25e-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 1.033 1          1.017
## Year            1.033 16          1.001

```

Residuals from last author



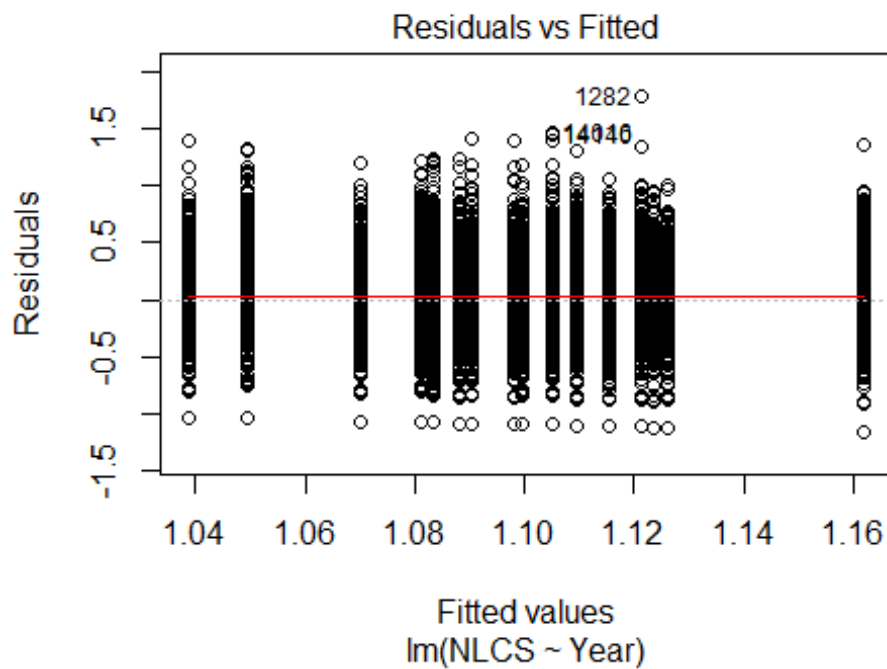
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19340 -0.22087 0.00284 0.21715 1.21502
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17153 0.02362 49.60 <2e-16 ***
## LastAuthorFemale1 -0.01030 0.01300 -0.79 0.4283
## Year1997 -0.03867 0.03289 -1.18 0.2398
## Year1998 -0.02227 0.03432 -0.65 0.5164
## Year1999 -0.04083 0.03292 -1.24 0.2148
## Year2000 -0.00995 0.03363 -0.30 0.7673
## Year2001 0.00331 0.03832 0.09 0.9312
## Year2002 0.02186 0.03242 0.67 0.5001
## Year2003 -0.03967 0.03186 -1.25 0.2132
## Year2004 -0.09130 0.03325 -2.75 0.0061 **
## Year2005 -0.05751 0.03124 -1.84 0.0657 .
## Year2006 -0.06742 0.03041 -2.22 0.0267 *
```

```

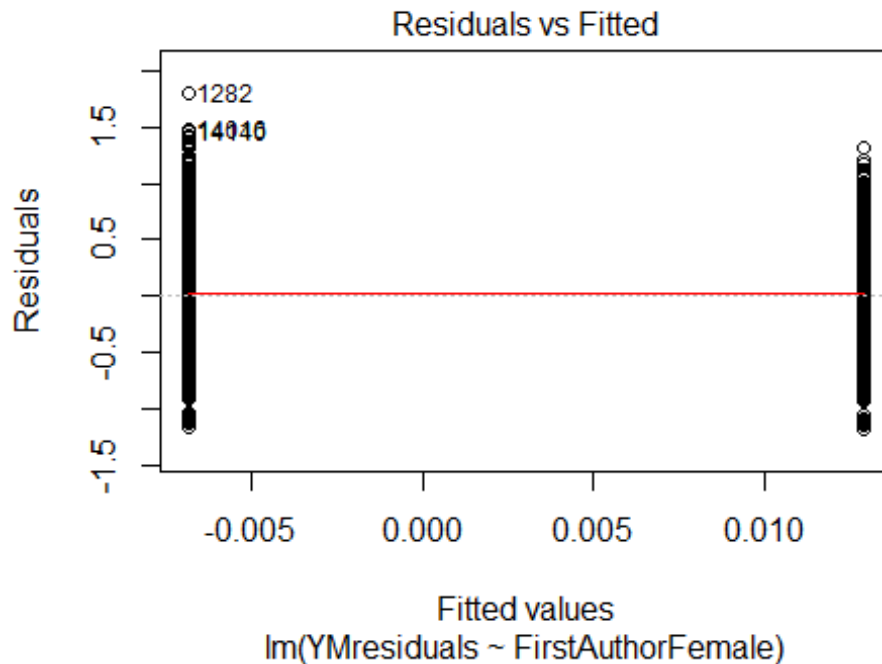
## Year2007          -0.08455      0.03033      -2.79      0.0053 **
## Year2008          -0.06236      0.02867      -2.18      0.0296 *
## Year2009          -0.05017      0.02919      -1.72      0.0858 .
## Year2010          -0.06334      0.03053      -2.07      0.0380 *
## Year2011          -0.05879      0.03250      -1.81      0.0705 .
## Year2012          -0.07809      0.03115      -2.51      0.0122 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.326
## Multiple R-squared:  0.00821,    Adjusted R-squared:  0.0044
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 353 weights are ~= 1. The remaining 4097 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.134  0.870  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      2.25e-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: 4450"
## [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
##  976  914  836  802  891  865  771  759  765  845  841  892 1087 1178 1125
## 2011 2012
## 1093 1238
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  555  517  523  525  413  288  553  539  541  617  604  653  816  850  842
## 2011 2012

```

```
## 823 953
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 499 457 471 459 371 256 494 484 483 550 532 594 729 772 766
## 2011 2012
## 757 871
## [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
```

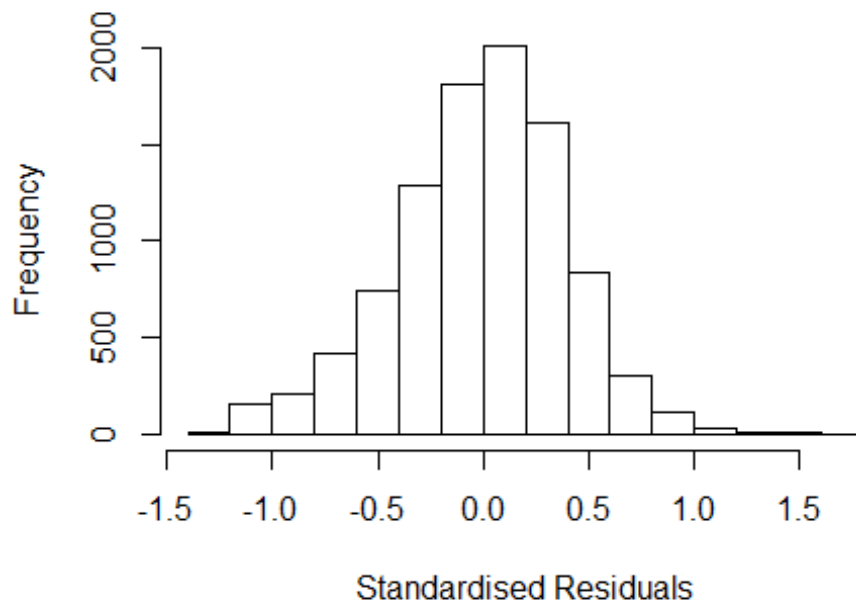


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 77, df = 1, p-value <2e-16
```



```
## [1] "Female first author team size 2018 geometric mean: 3.07769272388705"
## [1] "Male first author team size 2018 geometric mean: 2.92226933032813"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 72000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.8792897297959"
## [1] "Male last author team size 2018 geometric mean: 3.02785049379098"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 49000, 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 1.072 1          1.035
## LastAuthorFemale  1.055 1          1.027
## UniqueAuthors    1.091 4          1.011
## Year              1.127 16         1.004
```


Residuals from first and last author and team size



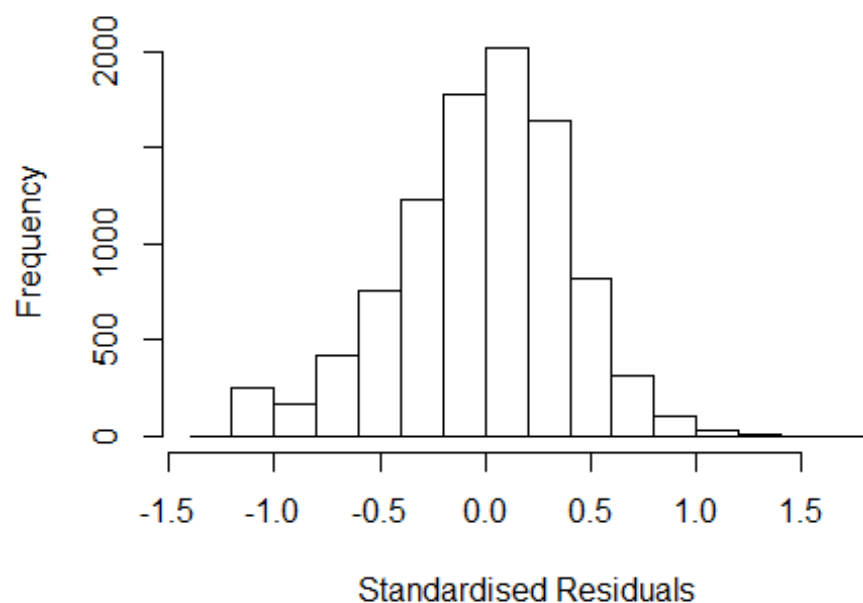
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3022 -0.2571  0.0148  0.2565  1.6869
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06206    0.02245   47.31 < 2e-16 ***
## FirstAuthorFemale1  0.01682    0.00857    1.96  0.04966 *
## LastAuthorFemale1 -0.02379    0.01010   -2.36  0.01852 *
## UniqueAuthors2     0.13515    0.01557    8.68 < 2e-16 ***
## UniqueAuthors3     0.14831    0.01627    9.12 < 2e-16 ***
## UniqueAuthors4     0.18730    0.01803   10.39 < 2e-16 ***
## UniqueAuthors5     0.24012    0.01887   12.72 < 2e-16 ***
## Year1997          -0.02428    0.02699   -0.90  0.36831
## Year1998          -0.06191    0.02607   -2.37  0.01758 *
## Year1999          -0.06653    0.02568   -2.59  0.00958 **
```

```

## Year2000      -0.05143    0.02757   -1.87  0.06213 .
## Year2001      -0.06081    0.03142   -1.94  0.05301 .
## Year2002      -0.08457    0.02542   -3.33  0.00088 ***
## Year2003      -0.09655    0.02526   -3.82  0.00013 ***
## Year2004      -0.14851    0.02771   -5.36  8.6e-08 ***
## Year2005      -0.08650    0.02549   -3.39  0.00069 ***
## Year2006      -0.09052    0.02504   -3.62  0.00030 ***
## Year2007      -0.10436    0.02514   -4.15  3.3e-05 ***
## Year2008      -0.09523    0.02398   -3.97  7.2e-05 ***
## Year2009      -0.08685    0.02400   -3.62  0.00030 ***
## Year2010      -0.10892    0.02432   -4.48  7.6e-06 ***
## Year2011      -0.11479    0.02453   -4.68  2.9e-06 ***
## Year2012      -0.14852    0.02460   -6.04  1.6e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.379
## Multiple R-squared:  0.0313, Adjusted R-squared:  0.0291
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 775 weights are ~= 1. The remaining 8770 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0092 0.8670 0.9500 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.05e-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 1.059 1 1.029
## LastAuthorFemale 1.045 1 1.022
## Year 1.044 16 1.001

```

Residuals from first and last author



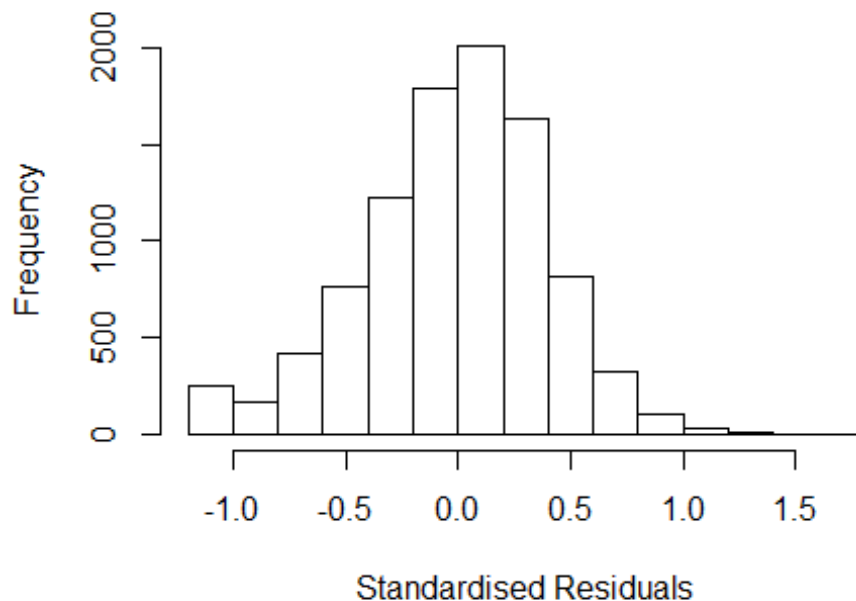
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2011 -0.2599  0.0167  0.2586  1.7595
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17437    0.01930   60.86 < 2e-16 ***
## FirstAuthorFemale1  0.02676    0.00861    3.11  0.0019 **
## LastAuthorFemale1 -0.02760    0.01020   -2.70  0.0069 **
## Year1997        -0.02192    0.02715   -0.81  0.4196
## Year1998        -0.05243    0.02637   -1.99  0.0468 *
## Year1999        -0.05126    0.02586   -1.98  0.0475 *
## Year2000        -0.03315    0.02721   -1.22  0.2232
## Year2001        -0.03571    0.03105   -1.15  0.2502
## Year2002        -0.07110    0.02571   -2.77  0.0057 **
## Year2003        -0.07825    0.02543   -3.08  0.0021 **
## Year2004        -0.12617    0.02775   -4.55 5.5e-06 ***
## Year2005        -0.06760    0.02577   -2.62  0.0087 **
```

```

## Year2006      -0.06737      0.02523      -2.67      0.0076 **
## Year2007      -0.07723      0.02526      -3.06      0.0022 **
## Year2008      -0.06918      0.02425      -2.85      0.0044 **
## Year2009      -0.05951      0.02416      -2.46      0.0138 *
## Year2010      -0.07831      0.02462      -3.18      0.0015 **
## Year2011      -0.07916      0.02467      -3.21      0.0013 **
## Year2012      -0.10843      0.02468      -4.39      1.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.00637,    Adjusted R-squared:  0.00449
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 787 weights are ~= 1. The remaining 8758 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0007 0.8660 0.9490 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      1.05e-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 1.031 1      1.016
## Year      1.031 16      1.001

```

Residuals from first author



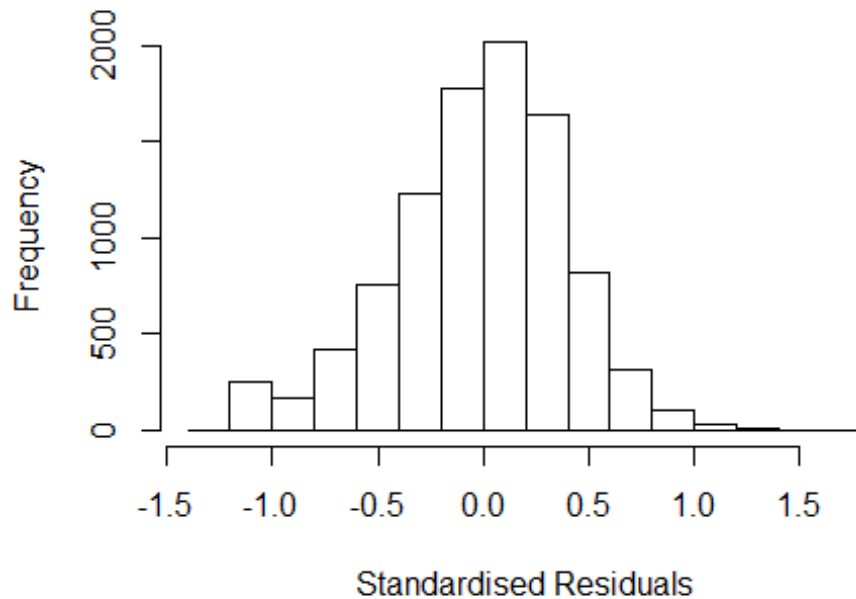
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1936 -0.2606 0.0168 0.2564 1.7642
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1708 0.0192 61.00 < 2e-16 ***
## FirstAuthorFemale1 0.0227 0.0085 2.67 0.0075 **
## Year1997 -0.0230 0.0271 -0.85 0.3957
## Year1998 -0.0532 0.0263 -2.02 0.0431 *
## Year1999 -0.0512 0.0258 -1.99 0.0471 *
## Year2000 -0.0337 0.0272 -1.24 0.2148
## Year2001 -0.0377 0.0310 -1.22 0.2244
## Year2002 -0.0711 0.0257 -2.77 0.0056 **
## Year2003 -0.0788 0.0254 -3.10 0.0020 **
## Year2004 -0.1284 0.0277 -4.63 3.7e-06 ***
## Year2005 -0.0684 0.0257 -2.66 0.0079 **
## Year2006 -0.0676 0.0252 -2.68 0.0073 **
```

```

## Year2007          -0.0774      0.0252   -3.07   0.0021 **
## Year2008          -0.0702      0.0242   -2.90   0.0037 **
## Year2009          -0.0602      0.0241   -2.50   0.0126 *
## Year2010          -0.0795      0.0246   -3.23   0.0012 **
## Year2011          -0.0809      0.0246   -3.29   0.0010 **
## Year2012          -0.1107      0.0246   -4.50   7.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.00561,    Adjusted R-squared:  0.00384
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 775 weights are ~= 1. The remaining 8770 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0004 0.8660 0.9500 0.8950 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.05e-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 1.017 1          1.008
## Year              1.017 16          1.001

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.180 -0.262 0.019 0.257 1.753
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1803 0.0192 61.42 < 2e-16 ***
## LastAuthorFemale1 -0.0223 0.0101 -2.21 0.0270 *
## Year1997 -0.0218 0.0272 -0.80 0.4232
## Year1998 -0.0516 0.0264 -1.95 0.0508 .
## Year1999 -0.0499 0.0259 -1.93 0.0537 .
## Year2000 -0.0319 0.0272 -1.18 0.2399
## Year2001 -0.0342 0.0311 -1.10 0.2714
## Year2002 -0.0696 0.0257 -2.71 0.0068 **
## Year2003 -0.0763 0.0255 -2.99 0.0028 **
## Year2004 -0.1229 0.0278 -4.43 9.6e-06 ***
## Year2005 -0.0651 0.0258 -2.53 0.0116 *
## Year2006 -0.0643 0.0253 -2.54 0.0110 *
```

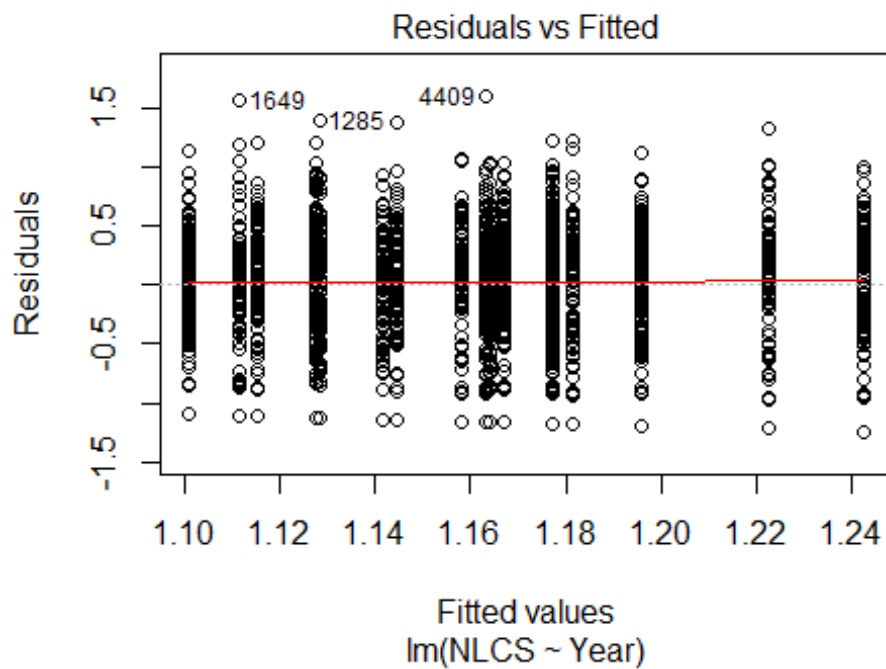
```

## Year2007          -0.0753      0.0253   -2.98   0.0029 **
## Year2008          -0.0654      0.0243   -2.69   0.0071 **
## Year2009          -0.0555      0.0242   -2.30   0.0217 *
## Year2010          -0.0740      0.0246   -3.00   0.0027 **
## Year2011          -0.0732      0.0246   -2.97   0.0030 **
## Year2012          -0.1041      0.0247   -4.22   2.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.00536,    Adjusted R-squared:  0.00359
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 771 weights are ~= 1. The remaining 8774 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.001  0.867  0.949  0.895  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.05e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 9545"
## [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
## 352 341 283 272 271 302 321 311 265 261 363 400 369 409 397
## 2011 2012
## 399 468
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 113 108 89 96 80 88 114 110 123 106 157 176 172 163 156
## 2011 2012

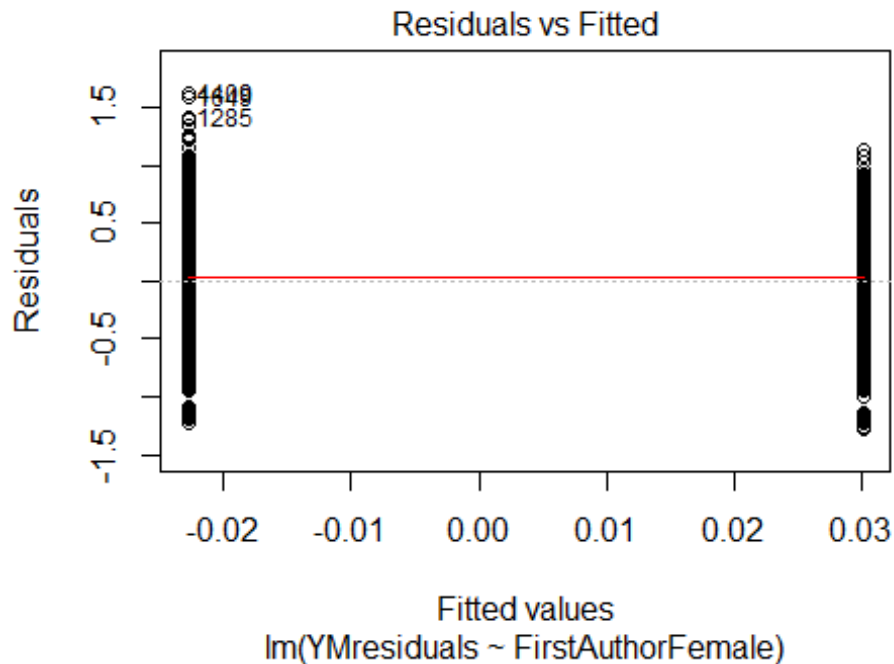
```



```
## 178 214
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 96 89 72 80 70 70 101 87 98 89 132 146 146 135 131
## 2011 2012
## 149 180
## [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
```

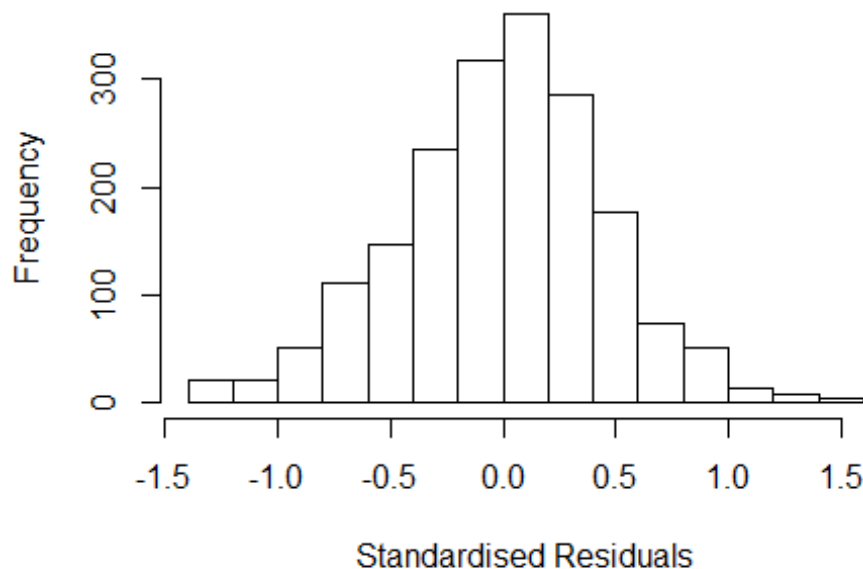


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 12, df = 1, p-value = 6e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 3.88044539228032"
## [1] "Male first author team size 2018 geometric mean: 3.61193082387329"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4900, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.4354737937472"
## [1] "Male last author team size 2018 geometric mean: 4.05756585031591"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4000, 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 1.033 1      1.017
## LastAuthorFemale  1.108 1      1.053
## UniqueAuthors    1.200 4      1.023
## Year             1.278 16      1.008
```

Residuals from first and last author and team size



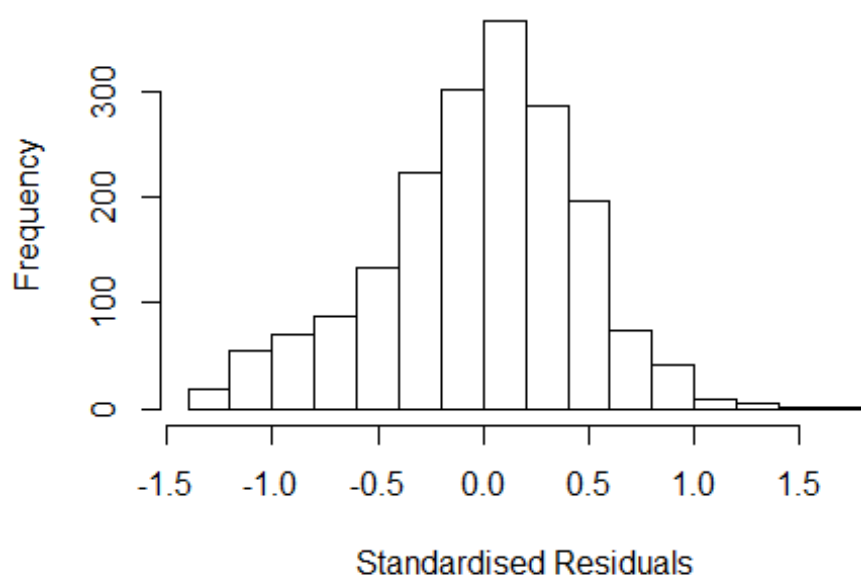
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3377 -0.2919 0.0157 0.2963 1.5883
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.87700 0.07508 11.68 < 2e-16 ***
## FirstAuthorFemale1 0.03986 0.02126 1.87 0.06096 .
## LastAuthorFemale1 -0.00261 0.02485 -0.11 0.91627
## UniqueAuthors2 0.44698 0.05853 7.64 3.5e-14 ***
## UniqueAuthors3 0.45570 0.05579 8.17 5.8e-16 ***
## UniqueAuthors4 0.46074 0.05751 8.01 2.0e-15 ***
## UniqueAuthors5 0.46739 0.05575 8.38 < 2e-16 ***
## Year1997 -0.09640 0.08281 -1.16 0.24453
## Year1998 -0.05962 0.08599 -0.69 0.48813
## Year1999 -0.19052 0.07927 -2.40 0.01633 *
```

```

## Year2000      -0.19890    0.07946   -2.50  0.01240 *
## Year2001      -0.12264    0.08004   -1.53  0.12566
## Year2002      -0.13322    0.07447   -1.79  0.07381 .
## Year2003      -0.18533    0.07455   -2.49  0.01301 *
## Year2004      -0.18086    0.06878   -2.63  0.00862 **
## Year2005      -0.20523    0.06614   -3.10  0.00194 **
## Year2006      -0.22470    0.06399   -3.51  0.00046 ***
## Year2007      -0.14931    0.06591   -2.27  0.02360 *
## Year2008      -0.09605    0.06532   -1.47  0.14160
## Year2009      -0.13542    0.06590   -2.05  0.04003 *
## Year2010      -0.12856    0.06586   -1.95  0.05108 .
## Year2011      -0.12750    0.06523   -1.95  0.05076 .
## Year2012      -0.16625    0.06423   -2.59  0.00972 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.431
## Multiple R-squared:  0.0971, Adjusted R-squared:  0.0863
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 169 weights are ~= 1. The remaining 1702 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.145  0.864  0.948  0.895  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.011 1 1.005
## LastAuthorFemale 1.092 1 1.045
## Year 1.102 16 1.003

```

Residuals from first and last author



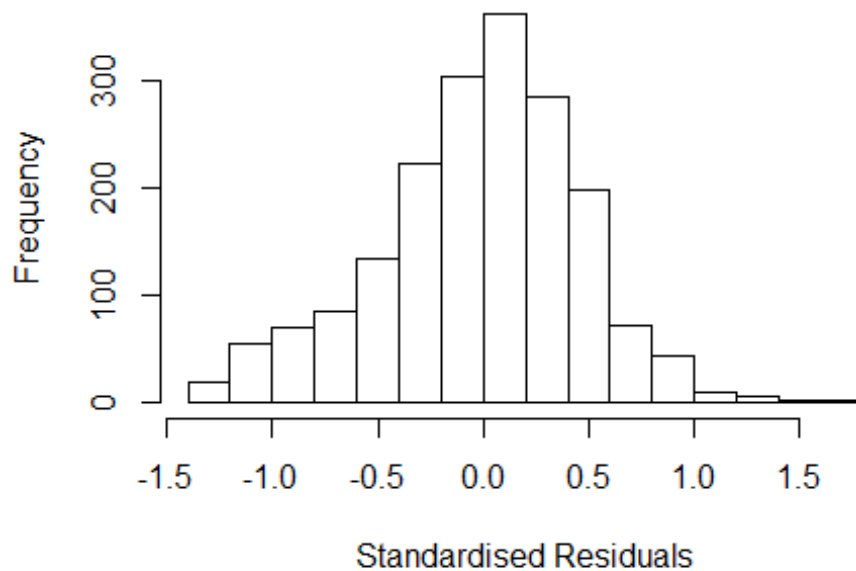
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3105 -0.2962  0.0273  0.3091  1.6051
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2674     0.0548   23.12  <2e-16 ***
## FirstAuthorFemale1  0.0598     0.0218    2.74  0.0062 **
## LastAuthorFemale1 -0.0175     0.0257   -0.68  0.4943
## Year1997          -0.0845     0.0831   -1.02  0.3099
## Year1998          -0.0167     0.0875   -0.19  0.8489
## Year1999          -0.1851     0.0785   -2.36  0.0185 *
## Year2000          -0.1927     0.0851   -2.26  0.0237 *
## Year2001          -0.0727     0.0803   -0.90  0.3656
## Year2002          -0.1125     0.0761   -1.48  0.1392
## Year2003          -0.1437     0.0782   -1.84  0.0661 .
## Year2004          -0.1659     0.0699   -2.37  0.0177 *
## Year2005          -0.1617     0.0670   -2.41  0.0160 *
```

```

## Year2006          -0.2077      0.0642   -3.23   0.0012 **
## Year2007          -0.1095      0.0672   -1.63   0.1035
## Year2008          -0.0807      0.0662   -1.22   0.2231
## Year2009          -0.0907      0.0663   -1.37   0.1714
## Year2010          -0.1092      0.0665   -1.64   0.1006
## Year2011          -0.0879      0.0653   -1.35   0.1787
## Year2012          -0.1308      0.0646   -2.03   0.0429 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0163, Adjusted R-squared:  0.00671
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1714 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.151  0.861  0.949  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      5.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.01 1      1.005
## Year              1.01 16      1.000

```

Residuals from first author



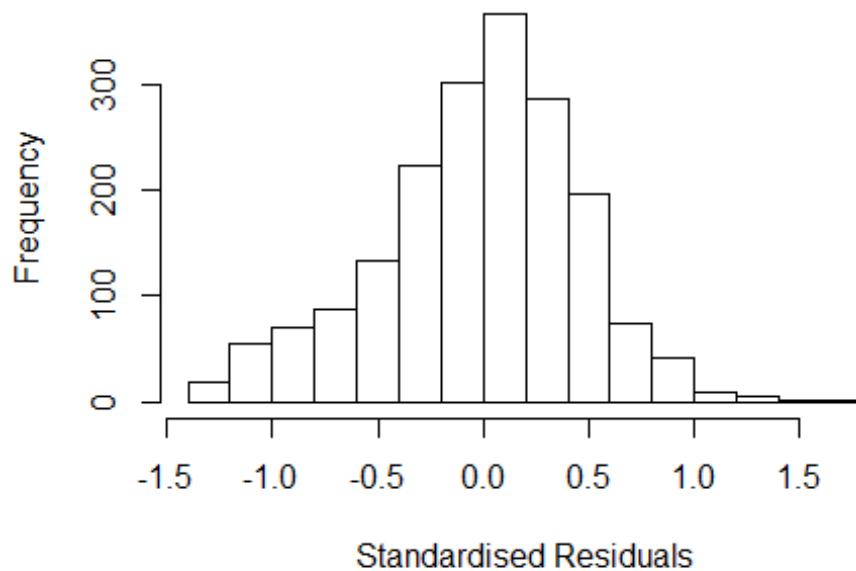
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3230 -0.2978 0.0249 0.3082 1.6096
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2655 0.0546 23.18 <2e-16 ***
## FirstAuthorFemale1 0.0575 0.0220 2.61 0.0090 **
## Year1997 -0.0852 0.0829 -1.03 0.3043
## Year1998 -0.0165 0.0874 -0.19 0.8507
## Year1999 -0.1848 0.0783 -2.36 0.0184 *
## Year2000 -0.1939 0.0852 -2.28 0.0229 *
## Year2001 -0.0754 0.0799 -0.94 0.3454
## Year2002 -0.1151 0.0758 -1.52 0.1290
## Year2003 -0.1437 0.0781 -1.84 0.0657 .
## Year2004 -0.1672 0.0698 -2.40 0.0166 *
## Year2005 -0.1640 0.0667 -2.46 0.0140 *
## Year2006 -0.2080 0.0641 -3.24 0.0012 **
```

```

## Year2007          -0.1121      0.0669   -1.68   0.0938 .
## Year2008          -0.0833      0.0660   -1.26   0.2071
## Year2009          -0.0934      0.0660   -1.42   0.1572
## Year2010          -0.1119      0.0662   -1.69   0.0913 .
## Year2011          -0.0916      0.0647   -1.42   0.1570
## Year2012          -0.1332      0.0644   -2.07   0.0388 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.016, Adjusted R-squared:  0.00701
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 158 weights are ~= 1. The remaining 1713 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.148  0.862  0.949  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      5.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.086 1          1.042
## Year            1.086 16          1.003

```


Residuals from last author



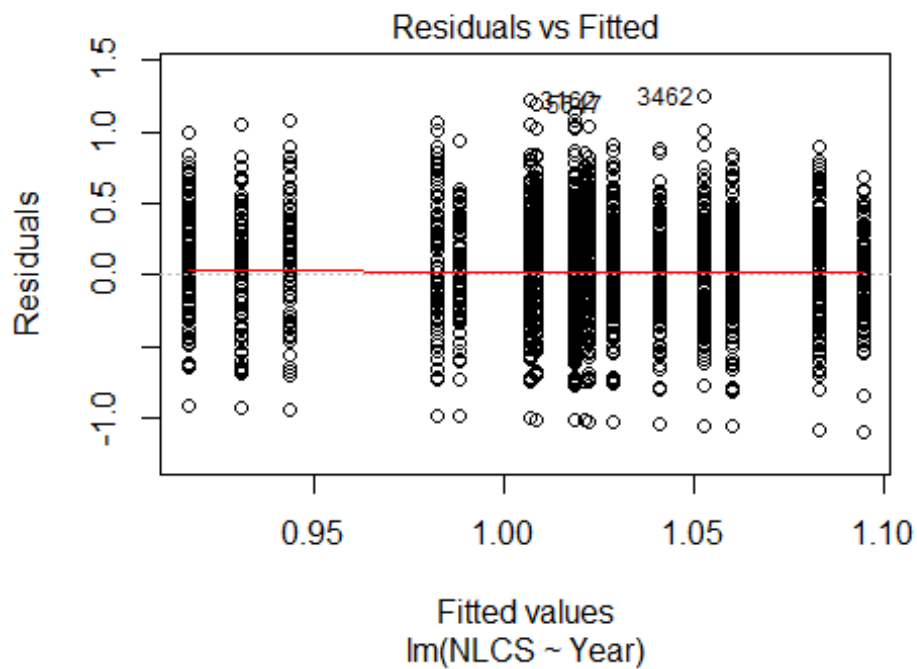
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2889 -0.2972 0.0276 0.3084 1.5834
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2889 0.0539 23.92 <2e-16 ***
## LastAuthorFemale1 -0.0077 0.0258 -0.30 0.7656
## Year1997 -0.0907 0.0830 -1.09 0.2743
## Year1998 -0.0155 0.0876 -0.18 0.8595
## Year1999 -0.1820 0.0786 -2.32 0.0207 *
## Year2000 -0.1911 0.0848 -2.25 0.0244 *
## Year2001 -0.0748 0.0802 -0.93 0.3509
## Year2002 -0.1119 0.0760 -1.47 0.1411
## Year2003 -0.1437 0.0785 -1.83 0.0674 .
## Year2004 -0.1653 0.0701 -2.36 0.0184 *
## Year2005 -0.1601 0.0667 -2.40 0.0164 *
## Year2006 -0.2043 0.0644 -3.17 0.0015 **
```

```

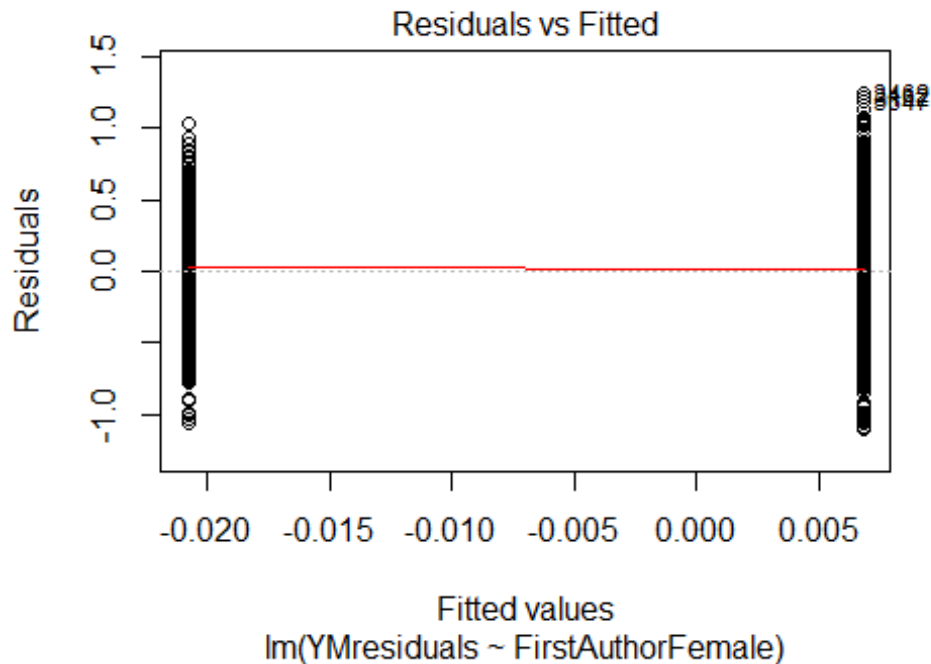
## Year2007          -0.1093      0.0671   -1.63    0.1033
## Year2008          -0.0748      0.0663   -1.13    0.2590
## Year2009          -0.0872      0.0664   -1.31    0.1894
## Year2010          -0.1027      0.0665   -1.54    0.1231
## Year2011          -0.0803      0.0654   -1.23    0.2203
## Year2012          -0.1253      0.0644   -1.95    0.0519 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0124, Adjusted R-squared:  0.00331
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 150 weights are ~= 1. The remaining 1721 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.163  0.861  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      5.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1871"
## [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
##  240  246  247  205  269  255  239  251  228  268  325  321  381  351  334
## 2011 2012
##  317  327
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  101   89  115   97   91   72  149  149  135  162  193  202  238  219  218
## 2011 2012

```

```
## 205 221
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 89 80 103 86 81 61 125 128 117 133 164 170 194 183 187
## 2011 2012
## 171 201
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 48, df = 16, p-value = 4e-05
```

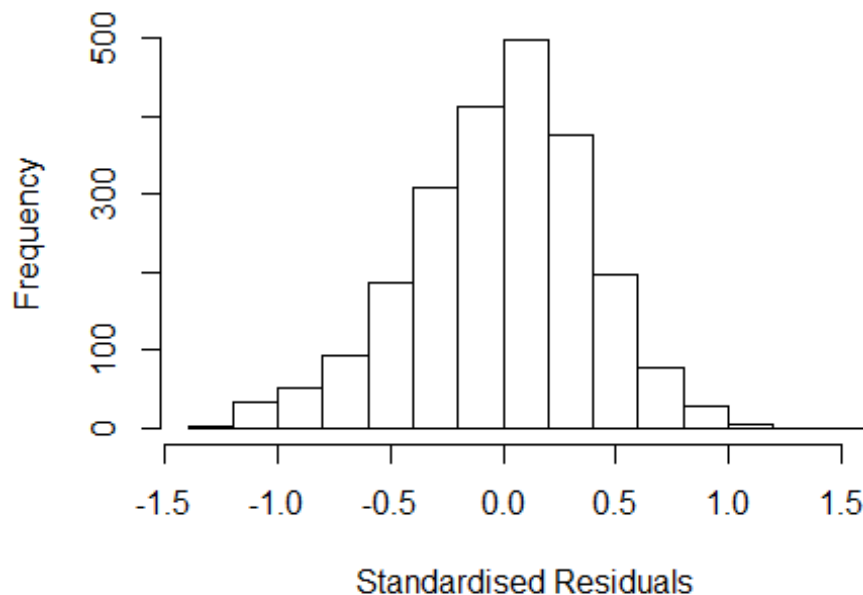


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.03, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 3.5756852389704"
## [1] "Male first author team size 2018 geometric mean: 3.00104741460463"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3400, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.58489407072957"
## [1] "Male last author team size 2018 geometric mean: 3.05111501161634"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2800, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.095 1 1.047
## LastAuthorFemale 1.047 1 1.023
## UniqueAuthors 1.316 4 1.035
## Year 1.351 16 1.009
```

Residuals from first and last author and team size



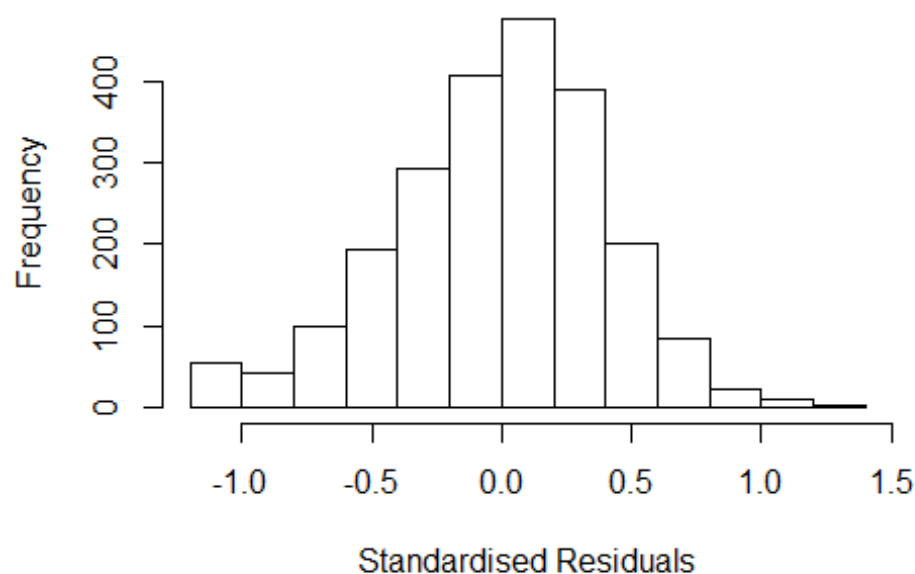
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2314 -0.2662  0.0181  0.2555  1.4362
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.853437   0.054822   15.57 < 2e-16 ***
## FirstAuthorFemale1 -0.035559   0.020052   -1.77  0.076 .
## LastAuthorFemale1 -0.006539   0.023919   -0.27  0.785
## UniqueAuthors2     0.181500   0.030815    5.89 4.4e-09 ***
## UniqueAuthors3     0.225546   0.031722    7.11 1.6e-12 ***
## UniqueAuthors4     0.263949   0.034183    7.72 1.7e-14 ***
## UniqueAuthors5     0.339733   0.036797    9.23 < 2e-16 ***
## Year1997          -0.073186   0.076224   -0.96  0.337
## Year1998          -0.032602   0.069338   -0.47  0.638
## Year1999           0.002598   0.070347    0.04  0.971
```

```

## Year2000      0.045813    0.062422    0.73    0.463
## Year2001      0.035571    0.075304    0.47    0.637
## Year2002      0.044790    0.061562    0.73    0.467
## Year2003      0.016359    0.059794    0.27    0.784
## Year2004     -0.038191    0.062903   -0.61    0.544
## Year2005      0.085540    0.064505    1.33    0.185
## Year2006     -0.040019    0.063691   -0.63    0.530
## Year2007      0.012396    0.059895    0.21    0.836
## Year2008     -0.036830    0.059049   -0.62    0.533
## Year2009     -0.000903    0.059223   -0.02    0.988
## Year2010     -0.131448    0.062977   -2.09    0.037 *
## Year2011     -0.047472    0.063141   -0.75    0.452
## Year2012     -0.038990    0.062346   -0.63    0.532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0669, Adjusted R-squared:  0.0578
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 214 weights are ~= 1. The remaining 2059 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.124  0.867  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      4.40e-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 1.065 1      1.032
## LastAuthorFemale  1.034 1      1.017
## Year              1.064 16      1.002

```

Residuals from first and last author



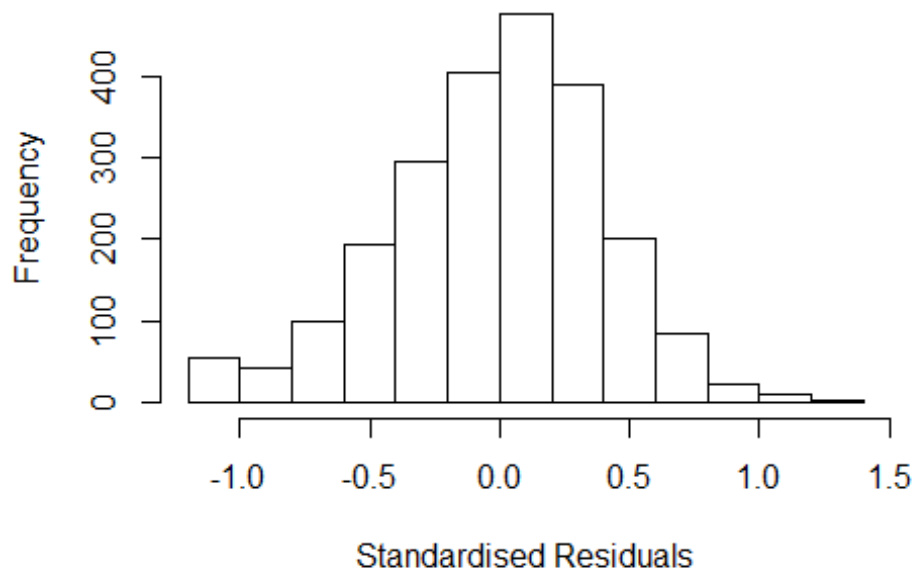
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1318 -0.2667  0.0202  0.2583  1.2228
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.99029    0.05408   18.31  <2e-16 ***
## FirstAuthorFemale1 -0.02131    0.02040   -1.04    0.296
## LastAuthorFemale1 -0.00241    0.02448   -0.10    0.921
## Year1997        -0.05277    0.07832   -0.67    0.501
## Year1998        -0.02964    0.07185   -0.41    0.680
## Year1999         0.01954    0.07343    0.27    0.790
## Year2000         0.07826    0.06484    1.21    0.228
## Year2001         0.06123    0.07688    0.80    0.426
## Year2002         0.08260    0.06350    1.30    0.193
## Year2003         0.06769    0.06138    1.10    0.270
## Year2004         0.02646    0.06433    0.41    0.681
## Year2005         0.14151    0.06618    2.14    0.033 *
```

```

## Year2006          0.02379    0.06365    0.37    0.709
## Year2007          0.08889    0.06052    1.47    0.142
## Year2008          0.03639    0.06088    0.60    0.550
## Year2009          0.06110    0.06056    1.01    0.313
## Year2010         -0.05460    0.06523   -0.84    0.403
## Year2011          0.02834    0.06455    0.44    0.661
## Year2012          0.03811    0.06375    0.60    0.550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.0149, Adjusted R-squared:  0.00704
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 188 weights are ~= 1. The remaining 2085 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.301  0.867  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      4.40e-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 1.048 1         1.024
## Year              1.048 16         1.001

```


Residuals from first author



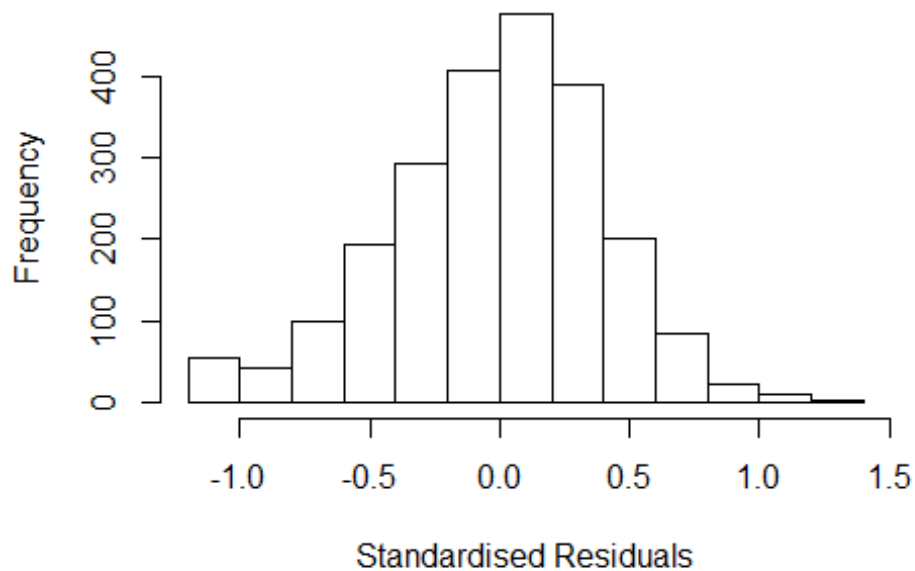
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1316 -0.2668 0.0198 0.2578 1.2231
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9901 0.0540 18.32 <2e-16 ***
## FirstAuthorFemale1 -0.0217 0.0203 -1.07 0.285
## Year1997 -0.0528 0.0784 -0.67 0.501
## Year1998 -0.0297 0.0719 -0.41 0.680
## Year1999 0.0195 0.0735 0.26 0.791
## Year2000 0.0781 0.0649 1.20 0.229
## Year2001 0.0611 0.0769 0.80 0.427
## Year2002 0.0826 0.0635 1.30 0.194
## Year2003 0.0676 0.0614 1.10 0.271
## Year2004 0.0263 0.0644 0.41 0.683
## Year2005 0.1415 0.0662 2.14 0.033 *
## Year2006 0.0236 0.0636 0.37 0.711
```

```

## Year2007          0.0888      0.0605      1.47      0.142
## Year2008          0.0363      0.0609      0.60      0.551
## Year2009          0.0610      0.0606      1.01      0.314
## Year2010         -0.0547      0.0652     -0.84      0.402
## Year2011          0.0283      0.0646      0.44      0.661
## Year2012          0.0379      0.0638      0.59      0.552
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.0149, Adjusted R-squared:  0.00747
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 2088 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.300  0.867  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      4.40e-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 1.018 1      1.009
## Year              1.018 16      1.001

```

Residuals from last author



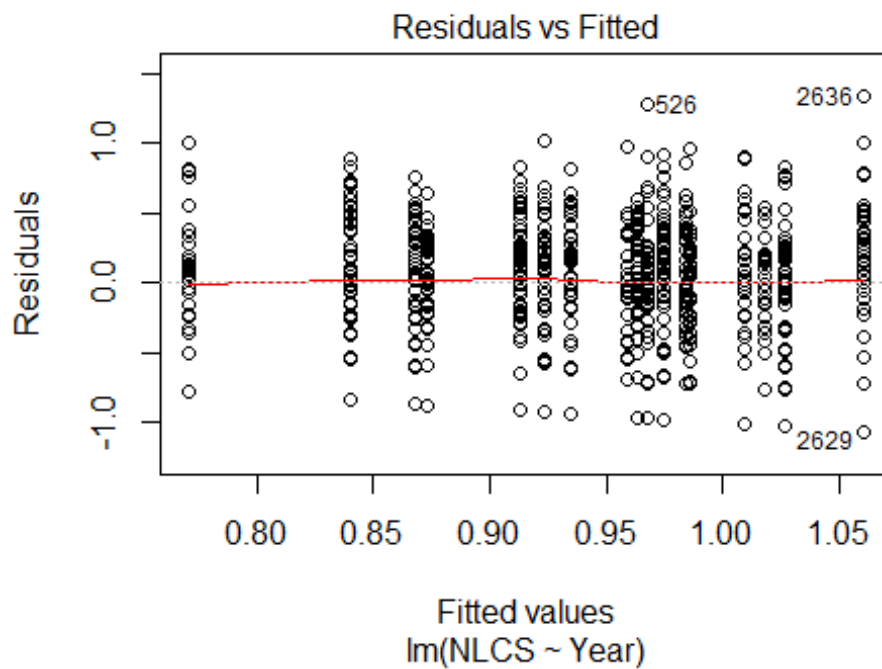
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1261 -0.2677 0.0228 0.2598 1.2281
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9899 0.0541 18.31 <2e-16 ***
## LastAuthorFemale1 -0.0069 0.0244 -0.28 0.777
## Year1997 -0.0538 0.0783 -0.69 0.492
## Year1998 -0.0315 0.0718 -0.44 0.661
## Year1999 0.0162 0.0732 0.22 0.825
## Year2000 0.0763 0.0647 1.18 0.238
## Year2001 0.0575 0.0771 0.75 0.455
## Year2002 0.0785 0.0633 1.24 0.215
## Year2003 0.0635 0.0612 1.04 0.299
## Year2004 0.0223 0.0642 0.35 0.728
## Year2005 0.1362 0.0658 2.07 0.039 *
## Year2006 0.0180 0.0632 0.29 0.776
```

```

## Year2007          0.0849      0.0604      1.41      0.160
## Year2008          0.0319      0.0606      0.53      0.599
## Year2009          0.0568      0.0604      0.94      0.347
## Year2010         -0.0595      0.0651     -0.91      0.361
## Year2011          0.0216      0.0640      0.34      0.736
## Year2012          0.0327      0.0635      0.52      0.606
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.0144, Adjusted R-squared:  0.00696
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 196 weights are ~= 1. The remaining 2077 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.298  0.866  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      4.40e-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: 2273"
## [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
##  192  149  160  131  167  189  163  143  167  143  143  144  121  165  126
## 2011 2012
##  121  134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   52   44   49   33   31   30   43   25   64   49   46   53   60   71   48
## 2011 2012

```

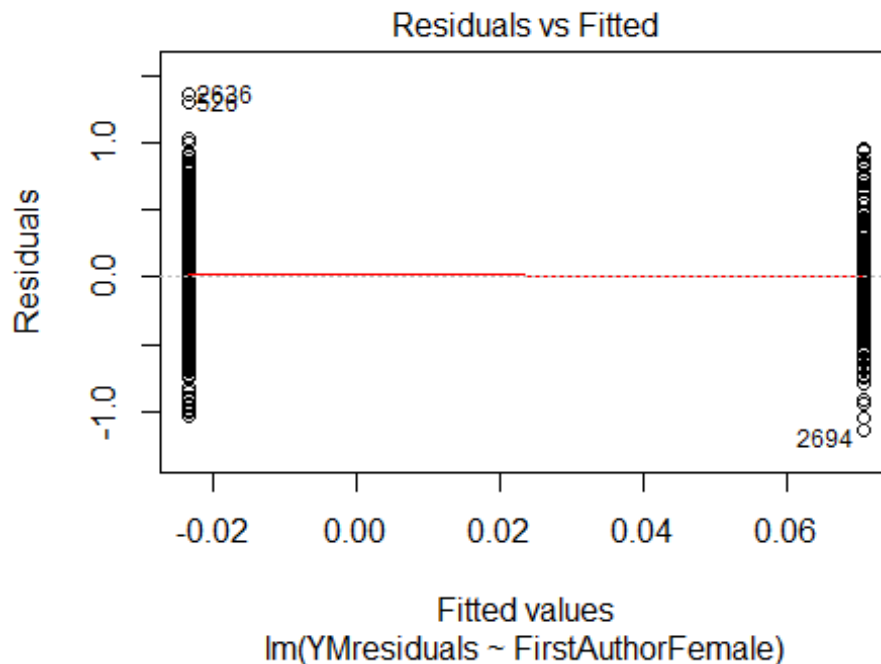
```
## 55 62
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 42 34 41 26 28 23 31 21 50 38 42 43 52 56 41
## 2011 2012
## 47 54
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 24, df = 16, p-value = 0.1
```



```
##
## 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: 3.66284150148471"
## [1] "Male first author team size 2018 geometric mean: 3.19043679542553"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

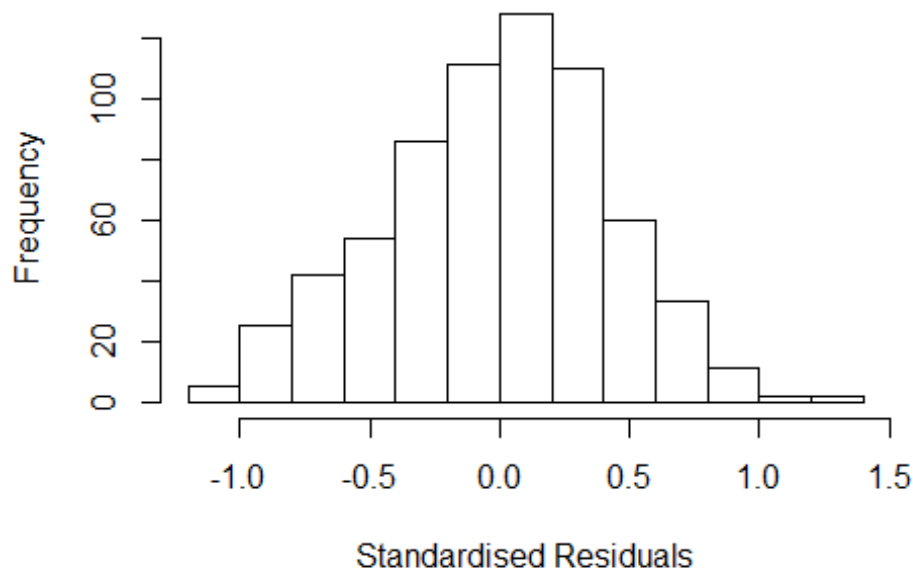
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 51, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.3658654363386"
## [1] "Male last author team size 2018 geometric mean: 3.23617568565357"

## Warning in wilcox.test.default(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 = 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  1.185  1      1.089
## UniqueAuthors    1.996  4      1.090
## Year              2.338 16      1.027
```

Residuals from first and last author and team size



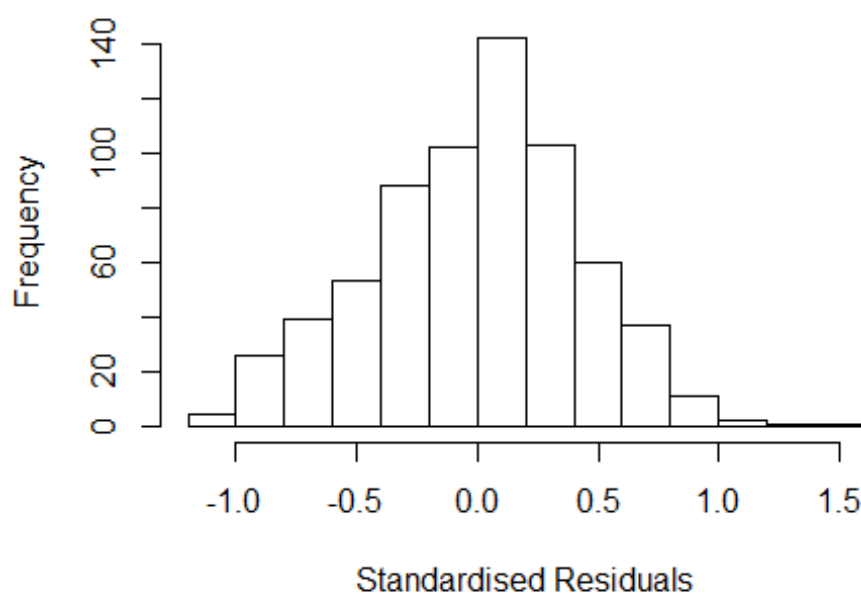
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0625 -0.3120  0.0189  0.2875  1.3972
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.033852   0.091693   11.28  <2e-16 ***
## FirstAuthorFemale1  0.095599   0.041332    2.31  0.0210 *
## LastAuthorFemale1  0.029551   0.042442    0.70  0.4865
## UniqueAuthors2   -0.030867   0.077920   -0.40  0.6921
## UniqueAuthors3    0.033551   0.079843    0.42  0.6745
## UniqueAuthors4    0.017572   0.084423    0.21  0.8352
## UniqueAuthors5   -0.054893   0.082466   -0.67  0.5059
## Year1997         -0.105339   0.098210   -1.07  0.2839
## Year1998         -0.129417   0.096441   -1.34  0.1801
## Year1999         -0.015693   0.096193   -0.16  0.8705
```

```

## Year2000      -0.098665    0.098965    -1.00    0.3192
## Year2001      0.000182    0.146941     0.00    0.9990
## Year2002     -0.243872    0.114139    -2.14    0.0330 *
## Year2003      0.035926    0.092795     0.39    0.6988
## Year2004     -0.125169    0.091203    -1.37    0.1704
## Year2005     -0.176576    0.109704    -1.61    0.1080
## Year2006     -0.213232    0.080718    -2.64    0.0084 **
## Year2007     -0.057194    0.088928    -0.64    0.5204
## Year2008     -0.247443    0.100093    -2.47    0.0137 *
## Year2009     -0.136770    0.088576    -1.54    0.1231
## Year2010     -0.169457    0.099762    -1.70    0.0899 .
## Year2011     -0.036062    0.107744    -0.33    0.7380
## Year2012     -0.149871    0.090799    -1.65    0.0993 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.044, Adjusted R-squared:  0.0114
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 59 weights are ~= 1. The remaining 610 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.283  0.866  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      1.49e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.129 1      1.062
## LastAuthorFemale  1.123 1      1.060
## Year              1.267 16      1.007

```


Residuals from first and last author



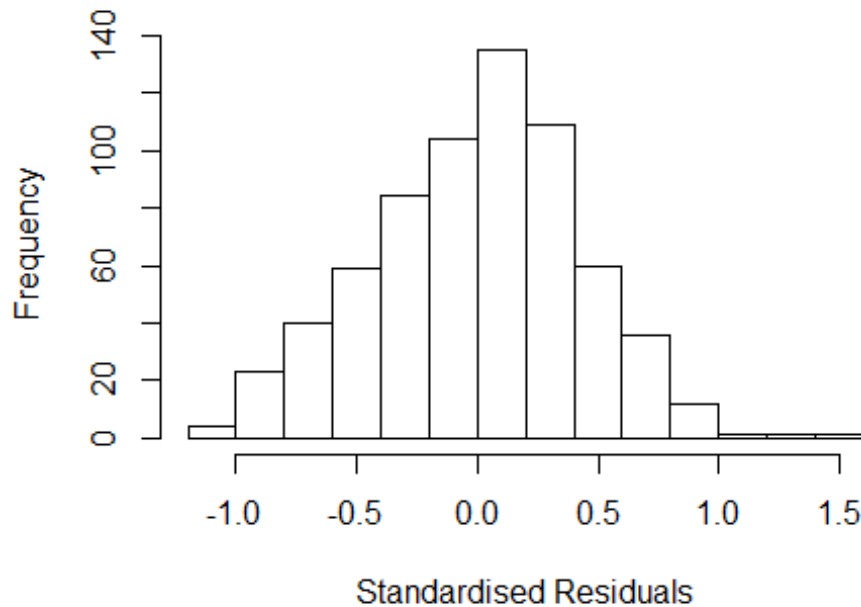
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0854 -0.3150 0.0281 0.2776 1.4119
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02774 0.06790 15.14 <2e-16 ***
## FirstAuthorFemale1 0.10224 0.04067 2.51 0.012 *
## LastAuthorFemale1 0.03207 0.04180 0.77 0.443
## Year1997 -0.10208 0.09762 -1.05 0.296
## Year1998 -0.11851 0.09408 -1.26 0.208
## Year1999 -0.02516 0.09630 -0.26 0.794
## Year2000 -0.09649 0.09956 -0.97 0.333
## Year2001 -0.00892 0.14126 -0.06 0.950
## Year2002 -0.25383 0.11399 -2.23 0.026 *
## Year2003 0.02912 0.09213 0.32 0.752
## Year2004 -0.12762 0.09064 -1.41 0.160
## Year2005 -0.17177 0.10826 -1.59 0.113
```

```

## Year2006      -0.21254    0.08300   -2.56    0.011 *
## Year2007      -0.06386    0.08742   -0.73    0.465
## Year2008      -0.24483    0.10389   -2.36    0.019 *
## Year2009      -0.14105    0.09032   -1.56    0.119
## Year2010      -0.16836    0.10120   -1.66    0.097 .
## Year2011      -0.04461    0.10817   -0.41    0.680
## Year2012      -0.15601    0.09093   -1.72    0.087 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.435
## Multiple R-squared:  0.0382, Adjusted R-squared:  0.0115
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 605 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.270  0.867   0.947   0.906   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.49e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.126 1      1.061
## Year      1.126 16      1.004

```

Residuals from first author



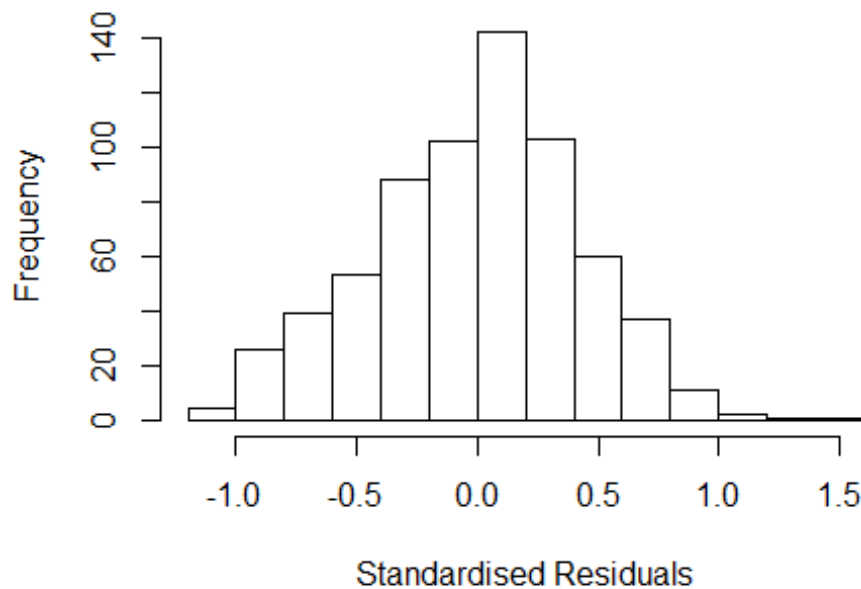
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0963 -0.3187  0.0273  0.2840  1.4023
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03460    0.06795   15.23  <2e-16 ***
## FirstAuthorFemale1 0.10369    0.04062    2.55   0.011 *
## Year1997       -0.10671    0.09734   -1.10   0.273
## Year1998       -0.12007    0.09416   -1.28   0.203
## Year1999       -0.01751    0.09543   -0.18   0.854
## Year2000       -0.09635    0.09974   -0.97   0.334
## Year2001       -0.00681    0.14097   -0.05   0.961
## Year2002       -0.25126    0.11406   -2.20   0.028 *
## Year2003        0.02532    0.09275    0.27   0.785
## Year2004       -0.12558    0.09107   -1.38   0.168
## Year2005       -0.17348    0.10852   -1.60   0.110
## Year2006       -0.20895    0.08313   -2.51   0.012 *
```

```

## Year2007          -0.06141    0.08778   -0.70    0.484
## Year2008          -0.24387    0.10412   -2.34    0.019 *
## Year2009          -0.14075    0.09058   -1.55    0.121
## Year2010          -0.16939    0.10112   -1.68    0.094 .
## Year2011          -0.04194    0.10846   -0.39    0.699
## Year2012          -0.15789    0.09118   -1.73    0.084 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.0372, Adjusted R-squared:  0.0121
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 601 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.279  0.866  0.945  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      1.49e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



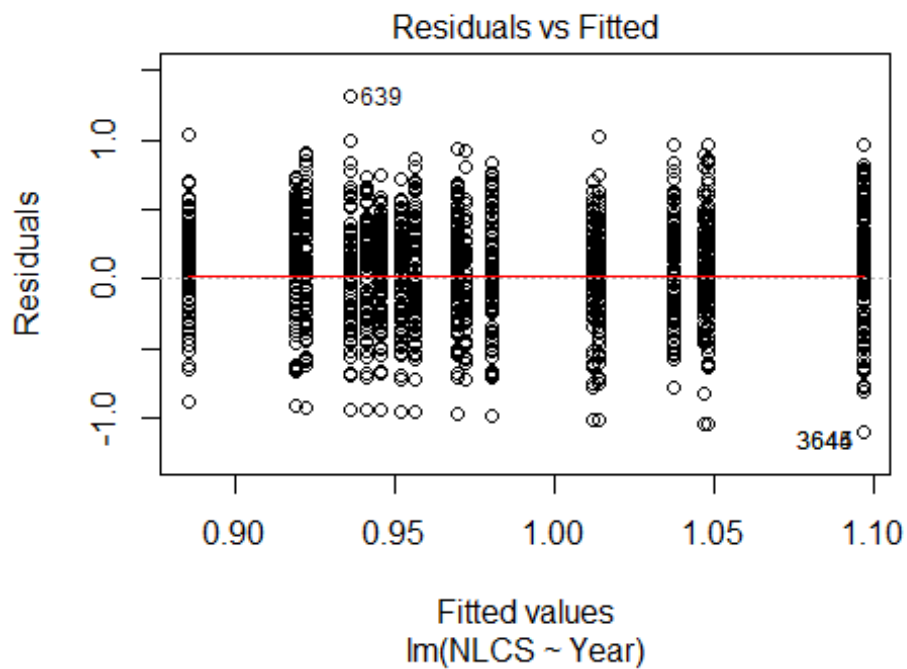
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0534 -0.3141 0.0267 0.2892 1.3812
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0534 0.0666 15.81 <2e-16 ***
## LastAuthorFemale1 0.0370 0.0417 0.89 0.375
## Year1997 -0.1197 0.0945 -1.27 0.206
## Year1998 -0.1238 0.0941 -1.32 0.188
## Year1999 -0.0484 0.0962 -0.50 0.615
## Year2000 -0.0976 0.1024 -0.95 0.341
## Year2001 -0.0191 0.1381 -0.14 0.890
## Year2002 -0.2545 0.1181 -2.15 0.032 *
## Year2003 0.0237 0.0931 0.25 0.799
## Year2004 -0.1335 0.0910 -1.47 0.143
## Year2005 -0.1727 0.1062 -1.63 0.105
## Year2006 -0.2016 0.0832 -2.42 0.016 *
```

```

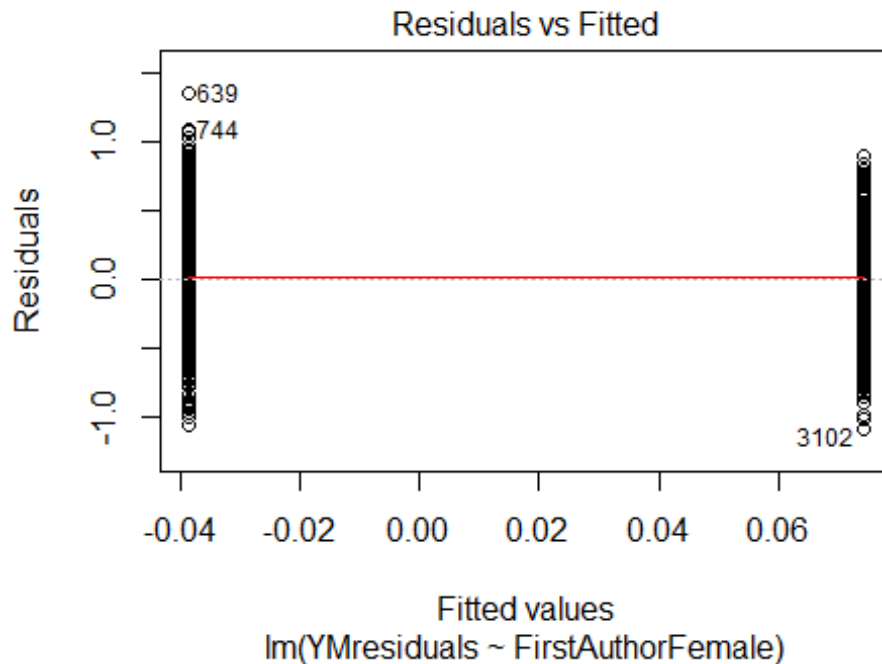
## Year2007          -0.0683      0.0859   -0.79    0.427
## Year2008          -0.2415      0.1062   -2.27    0.023 *
## Year2009          -0.1321      0.0916   -1.44    0.150
## Year2010          -0.1617      0.1009   -1.60    0.109
## Year2011          -0.0396      0.1071   -0.37    0.711
## Year2012          -0.1473      0.0908   -1.62    0.105
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0287, Adjusted R-squared:  0.00332
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 602 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.291  0.869  0.946  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      1.49e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 669"
## [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
##  208  214  179  176  201  183  182  180  170  191  193  190  190  182  198
## 2011 2012
##  186  213
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  106  104   92   97   76   56  108  103   87  105  104  115  125  115  112
## 2011 2012

```

```
## 130 127
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 101 88 83 85 66 51 95 87 79 92 89 101 111 102 99
## 2011 2012
## 113 110
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```

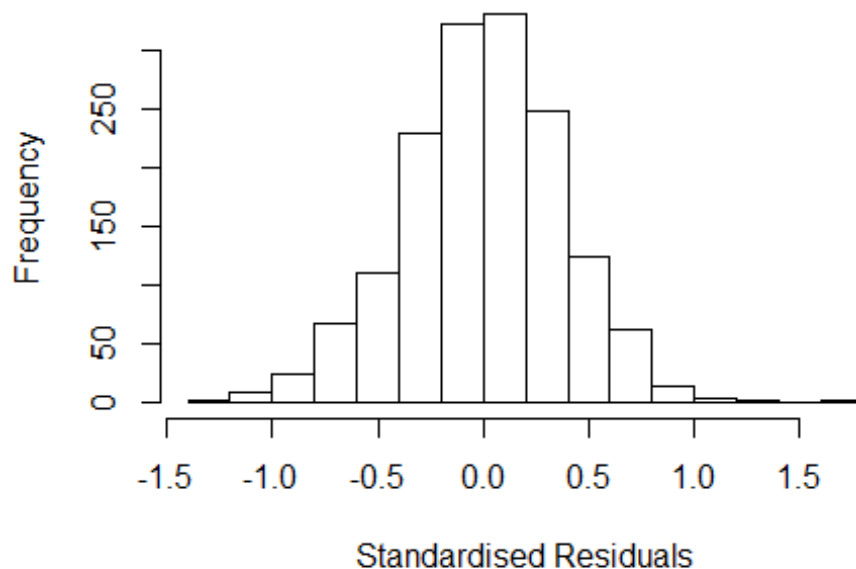


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 9e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 2.86253658700412"
## [1] "Male first author team size 2018 geometric mean: 2.76581403840538"
##
## 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: 2.62010969924662"
## [1] "Male last author team size 2018 geometric mean: 2.89394187333789"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.152 1      1.073
## LastAuthorFemale  1.075 1      1.037
## UniqueAuthors     1.257 4      1.029
## Year              1.410 16     1.011
```


Residuals from first and last author and team size



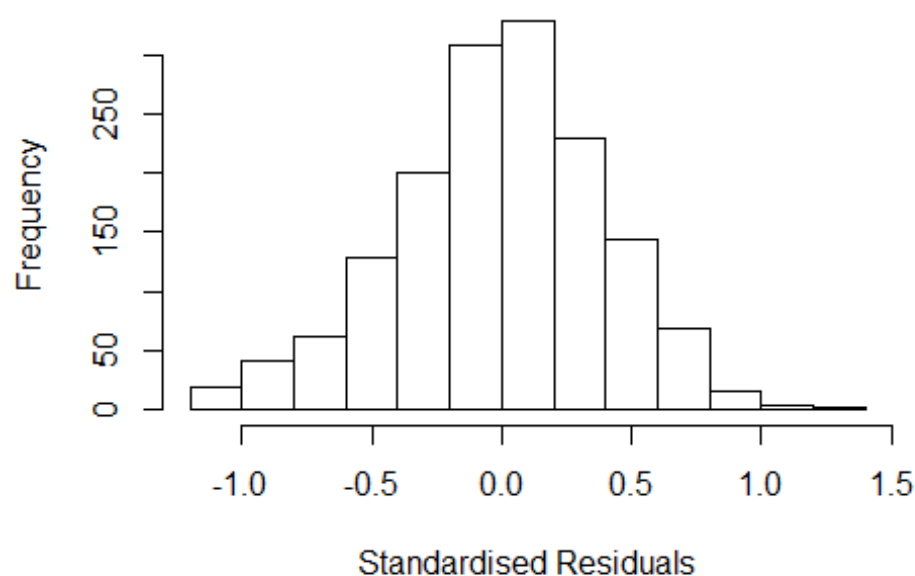
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29590 -0.23769  0.00475  0.24333  1.60655
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.77084    0.04929   15.64 < 2e-16 ***
## FirstAuthorFemale1 0.09602    0.02046    4.69 2.9e-06 ***
## LastAuthorFemale1 -0.01345    0.02355   -0.57  0.5679
## UniqueAuthors2    0.29093    0.03528    8.25 3.5e-16 ***
## UniqueAuthors3    0.28876    0.03682    7.84 8.2e-15 ***
## UniqueAuthors4    0.37628    0.04080    9.22 < 2e-16 ***
## UniqueAuthors5    0.49016    0.04251   11.53 < 2e-16 ***
## Year1997         -0.04783    0.05712   -0.84  0.4025
## Year1998         -0.12939    0.05319   -2.43  0.0151 *
## Year1999         -0.16936    0.05324   -3.18  0.0015 **
```

```

## Year2000      -0.00398    0.05822   -0.07    0.9455
## Year2001      -0.06158    0.06471   -0.95    0.3414
## Year2002      -0.04534    0.05341   -0.85    0.3960
## Year2003      -0.12815    0.05109   -2.51    0.0122 *
## Year2004      -0.05288    0.05352   -0.99    0.3232
## Year2005      -0.11936    0.05280   -2.26    0.0239 *
## Year2006      -0.14085    0.05441   -2.59    0.0097 **
## Year2007      -0.13732    0.05332   -2.58    0.0101 *
## Year2008      -0.11706    0.05250   -2.23    0.0259 *
## Year2009      -0.04767    0.05202   -0.92    0.3596
## Year2010      -0.11642    0.05600   -2.08    0.0378 *
## Year2011      -0.06150    0.05624   -1.09    0.2743
## Year2012      -0.18443    0.05868   -3.14    0.0017 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.363
## Multiple R-squared:  0.157, Adjusted R-squared:  0.145
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 146 weights are ~= 1. The remaining 1406 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0112 0.8680 0.9510 0.9000 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.44e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.130 1 1.063
## LastAuthorFemale 1.048 1 1.024
## Year 1.156 16 1.005

```

Residuals from first and last author



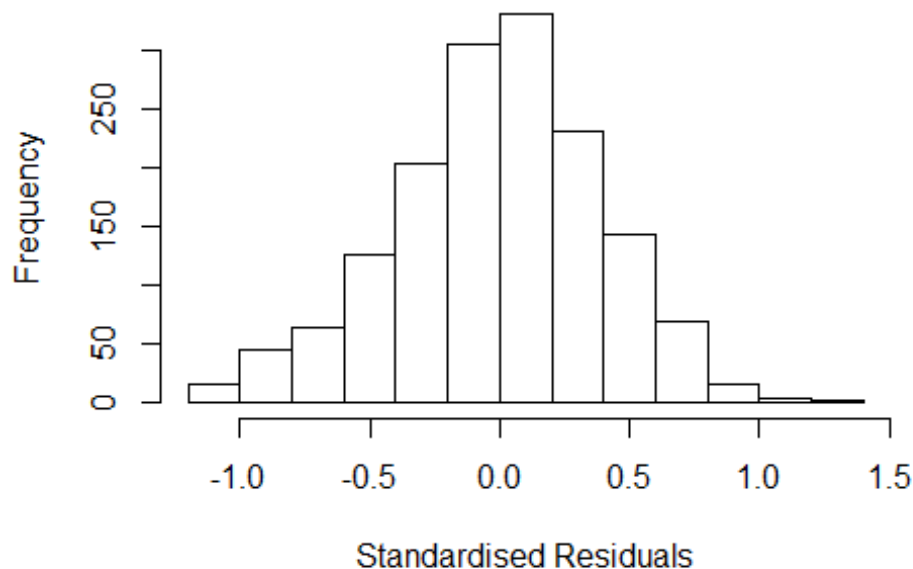
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1142 -0.2472  0.0114  0.2505  1.3582
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.04792    0.04235   24.75 < 2e-16 ***
## FirstAuthorFemale1 0.12223    0.02139    5.72 1.3e-08 ***
## LastAuthorFemale1 -0.01083    0.02488   -0.44 0.66352
## Year1997        -0.10322    0.06099   -1.69 0.09078 .
## Year1998        -0.15807    0.05946   -2.66 0.00793 **
## Year1999        -0.20904    0.05906   -3.54 0.00041 ***
## Year2000         0.01428    0.06098    0.23 0.81487
## Year2001        -0.09292    0.06698   -1.39 0.16556
## Year2002        -0.04402    0.05776   -0.76 0.44608
## Year2003        -0.12441    0.05611   -2.22 0.02674 *
## Year2004        -0.04686    0.05449   -0.86 0.38998
## Year2005        -0.13460    0.05810   -2.32 0.02064 *
```

```

## Year2006          -0.15016      0.05758      -2.61  0.00920 **
## Year2007          -0.11638      0.05915      -1.97  0.04931 *
## Year2008          -0.11544      0.05789      -1.99  0.04631 *
## Year2009          -0.04509      0.05521      -0.82  0.41423
## Year2010          -0.08036      0.05912      -1.36  0.17426
## Year2011           0.00241      0.05876       0.04  0.96728
## Year2012          -0.16820      0.06373      -2.64  0.00840 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0461, Adjusted R-squared:  0.0349
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 127 weights are ~= 1. The remaining 1425 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.178  0.863   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      6.44e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.113 1      1.055
## Year              1.113 16      1.003

```

Residuals from first author



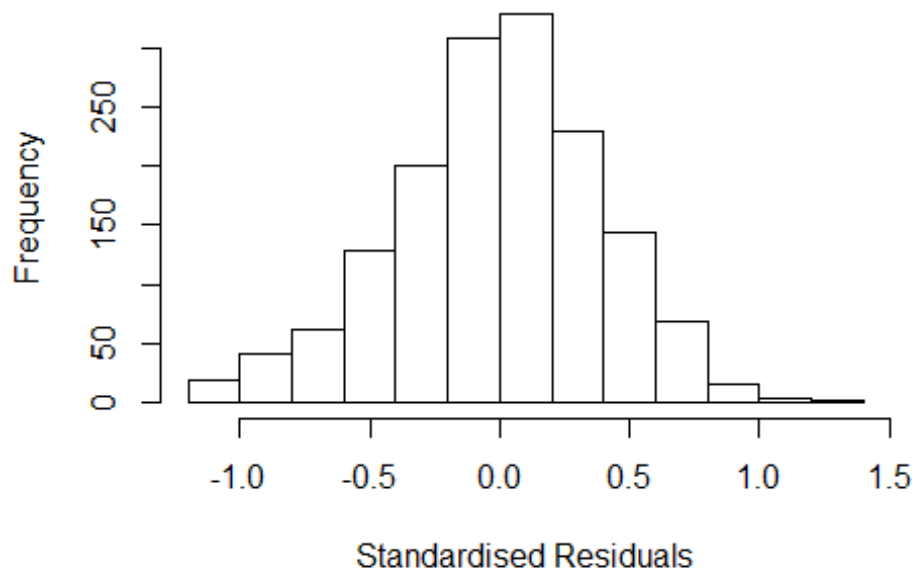
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1219 -0.2464 0.0107 0.2524 1.3598
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04628 0.04203 24.89 < 2e-16 ***
## FirstAuthorFemale1 0.12070 0.02129 5.67 1.7e-08 ***
## Year1997 -0.10308 0.06104 -1.69 0.09146 .
## Year1998 -0.15810 0.05945 -2.66 0.00791 **
## Year1999 -0.20925 0.05911 -3.54 0.00041 ***
## Year2000 0.01332 0.06087 0.22 0.82682
## Year2001 -0.09308 0.06683 -1.39 0.16387
## Year2002 -0.04409 0.05776 -0.76 0.44542
## Year2003 -0.12435 0.05613 -2.22 0.02689 *
## Year2004 -0.04680 0.05445 -0.86 0.39019
## Year2005 -0.13523 0.05808 -2.33 0.02001 *
## Year2006 -0.15062 0.05763 -2.61 0.00905 **
```

```

## Year2007          -0.11657    0.05912   -1.97  0.04881 *
## Year2008          -0.11515    0.05792   -1.99  0.04698 *
## Year2009          -0.04506    0.05520   -0.82  0.41443
## Year2010          -0.08123    0.05918   -1.37  0.17010
## Year2011           0.00321    0.05867    0.05  0.95636
## Year2012          -0.16881    0.06375   -2.65  0.00818 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0461, Adjusted R-squared:  0.0355
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 128 weights are ~= 1. The remaining 1424 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.176  0.863  0.952  0.901  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.44e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.035 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
## -1.1183 -0.2683 0.0126 0.2584 1.3367
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0825 0.0416 26.01 < 2e-16 ***
## LastAuthorFemale1 0.0119 0.0245 0.49 0.62728
## Year1997 -0.1168 0.0607 -1.93 0.05439 .
## Year1998 -0.1713 0.0600 -2.85 0.00440 **
## Year1999 -0.2049 0.0601 -3.41 0.00067 ***
## Year2000 0.0216 0.0599 0.36 0.71831
## Year2001 -0.0987 0.0676 -1.46 0.14465
## Year2002 -0.0546 0.0575 -0.95 0.34204
## Year2003 -0.1266 0.0580 -2.18 0.02913 *
## Year2004 -0.0536 0.0547 -0.98 0.32705
## Year2005 -0.1142 0.0580 -1.97 0.04911 *
## Year2006 -0.1402 0.0576 -2.43 0.01507 *
```

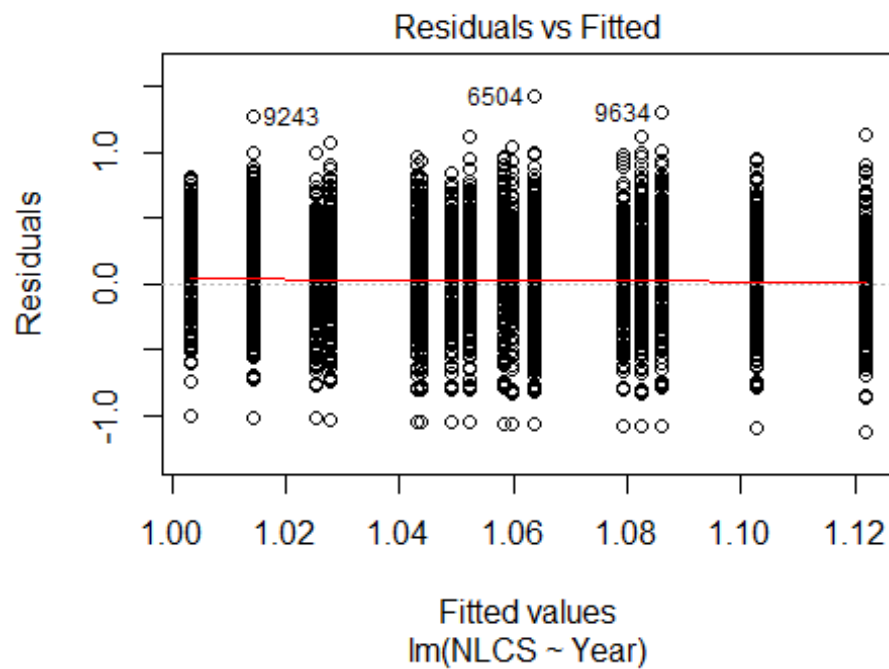
```

## Year2007          -0.1225      0.0590   -2.08  0.03806 *
## Year2008          -0.1052      0.0579   -1.82  0.06926 .
## Year2009          -0.0345      0.0545   -0.63  0.52663
## Year2010          -0.0717      0.0602   -1.19  0.23405
## Year2011           0.0358      0.0593    0.60  0.54606
## Year2012          -0.1567      0.0641   -2.44  0.01463 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.027, Adjusted R-squared:  0.0162
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1411 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.207  0.863  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      6.44e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1552"
## [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
##  679  619  545  456  545  570  529  492  514  544  506  540  507  468  474
## 2011 2012
##  430  423
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  258  199  227  193  169  136  259  231  266  282  259  300  312  267  291
## 2011 2012

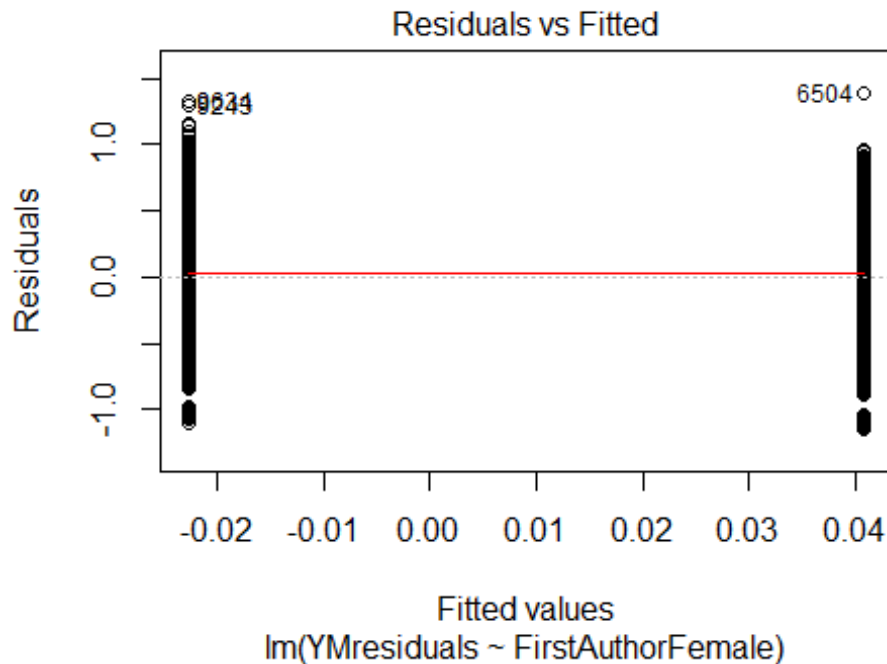
```



```
## 260 262
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 216 166 189 168 145 113 215 202 230 245 212 260 262 220 258
## 2011 2012
## 221 236
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 41, df = 16, p-value = 5e-04
```

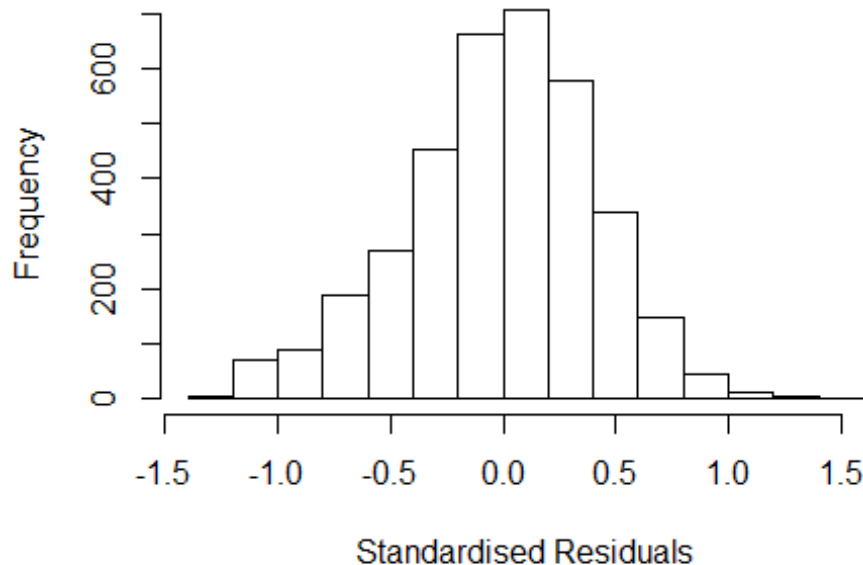


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 12, df = 1, p-value = 5e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 3.35727259823923"
## [1] "Male first author team size 2018 geometric mean: 3.34492967147475"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3900, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.43325188640052"
## [1] "Male last author team size 2018 geometric mean: 3.32476544871484"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1          1.029
## LastAuthorFemale  1.047 1          1.023
## UniqueAuthors    1.169 4          1.020
## Year              1.192 16         1.006
```

Residuals from first and last author and team size



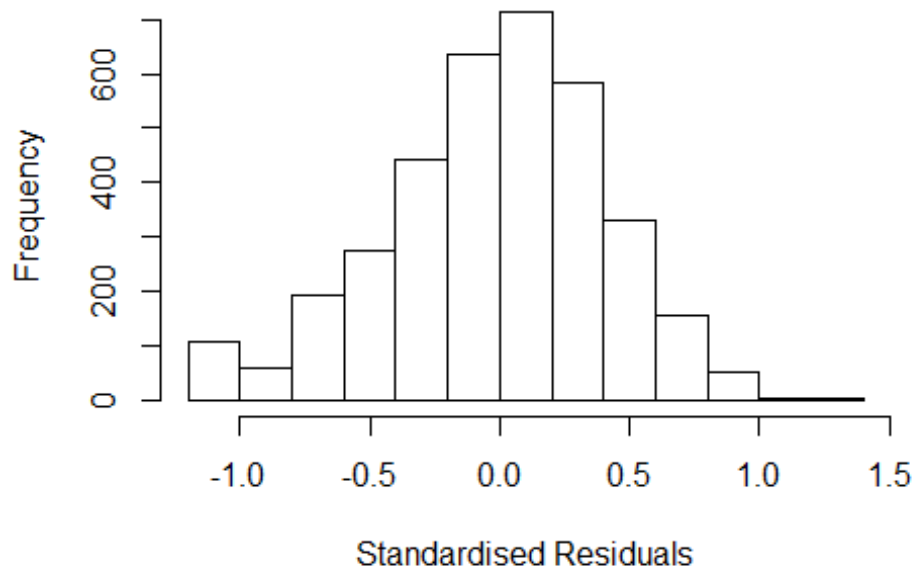
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2323 -0.2640  0.0147  0.2730  1.5786
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91431    0.03693   24.75 < 2e-16 ***
## FirstAuthorFemale1 0.05628    0.01464    3.84 0.00012 ***
## LastAuthorFemale1 0.02292    0.01727    1.33 0.18466
## UniqueAuthors2    0.18950    0.02886    6.57 5.9e-11 ***
## UniqueAuthors3    0.20604    0.02894    7.12 1.3e-12 ***
## UniqueAuthors4    0.25649    0.03153    8.13 5.7e-16 ***
## UniqueAuthors5    0.34243    0.03156   10.85 < 2e-16 ***
## Year1997         -0.00803    0.04150   -0.19 0.84653
## Year1998         -0.04420    0.04249   -1.04 0.29839
## Year1999         -0.03802    0.04221   -0.90 0.36770
```

```

## Year2000      -0.04961      0.04443      -1.12      0.26422
## Year2001      -0.05116      0.04647      -1.10      0.27102
## Year2002      -0.07548      0.04005      -1.88      0.05956 .
## Year2003      -0.09664      0.04042      -2.39      0.01687 *
## Year2004      -0.12654      0.04094      -3.09      0.00201 **
## Year2005      -0.10611      0.03863      -2.75      0.00605 **
## Year2006      -0.08279      0.04187      -1.98      0.04808 *
## Year2007      -0.05715      0.03750      -1.52      0.12762
## Year2008      -0.13337      0.03981      -3.35      0.00082 ***
## Year2009      -0.08073      0.04457      -1.81      0.07015 .
## Year2010      -0.11123      0.04416      -2.52      0.01182 *
## Year2011      -0.09786      0.04323      -2.26      0.02366 *
## Year2012      -0.06284      0.04194      -1.50      0.13411
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.398
## Multiple R-squared:  0.0581, Adjusted R-squared:  0.0522
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 288 weights are ~= 1. The remaining 3270 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0811 0.8640 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          2.81e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.038 1      1.019
## LastAuthorFemale 1.042 1      1.021
## Year 1.045 16      1.001

```

Residuals from first and last author



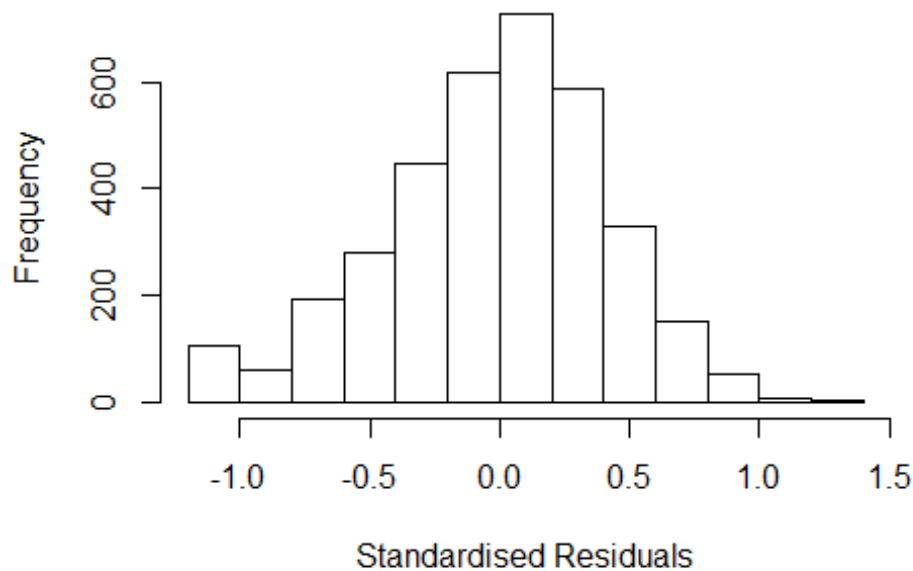
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1652 -0.2717 0.0229 0.2717 1.3826
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08515 0.02997 36.21 < 2e-16 ***
## FirstAuthorFemale1 0.06970 0.01479 4.71 2.5e-06 ***
## LastAuthorFemale1 0.01657 0.01759 0.94 0.346
## Year1997 0.00474 0.04188 0.11 0.910
## Year1998 -0.02966 0.04265 -0.70 0.487
## Year1999 -0.02464 0.04178 -0.59 0.555
## Year2000 -0.03138 0.04482 -0.70 0.484
## Year2001 -0.03246 0.04843 -0.67 0.503
## Year2002 -0.05691 0.04057 -1.40 0.161
## Year2003 -0.06546 0.04097 -1.60 0.110
## Year2004 -0.09297 0.04038 -2.30 0.021 *
## Year2005 -0.06512 0.03920 -1.66 0.097 .
```

```

## Year2006      -0.04945    0.04182   -1.18    0.237
## Year2007      -0.00624    0.03713   -0.17    0.866
## Year2008      -0.10033    0.04015   -2.50    0.013 *
## Year2009      -0.02686    0.04371   -0.61    0.539
## Year2010      -0.06431    0.04547   -1.41    0.157
## Year2011      -0.04215    0.04326   -0.97    0.330
## Year2012      -0.00699    0.04177   -0.17    0.867
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.0122, Adjusted R-squared:  0.00716
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 274 weights are ~= 1. The remaining 3284 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.219  0.864  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      2.81e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.02 1      1.010
## Year              1.02 16      1.001

```

Residuals from first author



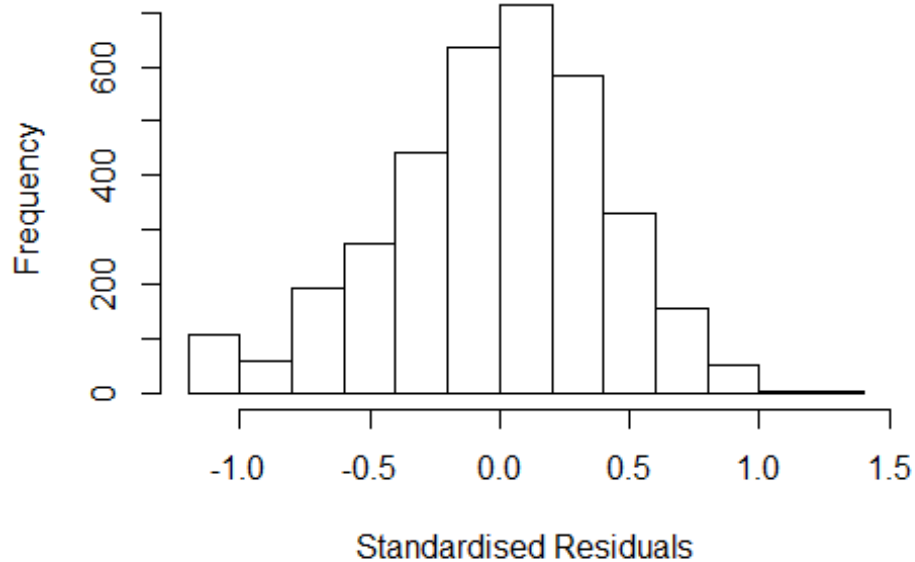
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.159 -0.273 0.023 0.274 1.378
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08716 0.02986 36.41 <2e-16 ***
## FirstAuthorFemale1 0.07181 0.01466 4.90 1e-06 ***
## Year1997 0.00451 0.04189 0.11 0.914
## Year1998 -0.02949 0.04262 -0.69 0.489
## Year1999 -0.02276 0.04168 -0.55 0.585
## Year2000 -0.02961 0.04474 -0.66 0.508
## Year2001 -0.03171 0.04842 -0.65 0.513
## Year2002 -0.05602 0.04056 -1.38 0.167
## Year2003 -0.06512 0.04096 -1.59 0.112
## Year2004 -0.09160 0.04033 -2.27 0.023 *
## Year2005 -0.06431 0.03921 -1.64 0.101
## Year2006 -0.04925 0.04182 -1.18 0.239
```

```

## Year2007          -0.00586    0.03713   -0.16    0.875
## Year2008          -0.09863    0.04012   -2.46    0.014 *
## Year2009          -0.02569    0.04379   -0.59    0.558
## Year2010          -0.06347    0.04541   -1.40    0.162
## Year2011          -0.04108    0.04328   -0.95    0.343
## Year2012          -0.00585    0.04174   -0.14    0.889
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.404
## Multiple R-squared:  0.0119, Adjusted R-squared:  0.00715
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 274 weights are ~= 1. The remaining 3284 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.222  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      2.81e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.024 1          1.012
## Year            1.024 16          1.001

```


Residuals from last author



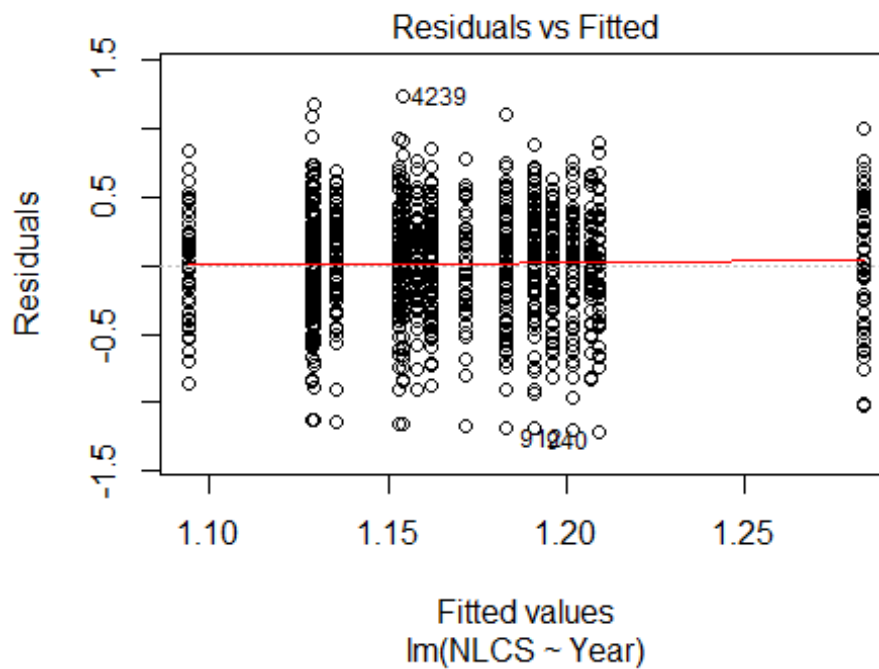
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1339 -0.2790 0.0191 0.2745 1.4229
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.105388 0.029776 37.12 <2e-16 ***
## LastAuthorFemale1 0.028738 0.017505 1.64 0.101
## Year1997 0.001851 0.042088 0.04 0.965
## Year1998 -0.030884 0.043077 -0.72 0.473
## Year1999 -0.026652 0.042417 -0.63 0.530
## Year2000 -0.032297 0.044982 -0.72 0.473
## Year2001 -0.030933 0.048217 -0.64 0.521
## Year2002 -0.052580 0.040861 -1.29 0.198
## Year2003 -0.062427 0.041539 -1.50 0.133
## Year2004 -0.090008 0.040741 -2.21 0.027 *
## Year2005 -0.061914 0.039483 -1.57 0.117
## Year2006 -0.040315 0.041966 -0.96 0.337
```

```

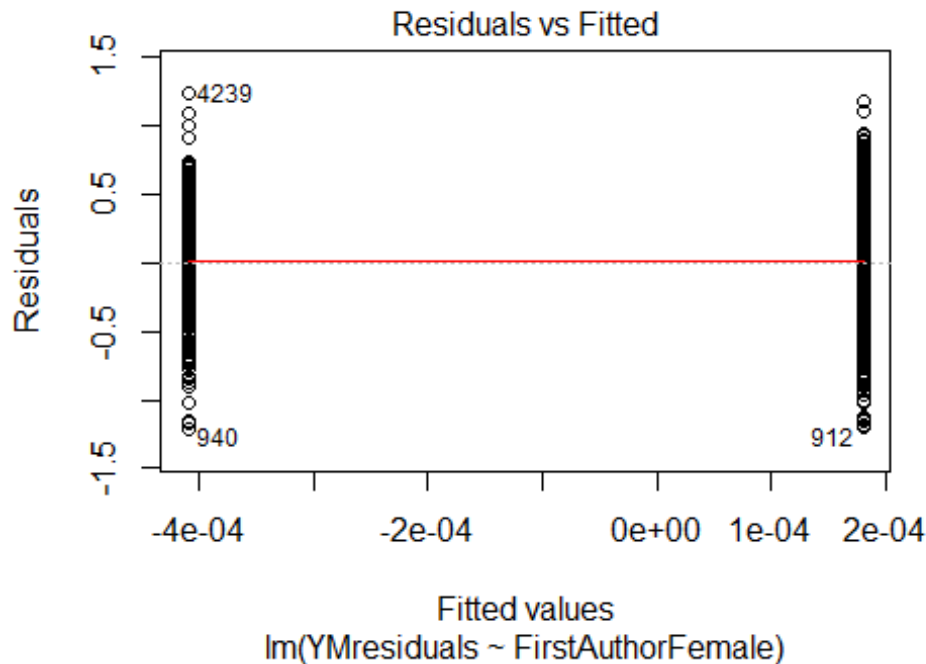
## Year2007          -0.000273    0.037397    -0.01    0.994
## Year2008          -0.097698    0.040405    -2.42    0.016 *
## Year2009          -0.022208    0.043771    -0.51    0.612
## Year2010          -0.059175    0.045892    -1.29    0.197
## Year2011          -0.030572    0.043291    -0.71    0.480
## Year2012           0.000730    0.042295     0.02    0.986
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.406
## Multiple R-squared:  0.00603,    Adjusted R-squared:  0.00126
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 288 weights are ~= 1. The remaining 3270 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.194  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      2.81e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3558"
## [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
## 184 191 203 171 183 178 177 161 160 140 206 242 240 207 214
## 2011 2012
## 241 212
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 44 40 59 53 40 47 56 39 59 55 87 99 103 89 95
## 2011 2012

```

```
## 105 105
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 38 33 52 41 33 38 45 30 44 41 72 76 82 75 80
## 2011 2012
## 79 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 = 19, df = 16, p-value = 0.3
```

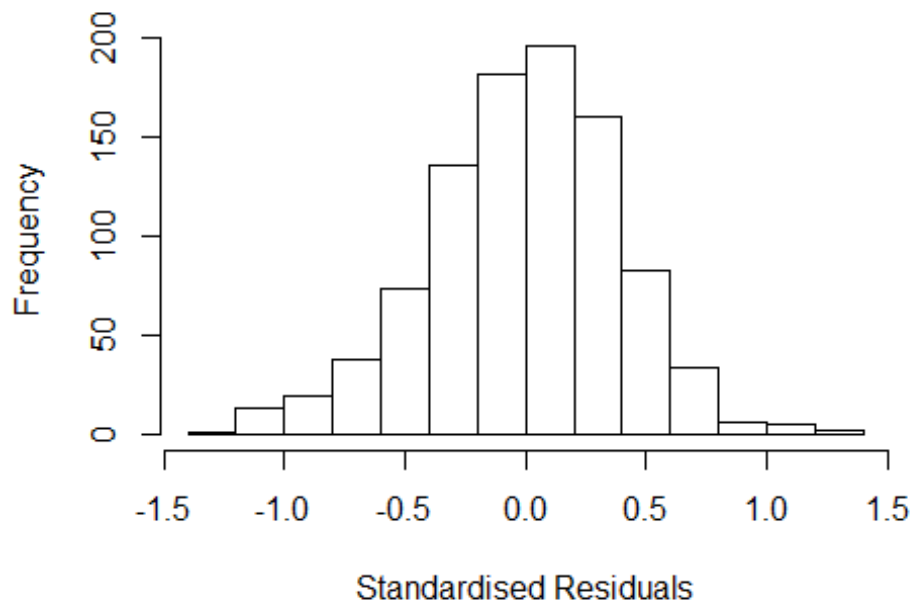


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.45, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 3.37251779841825"
## [1] "Male first author team size 2018 geometric mean: 3.46289283982039"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 720, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.51217338668086"
## [1] "Male last author team size 2018 geometric mean: 3.4093075847409"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 640, 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 1.126 1      1.061
## LastAuthorFemale  1.108 1      1.053
## UniqueAuthors    1.735 4      1.071
## Year             1.866 16      1.020
```

Residuals from first and last author and team size



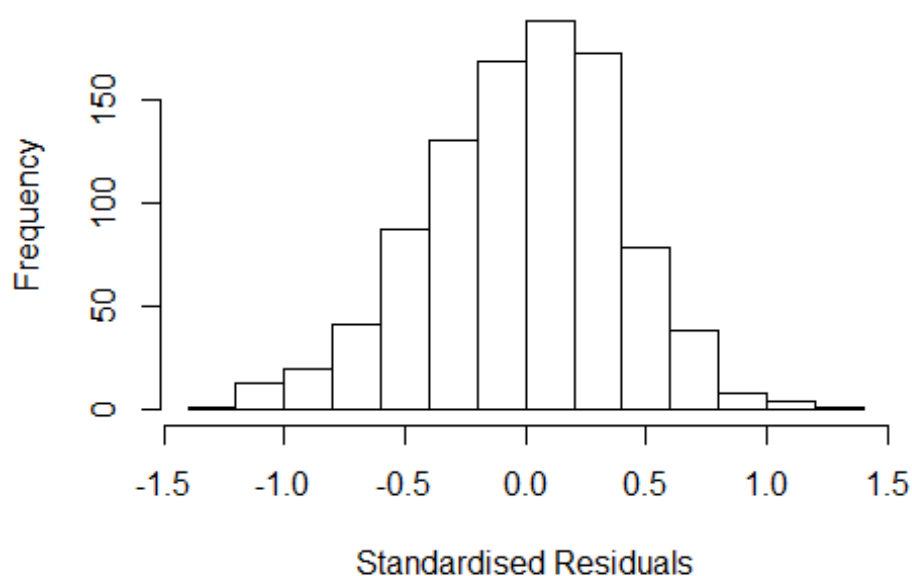
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2140 -0.2589  0.0179  0.2568  1.3299
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.15463    0.10933   10.56 < 2e-16 ***
## FirstAuthorFemale1 -0.00766    0.02901   -0.26  0.79185
## LastAuthorFemale1 -0.03558    0.03891   -0.91  0.36078
## UniqueAuthors2     0.15048    0.05524    2.72  0.00657 **
## UniqueAuthors3     0.12734    0.05618    2.27  0.02364 *
## UniqueAuthors4     0.20262    0.05810    3.49  0.00051 ***
## UniqueAuthors5     0.28567    0.05970    4.79  2e-06 ***
## Year1997         -0.11797    0.12321   -0.96  0.33861
## Year1998         -0.13775    0.10992   -1.25  0.21048
## Year1999         -0.06679    0.11549   -0.58  0.56319
```

```

## Year2000      -0.05918      0.12333      -0.48      0.63142
## Year2001      -0.13439      0.11654      -1.15      0.24912
## Year2002      -0.13103      0.11277      -1.16      0.24555
## Year2003      -0.13315      0.12985      -1.03      0.30542
## Year2004      -0.11064      0.11277      -0.98      0.32680
## Year2005      -0.04924      0.10967      -0.45      0.65353
## Year2006      -0.20601      0.10672      -1.93      0.05386 .
## Year2007      -0.18252      0.10729      -1.70      0.08924 .
## Year2008      -0.09115      0.10770      -0.85      0.39760
## Year2009      -0.11762      0.10681      -1.10      0.27111
## Year2010      -0.12112      0.10880      -1.11      0.26590
## Year2011      -0.19098      0.10513      -1.82      0.06958 .
## Year2012      -0.14606      0.10592      -1.38      0.16824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.383
## Multiple R-squared:  0.0467, Adjusted R-squared:  0.024
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 872 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.203  0.872  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          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 1.075 1          1.037
## LastAuthorFemale  1.076 1          1.037
## Year              1.153 16          1.004

```

Residuals from first and last author



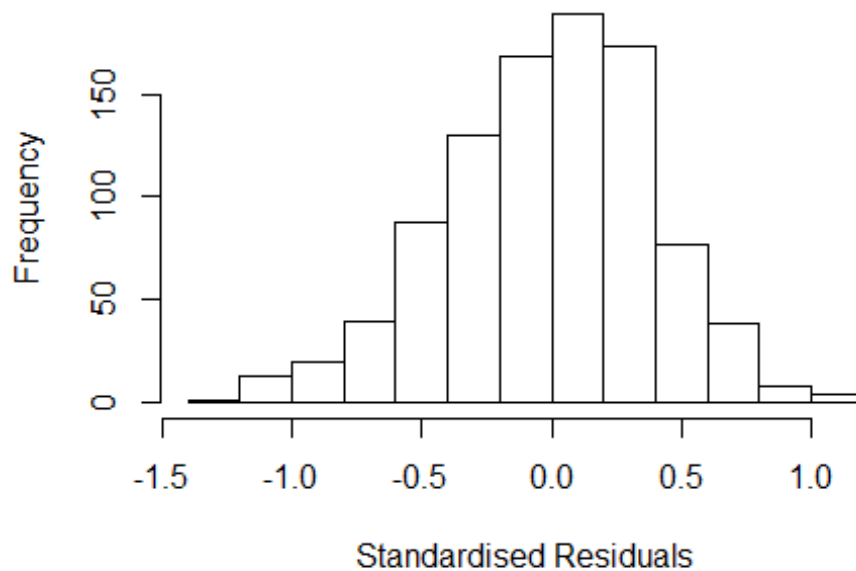
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2157 -0.2762  0.0133  0.2618  1.2192
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2759     0.0934   13.66  <2e-16 ***
## FirstAuthorFemale1  0.0174     0.0288    0.61    0.54
## LastAuthorFemale1 -0.0267     0.0388   -0.69    0.49
## Year1997          -0.1192     0.1214   -0.98    0.33
## Year1998          -0.1526     0.1073   -1.42    0.16
## Year1999          -0.0582     0.1151   -0.51    0.61
## Year2000          -0.0588     0.1260   -0.47    0.64
## Year2001          -0.1203     0.1138   -1.06    0.29
## Year2002          -0.1116     0.1109   -1.01    0.31
## Year2003          -0.0942     0.1298   -0.73    0.47
## Year2004          -0.0709     0.1114   -0.64    0.52
## Year2005          -0.0287     0.1086   -0.26    0.79
```

```

## Year2006          -0.1583      0.1042    -1.52      0.13
## Year2007          -0.1575      0.1064    -1.48      0.14
## Year2008          -0.0603      0.1054    -0.57      0.57
## Year2009          -0.0985      0.1054    -0.93      0.35
## Year2010          -0.0797      0.1067    -0.75      0.46
## Year2011          -0.1508      0.1043    -1.45      0.15
## Year2012          -0.0979      0.1039    -0.94      0.35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.398
## Multiple R-squared:  0.0124, Adjusted R-squared:  -0.00671
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 867 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.328  0.879   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      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 1.076 1      1.037
## Year              1.076 16      1.002

```


Residuals from first author



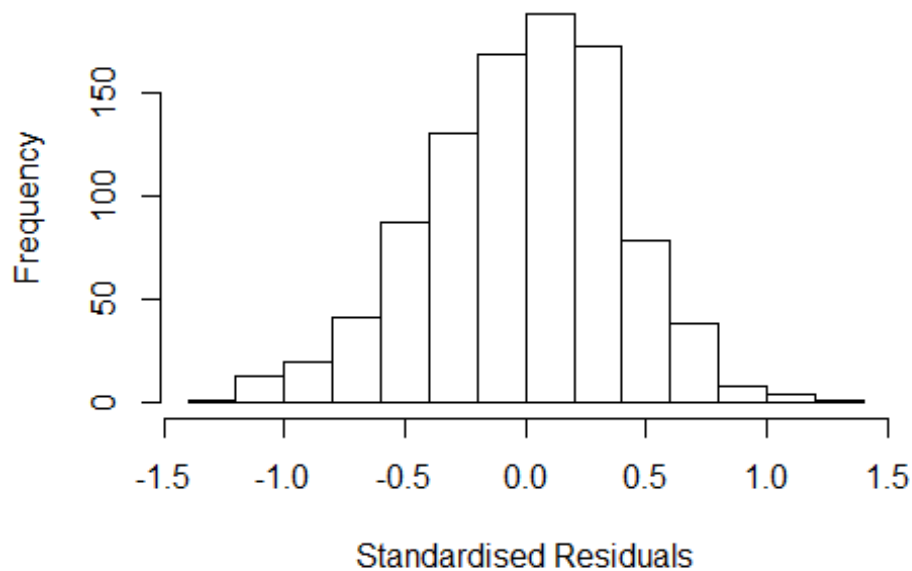
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2110 -0.2755 0.0142 0.2604 1.1978
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2760 0.0936 13.63 <2e-16 ***
## FirstAuthorFemale1 0.0159 0.0289 0.55 0.58
## Year1997 -0.1226 0.1210 -1.01 0.31
## Year1998 -0.1536 0.1073 -1.43 0.15
## Year1999 -0.0605 0.1153 -0.52 0.60
## Year2000 -0.0646 0.1256 -0.51 0.61
## Year2001 -0.1257 0.1135 -1.11 0.27
## Year2002 -0.1149 0.1107 -1.04 0.30
## Year2003 -0.0956 0.1297 -0.74 0.46
## Year2004 -0.0748 0.1112 -0.67 0.50
## Year2005 -0.0321 0.1083 -0.30 0.77
## Year2006 -0.1601 0.1043 -1.54 0.13
```

```

## Year2007          -0.1612      0.1062    -1.52      0.13
## Year2008          -0.0649      0.1050    -0.62      0.54
## Year2009          -0.1015      0.1054    -0.96      0.34
## Year2010          -0.0852      0.1062    -0.80      0.42
## Year2011          -0.1562      0.1039    -1.50      0.13
## Year2012          -0.1016      0.1038    -0.98      0.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.398
## Multiple R-squared:  0.0119, Adjusted R-squared:  -0.00618
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 869 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.334  0.877  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      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 1.075 1      1.037
## Year              1.075 16      1.002

```

Residuals from last author



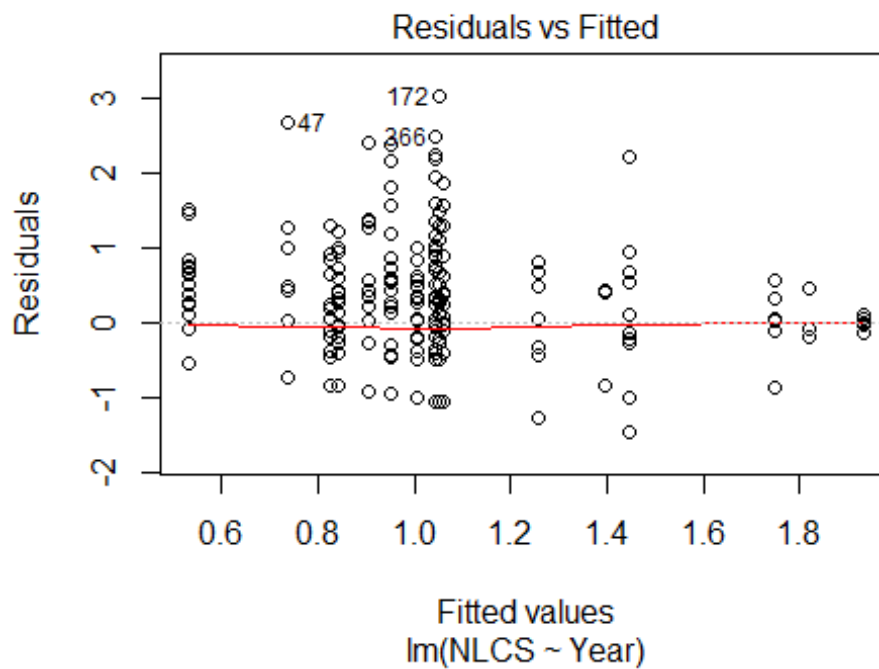
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2218 -0.2776 0.0146 0.2650 1.2280
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2772 0.0936 13.65 <2e-16 ***
## LastAuthorFemale1 -0.0250 0.0388 -0.64 0.52
## Year1997 -0.1174 0.1212 -0.97 0.33
## Year1998 -0.1499 0.1069 -1.40 0.16
## Year1999 -0.0561 0.1150 -0.49 0.63
## Year2000 -0.0549 0.1253 -0.44 0.66
## Year2001 -0.1173 0.1137 -1.03 0.30
## Year2002 -0.1082 0.1105 -0.98 0.33
## Year2003 -0.0911 0.1297 -0.70 0.48
## Year2004 -0.0686 0.1110 -0.62 0.54
## Year2005 -0.0255 0.1081 -0.24 0.81
## Year2006 -0.1526 0.1033 -1.48 0.14
```

```

## Year2007          -0.1542      0.1060    -1.45      0.15
## Year2008          -0.0554      0.1045    -0.53      0.60
## Year2009          -0.0951      0.1049    -0.91      0.36
## Year2010          -0.0731      0.1052    -0.69      0.49
## Year2011          -0.1453      0.1033    -1.41      0.16
## Year2012          -0.0922      0.1028    -0.90      0.37
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.399
## Multiple R-squared:  0.012, Adjusted R-squared:  -0.00603
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 871 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.322  0.879   0.951   0.907   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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: 948"
## [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 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    4    7    6    4    8   17   23   34   21   11   35   23   23   29   45
## 2012
##    60
##
## 1996 1997 1998 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    4    5    6    3    7   15   21   33   19   11   26   18   20   22   40
## 2012

```

```
## 50
##
## 1996 1997 1998 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
## 4 5 5 3 7 14 20 33 19 11 26 18 20 21 40
## 2012
## 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 = 40, df = 15, p-value = 4e-04
```



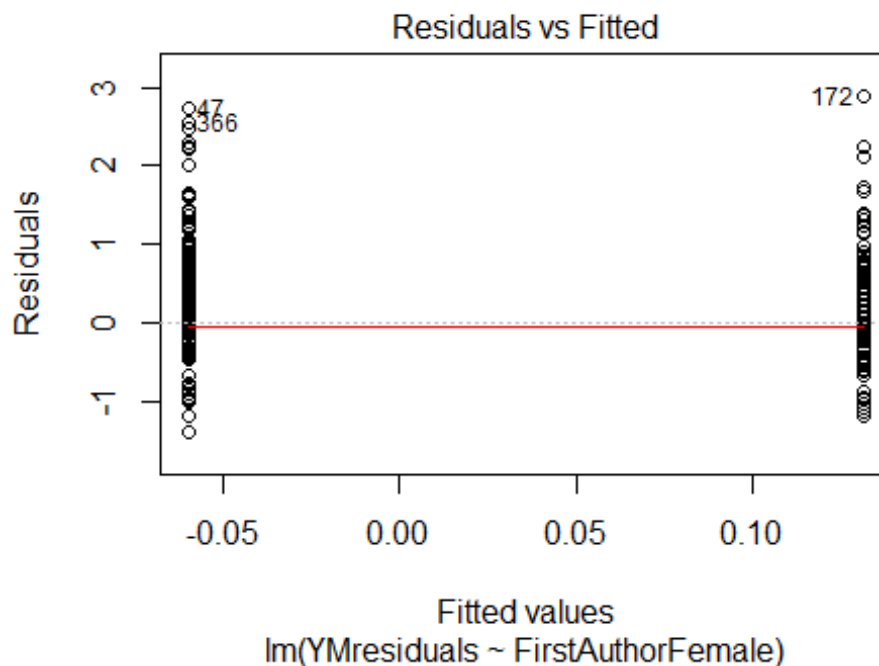
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.75, df = 1, p-value = 0.4

## [1] "Female first author team size 2018 geometric mean: 1.58050076202388"
## [1] "Male first author team size 2018 geometric mean: 1.29170834209075"

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

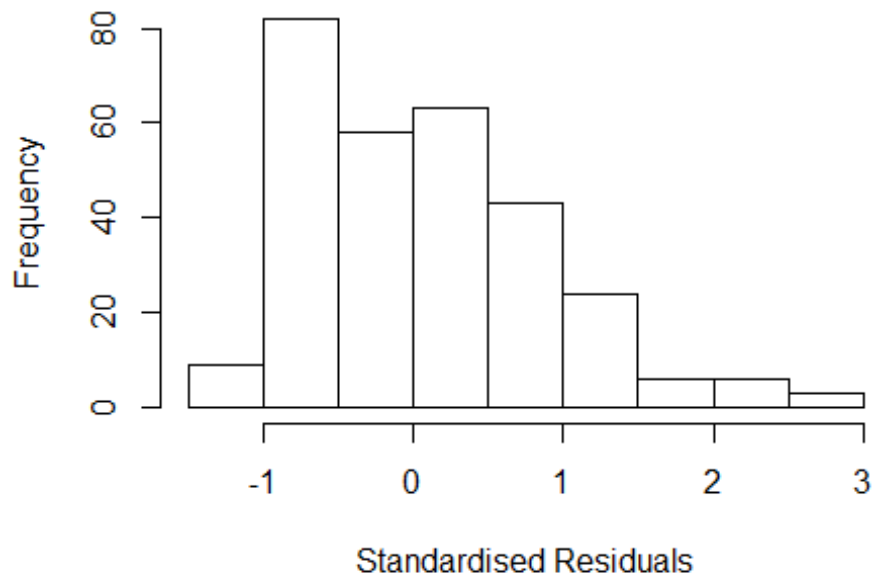
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.60730105586399"
## [1] "Male last author team size 2018 geometric mean: 1.24899132937904"

## Warning in wilcox.test.default(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 = 0.3
## 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 4.472  1      2.115
## LastAuthorFemale  3.773  1      1.942
## UniqueAuthors    6.001  4      1.251
## Year              8.688 15      1.075
```

Residuals from first and last author and team size



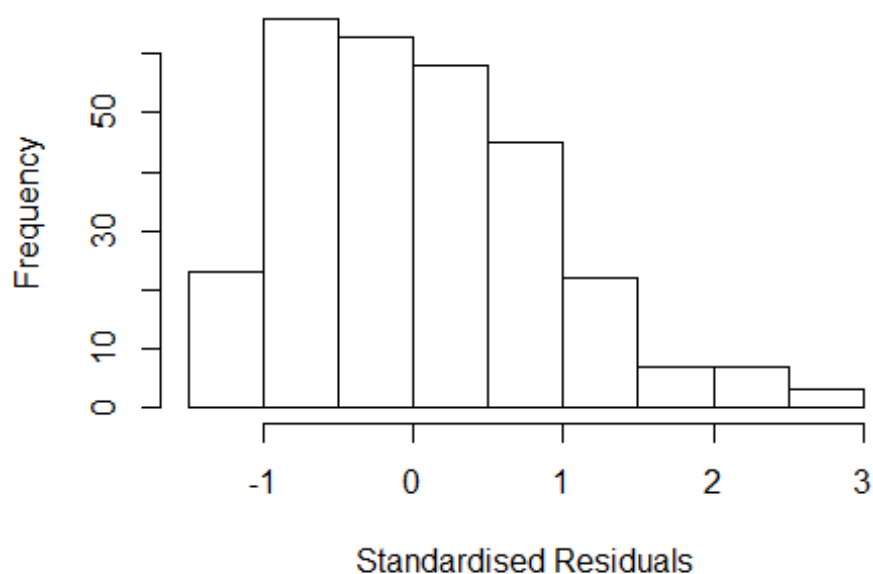
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 47   0141456190 3.402 2002    1200     1    2.867
## 172 61449340446 4.063 2007    1200     1    2.937
## 366 84872181712 3.529 2012    1200     2    2.800
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2411 -0.6685 -0.0123  0.6004  2.9365
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.6104    0.1180   13.64 < 2e-16 ***
## FirstAuthorFemale1  0.1683    0.1723    0.98  0.32952
## LastAuthorFemale1  0.0817    0.1744    0.47  0.63962
## UniqueAuthors2     0.2903    0.1844    1.57  0.11656
## UniqueAuthors3    -0.1259    0.2264   -0.56  0.57848
## UniqueAuthors4     1.0934    0.2355    4.64 5.4e-06 ***
## UniqueAuthors5     1.1236    0.6885    1.63  0.10384
## Year1997           0.2628    0.0920    2.86  0.00459 **
```

```

## Year1998          0.0296      0.2129      0.14  0.88943
## Year2000         -0.2013      0.3919     -0.51  0.60789
## Year2001         -0.4260      0.2994     -1.42  0.15594
## Year2002         -1.0757      0.2466     -4.36  1.8e-05 ***
## Year2003         -0.9567      0.2417     -3.96  9.6e-05 ***
## Year2004         -1.1565      0.1555     -7.44  1.4e-12 ***
## Year2005         -0.8944      0.1997     -4.48  1.1e-05 ***
## Year2006         -0.3693      0.3336     -1.11  0.26932
## Year2007         -0.7339      0.2307     -3.18  0.00164 **
## Year2008         -0.7140      0.1751     -4.08  6.0e-05 ***
## Year2009         -0.7769      0.2142     -3.63  0.00034 ***
## Year2010         -0.8729      0.1769     -4.94  1.4e-06 ***
## Year2011         -0.9369      0.1835     -5.11  6.2e-07 ***
## Year2012         -0.8815      0.2041     -4.32  2.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.8
## Multiple R-squared:  0.158, Adjusted R-squared:  0.0926
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 267 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.150  0.882  0.937  0.899  0.971  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.40e-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 4.231 1 2.057
## LastAuthorFemale 3.453 1 1.858
## Year 2.371 15 1.029

```


Residuals from first and last author



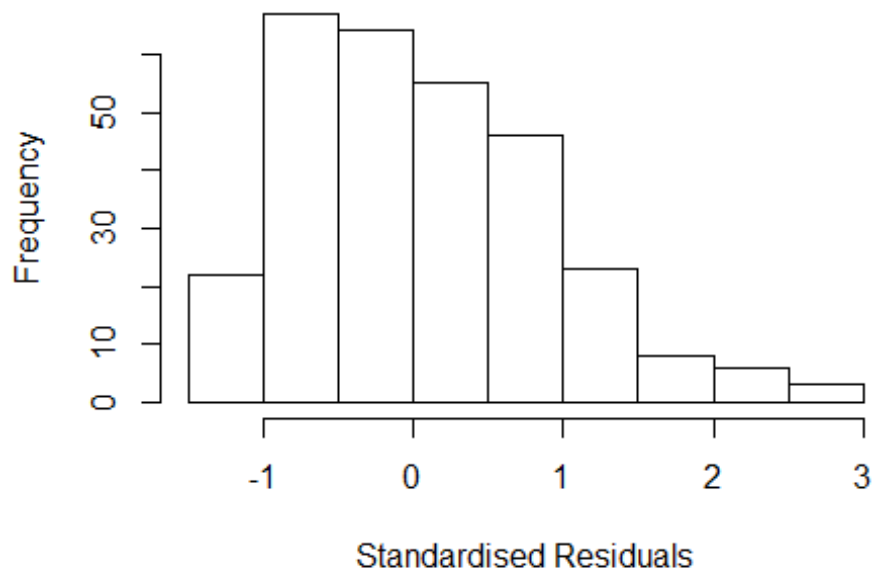
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 47   0141456190 3.402 2002    1200     1    2.869
## 172 61449340446 4.063 2007    1200     1    2.882
## 366 84872181712 3.529 2012    1200     2    2.726
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2706 -0.6540 -0.0206  0.6058  2.8823
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.6791    0.1360   12.34 < 2e-16 ***
## FirstAuthorFemale1 -0.0165    0.2018   -0.08  0.93494
## LastAuthorFemale1  0.3026    0.2025    1.49  0.13623
## Year1997          0.1988    0.1229    1.62  0.10681
## Year1998         -0.0396    0.2230   -0.18  0.85924
## Year2000         -0.2554    0.3792   -0.67  0.50121
## Year2001         -0.4817    0.2903   -1.66  0.09820 .
## Year2002         -1.1460    0.2542   -4.51  9.7e-06 ***
## Year2003         -1.0250    0.2503   -4.10  5.5e-05 ***
## Year2004         -1.2305    0.1683   -7.31  2.9e-12 ***
```

```

## Year2005          -0.9553      0.2078   -4.60   6.5e-06 ***
## Year2006          -0.4084      0.3584   -1.14   0.25547
## Year2007          -0.7845      0.2444   -3.21   0.00148 **
## Year2008          -0.7275      0.1937   -3.76   0.00021 ***
## Year2009          -0.7546      0.2381   -3.17   0.00170 **
## Year2010          -0.9246      0.1880   -4.92   1.5e-06 ***
## Year2011          -0.9389      0.2313   -4.06   6.4e-05 ***
## Year2012          -0.8758      0.2146   -4.08   5.9e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.822
## Multiple R-squared:  0.126, Adjusted R-squared:  0.0721
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 271 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.194  0.884   0.939   0.902   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      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 1.531 1         1.237
## Year              1.531 15         1.014

```

Residuals from first author



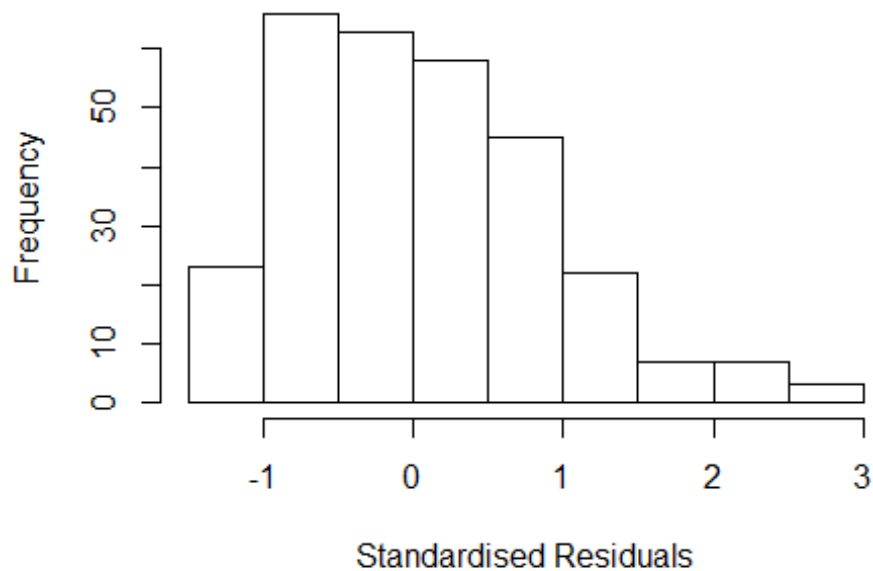
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 47   0141456190 3.402 2002    1200     1    2.869
## 172 61449340446 4.063 2007    1200     1    2.882
## 366 84872181712 3.529 2012    1200     2    2.726
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2876 -0.6652 -0.0336  0.6024  2.9221
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.63641    0.15205   10.76 < 2e-16 ***
## FirstAuthorFemale1 0.24091    0.12689    1.90 0.05865 .
## Year1997        0.19874    0.14346    1.39 0.16706
## Year1998        0.00294    0.23330    0.01 0.98994
## Year2000       -0.28100    0.47457   -0.59 0.55426
## Year2001       -0.42739    0.29797   -1.43 0.15261
## Year2002       -1.08856    0.26476   -4.11 5.2e-05 ***
## Year2003       -0.97123    0.25996   -3.74 0.00023 ***
## Year2004       -1.17916    0.18138   -6.50 3.7e-10 ***
## Year2005       -0.90031    0.21830   -4.12 4.9e-05 ***
```

```

## Year2006      -0.34882    0.36036   -0.97   0.33390
## Year2007      -0.73645    0.25346   -2.91   0.00396 **
## Year2008      -0.67408    0.20789   -3.24   0.00133 **
## Year2009      -0.70826    0.25435   -2.78   0.00573 **
## Year2010      -0.84887    0.19490   -4.36   1.9e-05 ***
## Year2011      -0.86503    0.22237   -3.89   0.00013 ***
## Year2012      -0.82717    0.22025   -3.76   0.00021 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.828
## Multiple R-squared:  0.119, Adjusted R-squared:  0.0684
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 276 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.187  0.889   0.941   0.905   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      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 1.34 1      1.157
## Year              1.34 15      1.010

```

Residuals from last author



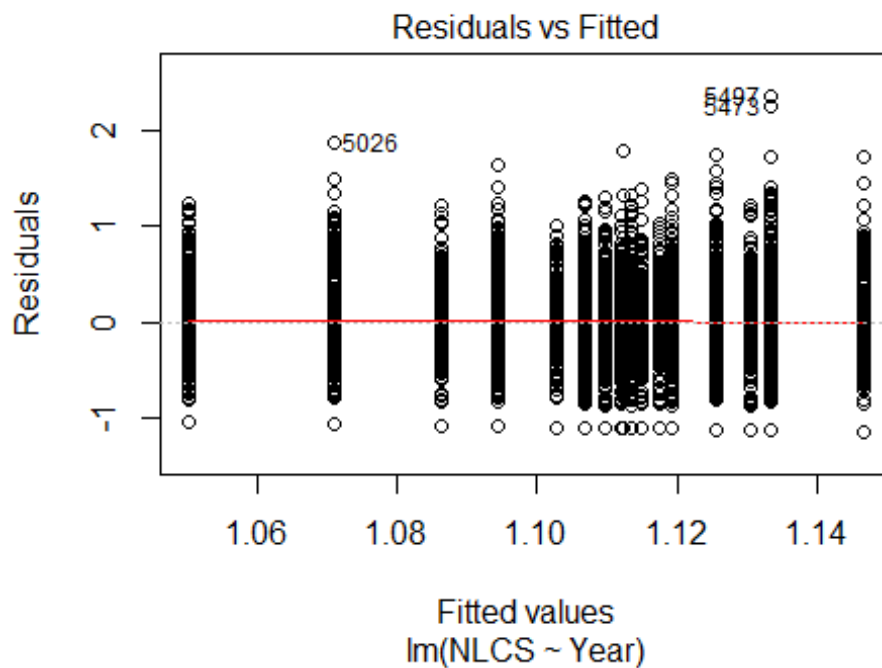
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 47   0141456190 3.402 2002    1200     1    2.869
## 172 61449340446 4.063 2007    1200     1    2.882
## 366 84872181712 3.529 2012    1200     2    2.726
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2705 -0.6540 -0.0225  0.6063  2.8799
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.6739     0.1139   14.69 < 2e-16 ***
## LastAuthorFemale1  0.2882     0.1267    2.27 0.02371 *
## Year1997          0.2003     0.1205    1.66 0.09761 .
## Year1998         -0.0344     0.2105   -0.16 0.87018
## Year2000         -0.2549     0.3833   -0.67 0.50655
## Year2001         -0.4776     0.2900   -1.65 0.10067
## Year2002         -1.1405     0.2434   -4.69 4.4e-06 ***
## Year2003         -1.0199     0.2393   -4.26 2.8e-05 ***
## Year2004         -1.2256     0.1525   -8.04 2.7e-14 ***
## Year2005         -0.9506     0.1957   -4.86 2.0e-06 ***
```

```

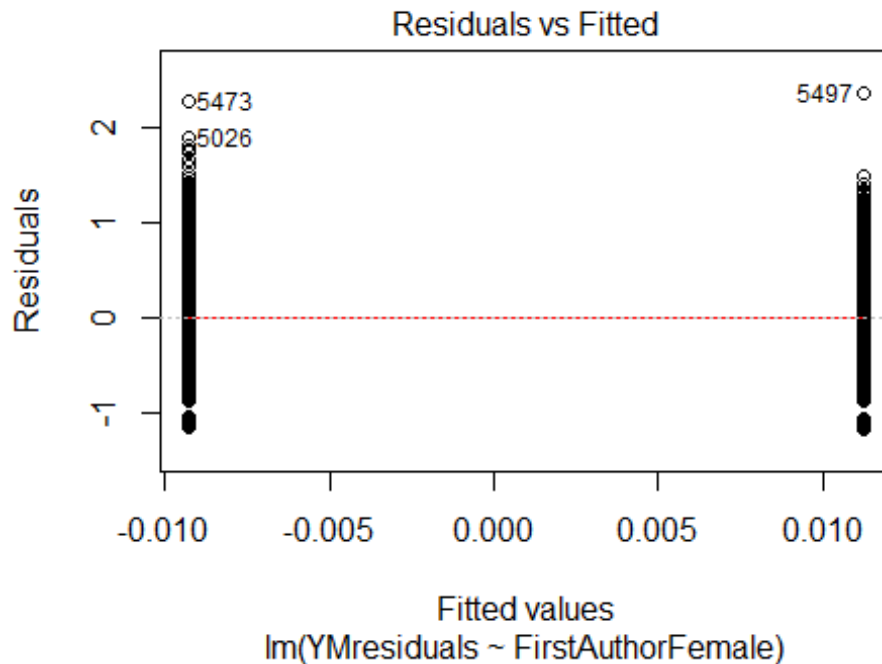
## Year2006          -0.4034      0.3538    -1.14   0.25530
## Year2007          -0.7790      0.2256    -3.45   0.00064 ***
## Year2008          -0.7228      0.1817    -3.98   8.9e-05 ***
## Year2009          -0.7502      0.2391    -3.14   0.00188 **
## Year2010          -0.9194      0.1745    -5.27   2.8e-07 ***
## Year2011          -0.9330      0.2048    -4.56   7.8e-06 ***
## Year2012          -0.8713      0.2101    -4.15   4.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.824
## Multiple R-squared:  0.126, Adjusted R-squared:  0.0752
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 271 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.197  0.885   0.939   0.903   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      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 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
## 244 245 191 194 231 283 222 217 237 256 330 334 343 395 407
## 2011 2012
## 445 466
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 195 177 155 151 162 173 182 177 194 215 266 276 279 336 338

```

```
## 2011 2012
## 374 400
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 177 165 140 146 154 161 168 162 178 194 236 263 257 303 313
## 2011 2012
## 348 363
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 68, df = 16, p-value = 2e-08
```

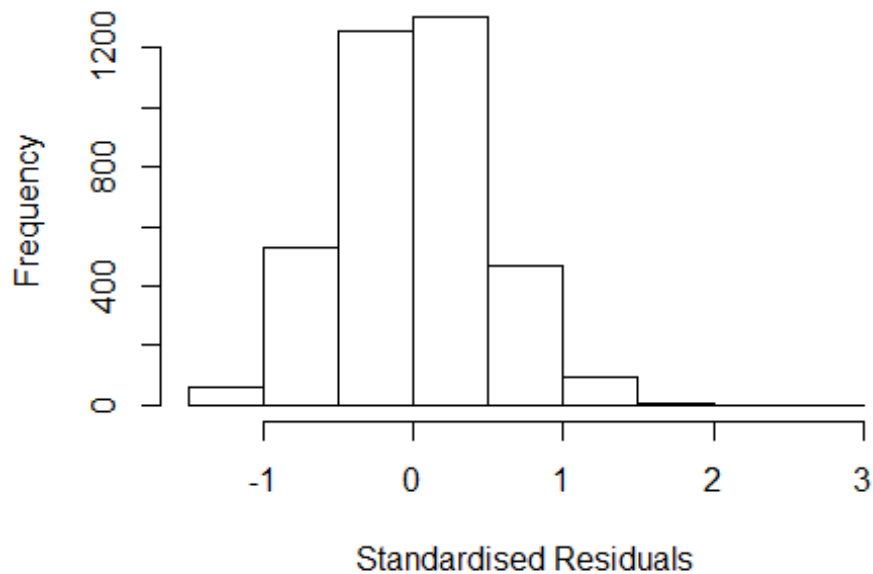


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 12, df = 1, p-value = 5e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 2.04167089500667"
## [1] "Male first author team size 2018 geometric mean: 1.83767848247501"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.97727069801264"
## [1] "Male last author team size 2018 geometric mean: 1.90346252236055"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14000, 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 1.203 1      1.097
## LastAuthorFemale 1.200 1      1.095
## UniqueAuthors    1.115 4      1.014
## Year             1.141 16     1.004
```


Residuals from first and last author and team size



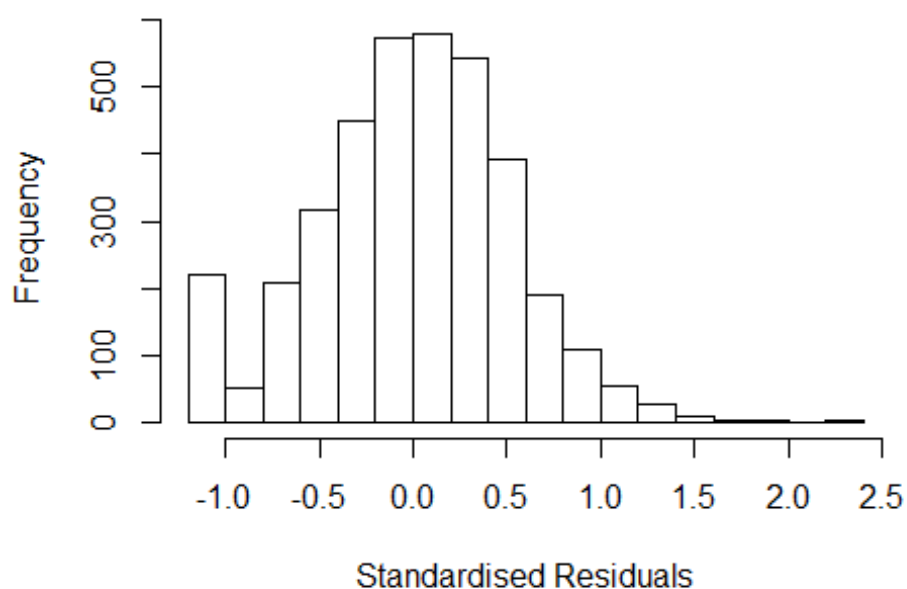
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 5497 79953204979 3.499 2011    1201      3      2.523
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28209 -0.32625  0.00394  0.33478  2.52264
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9231    0.0493   18.74 < 2e-16 ***
## FirstAuthorFemale1 0.0139    0.0183    0.76   0.45
## LastAuthorFemale1 0.0133    0.0188    0.71   0.48
## UniqueAuthors2    0.2111    0.0211   10.02 < 2e-16 ***
## UniqueAuthors3    0.2546    0.0244   10.43 < 2e-16 ***
## UniqueAuthors4    0.2119    0.0325    6.53 7.5e-11 ***
## UniqueAuthors5    0.3064    0.0312    9.83 < 2e-16 ***
## Year1997          0.0724    0.0652    1.11   0.27
## Year1998          0.0794    0.0618    1.29   0.20
## Year1999          0.0612    0.0616    0.99   0.32
```

```

## Year2000      0.0424      0.0611      0.69      0.49
## Year2001      0.0587      0.0616      0.95      0.34
## Year2002      0.0297      0.0626      0.48      0.63
## Year2003      0.0129      0.0594      0.22      0.83
## Year2004      0.0270      0.0628      0.43      0.67
## Year2005      0.0435      0.0569      0.77      0.44
## Year2006      0.0116      0.0571      0.20      0.84
## Year2007      0.0435      0.0553      0.79      0.43
## Year2008      0.0386      0.0584      0.66      0.51
## Year2009      0.0139      0.0559      0.25      0.80
## Year2010     -0.0161      0.0580     -0.28      0.78
## Year2011      0.0260      0.0566      0.46      0.65
## Year2012      0.0401      0.0551      0.73      0.47
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.0549, Adjusted R-squared:  0.0492
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(3258,3274) are outliers with |weight| = 0 ( < 2.7e-05);
## 291 weights are ~= 1. The remaining 3435 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0475 0.8710 0.9530 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      2.68e-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 1.180 1      1.086
## LastAuthorFemale 1.191 1      1.091
## Year      1.059 16      1.002

```

Residuals from first and last author



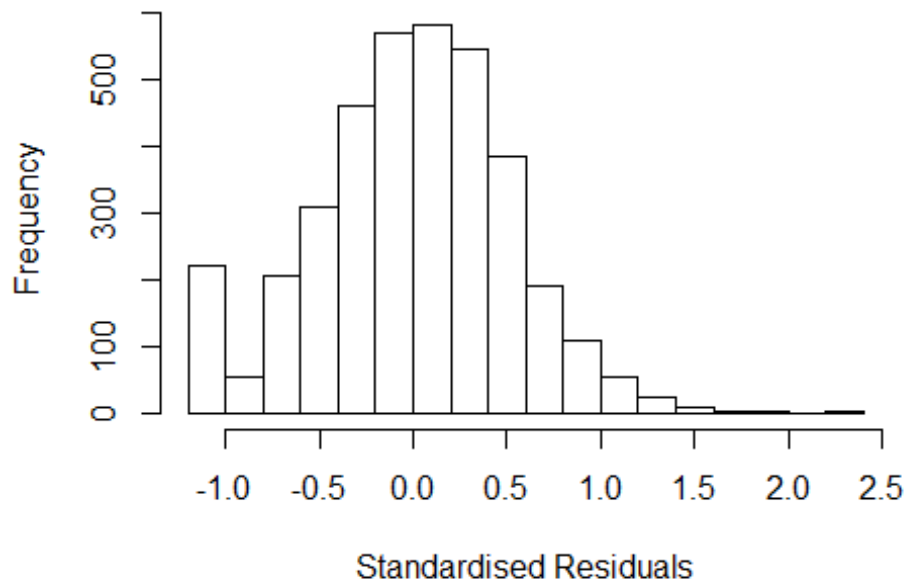
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1693 -0.3333 0.0136 0.3357 2.3789
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0486 0.0486 21.55 <2e-16 ***
## FirstAuthorFemale1 0.0399 0.0186 2.14 0.032 *
## LastAuthorFemale1 -0.0115 0.0192 -0.60 0.549
## Year1997 0.0690 0.0664 1.04 0.299
## Year1998 0.0924 0.0631 1.46 0.143
## Year1999 0.0731 0.0630 1.16 0.246
## Year2000 0.0623 0.0632 0.99 0.324
## Year2001 0.0690 0.0621 1.11 0.267
## Year2002 0.0421 0.0628 0.67 0.503
## Year2003 0.0363 0.0598 0.61 0.544
## Year2004 0.0459 0.0634 0.72 0.469
## Year2005 0.0468 0.0570 0.82 0.412
```

```

## Year2006          0.0125      0.0581      0.22      0.829
## Year2007          0.0503      0.0557      0.90      0.367
## Year2008          0.0573      0.0585      0.98      0.327
## Year2009          0.0211      0.0569      0.37      0.711
## Year2010          0.0020      0.0593      0.03      0.973
## Year2011          0.0432      0.0576      0.75      0.453
## Year2012          0.0656      0.0559      1.17      0.241
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.499
## Multiple R-squared:  0.00334,    Adjusted R-squared:  -0.0015
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 3274 is an outlier with |weight| = 0 ( < 2.7e-05);
## 329 weights are ~= 1. The remaining 3398 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0013 0.8660 0.9500 0.9000 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.68e-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 1.027 1          1.014
## Year              1.027 16          1.001

```

Residuals from first author



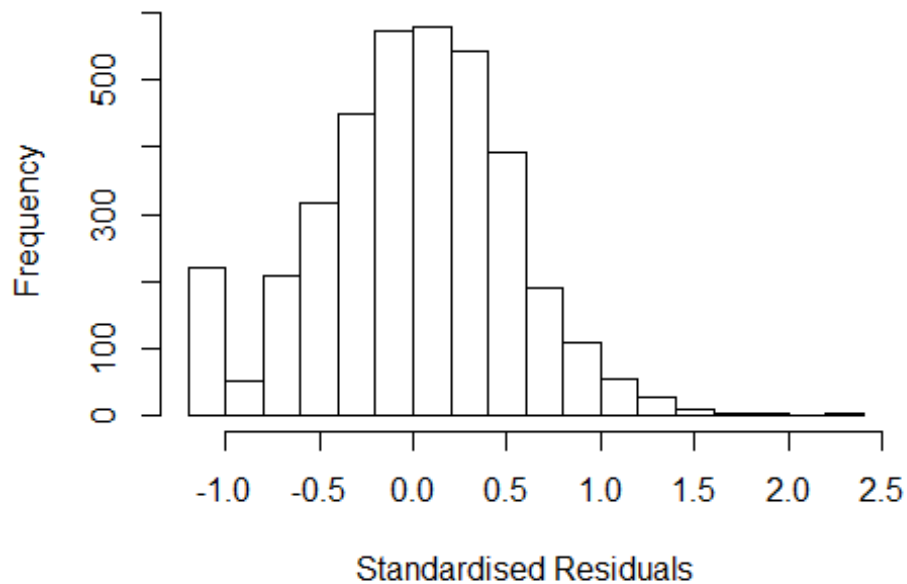
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1724 -0.3341 0.0121 0.3346 2.3749
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04706 0.04860 21.54 <2e-16 ***
## FirstAuthorFemale1 0.03443 0.01749 1.97 0.049 *
## Year1997 0.06868 0.06641 1.03 0.301
## Year1998 0.09091 0.06298 1.44 0.149
## Year1999 0.07319 0.06291 1.16 0.245
## Year2000 0.06217 0.06322 0.98 0.325
## Year2001 0.06858 0.06215 1.10 0.270
## Year2002 0.04196 0.06283 0.67 0.504
## Year2003 0.03488 0.05966 0.58 0.559
## Year2004 0.04636 0.06333 0.73 0.464
## Year2005 0.04579 0.05696 0.80 0.422
## Year2006 0.01138 0.05798 0.20 0.844
```

```

## Year2007          0.04966      0.05565      0.89      0.372
## Year2008          0.05655      0.05847      0.97      0.334
## Year2009          0.02080      0.05687      0.37      0.715
## Year2010          0.00143      0.05925      0.02      0.981
## Year2011          0.04264      0.05753      0.74      0.459
## Year2012          0.06525      0.05585      1.17      0.243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.499
## Multiple R-squared:  0.00327,    Adjusted R-squared:  -0.0013
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 3274 is an outlier with |weight| = 0 ( < 2.7e-05);
## 327 weights are ~= 1. The remaining 3400 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0012 0.8670 0.9500 0.9000 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.68e-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 1.034 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
## -1.1560 -0.3384 0.0152 0.3390 2.3857
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05738 0.04835 21.87 <2e-16 ***
## LastAuthorFemale1 0.00797 0.01806 0.44 0.66
## Year1997 0.07101 0.06646 1.07 0.29
## Year1998 0.09069 0.06321 1.43 0.15
## Year1999 0.07317 0.06300 1.16 0.25
## Year2000 0.06369 0.06327 1.01 0.31
## Year2001 0.06866 0.06223 1.10 0.27
## Year2002 0.04431 0.06300 0.70 0.48
## Year2003 0.03648 0.06003 0.61 0.54
## Year2004 0.04744 0.06332 0.75 0.45
## Year2005 0.04718 0.05702 0.83 0.41
## Year2006 0.01534 0.05819 0.26 0.79
```

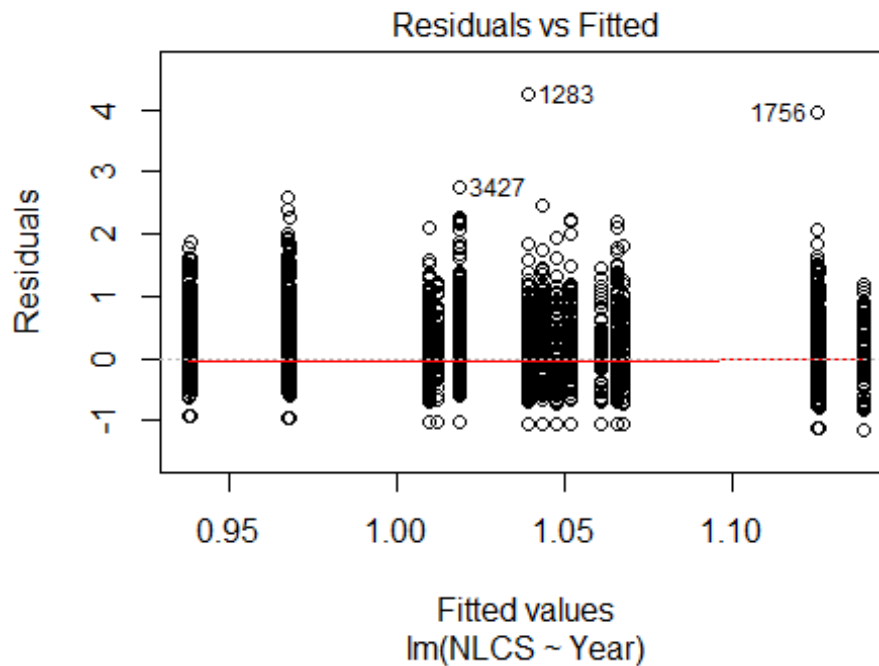
```

## Year2007      0.05380      0.05579      0.96      0.34
## Year2008      0.05862      0.05869      1.00      0.32
## Year2009      0.02177      0.05697      0.38      0.70
## Year2010      0.00598      0.05944      0.10      0.92
## Year2011      0.04797      0.05761      0.83      0.41
## Year2012      0.06874      0.05596      1.23      0.22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.499
## Multiple R-squared:  0.00222,    Adjusted R-squared:  -0.00235
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 3274 is an outlier with |weight| = 0 ( < 2.7e-05);
## 326 weights are ~= 1. The remaining 3401 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0021 0.8670 0.9500 0.9000 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.68e-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: 3728"
## [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
##   88   96  103  126  133  145  146  166  166  206  214  259  245  237  268
## 2011 2012
##  352  336
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   72   80   89  101  100  103  126  147  142  178  182  224  212  207  222

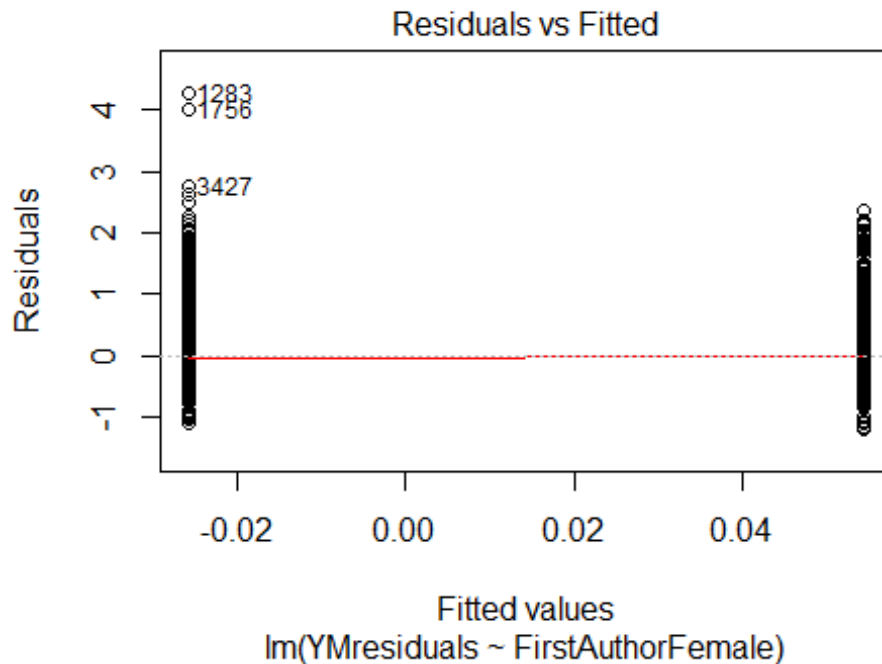
```



```
## 2011 2012
## 311 289
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 71 78 88 100 99 100 124 146 140 178 179 220 207 204 216
## 2011 2012
## 304 284
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 56, df = 16, p-value = 2e-06
```

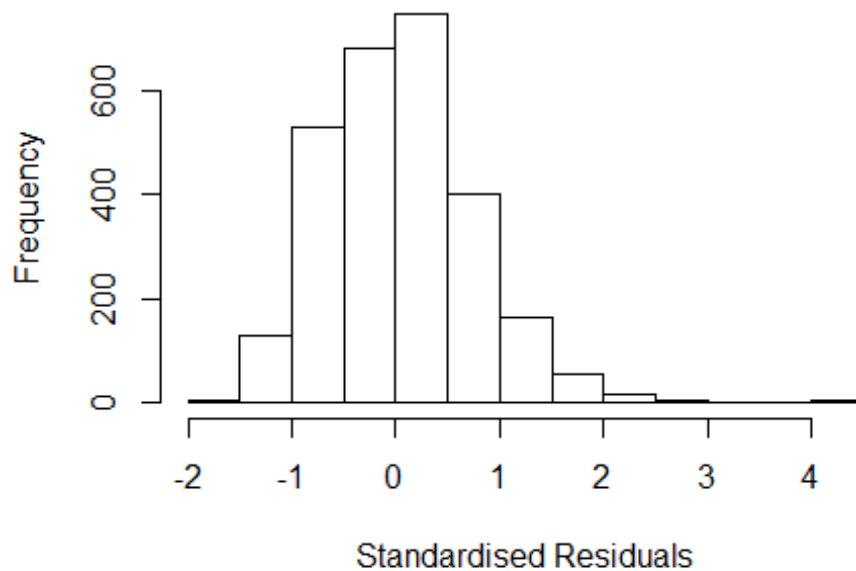


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.48, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 1.17408115948949"
## [1] "Male first author team size 2018 geometric mean: 1.11731896912902"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.14182582868466"
## [1] "Male last author team size 2018 geometric mean: 1.14263439947345"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, 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 3.711 1          1.926
## LastAuthorFemale  3.736 1          1.933
## UniqueAuthors    1.112 4          1.013
## Year              1.115 16         1.003
```

Residuals from first and last author and team size



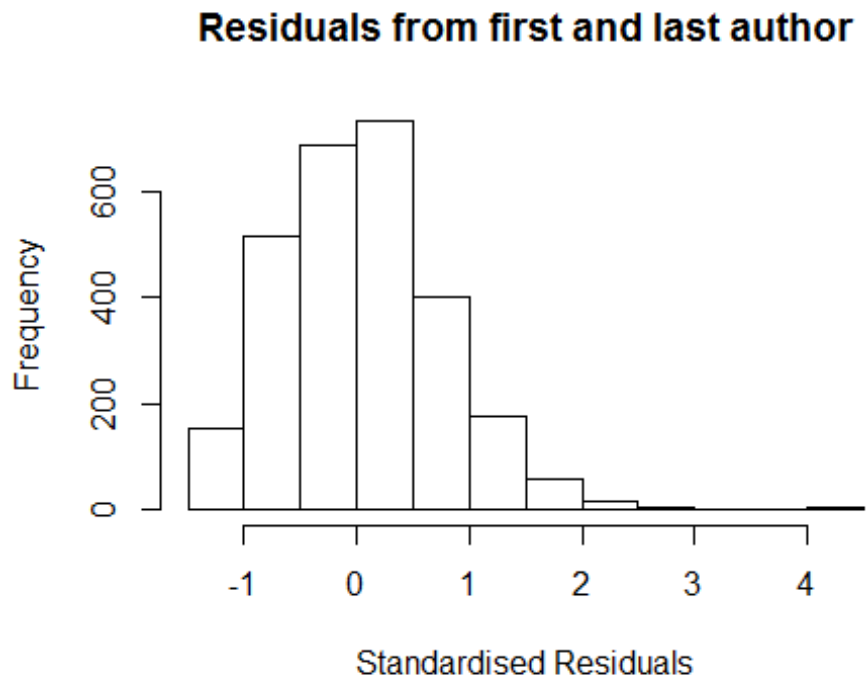
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 839  29644432773 3.510 2002    1202     2    2.538
## 1283 34247665578 5.284 2005    1202     4    4.323
## 1756 62249126614 5.098 2007    1202     3    4.082
## 2920 80052368976 3.559 2011    1202     2    2.698
## 3427 84859871117 3.764 2012    1202     3    2.899
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## 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 -0.4799  0.0119  0.4686  4.3229
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.02500    0.08442   12.14 < 2e-16 ***
## FirstAuthorFemale1 0.10360    0.05742    1.80  0.07131 .
## LastAuthorFemale1 -0.02930    0.05737   -0.51  0.60956
## UniqueAuthors2     0.18325    0.04807    3.81  0.00014 ***
## UniqueAuthors3     0.57879    0.09032    6.41  1.7e-10 ***
## UniqueAuthors4     0.84502    0.24908    3.39  0.00070 ***
```

```

## UniqueAuthors5      0.94513      0.11109      8.51 < 2e-16 ***
## Year1997             0.07221      0.11419      0.63 0.52718
## Year1998            -0.05485      0.10840     -0.51 0.61291
## Year1999             0.03691      0.10224      0.36 0.71814
## Year2000            -0.02693      0.10284     -0.26 0.79343
## Year2001            -0.02610      0.10742     -0.24 0.80804
## Year2002            -0.05254      0.11166     -0.47 0.63800
## Year2003            -0.16850      0.10679     -1.58 0.11473
## Year2004            -0.05668      0.10500     -0.54 0.58937
## Year2005            -0.06389      0.09840     -0.65 0.51623
## Year2006            -0.08563      0.09970     -0.86 0.39052
## Year2007            -0.00874      0.09677     -0.09 0.92801
## Year2008            -0.06264      0.09897     -0.63 0.52687
## Year2009            -0.15028      0.09696     -1.55 0.12126
## Year2010            -0.13770      0.09417     -1.46 0.14380
## Year2011            -0.16391      0.09535     -1.72 0.08570 .
## Year2012            -0.16009      0.10096     -1.59 0.11292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.725
## Multiple R-squared:  0.0364, Adjusted R-squared:  0.0286
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 2 observations c(987,1357) are outliers with |weight| = 0 ( < 3.7e-05);
## 229 weights are ~= 1. The remaining 2507 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0732 0.8680 0.9530 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      3.65e-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 4.727 1          2.174

```

## LastAuthorFemale	4.747	1	2.179
## Year	1.032	16	1.001



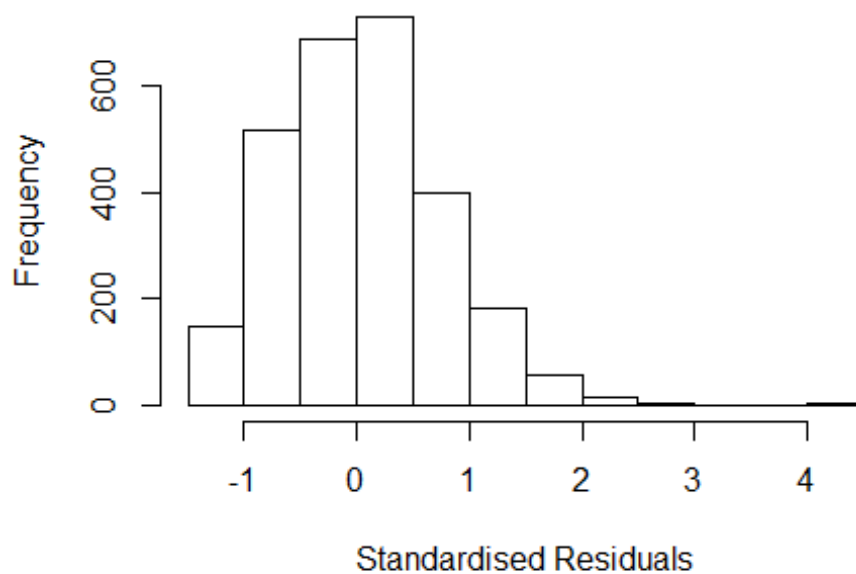
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 839  29644432773 3.510 2002    1202     2    2.521
## 1283 34247665578 5.284 2005    1202     4    4.309
## 1756 62249126614 5.098 2007    1202     3    4.056
## 2920 80052368976 3.559 2011    1202     2    2.666
## 3427 84859871117 3.764 2012    1202     3    2.857
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2086 -0.4842  0.0157  0.4698  4.3087
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03779    0.08441   12.29  <2e-16 ***
## FirstAuthorFemale1 0.13415    0.06625    2.03   0.043 *
## LastAuthorFemale1 -0.04639    0.06638   -0.70   0.485
## Year1997         0.08306    0.11384    0.73   0.466
## Year1998        -0.04788    0.10996   -0.44   0.663
```

```

## Year1999      0.05977    0.10392    0.58    0.565
## Year2000     -0.01322    0.10345   -0.13    0.898
## Year2001     -0.00986    0.10677   -0.09    0.926
## Year2002     -0.04879    0.11252   -0.43    0.665
## Year2003     -0.16837    0.10645   -1.58    0.114
## Year2004     -0.05591    0.10584   -0.53    0.597
## Year2005     -0.06245    0.09837   -0.63    0.526
## Year2006     -0.08311    0.10022   -0.83    0.407
## Year2007      0.00390    0.09716    0.04    0.968
## Year2008     -0.04172    0.09927   -0.42    0.674
## Year2009     -0.11948    0.09726   -1.23    0.219
## Year2010     -0.12362    0.09480   -1.30    0.192
## Year2011     -0.14528    0.09571   -1.52    0.129
## Year2012     -0.13044    0.10091   -1.29    0.196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.735
## Multiple R-squared:  0.0115, Adjusted R-squared:  0.00499
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 2 observations c(987,1357) are outliers with |weight| = 0 ( < 3.7e-05);
## 242 weights are ~ = 1. The remaining 2494 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0976 0.8630 0.9530 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      3.65e-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 1.016 1          1.008
## Year              1.016 16          1.000

```

Residuals from first author



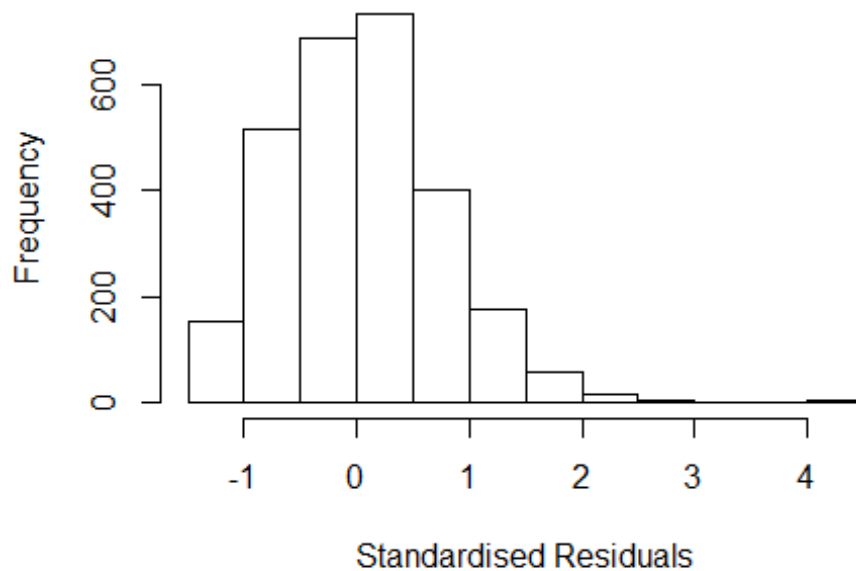
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 839  29644432773 3.510 2002    1202     2    2.521
## 1283 34247665578 5.284 2005    1202     4    4.309
## 1756 62249126614 5.098 2007    1202     3    4.056
## 2920 80052368976 3.559 2011    1202     2    2.666
## 3427 84859871117 3.764 2012    1202     3    2.857
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2132 -0.4818  0.0156  0.4717  4.3107
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03806    0.08440   12.30  <2e-16 ***
## FirstAuthorFemale1 0.09269    0.03077    3.01  0.0026 **
## Year1997         0.08245    0.11394    0.72  0.4693
## Year1998        -0.04950    0.10989   -0.45  0.6525
## Year1999         0.05840    0.10391    0.56  0.5741
## Year2000        -0.01406    0.10346   -0.14  0.8919
## Year2001        -0.01222    0.10670   -0.11  0.9088
## Year2002        -0.05214    0.11230   -0.46  0.6425
```

```

## Year2003          -0.16932      0.10644     -1.59      0.1118
## Year2004          -0.05816      0.10568     -0.55      0.5821
## Year2005          -0.06471      0.09828     -0.66      0.5103
## Year2006          -0.08437      0.10022     -0.84      0.3999
## Year2007           0.00232      0.09712       0.02      0.9810
## Year2008          -0.04372      0.09930     -0.44      0.6598
## Year2009          -0.12182      0.09709     -1.25      0.2097
## Year2010          -0.12576      0.09476     -1.33      0.1846
## Year2011          -0.14732      0.09564     -1.54      0.1236
## Year2012          -0.13200      0.10084     -1.31      0.1906
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.735
## Multiple R-squared:  0.0114, Adjusted R-squared:  0.00521
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 2 observations c(987,1357) are outliers with |weight| = 0 ( < 3.7e-05);
## 241 weights are ~= 1. The remaining 2495 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0968 0.8640 0.9530 0.9130 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          3.65e-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



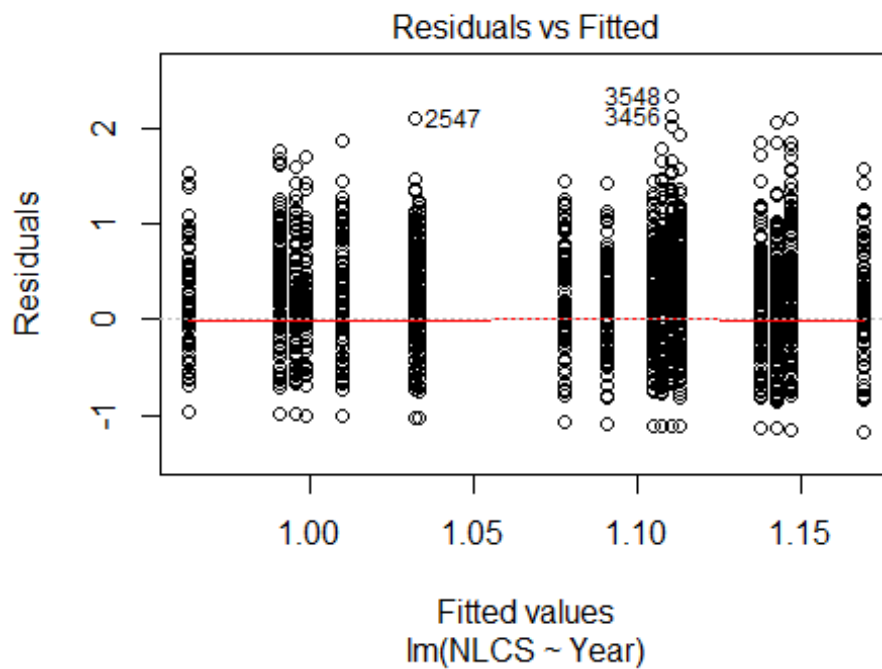
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 839  29644432773 3.510 2002    1202     2    2.521
## 1283 34247665578 5.284 2005    1202     4    4.309
## 1756 62249126614 5.098 2007    1202     3    4.056
## 2920 80052368976 3.559 2011    1202     2    2.666
## 3427 84859871117 3.764 2012    1202     3    2.857
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2023 -0.4815  0.0154  0.4674  4.3041
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.04403    0.08449   12.36  <2e-16 ***
## LastAuthorFemale1 0.07375    0.03065    2.41   0.016 *
## Year1997         0.08455    0.11423    0.74   0.459
## Year1998        -0.05186    0.11004   -0.47   0.637
## Year1999         0.05783    0.10396    0.56   0.578
## Year2000        -0.01551    0.10365   -0.15   0.881
## Year2001        -0.01266    0.10677   -0.12   0.906
## Year2002        -0.05435    0.11250   -0.48   0.629
```

```

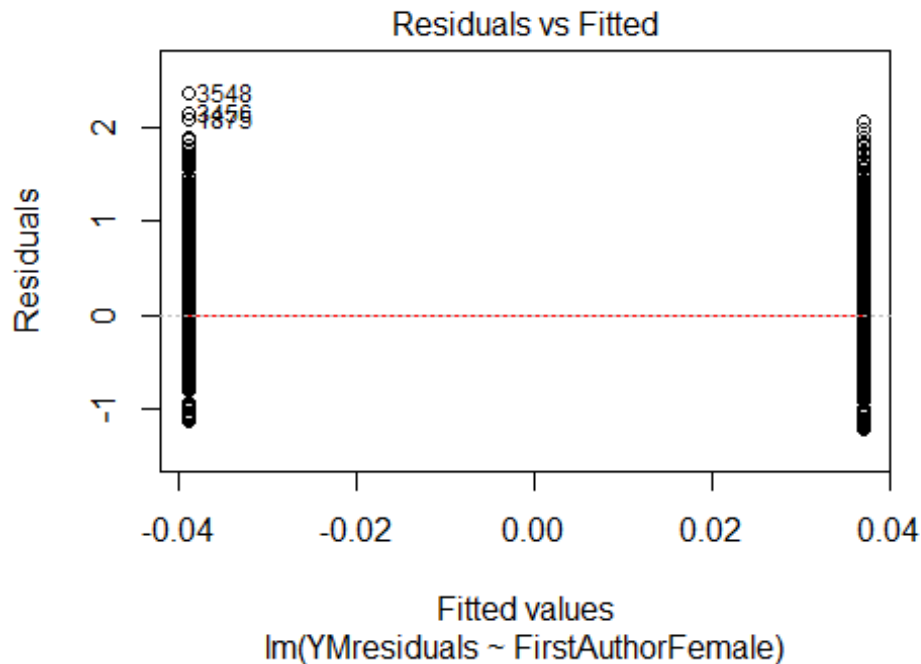
## Year2003      -0.16755      0.10662      -1.57      0.116
## Year2004      -0.05967      0.10567      -0.56      0.572
## Year2005      -0.06415      0.09837      -0.65      0.514
## Year2006      -0.08256      0.10047      -0.82      0.411
## Year2007       0.00477      0.09737       0.05      0.961
## Year2008      -0.04449      0.09944      -0.45      0.655
## Year2009      -0.12333      0.09718      -1.27      0.205
## Year2010      -0.12555      0.09494      -1.32      0.186
## Year2011      -0.14697      0.09589      -1.53      0.125
## Year2012      -0.13043      0.10172      -1.28      0.200
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.737
## Multiple R-squared:  0.0101, Adjusted R-squared:  0.00389
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 2 observations c(987,1357) are outliers with |weight| = 0 ( < 3.7e-05);
## 244 weights are ~ = 1. The remaining 2492 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.101  0.863   0.953   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      3.65e-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 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
## 115 135 120 129 133 144 169 147 156 190 196 203 240 252 210
## 2011 2012

```

```
## 271 265
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 95 108 99 103 102 107 149 124 132 156 168 184 209 213 171
## 2011 2012
## 212 214
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 93 105 95 98 95 103 144 118 122 150 160 178 198 199 159
## 2011 2012
## 205 200
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 33, df = 16, p-value = 0.007
```

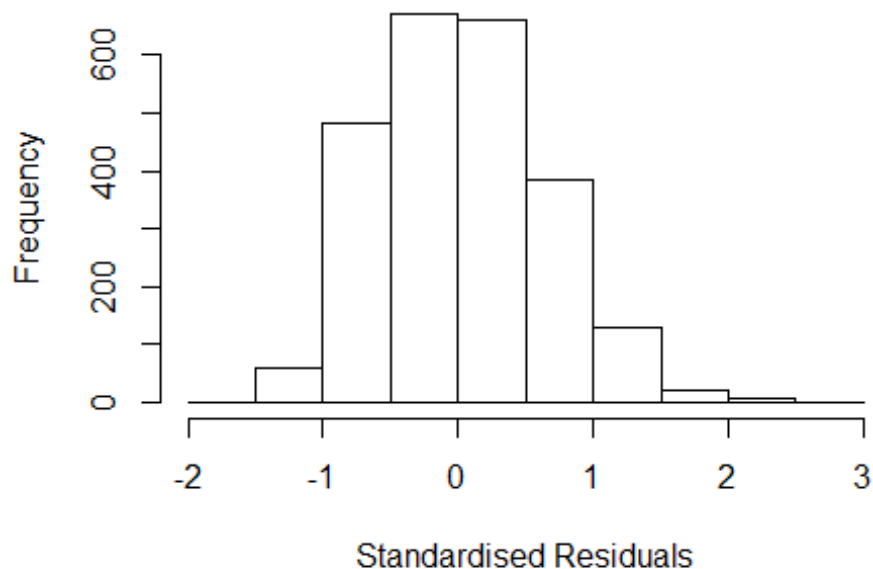


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.9, df = 1, p-value = 0.005
```



```
## [1] "Female first author team size 2018 geometric mean: 1.62241832667836"
## [1] "Male first author team size 2018 geometric mean: 1.6166256979647"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4200, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.68232702500508"
## [1] "Male last author team size 2018 geometric mean: 1.54177544339418"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4500, 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 1.786 1      1.336
## LastAuthorFemale  1.819 1      1.349
## UniqueAuthors    1.214 4      1.025
## Year              1.203 16     1.006
```

Residuals from first and last author and team size



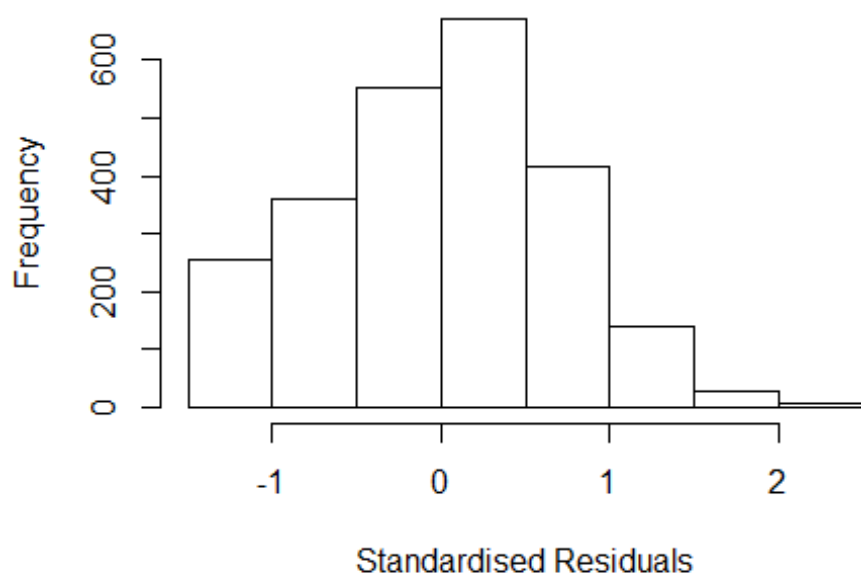
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3548 84856557156 3.439 2012    1203      3      2.63
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.51591 -0.44380 -0.00146  0.45163  2.62982
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8426    0.0697   12.09 < 2e-16 ***
## FirstAuthorFemale1 -0.0185    0.0359   -0.52  0.606
## LastAuthorFemale1  0.0674    0.0360    1.87  0.061 .
## UniqueAuthors2    0.4692    0.0326   14.37 < 2e-16 ***
## UniqueAuthors3    0.5053    0.0439   11.51 < 2e-16 ***
## UniqueAuthors4    0.5784    0.0600    9.64 < 2e-16 ***
## UniqueAuthors5    0.5260    0.0854    6.16 8.4e-10 ***
## Year1997         0.1138    0.0914    1.25  0.213
## Year1998         0.0257    0.0979    0.26  0.793
## Year1999        -0.0816    0.0937   -0.87  0.384
```

```

## Year2000          -0.0175      0.0867   -0.20      0.840
## Year2001          -0.0186      0.0930   -0.20      0.841
## Year2002          -0.0915      0.0907   -1.01      0.314
## Year2003          -0.0420      0.0863   -0.49      0.627
## Year2004          -0.0226      0.0933   -0.24      0.809
## Year2005           0.0767      0.0868     0.88      0.377
## Year2006           0.0460      0.0840     0.55      0.584
## Year2007           0.0184      0.0828     0.22      0.825
## Year2008           0.0181      0.0832     0.22      0.827
## Year2009          -0.0535      0.0819   -0.65      0.513
## Year2010          -0.0238      0.0815   -0.29      0.770
## Year2011           0.0101      0.0867     0.12      0.907
## Year2012          -0.0334      0.0889   -0.38      0.707
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.672
## Multiple R-squared:  0.127, Adjusted R-squared:  0.119
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 222 weights are ~= 1. The remaining 2200 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0914 0.8690 0.9520 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      4.13e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.929 1      1.389
## LastAuthorFemale 1.930 1      1.389
## Year              1.040 16      1.001

```

Residuals from first and last author



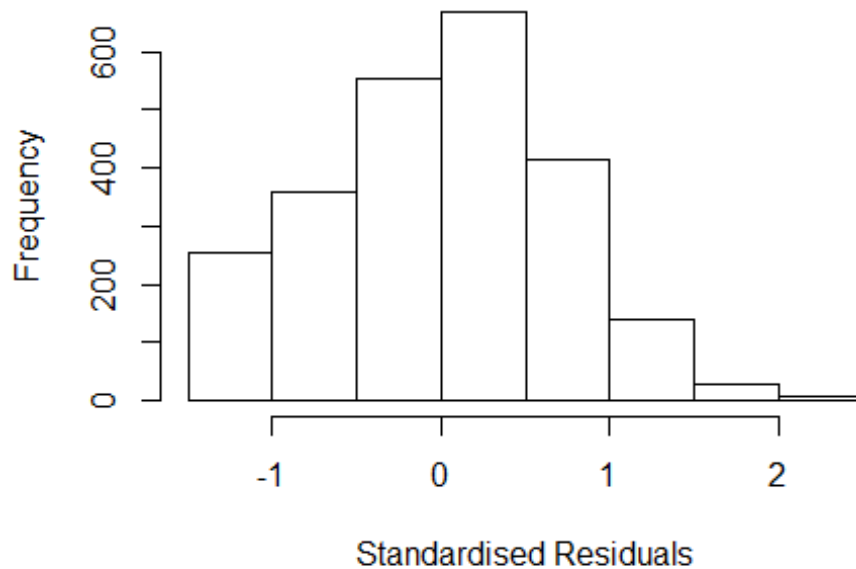
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2247 -0.5067 0.0435 0.4877 2.4303
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9553 0.0712 13.42 <2e-16 ***
## FirstAuthorFemale1 0.0744 0.0405 1.84 0.066 .
## LastAuthorFemale1 0.0136 0.0403 0.34 0.737
## Year1997 0.1814 0.0965 1.88 0.060 .
## Year1998 0.0418 0.1030 0.41 0.685
## Year1999 -0.0729 0.0968 -0.75 0.451
## Year2000 0.0850 0.0921 0.92 0.356
## Year2001 0.0222 0.0969 0.23 0.819
## Year2002 -0.0483 0.0950 -0.51 0.612
## Year2003 -0.0201 0.0937 -0.21 0.830
## Year2004 -0.0278 0.0957 -0.29 0.772
## Year2005 0.1028 0.0911 1.13 0.259
```

```

## Year2006          0.1332      0.0876      1.52      0.128
## Year2007          0.0691      0.0858      0.81      0.421
## Year2008          0.1056      0.0860      1.23      0.220
## Year2009          0.0213      0.0845      0.25      0.801
## Year2010          0.0922      0.0852      1.08      0.279
## Year2011          0.1210      0.0905      1.34      0.181
## Year2012          0.0534      0.0950      0.56      0.574
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.725
## Multiple R-squared:  0.0117, Adjusted R-squared:  0.00433
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 196 weights are ~= 1. The remaining 2226 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.239  0.860  0.950  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      4.13e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.011 1      1.006
## Year              1.011 16      1.000

```


Residuals from first author



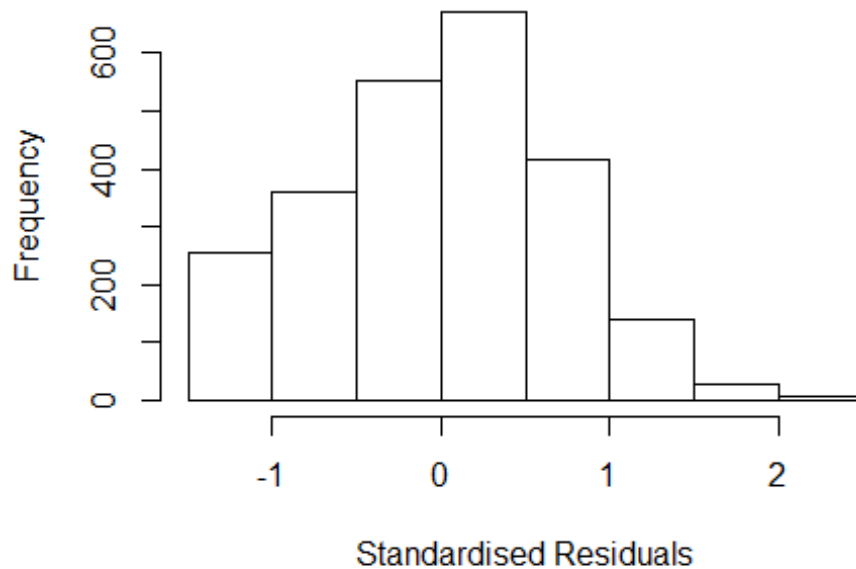
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2226 -0.5106 0.0437 0.4872 2.4289
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9569 0.0709 13.50 <2e-16 ***
## FirstAuthorFemale1 0.0840 0.0293 2.87 0.0042 **
## Year1997 0.1818 0.0966 1.88 0.0601 .
## Year1998 0.0423 0.1031 0.41 0.6815
## Year1999 -0.0728 0.0968 -0.75 0.4520
## Year2000 0.0845 0.0920 0.92 0.3588
## Year2001 0.0227 0.0970 0.23 0.8152
## Year2002 -0.0485 0.0950 -0.51 0.6100
## Year2003 -0.0196 0.0937 -0.21 0.8339
## Year2004 -0.0279 0.0957 -0.29 0.7707
## Year2005 0.1024 0.0910 1.13 0.2605
## Year2006 0.1330 0.0875 1.52 0.1289
```

```

## Year2007          0.0689      0.0858      0.80      0.4225
## Year2008          0.1055      0.0859      1.23      0.2199
## Year2009          0.0217      0.0845      0.26      0.7973
## Year2010          0.0920      0.0852      1.08      0.2798
## Year2011          0.1207      0.0905      1.33      0.1822
## Year2012          0.0532      0.0950      0.56      0.5757
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.725
## Multiple R-squared:  0.0117, Adjusted R-squared:  0.00469
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 197 weights are ~= 1. The remaining 2225 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.239  0.859  0.950  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      4.13e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.011 1      1.006
## Year      1.011 16      1.000

```

Residuals from last author



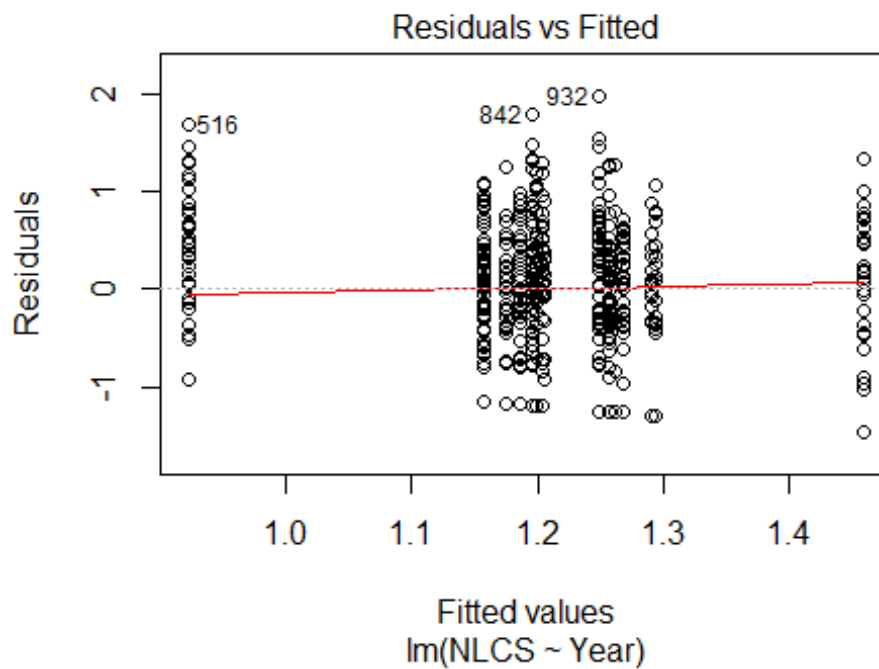
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2119 -0.5108 0.0403 0.4834 2.4188
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9665 0.0709 13.63 <2e-16 ***
## LastAuthorFemale1 0.0656 0.0292 2.25 0.025 *
## Year1997 0.1798 0.0962 1.87 0.062 .
## Year1998 0.0422 0.1028 0.41 0.681
## Year1999 -0.0722 0.0966 -0.75 0.455
## Year2000 0.0859 0.0923 0.93 0.352
## Year2001 0.0243 0.0969 0.25 0.802
## Year2002 -0.0412 0.0952 -0.43 0.665
## Year2003 -0.0231 0.0934 -0.25 0.805
## Year2004 -0.0238 0.0955 -0.25 0.803
## Year2005 0.1076 0.0912 1.18 0.238
## Year2006 0.1357 0.0874 1.55 0.121
```

```

## Year2007          0.0735      0.0859      0.86      0.392
## Year2008          0.1088      0.0858      1.27      0.205
## Year2009          0.0218      0.0847      0.26      0.797
## Year2010          0.0955      0.0850      1.12      0.261
## Year2011          0.1258      0.0905      1.39      0.164
## Year2012          0.0537      0.0950      0.57      0.572
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.725
## Multiple R-squared:  0.0104, Adjusted R-squared:  0.00339
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 206 weights are ~= 1. The remaining 2216 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.243  0.857  0.949  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      4.13e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2422"
## [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
##   13   14   25   33   23   23   37   39   43   37   42   67   58   50   50
## 2011 2012
##   58   59
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   10    9   21   15   17   12   32   30   31   31   32   58   47   39   41
## 2011 2012

```

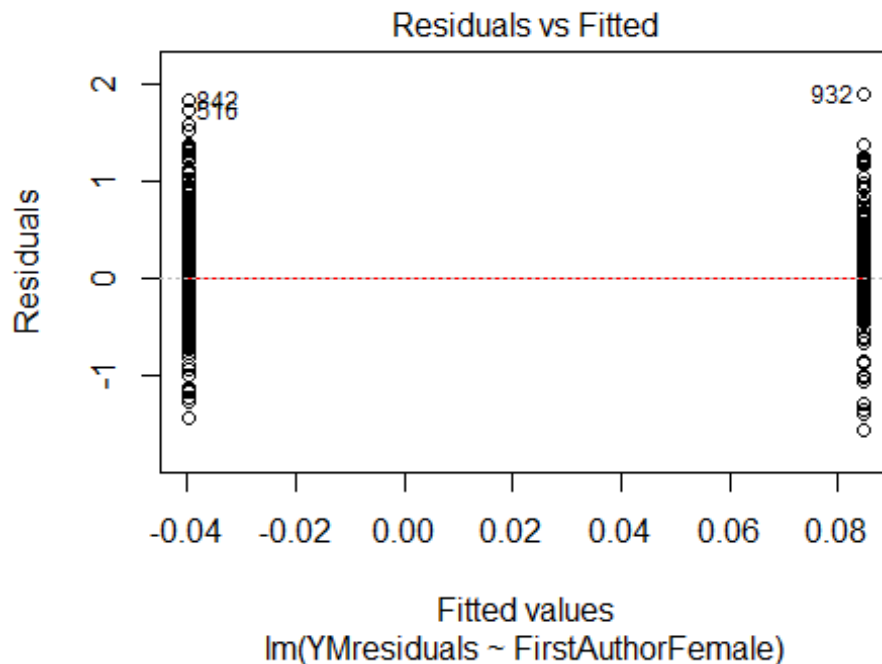
```
## 46 54
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 9 9 17 13 16 11 31 25 29 29 28 57 47 35 41
## 2011 2012
## 45 52
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 41, df = 16, p-value = 5e-04
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.19, df = 1, p-value = 0.7
## [1] "Female first author team size 2018 geometric mean: 2.08723661325224"
## [1] "Male first author team size 2018 geometric mean: 1.5028244828102"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

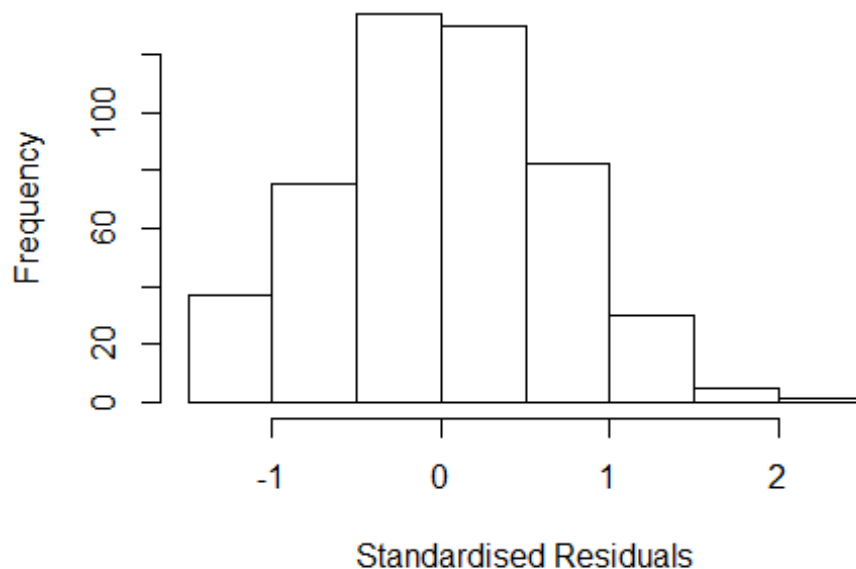
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 600, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.8937961742363"
## [1] "Male last author team size 2018 geometric mean: 1.62329836228009"

## Warning in wilcox.test.default(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 = 0.4
## 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  1.263  1      1.124
## UniqueAuthors    1.988  4      1.090
## Year              2.301 16      1.026
```

Residuals from first and last author and team size



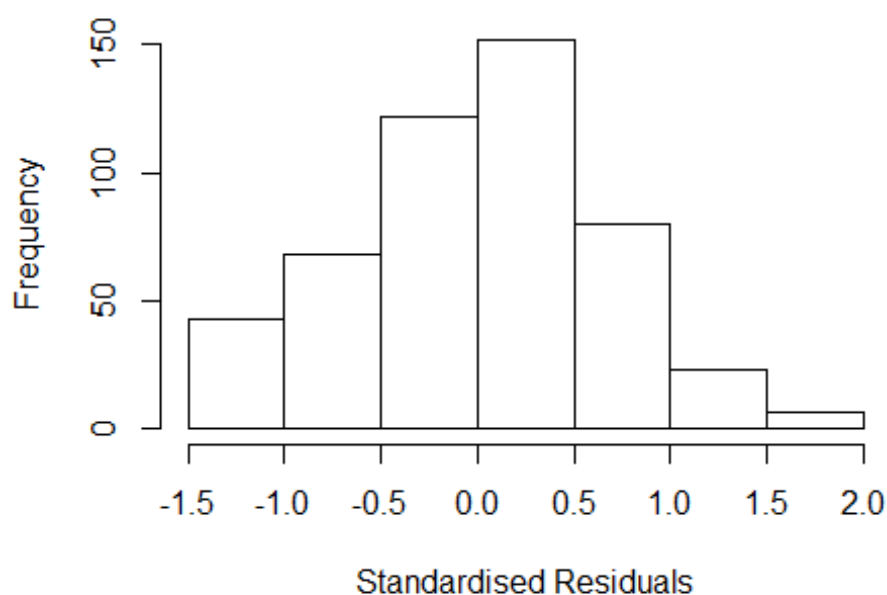
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34364 -0.44661 0.00184 0.46583 2.05753
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18255 0.25818 4.58 6.0e-06 ***
## FirstAuthorFemale1 0.17000 0.06643 2.56 0.01081 *
## LastAuthorFemale1 -0.08728 0.07318 -1.19 0.23360
## UniqueAuthors2 0.11912 0.08084 1.47 0.14128
## UniqueAuthors3 0.44387 0.08704 5.10 4.9e-07 ***
## UniqueAuthors4 0.51442 0.11284 4.56 6.6e-06 ***
## UniqueAuthors5 0.33646 0.10107 3.33 0.00094 ***
## Year1997 -0.02420 0.28385 -0.09 0.93210
## Year1998 -0.15826 0.29370 -0.54 0.59026
## Year1999 -0.06701 0.28777 -0.23 0.81598
```

```

## Year2000      0.04197      0.30863      0.14      0.89190
## Year2001     -0.15985      0.30335     -0.53      0.59848
## Year2002     -0.13599      0.28779     -0.47      0.63676
## Year2003     -0.12144      0.29472     -0.41      0.68048
## Year2004      0.04562      0.29208      0.16      0.87595
## Year2005     -0.10560      0.26774     -0.39      0.69347
## Year2006      0.00164      0.27831      0.01      0.99530
## Year2007     -0.39235      0.28354     -1.38      0.16710
## Year2008     -0.16813      0.27027     -0.62      0.53419
## Year2009     -0.25884      0.27667     -0.94      0.34999
## Year2010     -0.10707      0.26722     -0.40      0.68883
## Year2011     -0.26109      0.28112     -0.93      0.35350
## Year2012     -0.15651      0.28243     -0.55      0.57972
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.667
## Multiple R-squared:  0.111, Adjusted R-squared:  0.069
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 458 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.321  0.876  0.949  0.913  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      2.02e-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 1.369 1      1.170
## LastAuthorFemale  1.333 1      1.154
## Year              1.283 16      1.008

```


Residuals from first and last author



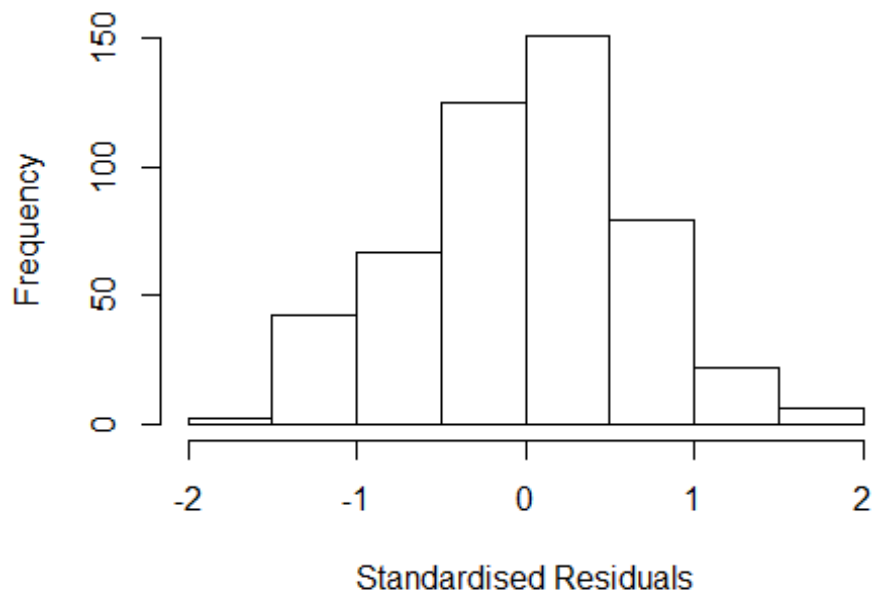
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4770 -0.4156 0.0418 0.4637 1.8877
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23362 0.25204 4.89 1.4e-06 ***
## FirstAuthorFemale1 0.24212 0.07179 3.37 0.00081 ***
## LastAuthorFemale1 -0.16708 0.07987 -2.09 0.03698 *
## Year1997 0.01503 0.26681 0.06 0.95510
## Year1998 -0.13783 0.29279 -0.47 0.63803
## Year1999 0.02983 0.28401 0.11 0.91640
## Year2000 0.11512 0.30709 0.37 0.70792
## Year2001 -0.08132 0.28340 -0.29 0.77428
## Year2002 -0.07589 0.28904 -0.26 0.79300
## Year2003 -0.09995 0.29239 -0.34 0.73262
## Year2004 0.16838 0.29187 0.58 0.56426
## Year2005 -0.08104 0.26429 -0.31 0.75928
```

```

## Year2006          0.03790    0.27152    0.14  0.88904
## Year2007          -0.39523    0.28445   -1.39  0.16534
## Year2008          -0.09529    0.26744   -0.36  0.72178
## Year2009          -0.07423    0.27477   -0.27  0.78715
## Year2010          -0.00769    0.26369   -0.03  0.97673
## Year2011          -0.14236    0.28023   -0.51  0.61168
## Year2012          -0.07553    0.28004   -0.27  0.78749
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.664
## Multiple R-squared:  0.0529, Adjusted R-squared:  0.017
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 444 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.400  0.860  0.950  0.904  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.02e-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 1.127 1      1.062
## Year              1.127 16      1.004

```

Residuals from first author

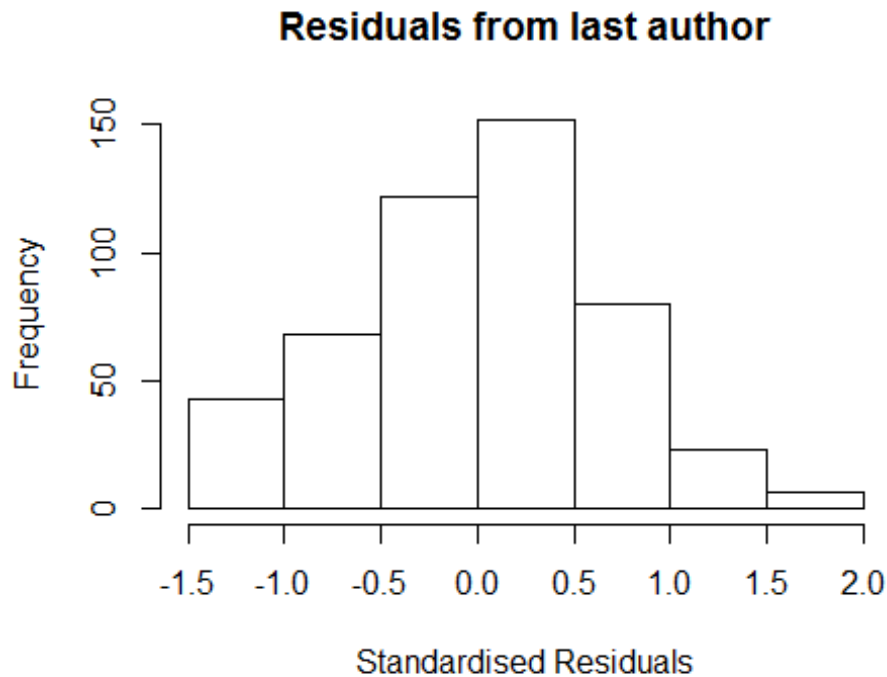


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5345 -0.4390  0.0378  0.4454  1.9140
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2015     0.2599   4.62 4.9e-06 ***
## FirstAuthorFemale1  0.1512     0.0673   2.25  0.025 *
## Year1997          0.0385     0.2760   0.14  0.889
## Year1998         -0.1080     0.2999  -0.36  0.719
## Year1999          0.0346     0.2976   0.12  0.907
## Year2000          0.1362     0.3172   0.43  0.668
## Year2001         -0.0668     0.2975  -0.22  0.822
## Year2002         -0.0678     0.2994  -0.23  0.821
## Year2003         -0.0830     0.3034  -0.27  0.785
## Year2004          0.1817     0.3026   0.60  0.548
## Year2005         -0.0704     0.2734  -0.26  0.797
## Year2006          0.0537     0.2801   0.19  0.848
```

```

## Year2007          -0.3805      0.2915   -1.31    0.192
## Year2008          -0.0699      0.2751   -0.25    0.800
## Year2009          -0.0503      0.2814   -0.18    0.858
## Year2010          -0.0144      0.2739   -0.05    0.958
## Year2011          -0.1122      0.2878   -0.39    0.697
## Year2012          -0.0448      0.2885   -0.16    0.877
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.663
## Multiple R-squared:  0.0472, Adjusted R-squared:  0.0132
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 454 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.384  0.865   0.952   0.905   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.02e-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 1.094 1          1.046
## Year              1.094 16          1.003

```



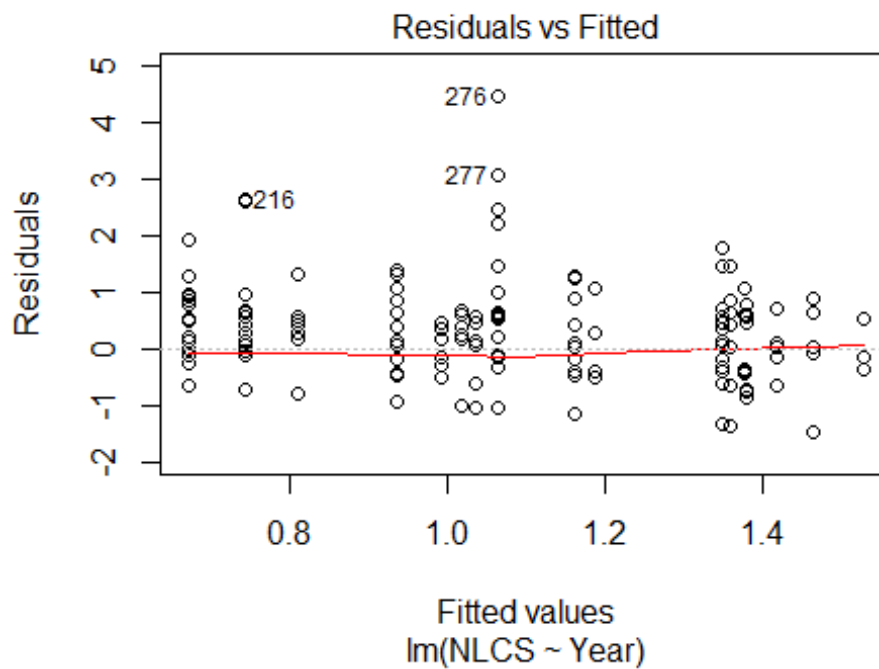
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4473 -0.4307  0.0438  0.4410  2.0217
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2342     0.2738   4.51 8.3e-06 ***
## LastAuthorFemale1 -0.0157     0.0737  -0.21   0.83
## Year1997          0.0249     0.2865   0.09   0.93
## Year1998         -0.1043     0.3123  -0.33   0.74
## Year1999          0.0860     0.3017   0.28   0.78
## Year2000          0.1291     0.3221   0.40   0.69
## Year2001         -0.0490     0.3085  -0.16   0.87
## Year2002         -0.0342     0.3074  -0.11   0.91
## Year2003         -0.0720     0.3105  -0.23   0.82
## Year2004          0.2131     0.3128   0.68   0.50
## Year2005         -0.0482     0.2855  -0.17   0.87
## Year2006          0.0697     0.2898   0.24   0.81
```

```

## Year2007          -0.3566      0.3054    -1.17      0.24
## Year2008          -0.0591      0.2865    -0.21      0.84
## Year2009          -0.0235      0.2917    -0.08      0.94
## Year2010           0.0201      0.2831     0.07      0.94
## Year2011          -0.0869      0.2966    -0.29      0.77
## Year2012          -0.0339      0.2986    -0.11      0.91
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.662
## Multiple R-squared:  0.0364, Adjusted R-squared:  0.002
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 447 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.331  0.849  0.949  0.903  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 494"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
##    3    8    7    9    7   11   13    9   10   19   24   45   21    8   10
## 2011 2012
##   41   39
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    7    5    7    5    5    8    6    7   13   17   38   19    6   10
## 2011 2012

```

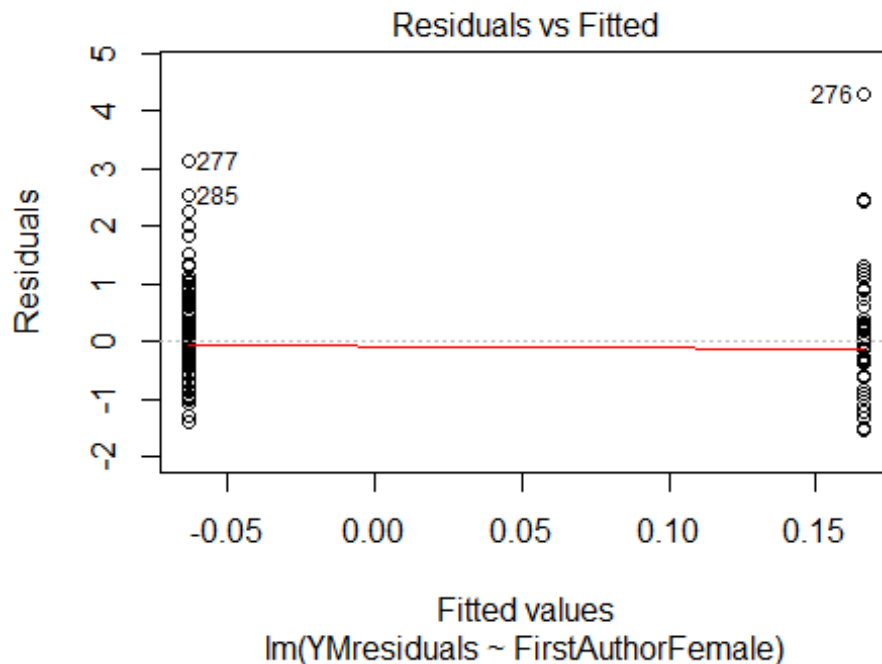
```
## 33 32
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3 7 5 7 5 5 8 6 7 13 17 38 19 6 10
## 2011 2012
## 33 32
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 31, df = 16, p-value = 0.01
```



```
##
## 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"
## [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 = 4.5, 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"

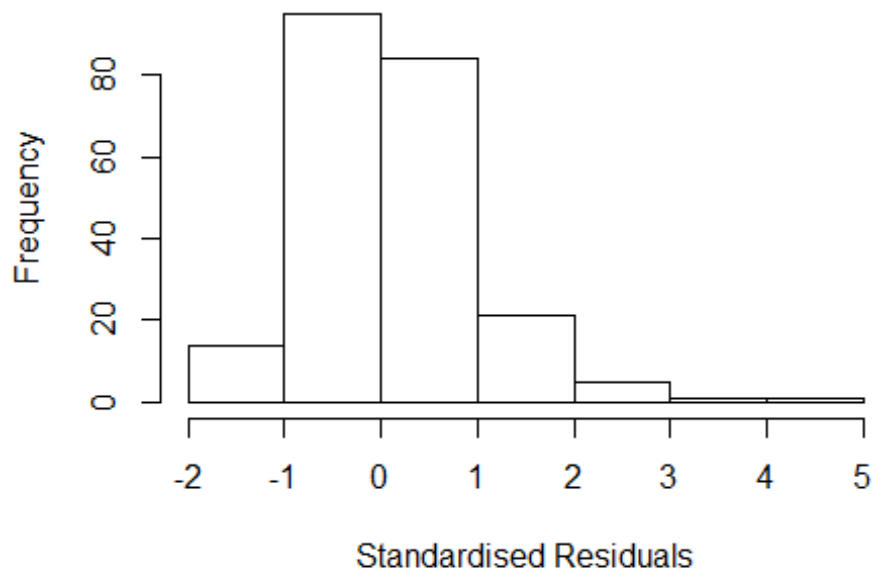
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4.5, 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	33.730	1	5.808
LastAuthorFemale	33.548	1	5.792
UniqueAuthors	2.775	2	1.291
Year	3.530	16	1.040

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
## 216 80052710834 3.371 2011    1205     1    2.661
## 241 84856751841 3.319 2011    1200     2    2.609
## 276 84871046337 5.507 2012    1205     1    4.833
## 277 84871056729 4.115 2012    1205     1    3.568
## 285 84872181712 3.529 2012    1200     2    2.982
## 286 84874030859 3.264 2012    1202     4    2.717
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5480 -0.5832  0.0107  0.5802  4.8332
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.4788    0.2504    5.91 1.5e-08 ***
## FirstAuthorFemale1  0.2501    0.7238    0.35  0.7300
## LastAuthorFemale1 -0.1231    0.7190   -0.17  0.8642
## UniqueAuthors2     0.6273    0.5478    1.15  0.2535
## UniqueAuthors3    -0.1940    0.2507   -0.77  0.4400
```

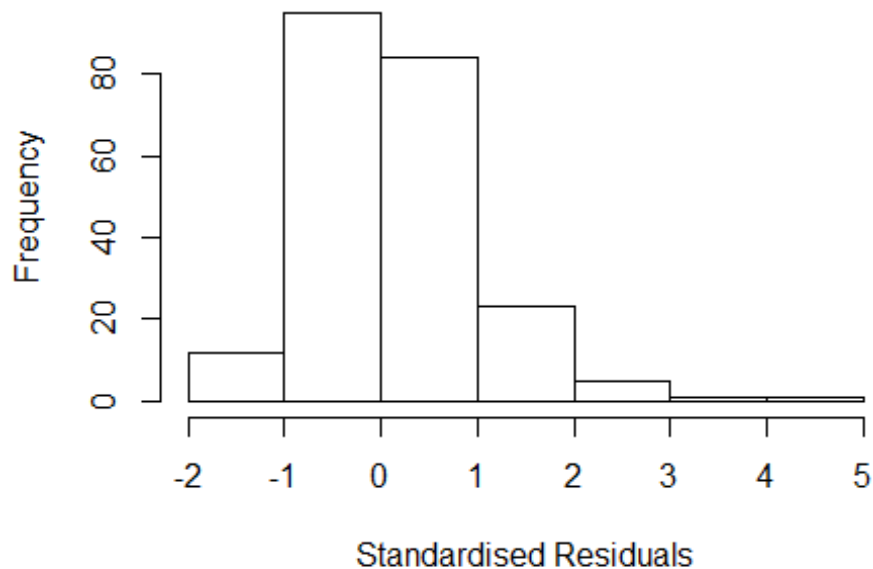
```

## Year1997          -0.4596      0.3236    -1.42    0.1572
## Year1998          -0.3524      0.3643    -0.97    0.3345
## Year1999          -0.4209      0.3686    -1.14    0.2549
## Year2000           0.0692      0.4556     0.15    0.8794
## Year2001          -0.1170      0.3175    -0.37    0.7128
## Year2002          -0.1185      0.4974    -0.24    0.8119
## Year2003          -0.1689      0.3513    -0.48    0.6312
## Year2004          -0.1992      0.4424    -0.45    0.6530
## Year2005          -0.3724      0.3381    -1.10    0.2720
## Year2006          -0.1957      0.3144    -0.62    0.5344
## Year2007          -0.8888      0.2853    -3.12    0.0021 **
## Year2008          -0.6009      0.3177    -1.89    0.0600 .
## Year2009          -0.6158      0.2968    -2.07    0.0393 *
## Year2010          -0.7425      0.3660    -2.03    0.0438 *
## Year2011          -0.8956      0.2796    -3.20    0.0016 **
## Year2012          -0.9320      0.3214    -2.90    0.0041 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.769
## Multiple R-squared:  0.167, Adjusted R-squared:  0.0841
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 2 observations c(208,209)
## are outliers with |weight| <= 0.00041 ( < 0.00045);
## 8 weights are ~= 1. The remaining 211 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0996 0.9010 0.9480 0.9070 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          4.52e-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 15.762 1          3.970

```

```
## LastAuthorFemale 15.917 1          3.990
## Year             1.552 16          1.014
```

Residuals from first and last author



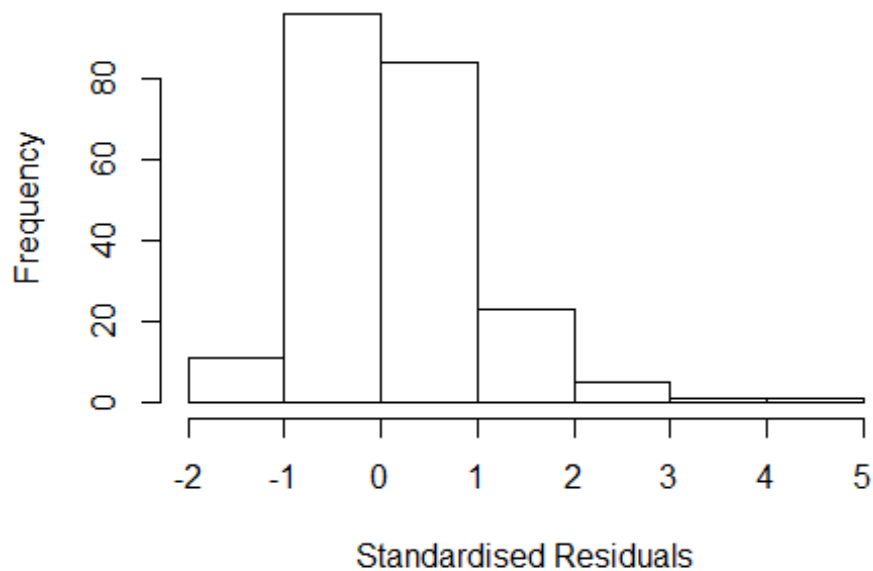
```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 216 80052710834 3.371 2011    1205     1    2.676
## 241 84856751841 3.319 2011    1200     2    2.624
## 276 84871046337 5.507 2012    1205     1    4.832
## 277 84871056729 4.115 2012    1205     1    3.559
## 285 84872181712 3.529 2012    1200     2    2.973
## 286 84874030859 3.264 2012    1202     4    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
## -1.5561 -0.5760  0.0246  0.5920  4.8318
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.4810     0.2502    5.92 1.4e-08 ***
## FirstAuthorFemale1  0.3444     0.4966    0.69  0.4887
## LastAuthorFemale1 -0.2256     0.4987   -0.45  0.6515
## Year1997        -0.4574     0.3242   -1.41  0.1597
```

```

## Year1998          -0.3559      0.3649   -0.98    0.3307
## Year1999          -0.4194      0.3721   -1.13    0.2610
## Year2000           0.0751      0.4554    0.16    0.8692
## Year2001          -0.1162      0.3172   -0.37    0.7145
## Year2002          -0.1128      0.5065   -0.22    0.8239
## Year2003          -0.1711      0.3525   -0.49    0.6280
## Year2004          -0.1614      0.4028   -0.40    0.6891
## Year2005          -0.3740      0.3390   -1.10    0.2712
## Year2006          -0.1768      0.3254   -0.54    0.5875
## Year2007          -0.9023      0.2855   -3.16    0.0018 **
## Year2008          -0.6195      0.3115   -1.99    0.0481 *
## Year2009          -0.5085      0.2864   -1.78    0.0773 .
## Year2010          -0.7759      0.3570   -2.17    0.0309 *
## Year2011          -0.9051      0.2794   -3.24    0.0014 **
## Year2012          -0.9247      0.3246   -2.85    0.0048 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.741
## Multiple R-squared:  0.162, Adjusted R-squared:  0.0877
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(208,209) are outliers with |weight| = 0 ( < 0.00045);
## 11 weights are ~= 1. The remaining 208 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0708 0.8900 0.9450 0.8980 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          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 1.24 1          1.114
## Year              1.24 16          1.007

```

Residuals from first author



```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 216 80052710834 3.371 2011    1205     1    2.676
## 241 84856751841 3.319 2011    1200     2    2.624
## 276 84871046337 5.507 2012    1205     1    4.832
## 277 84871056729 4.115 2012    1205     1    3.559
## 285 84872181712 3.529 2012    1200     2    2.973
## 286 84874030859 3.264 2012    1202     4    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
## -1.5570 -0.5738  0.0322  0.5942  4.8265
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.4779     0.2512    5.88 1.6e-08 ***
## FirstAuthorFemale1  0.1275     0.1386    0.92  0.3584
## Year1997       -0.4569     0.3250   -1.41  0.1612
## Year1998       -0.3551     0.3651   -0.97  0.3319
## Year1999       -0.4159     0.3731   -1.11  0.2663
## Year2000        0.0791     0.4564    0.17  0.8626
## Year2001       -0.1170     0.3185   -0.37  0.7139
```

```

## Year2002          -0.1114      0.5079   -0.22   0.8266
## Year2003          -0.1707      0.3531   -0.48   0.6293
## Year2004          -0.1363      0.3919   -0.35   0.7284
## Year2005          -0.3733      0.3402   -1.10   0.2737
## Year2006          -0.1760      0.3261   -0.54   0.5900
## Year2007          -0.9014      0.2862   -3.15   0.0019 **
## Year2008          -0.6327      0.3143   -2.01   0.0454 *
## Year2009          -0.5069      0.2877   -1.76   0.0796 .
## Year2010          -0.7504      0.3504   -2.14   0.0334 *
## Year2011          -0.9041      0.2801   -3.23   0.0015 **
## Year2012          -0.9250      0.3252   -2.84   0.0049 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.738
## Multiple R-squared:  0.162, Adjusted R-squared:  0.0917
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(208,209) are outliers with |weight| = 0 ( < 0.00045);
## 11 weights are ~= 1. The remaining 208 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.067  0.884   0.945   0.897   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          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 1.22  1           1.105
## Year             1.22 16           1.006
##
## [1] "List of 6 outliers with residuals above 2.5"
##           ScopusId NLCS Year OneField Fields residuals
## 216 80052710834 3.371 2011      1205      1      2.676
## 241 84856751841 3.319 2011      1200      2      2.624
## 276 84871046337 5.507 2012      1205      1      4.832
## 277 84871056729 4.115 2012      1205      1      3.559

```

```

## 285 84872181712 3.529 2012      1200      2      2.973
## 286 84874030859 3.264 2012      1202      4      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
## -1.5565 -0.5774  0.0472  0.5852  4.8421
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4852     0.2487   5.97   1e-08 ***
## LastAuthorFemale1  0.1069     0.1387   0.77   0.4419
## Year1997         -0.4577     0.3234  -1.42   0.1585
## Year1998         -0.3574     0.3647  -0.98   0.3282
## Year1999         -0.4234     0.3712  -1.14   0.2553
## Year2000          0.0712     0.4548   0.16   0.8757
## Year2001         -0.1152     0.3156  -0.37   0.7154
## Year2002         -0.1136     0.5066  -0.22   0.8227
## Year2003         -0.1720     0.3520  -0.49   0.6256
## Year2004         -0.1204     0.3817  -0.32   0.7528
## Year2005         -0.3751     0.3375  -1.11   0.2677
## Year2006         -0.1780     0.3244  -0.55   0.5839
## Year2007         -0.9045     0.2844  -3.18   0.0017 **
## Year2008         -0.6344     0.3134  -2.02   0.0443 *
## Year2009         -0.5107     0.2847  -1.79   0.0744 .
## Year2010         -0.7338     0.3521  -2.08   0.0384 *
## Year2011         -0.9078     0.2777  -3.27   0.0013 **
## Year2012         -0.9272     0.3229  -2.87   0.0045 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.74
## Multiple R-squared:  0.16, Adjusted R-squared:  0.0902
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(208,209) are outliers with |weight| = 0 ( < 0.00045);
## 10 weights are ~ 1. The remaining 209 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0701 0.8830 0.9450 0.8980 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      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 1206"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    5    5    8    2    2    6    3    3    4    6    5   13    6   10
## 2011 2012
##   14    5
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2    5    3    7    1    2    5    2    3    3    5    4   12    4    8
## 2011 2012
##    8    4
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2    5    3    7    1    2    5    2    3    3    5    4   12    4    8
## 2011 2012
##    8    4
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1.16993081275869"
## [1] "Male first author team size 2018 geometric mean: 1.18920711500272"

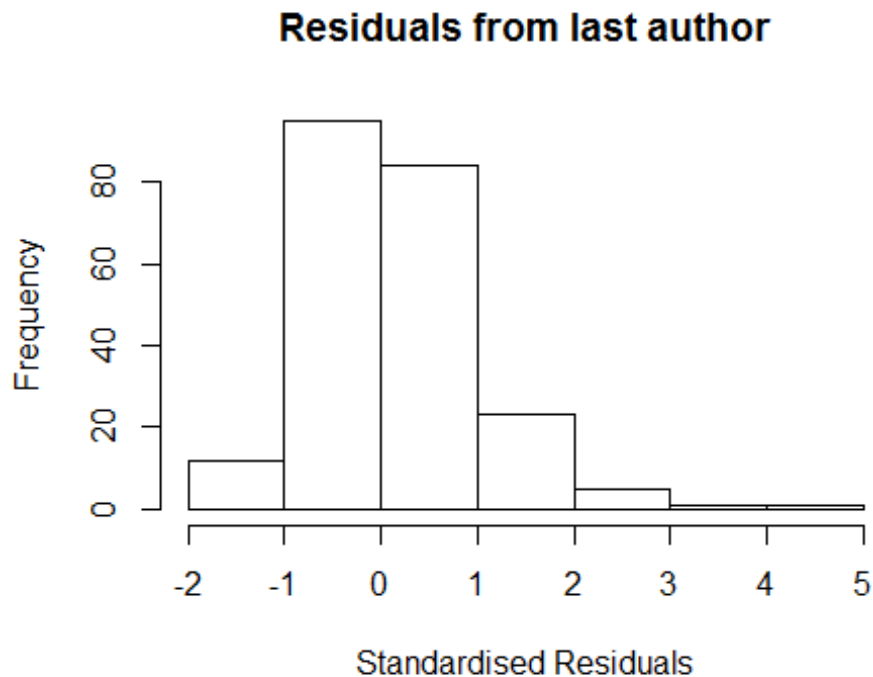
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 13, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.10408951367381"
## [1] "Male last author team size 2018 geometric mean: 1.31607401295249"

```



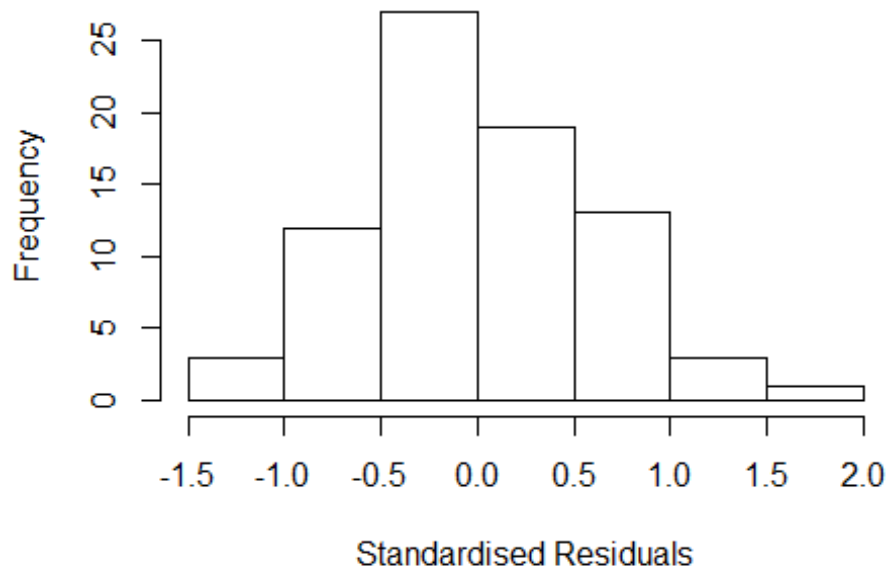
```
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12, 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	3.423e+01	1	5.851e+00
## LastAuthorFemale	8.823e+13	1	9.393e+06
## UniqueAuthors	4.385e+16	4	1.203e+02
## Year	4.091e+16	16	3.305e+00

Residuals from first and last author and team size



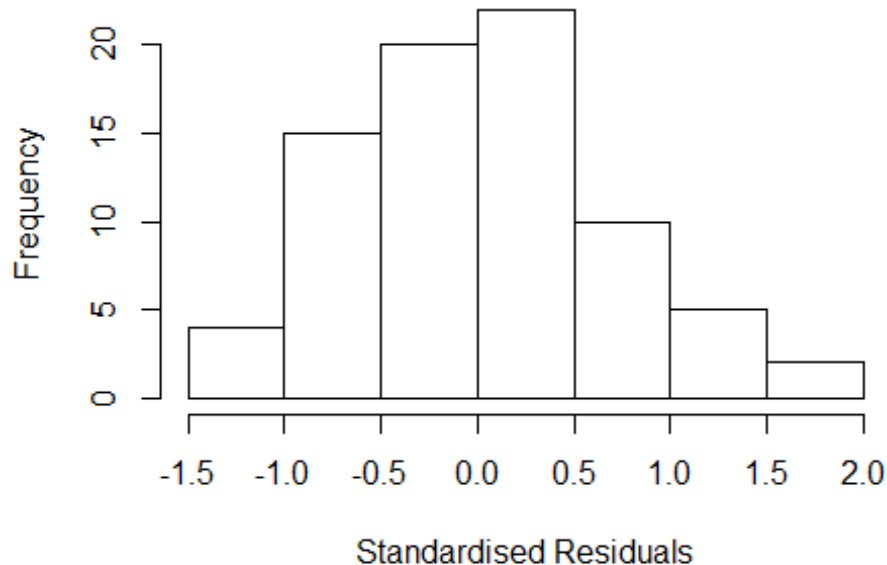
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1848 -0.3788 -0.0222 0.3464 1.5332
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.236 0.171 1.38 0.17345
## FirstAuthorFemale1 -0.539 0.450 -1.20 0.23593
## LastAuthorFemale1 0.669 0.414 1.62 0.11143
## UniqueAuthors2 0.406 0.287 1.42 0.16225
## UniqueAuthors3 1.555 0.314 4.96 7.2e-06 ***
## UniqueAuthors4 -0.395 0.466 -0.85 0.39970
## UniqueAuthors5 0.625 0.288 2.16 0.03476 *
## Year1997 0.698 0.599 1.17 0.24847
## Year1998 0.844 0.693 1.22 0.22898
## Year1999 0.819 0.460 1.78 0.08052 .
```

```

## Year2000          -0.139      0.357    -0.39  0.69900
## Year2001          0.394      0.307      1.28  0.20428
## Year2002          0.702      0.405      1.73  0.08852 .
## Year2003          0.797      0.340      2.35  0.02257 *
## Year2004          0.969      0.249      3.89  0.00027 ***
## Year2005          0.562      0.431      1.30  0.19835
## Year2006          0.513      0.333      1.54  0.12953
## Year2007          0.690      0.268      2.57  0.01283 *
## Year2008          0.392      0.283      1.38  0.17193
## Year2009          0.374      0.345      1.09  0.28221
## Year2010          0.162      0.311      0.52  0.60348
## Year2011          0.265      0.280      0.94  0.34950
## Year2012          0.369      0.437      0.84  0.40192
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.64
## Multiple R-squared:  0.414, Adjusted R-squared:  0.18
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 71 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.545  0.914  0.960  0.924  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      1.28e-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 11.86 1      3.444
## LastAuthorFemale  12.87 1      3.588
## Year              15.98 16      1.090

```

Residuals from first and last author



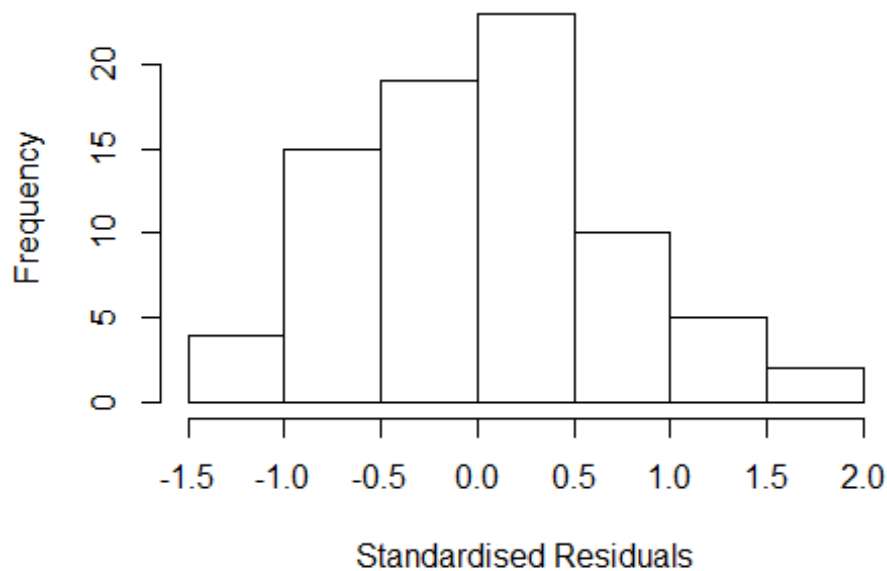
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27524 -0.45715 0.00745 0.41592 1.56990
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.235 0.170 1.39 0.171
## FirstAuthorFemale1 -0.212 0.330 -0.64 0.523
## LastAuthorFemale1 0.360 0.360 1.00 0.322
## Year1997 0.857 0.734 1.17 0.248
## Year1998 0.937 0.502 1.87 0.067 .
## Year1999 0.892 0.390 2.28 0.026 *
## Year2000 1.417 0.170 8.34 1.5e-11 ***
## Year2001 0.579 0.447 1.30 0.200
## Year2002 1.101 0.592 1.86 0.068 .
## Year2003 1.109 0.578 1.92 0.060 .
## Year2004 1.304 0.761 1.71 0.092 .
## Year2005 0.564 0.431 1.31 0.195
```

```

## Year2006          0.710      0.611      1.16      0.250
## Year2007          0.809      0.262      3.09      0.003 **
## Year2008          0.527      0.290      1.82      0.075 .
## Year2009          0.550      0.302      1.82      0.074 .
## Year2010          0.294      0.314      0.93      0.354
## Year2011          0.241      0.299      0.80      0.425
## Year2012          0.516      0.398      1.29      0.201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.716
## Multiple R-squared:  0.184, Adjusted R-squared:  -0.0651
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 71 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##   0.610  0.890  0.955  0.920  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      1.28e-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 9.031 1          3.005
## Year              9.031 16          1.071

```

Residuals from first author

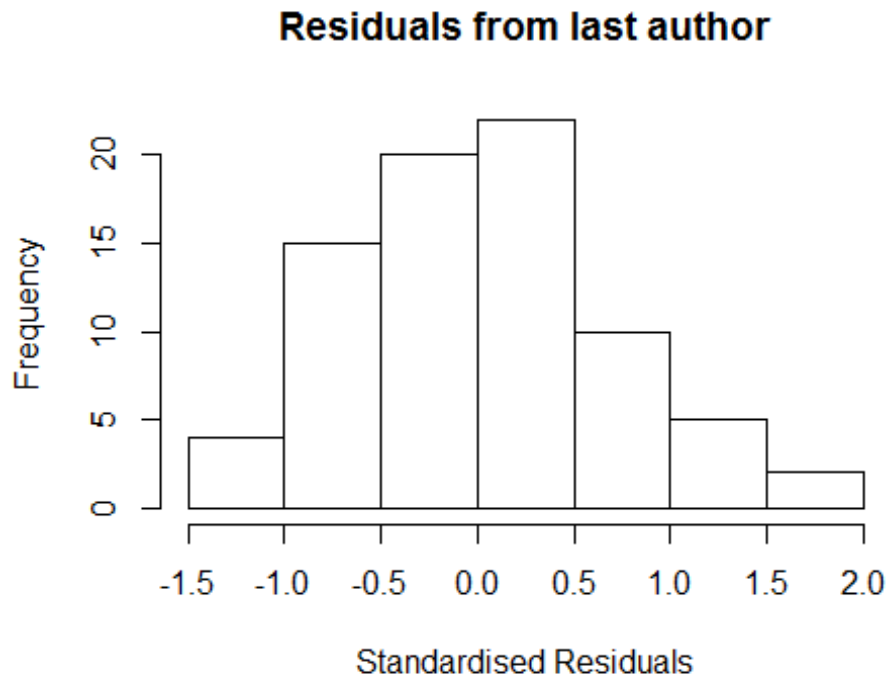


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.16059 -0.40195  0.00855  0.41992  1.66998
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.2355    0.1701   1.38  0.1713
## FirstAuthorFemale1 0.0582    0.2803   0.21  0.8361
## Year1997        0.8455    0.9802   0.86  0.3918
## Year1998        0.9753    0.5029   1.94  0.0572 .
## Year1999        0.8669    0.3860   2.25  0.0284 *
## Year2000        1.4165    0.1701   8.33 1.3e-11 ***
## Year2001        0.6693    0.4378   1.53  0.1316
## Year2002        1.1688    0.6158   1.90  0.0625 .
## Year2003        1.1090    0.5831   1.90  0.0620 .
## Year2004        1.2797    0.8200   1.56  0.1239
## Year2005        0.5950    0.4149   1.43  0.1567
## Year2006        0.7962    0.5834   1.36  0.1774
```

```

## Year2007          0.7624      0.2712      2.81    0.0067 **
## Year2008          0.5399      0.2917      1.85    0.0692 .
## Year2009          0.5504      0.3022      1.82    0.0736 .
## Year2010          0.3506      0.3274      1.07    0.2886
## Year2011          0.3434      0.2689      1.28    0.2065
## Year2012          0.5258      0.4200      1.25    0.2154
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.698
## Multiple R-squared:  0.165, Adjusted R-squared:  -0.0716
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 71 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.547  0.887   0.953   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      1.28e-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 8.029 1          2.834
## Year            8.029 16          1.067

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2751 -0.4768 0.0102 0.4475 1.6246
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.236 0.170 1.39 0.1712
## LastAuthorFemale1 0.197 0.290 0.68 0.5003
## Year1997 0.853 0.806 1.06 0.2940
## Year1998 0.916 0.503 1.82 0.0738 .
## Year1999 0.843 0.372 2.27 0.0270 *
## Year2000 1.416 0.170 8.33 1.3e-11 ***
## Year2001 0.531 0.444 1.20 0.2363
## Year2002 1.060 0.585 1.81 0.0750 .
## Year2003 1.109 0.582 1.91 0.0615 .
## Year2004 1.294 0.744 1.74 0.0871 .
## Year2005 0.545 0.440 1.24 0.2205
## Year2006 0.655 0.616 1.06 0.2917
```

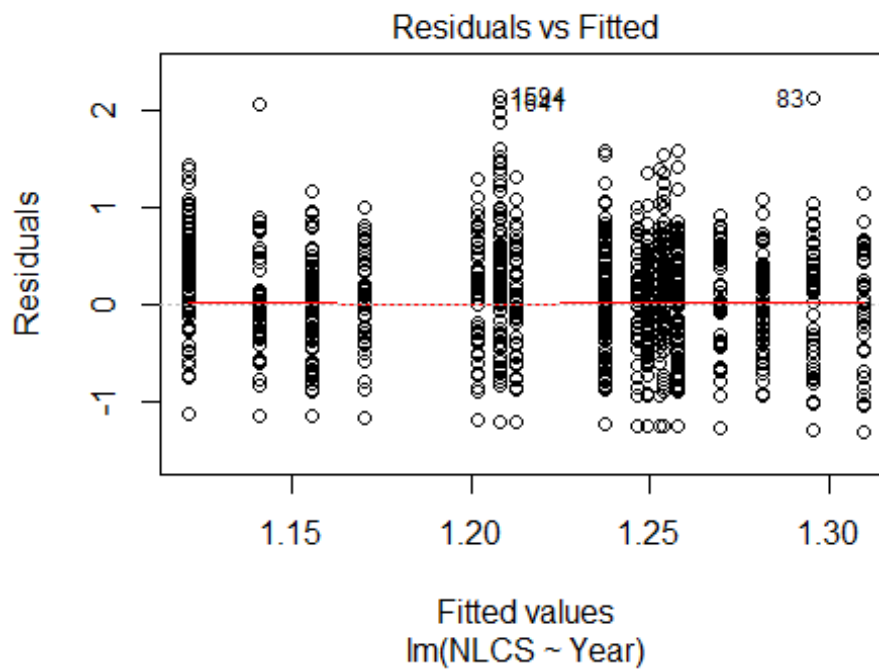


```

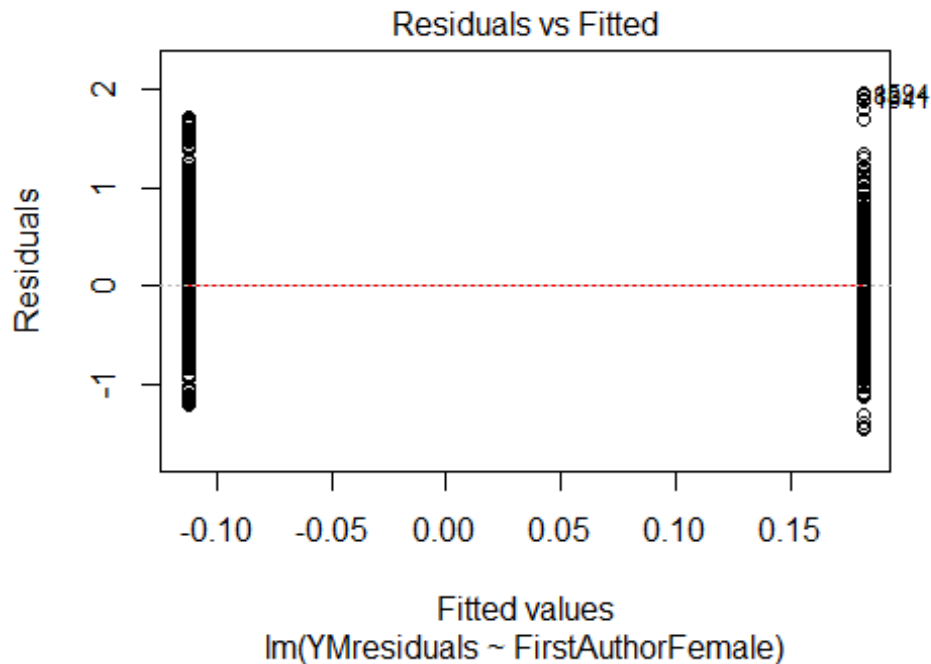
## Year2007          0.742      0.228      3.25      0.0019 **
## Year2008          0.468      0.277      1.69      0.0970 .
## Year2009          0.550      0.302      1.82      0.0736 .
## Year2010          0.300      0.328      0.91      0.3642
## Year2011          0.234      0.307      0.76      0.4488
## Year2012          0.507      0.390      1.30      0.1991
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.702
## Multiple R-squared:  0.178, Adjusted R-squared:  -0.0547
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 72 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.572  0.884  0.954  0.917  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.28e-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: 78"
## [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
##   52   67   71   58   66   76   51   65   55   59   92   94   96  140  145
## 2011 2012
##  137  149
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   42   51   58   53   54   59   42   53   47   49   80   78   87  127  128
## 2011 2012

```

```
## 121 134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 40 51 55 51 48 53 41 44 43 46 73 76 82 125 127
## 2011 2012
## 117 129
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 62, df = 16, p-value = 3e-07
```

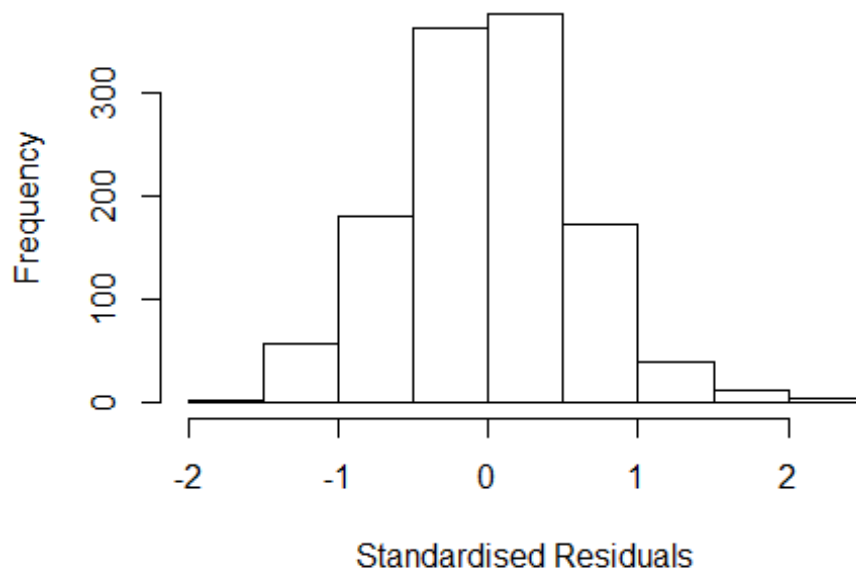


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.3, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 1.75742482367865"
## [1] "Male first author team size 2018 geometric mean: 1.44687457470993"
##
## 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: 2.01527791181"
## [1] "Male last author team size 2018 geometric mean: 1.25401997808435"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, 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 1.657 1      1.287
## LastAuthorFemale  1.563 1      1.250
## UniqueAuthors    1.673 4      1.066
## Year              1.521 16     1.013
```

Residuals from first and last author and team size



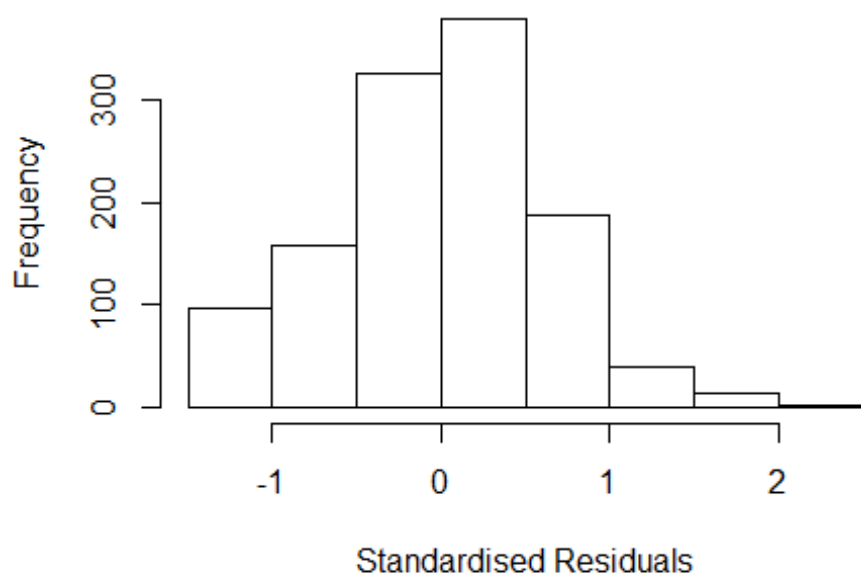
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.67744 -0.38648 0.00156 0.37590 2.20356
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1639 0.0951 12.24 < 2e-16 ***
## FirstAuthorFemale1 0.1339 0.0437 3.06 0.0022 **
## LastAuthorFemale1 0.0890 0.0437 2.04 0.0417 *
## UniqueAuthors2 0.2917 0.0483 6.04 2.0e-09 ***
## UniqueAuthors3 0.4905 0.0543 9.03 < 2e-16 ***
## UniqueAuthors4 0.4697 0.0632 7.43 2.1e-13 ***
## UniqueAuthors5 0.5400 0.0569 9.49 < 2e-16 ***
## Year1997 -0.1309 0.1352 -0.97 0.3332
## Year1998 -0.1482 0.1227 -1.21 0.2271
## Year1999 -0.2395 0.1192 -2.01 0.0448 *
```

```

## Year2000          -0.1706      0.1300    -1.31    0.1897
## Year2001          -0.1137      0.1145    -0.99    0.3208
## Year2002          -0.0523      0.1224    -0.43    0.6692
## Year2003          -0.2003      0.1428    -1.40    0.1611
## Year2004          -0.1303      0.1323    -0.99    0.3248
## Year2005          -0.2502      0.1197    -2.09    0.0368 *
## Year2006          -0.2864      0.1104    -2.60    0.0096 **
## Year2007          -0.1699      0.1127    -1.51    0.1319
## Year2008          -0.1925      0.1233    -1.56    0.1186
## Year2009          -0.1966      0.1064    -1.85    0.0650 .
## Year2010          -0.1413      0.1083    -1.30    0.1922
## Year2011          -0.2414      0.1173    -2.06    0.0397 *
## Year2012          -0.2494      0.1262    -1.98    0.0485 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.155, Adjusted R-squared:  0.139
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 106 weights are ~= 1. The remaining 1095 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0944 0.8540 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          8.33e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.788 1          1.337
## LastAuthorFemale 1.825 1          1.351
## Year          1.104 16          1.003

```

Residuals from first and last author



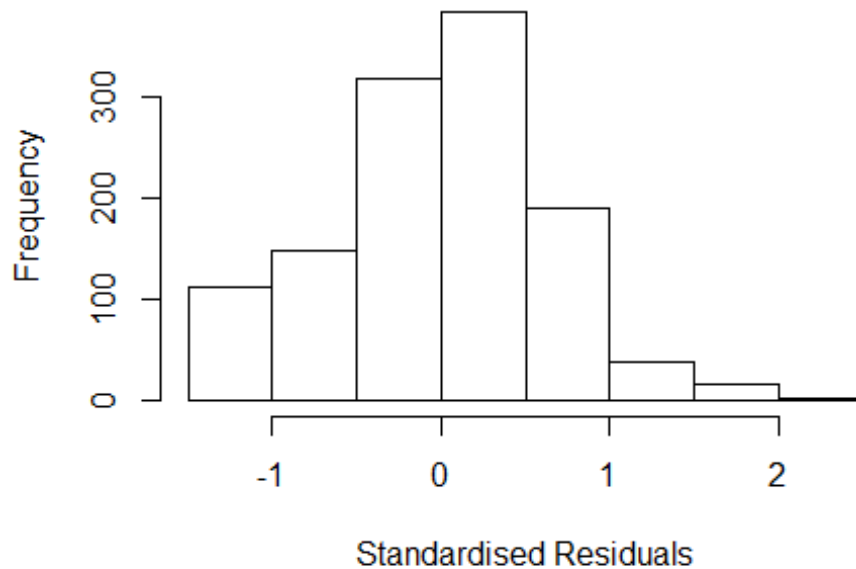
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4646 -0.4139 0.0434 0.4244 2.0308
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1633 0.0922 12.62 <2e-16 ***
## FirstAuthorFemale1 0.2494 0.0489 5.10 4e-07 ***
## LastAuthorFemale1 0.0899 0.0510 1.76 0.078 .
## Year1997 -0.0384 0.1278 -0.30 0.764
## Year1998 -0.0799 0.1253 -0.64 0.523
## Year1999 -0.1656 0.1209 -1.37 0.171
## Year2000 -0.0965 0.1304 -0.74 0.460
## Year2001 -0.0234 0.1103 -0.21 0.832
## Year2002 0.0752 0.1283 0.59 0.558
## Year2003 -0.0998 0.1447 -0.69 0.490
## Year2004 -0.0326 0.1318 -0.25 0.805
## Year2005 -0.1272 0.1205 -1.06 0.291
```

```

## Year2006          -0.1447      0.1119   -1.29    0.196
## Year2007          -0.0407      0.1090   -0.37    0.709
## Year2008          -0.0380      0.1194   -0.32    0.750
## Year2009          -0.0664      0.1056   -0.63    0.530
## Year2010          -0.0517      0.1071   -0.48    0.629
## Year2011          -0.1637      0.1145   -1.43    0.153
## Year2012          -0.1481      0.1257   -1.18    0.239
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.601
## Multiple R-squared:  0.0649, Adjusted R-squared:  0.0506
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 87 weights are ~= 1. The remaining 1114 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.230  0.868  0.948  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      8.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.043 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
## -1.4417 -0.4189 0.0318 0.4093 2.0073
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1655 0.0925 12.61 < 2e-16 ***
## FirstAuthorFemale1 0.3101 0.0373 8.30 2.8e-16 ***
## Year1997 -0.0353 0.1285 -0.27 0.78
## Year1998 -0.0716 0.1258 -0.57 0.57
## Year1999 -0.1566 0.1201 -1.30 0.19
## Year2000 -0.0894 0.1304 -0.69 0.49
## Year2001 -0.0206 0.1113 -0.19 0.85
## Year2002 0.0734 0.1283 0.57 0.57
## Year2003 -0.0942 0.1453 -0.65 0.52
## Year2004 -0.0275 0.1320 -0.21 0.83
## Year2005 -0.1285 0.1205 -1.07 0.29
## Year2006 -0.1389 0.1120 -1.24 0.22
```

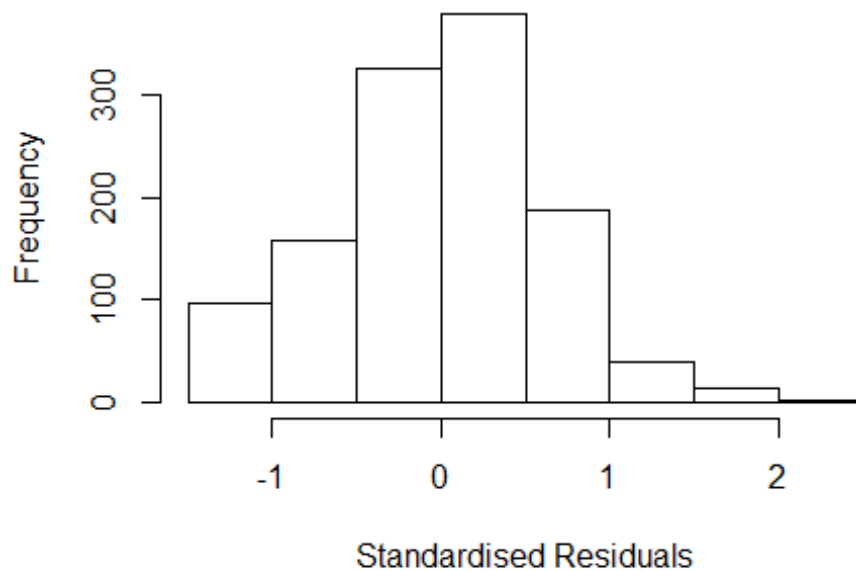


```

## Year2007          -0.0308      0.1095   -0.28      0.78
## Year2008          -0.0338      0.1194   -0.28      0.78
## Year2009          -0.0546      0.1055   -0.52      0.61
## Year2010          -0.0448      0.1072   -0.42      0.68
## Year2011          -0.1572      0.1149   -1.37      0.17
## Year2012          -0.1419      0.1266   -1.12      0.26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.603
## Multiple R-squared:  0.0623, Adjusted R-squared:  0.0488
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 96 weights are ~= 1. The remaining 1105 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.245  0.865   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      8.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.067 1          1.033
## Year            1.067 16          1.002

```

Residuals from last author



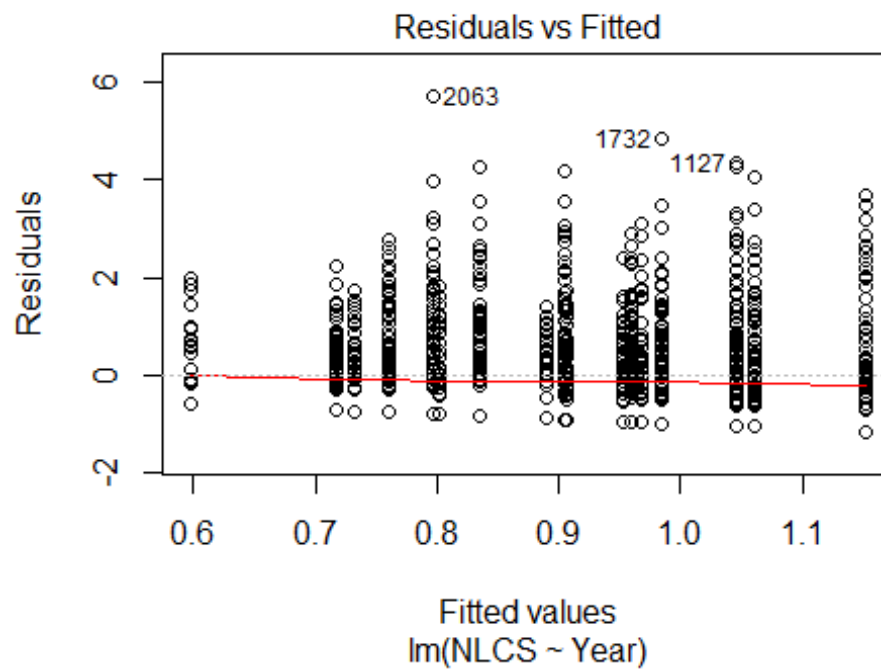
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4351 -0.4143 0.0207 0.4290 2.2301
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1923 0.0922 12.93 < 2e-16 ***
## LastAuthorFemale1 0.2650 0.0387 6.84 1.2e-11 ***
## Year1997 -0.0174 0.1279 -0.14 0.89
## Year1998 -0.0814 0.1262 -0.65 0.52
## Year1999 -0.1745 0.1234 -1.41 0.16
## Year2000 -0.1044 0.1305 -0.80 0.42
## Year2001 -0.0266 0.1094 -0.24 0.81
## Year2002 0.0807 0.1318 0.61 0.54
## Year2003 -0.0881 0.1440 -0.61 0.54
## Year2004 -0.0223 0.1354 -0.16 0.87
## Year2005 -0.1135 0.1206 -0.94 0.35
## Year2006 -0.1426 0.1136 -1.26 0.21
```

```

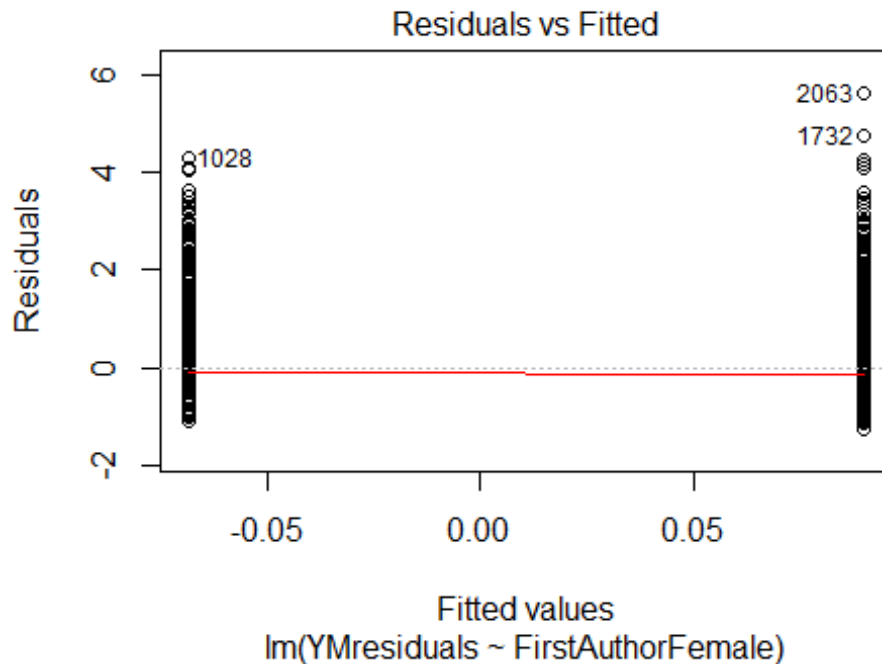
## Year2007          -0.0458      0.1086   -0.42      0.67
## Year2008          -0.0222      0.1221   -0.18      0.86
## Year2009          -0.0632      0.1064   -0.59      0.55
## Year2010          -0.0532      0.1079   -0.49      0.62
## Year2011          -0.1582      0.1154   -1.37      0.17
## Year2012          -0.1452      0.1255   -1.16      0.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.0466, Adjusted R-squared:  0.0329
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 101 weights are ~= 1. The remaining 1100 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.148  0.865  0.946  0.902  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1201"
## [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
##   56   53   61   65   50   67  192  197  178  224  199  217  187  132  102
## 2011 2012
##   203  236
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   47   42   49   58   39   53  167  178  162  194  172  185  170  116   93
## 2011 2012

```

```
## 175 197
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 47 42 49 57 38 52 167 178 161 192 171 184 170 115 92
## 2011 2012
## 173 196
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 85, df = 16, p-value = 2e-11
```

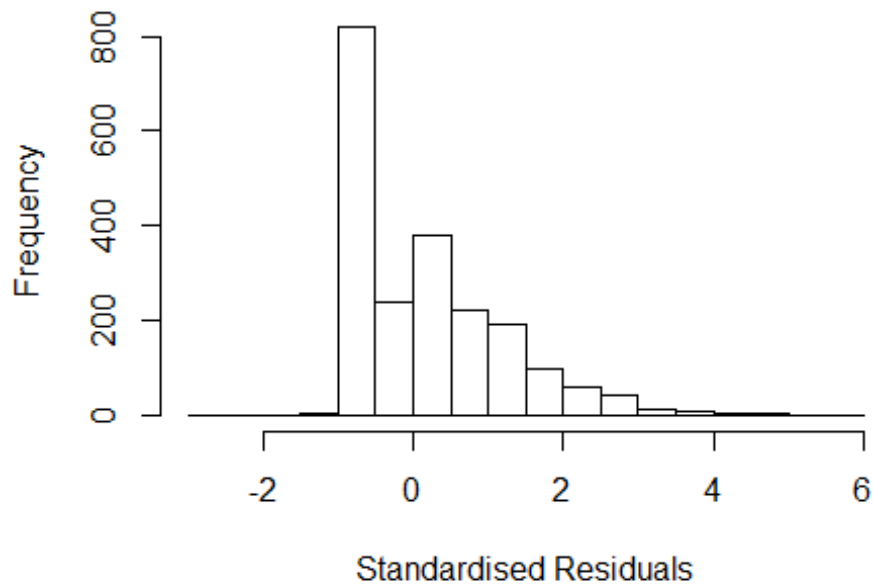


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.4, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 1.074327074737"
## [1] "Male first author team size 2018 geometric mean: 1"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3500, p-value = 0.006
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.0579678840895"
## [1] "Male last author team size 2018 geometric mean: 1.01793188433739"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3300, 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 30.52 1          5.524
## LastAuthorFemale  20.48 1          4.526
## UniqueAuthors    64.35 4          1.683
## Year              14.23 16          1.087
```

Residuals from first and last author and team size



```
## [1] "List of 71 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 118 61149294483 0.557 1998    1208      2    -2.632
## 386 61149478115 3.955 2002    1208      1     3.200
## 412 65849236676 3.746 2002    1208      1     2.991
## 414 67649953019 4.447 2002    1208      1     3.826
## 415 67650069906 5.062 2002    1208      1     4.307
## 473 22544438936 3.629 2002    1208      1     3.008
## 480 28744439229 3.363 2002    1208      1     2.742
## 492 60950350849 3.209 2002    1208      1     2.588
## 493 60950635288 3.855 2002    1208      1     3.234
## 504 62949205689 3.363 2002    1208      1     2.742
## 553 57049183828 3.214 2003    1208      1     2.515
## 696 60950496629 3.214 2003    1208      1     2.515
## 706 62149099241 3.347 2003    1208      1     2.648
## 707 62149142122 3.214 2003    1208      1     2.515
## 724 63849171517 3.537 2003    1208      2     2.972
## 761 60950688012 3.847 2004    1208      1     3.038
## 817 79958903992 3.622 2004    1208      1     2.663
## 891 62249219872 3.356 2004    1208      1     2.547
## 934 34347301294 3.822 2005    1208      1     2.968
## 941 60949418026 3.688 2005    1208      1     2.968
## 964 60950493221 3.688 2005    1208      1     2.968
## 985 61049443040 3.379 2005    1208      1     2.525
## 1001 67650091133 3.945 2005    1208      1     3.091
## 1028 34247665578 5.284 2005    1202      4     4.564
## 1115 64949099011 3.379 2005    1208      1     2.525
```

```

## 1120 70449794868 4.264 2005      1208      1      3.410
## 1127 60949545379 5.366 2005      1208      2      4.512
## 1128 60950284430 3.788 2005      1208      2      3.068
## 1130 61049203364 4.342 2005      1208      2      3.622
## 1157 34250638520 4.317 2006      1208      1      3.455
## 1163 46249116302 4.842 2006      1208      1      3.846
## 1183 60950449659 3.786 2006      1208      1      2.790
## 1185 60950481398 3.653 2006      1208      1      2.791
## 1195 60950581024 3.508 2006      1208      1      2.512
## 1211 61549109833 3.786 2006      1208      1      2.924
## 1291 60949833080 4.021 2006      1208      1      3.025
## 1309 60949152319 3.508 2006      1208      1      2.512
## 1327 67650021804 3.786 2006      1208      1      2.924
## 1357 57749162775 3.610 2007      1208      1      2.632
## 1422 64249150259 3.610 2007      1208      1      2.632
## 1428 67649497351 3.783 2007      1208      1      2.940
## 1475 62249126614 5.098 2007      1202      3      4.255
## 1506 61049173721 4.449 2007      1208      1      3.471
## 1538 61149390280 3.610 2007      1208      1      2.767
## 1579 60950570859 3.367 2008      1208      1      2.649
## 1589 57549107189 4.438 2008      1208      1      3.720
## 1673 70449970743 3.984 2008      1208      1      3.132
## 1728 61249742003 3.367 2008      1208      1      2.515
## 1732 62749100629 5.827 2008      1208      1      4.975
## 1777 77952081635 3.837 2009      1208      1      3.081
## 1838 68949194142 3.837 2009      1208      1      2.947
## 1858 61949203832 4.054 2009      1208      1      3.164
## 2015 84860777255 3.491 2011      1208      1      2.994
## 2036 84857704004 4.034 2011      1208      2      3.537
## 2063 80054701948 6.498 2011      1208      1      5.867
## 2138 79955378329 3.491 2011      1208      1      2.860
## 2159 79955684722 4.766 2011      1208      1      4.269
## 2191 82455174320 3.290 2011      1208      2      2.659
## 2268 84868374184 3.049 2012      1208      2      2.544
## 2290 84870978804 3.391 2012      1208      1      2.886
## 2305 84885600847 4.382 2012      1208      1      3.742
## 2306 84885627769 3.937 2012      1208      1      3.297
## 2350 84871197395 3.391 2012      1208      1      2.751
## 2361 84862275410 3.233 2012      1203      3      2.728
## 2367 84859360916 4.382 2012      1208      1      3.742
## 2370 84861758622 3.486 2012      1208      2      2.670
## 2378 84862085948 3.065 2012      1208      2      2.560
## 2383 84874030859 3.264 2012      1202      4      2.759
## 2398 84877706298 5.086 2012      1208      1      3.704
## 2421 84860705962 3.391 2012      1208      1      2.751
## 2427 84896996092 3.391 2012      1208      1      2.886

```

```
##
```

```
## Call:
```

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

```

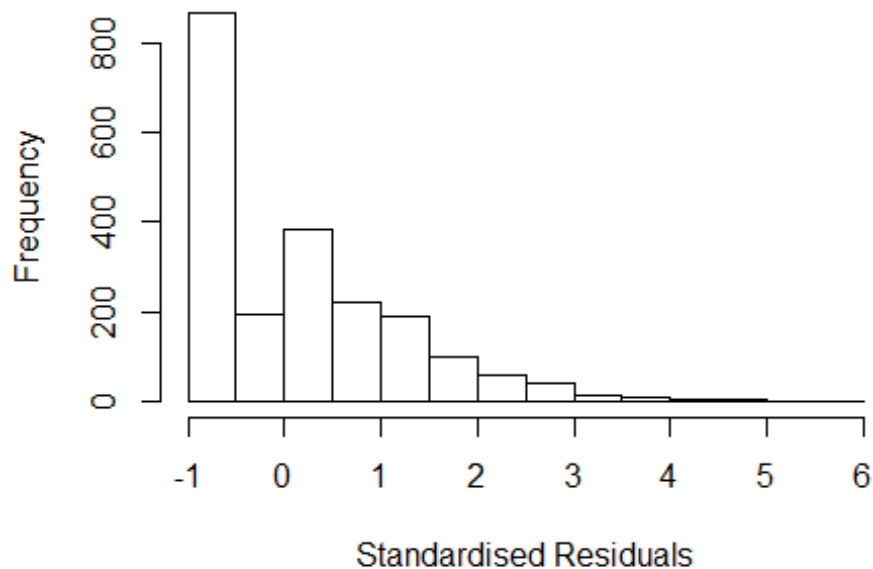
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -2.6323 -0.6992 -0.0447  0.7500  5.8671
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.4719      0.1034   4.56 5.3e-06 ***
## FirstAuthorFemale1 -0.0263      0.1731  -0.15  0.8791
## LastAuthorFemale1  0.1605      0.1746   0.92  0.3581
## UniqueAuthors2     0.1497      0.1133   1.32  0.1864
## UniqueAuthors3     0.9030      0.1766   5.11 3.5e-07 ***
## UniqueAuthors4     2.5146      1.4249   1.76  0.0778 .
## UniqueAuthors5     0.8530      0.1926   4.43 1.0e-05 ***
## Year1997            0.3629      0.1649   2.20  0.0278 *
## Year1998            0.2027      0.1702   1.19  0.2338
## Year1999            0.3799      0.1477   2.57  0.0102 *
## Year2000            0.3792      0.1712   2.21  0.0269 *
## Year2001            0.1711      0.1486   1.15  0.2495
## Year2002            0.1489      0.1271   1.17  0.2415
## Year2003            0.0931      0.1231   0.76  0.4494
## Year2004            0.3372      0.1265   2.67  0.0077 **
## Year2005            0.2484      0.1321   1.88  0.0601 .
## Year2006            0.3899      0.1323   2.95  0.0032 **
## Year2007            0.3715      0.1291   2.88  0.0040 **
## Year2008            0.2462      0.1318   1.87  0.0619 .
## Year2009            0.2843      0.1396   2.04  0.0418 *
## Year2010            0.0975      0.1277   0.76  0.4453
## Year2011            0.0249      0.1252   0.20  0.8426
## Year2012            0.0336      0.1245   0.27  0.7875
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.817
## Multiple R-squared:  0.0401, Adjusted R-squared:  0.0299
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## 9 observations c(337,338,877,958,989,1250,1480,1766,1847)
## are outliers with |weight| = 0 ( < 4.8e-05);
## 108 weights are ~ = 1. The remaining 1967 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0018 0.8760 0.9310 0.8760 0.9670 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x

```



```
##          1.00e-07          1.00e-07          4.80e-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 15.418 1          3.927
## LastAuthorFemale 15.483 1          3.935
## Year              1.085 16          1.003
```

Residuals from first and last author



```
## [1] "List of 71 outliers with residuals above 2.5"
##          ScopusId NLCS Year OneField Fields residuals
## 386 61149478115 3.955 2002 1208 1 3.194
## 412 65849236676 3.746 2002 1208 1 2.985
## 414 67649953019 4.447 2002 1208 1 3.826
## 415 67650069906 5.062 2002 1208 1 4.301
## 473 22544438936 3.629 2002 1208 1 3.008
## 480 28744439229 3.363 2002 1208 1 2.742
```

##	492	60950350849	3.209	2002	1208	1	2.588
##	493	60950635288	3.855	2002	1208	1	3.234
##	504	62949205689	3.363	2002	1208	1	2.742
##	553	57049183828	3.214	2003	1208	1	2.504
##	696	60950496629	3.214	2003	1208	1	2.504
##	706	62149099241	3.347	2003	1208	1	2.637
##	707	62149142122	3.214	2003	1208	1	2.504
##	724	63849171517	3.537	2003	1208	2	2.966
##	761	60950688012	3.847	2004	1208	1	3.039
##	817	79958903992	3.622	2004	1208	1	2.814
##	891	62249219872	3.356	2004	1208	1	2.548
##	934	34347301294	3.822	2005	1208	1	2.950
##	941	60949418026	3.688	2005	1208	1	2.956
##	964	60950493221	3.688	2005	1208	1	2.956
##	985	61049443040	3.379	2005	1208	1	2.507
##	1001	67650091133	3.945	2005	1208	1	3.073
##	1028	34247665578	5.284	2005	1202	4	4.552
##	1115	64949099011	3.379	2005	1208	1	2.507
##	1120	70449794868	4.264	2005	1208	1	3.392
##	1127	60949545379	5.366	2005	1208	2	4.494
##	1128	60950284430	3.788	2005	1208	2	3.056
##	1130	61049203364	4.342	2005	1208	2	3.610
##	1157	34250638520	4.317	2006	1208	1	3.459
##	1163	46249116302	4.842	2006	1208	1	3.845
##	1183	60950449659	3.786	2006	1208	1	2.789
##	1185	60950481398	3.653	2006	1208	1	2.795
##	1195	60950581024	3.508	2006	1208	1	2.511
##	1211	61549109833	3.786	2006	1208	1	2.928
##	1291	60949833080	4.021	2006	1208	1	3.024
##	1309	60949152319	3.508	2006	1208	1	2.511
##	1317	46049100517	4.639	2006	1208	1	3.642
##	1327	67650021804	3.786	2006	1208	1	2.928
##	1357	57749162775	3.610	2007	1208	1	2.620
##	1422	64249150259	3.610	2007	1208	1	2.620
##	1428	67649497351	3.783	2007	1208	1	2.932
##	1475	62249126614	5.098	2007	1202	3	4.247
##	1506	61049173721	4.449	2007	1208	1	3.459
##	1538	61149390280	3.610	2007	1208	1	2.759
##	1579	60950570859	3.367	2008	1208	1	2.637
##	1589	57549107189	4.438	2008	1208	1	3.708
##	1673	70449970743	3.984	2008	1208	1	3.115
##	1732	62749100629	5.827	2008	1208	1	4.958
##	1777	77952081635	3.837	2009	1208	1	3.070
##	1838	68949194142	3.837	2009	1208	1	2.930
##	1858	61949203832	4.054	2009	1208	1	3.147
##	2015	84860777255	3.491	2011	1208	1	2.989
##	2036	84857704004	4.034	2011	1208	2	3.532
##	2063	80054701948	6.498	2011	1208	1	5.856
##	2082	84857889273	3.887	2011	1208	1	3.304
##	2138	79955378329	3.491	2011	1208	1	2.849

```

## 2159 79955684722 4.766 2011      1208      1      4.264
## 2191 82455174320 3.290 2011      1208      2      2.648
## 2268 84868374184 3.049 2012      1208      2      2.516
## 2290 84870978804 3.391 2012      1208      1      2.858
## 2305 84885600847 4.382 2012      1208      1      3.709
## 2306 84885627769 3.937 2012      1208      1      3.264
## 2350 84871197395 3.391 2012      1208      1      2.718
## 2361 84862275410 3.233 2012      1203      3      2.700
## 2367 84859360916 4.382 2012      1208      1      3.709
## 2370 84861758622 3.486 2012      1208      2      2.872
## 2378 84862085948 3.065 2012      1208      2      2.532
## 2383 84874030859 3.264 2012      1202      4      2.731
## 2398 84877706298 5.086 2012      1208      1      4.494
## 2421 84860705962 3.391 2012      1208      1      2.718
## 2427 84896996092 3.391 2012      1208      1      2.858
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9975 -0.6800 -0.0443  0.7557  5.8562
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.4791     0.1033   4.64 3.7e-06 ***
## FirstAuthorFemale1 0.0586     0.1713   0.34  0.7324
## LastAuthorFemale1 0.0810     0.1727   0.47  0.6389
## Year1997          0.3543     0.1646   2.15  0.0315 *
## Year1998          0.2230     0.1593   1.40  0.1617
## Year1999          0.3748     0.1480   2.53  0.0114 *
## Year2000          0.3764     0.1709   2.20  0.0277 *
## Year2001          0.1620     0.1483   1.09  0.2749
## Year2002          0.1424     0.1267   1.12  0.2614
## Year2003          0.0917     0.1230   0.75  0.4561
## Year2004          0.3294     0.1259   2.62  0.0090 **
## Year2005          0.2531     0.1316   1.92  0.0546 .
## Year2006          0.3787     0.1325   2.86  0.0043 **
## Year2007          0.3717     0.1290   2.88  0.0040 **
## Year2008          0.2506     0.1319   1.90  0.0576 .
## Year2009          0.2880     0.1398   2.06  0.0394 *
## Year2010          0.1118     0.1289   0.87  0.3857
## Year2011          0.0231     0.1252   0.18  0.8537
## Year2012          0.0539     0.1248   0.43  0.6661
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.822

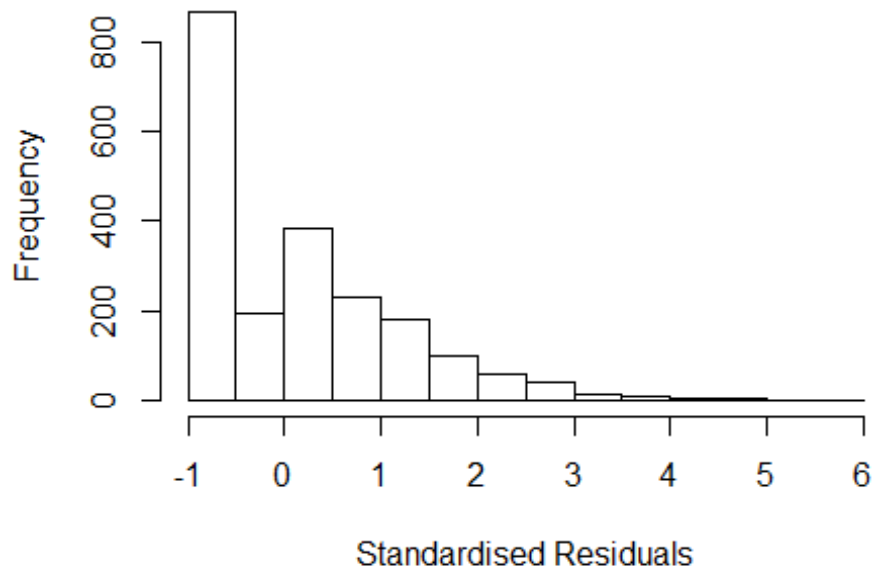
```

```

## Multiple R-squared:  0.0248, Adjusted R-squared:  0.0163
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 9 observations c(338,877,958,989,1250,1480,1766,1847,2043)
## are outliers with |weight| <= 1.9e-05 ( < 4.8e-05);
## 111 weights are ~= 1. The remaining 1964 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0002 0.8800 0.9300 0.8770 0.9660 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.80e-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 1.044 1          1.022
## Year              1.044 16          1.001

```

Residuals from first author



```
## [1] "List of 71 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 386  61149478115 3.955 2002    1208      1    3.194
## 412  65849236676 3.746 2002    1208      1    2.985
## 414  67649953019 4.447 2002    1208      1    3.826
## 415  67650069906 5.062 2002    1208      1    4.301
## 473  22544438936 3.629 2002    1208      1    3.008
## 480  28744439229 3.363 2002    1208      1    2.742
## 492  60950350849 3.209 2002    1208      1    2.588
## 493  60950635288 3.855 2002    1208      1    3.234
## 504  62949205689 3.363 2002    1208      1    2.742
## 553  57049183828 3.214 2003    1208      1    2.504
## 696  60950496629 3.214 2003    1208      1    2.504
## 706  62149099241 3.347 2003    1208      1    2.637
## 707  62149142122 3.214 2003    1208      1    2.504
## 724  63849171517 3.537 2003    1208      2    2.966
## 761  60950688012 3.847 2004    1208      1    3.039
## 817  79958903992 3.622 2004    1208      1    2.814
## 891  62249219872 3.356 2004    1208      1    2.548
## 934  34347301294 3.822 2005    1208      1    2.950
## 941  60949418026 3.688 2005    1208      1    2.956
## 964  60950493221 3.688 2005    1208      1    2.956
## 985  61049443040 3.379 2005    1208      1    2.507
## 1001 67650091133 3.945 2005    1208      1    3.073
## 1028 34247665578 5.284 2005    1202      4    4.552
## 1115 64949099011 3.379 2005    1208      1    2.507
## 1120 70449794868 4.264 2005    1208      1    3.392
```

```

## 1127 60949545379 5.366 2005      1208      2      4.494
## 1128 60950284430 3.788 2005      1208      2      3.056
## 1130 61049203364 4.342 2005      1208      2      3.610
## 1157 34250638520 4.317 2006      1208      1      3.459
## 1163 46249116302 4.842 2006      1208      1      3.845
## 1183 60950449659 3.786 2006      1208      1      2.789
## 1185 60950481398 3.653 2006      1208      1      2.795
## 1195 60950581024 3.508 2006      1208      1      2.511
## 1211 61549109833 3.786 2006      1208      1      2.928
## 1291 60949833080 4.021 2006      1208      1      3.024
## 1309 60949152319 3.508 2006      1208      1      2.511
## 1317 46049100517 4.639 2006      1208      1      3.642
## 1327 67650021804 3.786 2006      1208      1      2.928
## 1357 57749162775 3.610 2007      1208      1      2.620
## 1422 64249150259 3.610 2007      1208      1      2.620
## 1428 67649497351 3.783 2007      1208      1      2.932
## 1475 62249126614 5.098 2007      1202      3      4.247
## 1506 61049173721 4.449 2007      1208      1      3.459
## 1538 61149390280 3.610 2007      1208      1      2.759
## 1579 60950570859 3.367 2008      1208      1      2.637
## 1589 57549107189 4.438 2008      1208      1      3.708
## 1673 70449970743 3.984 2008      1208      1      3.115
## 1732 62749100629 5.827 2008      1208      1      4.958
## 1777 77952081635 3.837 2009      1208      1      3.070
## 1838 68949194142 3.837 2009      1208      1      2.930
## 1858 61949203832 4.054 2009      1208      1      3.147
## 2015 84860777255 3.491 2011      1208      1      2.989
## 2036 84857704004 4.034 2011      1208      2      3.532
## 2063 80054701948 6.498 2011      1208      1      5.856
## 2082 84857889273 3.887 2011      1208      1      3.304
## 2138 79955378329 3.491 2011      1208      1      2.849
## 2159 79955684722 4.766 2011      1208      1      4.264
## 2191 82455174320 3.290 2011      1208      2      2.648
## 2268 84868374184 3.049 2012      1208      2      2.516
## 2290 84870978804 3.391 2012      1208      1      2.858
## 2305 84885600847 4.382 2012      1208      1      3.709
## 2306 84885627769 3.937 2012      1208      1      3.264
## 2350 84871197395 3.391 2012      1208      1      2.718
## 2361 84862275410 3.233 2012      1203      3      2.700
## 2367 84859360916 4.382 2012      1208      1      3.709
## 2370 84861758622 3.486 2012      1208      2      2.872
## 2378 84862085948 3.065 2012      1208      2      2.532
## 2383 84874030859 3.264 2012      1202      4      2.731
## 2398 84877706298 5.086 2012      1208      1      4.494
## 2421 84860705962 3.391 2012      1208      1      2.718
## 2427 84896996092 3.391 2012      1208      1      2.858

```

```
##
```

```
## Call:
```

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

```

##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -0.9941 -0.6796 -0.0459  0.7537  5.8562
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.4752      0.1032   4.61 4.4e-06 ***
## FirstAuthorFemale1 0.1367      0.0446   3.07  0.0022 **
## Year1997          0.3591      0.1645   2.18  0.0292 *
## Year1998          0.2295      0.1589   1.44  0.1488
## Year1999          0.3768      0.1484   2.54  0.0112 *
## Year2000          0.3809      0.1708   2.23  0.0259 *
## Year2001          0.1672      0.1482   1.13  0.2594
## Year2002          0.1475      0.1266   1.16  0.2442
## Year2003          0.0953      0.1232   0.77  0.4394
## Year2004          0.3335      0.1261   2.65  0.0082 **
## Year2005          0.2573      0.1317   1.95  0.0510 .
## Year2006          0.3822      0.1328   2.88  0.0040 **
## Year2007          0.3770      0.1288   2.93  0.0034 **
## Year2008          0.2558      0.1320   1.94  0.0527 .
## Year2009          0.2922      0.1399   2.09  0.0369 *
## Year2010          0.1163      0.1290   0.90  0.3674
## Year2011          0.0299      0.1246   0.24  0.8104
## Year2012          0.0594      0.1250   0.47  0.6349
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.822
## Multiple R-squared:  0.0247, Adjusted R-squared:  0.0166
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 9 observations c(338,877,958,989,1250,1480,1766,1847,2043)
## are outliers with |weight| <= 3.5e-06 ( < 4.8e-05);
## 112 weights are ~1. The remaining 1963 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0002 0.8800 0.9300 0.8770 0.9660 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.80e-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 1.048 1          1.024
## Year             1.048 16          1.001

## [1] "List of 71 outliers with residuals above 2.5"
##          ScopusId NLCS Year OneField Fields residuals
## 386 61149478115 3.955 2002 1208 1 3.194
## 412 65849236676 3.746 2002 1208 1 2.985
## 414 67649953019 4.447 2002 1208 1 3.826
## 415 67650069906 5.062 2002 1208 1 4.301
## 473 22544438936 3.629 2002 1208 1 3.008
## 480 28744439229 3.363 2002 1208 1 2.742
## 492 60950350849 3.209 2002 1208 1 2.588
## 493 60950635288 3.855 2002 1208 1 3.234
## 504 62949205689 3.363 2002 1208 1 2.742
## 553 57049183828 3.214 2003 1208 1 2.504
## 696 60950496629 3.214 2003 1208 1 2.504
## 706 62149099241 3.347 2003 1208 1 2.637
## 707 62149142122 3.214 2003 1208 1 2.504
## 724 63849171517 3.537 2003 1208 2 2.966
## 761 60950688012 3.847 2004 1208 1 3.039
## 817 79958903992 3.622 2004 1208 1 2.814
## 891 62249219872 3.356 2004 1208 1 2.548
## 934 34347301294 3.822 2005 1208 1 2.950
## 941 60949418026 3.688 2005 1208 1 2.956
## 964 60950493221 3.688 2005 1208 1 2.956
## 985 61049443040 3.379 2005 1208 1 2.507
## 1001 67650091133 3.945 2005 1208 1 3.073
## 1028 34247665578 5.284 2005 1202 4 4.552
## 1115 64949099011 3.379 2005 1208 1 2.507
## 1120 70449794868 4.264 2005 1208 1 3.392
## 1127 60949545379 5.366 2005 1208 2 4.494
## 1128 60950284430 3.788 2005 1208 2 3.056
## 1130 61049203364 4.342 2005 1208 2 3.610
## 1157 34250638520 4.317 2006 1208 1 3.459
## 1163 46249116302 4.842 2006 1208 1 3.845
## 1183 60950449659 3.786 2006 1208 1 2.789
## 1185 60950481398 3.653 2006 1208 1 2.795
## 1195 60950581024 3.508 2006 1208 1 2.511
## 1211 61549109833 3.786 2006 1208 1 2.928
## 1291 60949833080 4.021 2006 1208 1 3.024
## 1309 60949152319 3.508 2006 1208 1 2.511
## 1317 46049100517 4.639 2006 1208 1 3.642
## 1327 67650021804 3.786 2006 1208 1 2.928

```



```

## 1357 57749162775 3.610 2007      1208      1      2.620
## 1422 64249150259 3.610 2007      1208      1      2.620
## 1428 67649497351 3.783 2007      1208      1      2.932
## 1475 62249126614 5.098 2007      1202      3      4.247
## 1506 61049173721 4.449 2007      1208      1      3.459
## 1538 61149390280 3.610 2007      1208      1      2.759
## 1579 60950570859 3.367 2008      1208      1      2.637
## 1589 57549107189 4.438 2008      1208      1      3.708
## 1673 70449970743 3.984 2008      1208      1      3.115
## 1732 62749100629 5.827 2008      1208      1      4.958
## 1777 77952081635 3.837 2009      1208      1      3.070
## 1838 68949194142 3.837 2009      1208      1      2.930
## 1858 61949203832 4.054 2009      1208      1      3.147
## 2015 84860777255 3.491 2011      1208      1      2.989
## 2036 84857704004 4.034 2011      1208      2      3.532
## 2063 80054701948 6.498 2011      1208      1      5.856
## 2082 84857889273 3.887 2011      1208      1      3.304
## 2138 79955378329 3.491 2011      1208      1      2.849
## 2159 79955684722 4.766 2011      1208      1      4.264
## 2191 82455174320 3.290 2011      1208      2      2.648
## 2268 84868374184 3.049 2012      1208      2      2.516
## 2290 84870978804 3.391 2012      1208      1      2.858
## 2305 84885600847 4.382 2012      1208      1      3.709
## 2306 84885627769 3.937 2012      1208      1      3.264
## 2350 84871197395 3.391 2012      1208      1      2.718
## 2361 84862275410 3.233 2012      1203      3      2.700
## 2367 84859360916 4.382 2012      1208      1      3.709
## 2370 84861758622 3.486 2012      1208      2      2.872
## 2378 84862085948 3.065 2012      1208      2      2.532
## 2383 84874030859 3.264 2012      1202      4      2.731
## 2398 84877706298 5.086 2012      1208      1      4.494
## 2421 84860705962 3.391 2012      1208      1      2.718
## 2427 84896996092 3.391 2012      1208      1      2.858
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9984 -0.6795 -0.0445  0.7549  5.8574
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4836    0.1028   4.70 2.7e-06 ***
## LastAuthorFemale1 0.1379    0.0449   3.07  0.0022 **
## Year1997        0.3504    0.1650   2.12  0.0338 *
## Year1998        0.2182    0.1591   1.37  0.1705
## Year1999        0.3732    0.1482   2.52  0.0118 *

```

```

## Year2000          0.3725      0.1707      2.18      0.0292 *
## Year2001          0.1584      0.1481      1.07      0.2850
## Year2002          0.1395      0.1267      1.10      0.2712
## Year2003          0.0893      0.1231      0.73      0.4683
## Year2004          0.3269      0.1259      2.60      0.0095 **
## Year2005          0.2505      0.1316      1.90      0.0571 .
## Year2006          0.3769      0.1328      2.84      0.0046 **
## Year2007          0.3684      0.1289      2.86      0.0043 **
## Year2008          0.2475      0.1317      1.88      0.0603 .
## Year2009          0.2855      0.1397      2.04      0.0411 *
## Year2010          0.1087      0.1290      0.84      0.3997
## Year2011          0.0191      0.1248      0.15      0.8786
## Year2012          0.0506      0.1248      0.41      0.6853
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.824
## Multiple R-squared:  0.0247, Adjusted R-squared:  0.0166
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 8 observations c(338,877,958,1250,1480,1766,1847,2043)
## are outliers with |weight| = 0 ( < 4.8e-05);
## 110 weights are ~ = 1. The remaining 1966 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8810 0.9300 0.8770 0.9660 0.9990
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          4.80e-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: 2084"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1209"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"

```

```

##
## 1996 1997 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    3    7   11    3    7    3    1    3    3    9    6   18    6    7   10
## 2012
##    3
##
## 1996 1997 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    2    7   10    1    5    3    1    3    2    8    6   16    5    6    9
## 2012
##    3
##
## 1996 1997 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    2    6   10    1    4    3    1    3    2    8    5   16    5    6    9
## 2012
##    3
## [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"

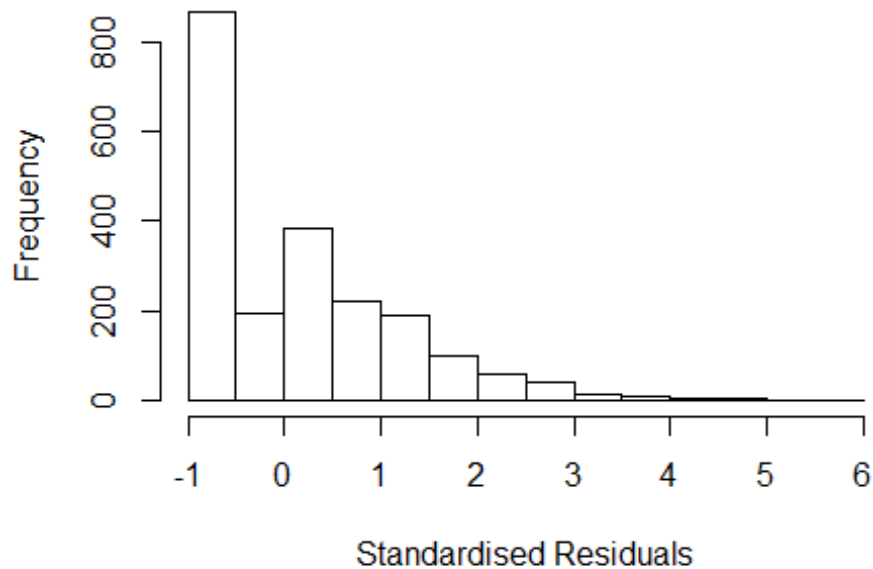
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9, 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

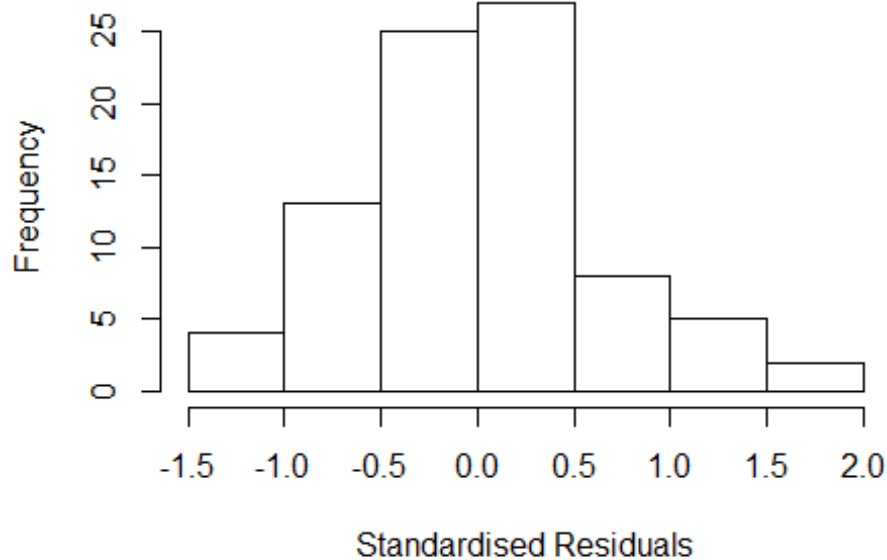
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9, 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 10.943  1      3.308
## LastAuthorFemale  7.148  1      2.674
## UniqueAuthors    12.420  2      1.877
## Year              78.057 15      1.156
```

Residuals from first and last author and team size



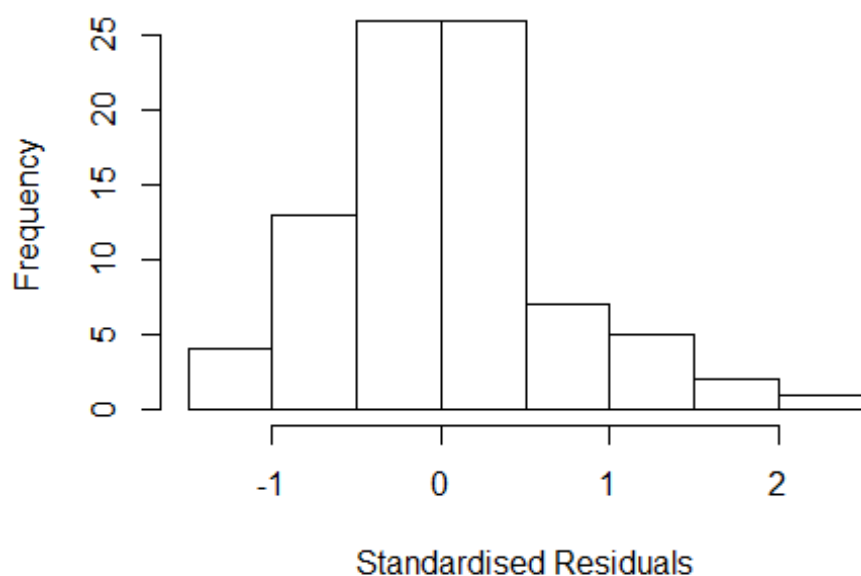
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.239204 -0.316543 0.000376 0.365972 1.765139
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3233 0.2718 1.19 0.2387
## FirstAuthorFemale1 -0.2318 0.2624 -0.88 0.3803
## LastAuthorFemale1 -0.0934 0.2069 -0.45 0.6532
## UniqueAuthors2 0.5322 0.1943 2.74 0.0080 **
## UniqueAuthors3 0.6772 0.3244 2.09 0.0408 *
## Year1997 0.5310 0.3938 1.35 0.1823
## Year1999 0.9159 0.3418 2.68 0.0094 **
## Year2000 0.4227 0.2718 1.55 0.1249
## Year2001 0.9469 0.2960 3.20 0.0021 **
## Year2002 0.8831 0.7827 1.13 0.2634
```

```

## Year2003          0.3647      0.2718      1.34      0.1845
## Year2004          1.1550      0.7714      1.50      0.1393
## Year2005          0.3263      0.2930      1.11      0.2697
## Year2006          1.0055      0.3602      2.79      0.0069 **
## Year2007          0.5275      0.3062      1.72      0.0898 .
## Year2008          0.8243      0.3240      2.54      0.0134 *
## Year2009          0.6993      0.3976      1.76      0.0834 .
## Year2010          0.2832      0.3453      0.82      0.4151
## Year2011          0.5297      0.3455      1.53      0.1302
## Year2012          1.9716      1.1013      1.79      0.0782 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.631
## Multiple R-squared:  0.358, Adjusted R-squared:  0.167
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 76 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.414  0.894  0.966  0.908  0.986  0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.19e-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 7.282 1      2.698
## LastAuthorFemale  7.285 1      2.699
## Year              12.104 15      1.087

```

Residuals from first and last author



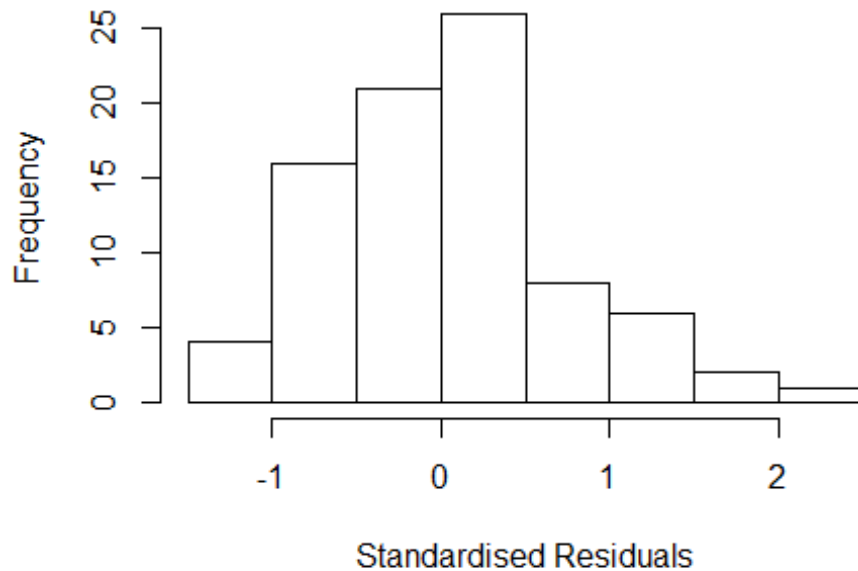
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.42e+00 -4.33e-01 5.97e-16 3.70e-01 2.30e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4987 0.4213 1.18 0.2407
## FirstAuthorFemale1 -0.0504 0.2952 -0.17 0.8649
## LastAuthorFemale1 -0.3221 0.2787 -1.16 0.2519
## Year1997 0.4742 0.5297 0.90 0.3739
## Year1999 0.9257 0.4439 2.09 0.0409 *
## Year2000 0.2473 0.4213 0.59 0.5592
## Year2001 0.9318 0.4383 2.13 0.0373 *
## Year2002 0.7788 0.8520 0.91 0.3640
## Year2003 0.1893 0.4213 0.45 0.6547
## Year2004 0.8143 0.9772 0.83 0.4076
## Year2005 0.1746 0.4336 0.40 0.6886
## Year2006 1.0224 0.4753 2.15 0.0352 *
```

```

## Year2007          0.7061      0.4348      1.62      0.1091
## Year2008          0.7957      0.4472      1.78      0.0798 .
## Year2009          0.6989      0.5443      1.28      0.2036
## Year2010          0.1620      0.4937      0.33      0.7439
## Year2011          0.5016      0.4874      1.03      0.3072
## Year2012          2.0865      0.7143      2.92      0.0048 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.685
## Multiple R-squared:  0.28,   Adjusted R-squared:  0.0942
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 75 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.238  0.908  0.960  0.909  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.19e-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 7.43 1      2.726
## Year              7.43 15      1.069

```


Residuals from first author



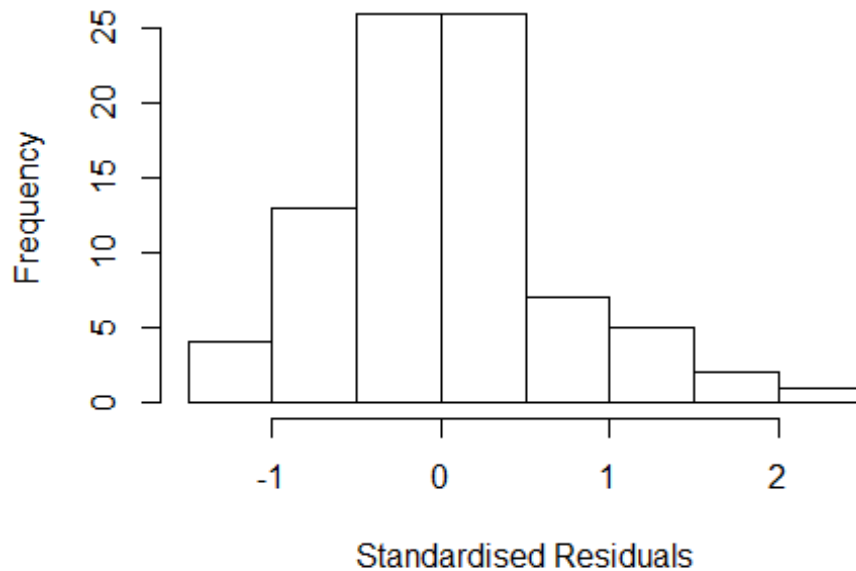
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.41889 -0.46810 0.00746 0.36152 2.16766
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6183 0.5263 1.17 0.244
## FirstAuthorFemale1 -0.2896 0.2315 -1.25 0.215
## Year1997 0.3546 0.6132 0.58 0.565
## Year1999 0.8006 0.5620 1.42 0.159
## Year2000 0.1277 0.5263 0.24 0.809
## Year2001 0.7726 0.5422 1.42 0.159
## Year2002 0.6085 0.8610 0.71 0.482
## Year2003 0.0697 0.5263 0.13 0.895
## Year2004 0.7406 1.0085 0.73 0.465
## Year2005 0.0135 0.5278 0.03 0.980
## Year2006 0.8880 0.5897 1.51 0.137
## Year2007 0.6040 0.5419 1.11 0.269
```

```

## Year2008          0.6640      0.5496      1.21      0.231
## Year2009          0.5250      0.6149      0.85      0.396
## Year2010          0.0200      0.5831      0.03      0.973
## Year2011          0.3202      0.5719      0.56      0.577
## Year2012          1.9260      0.7700      2.50      0.015 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.682
## Multiple R-squared:  0.265, Adjusted R-squared:  0.09
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 74 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.292  0.913   0.952   0.906   0.983   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.19e-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 3.307 1          1.819
## Year            3.307 15          1.041

```

Residuals from last author



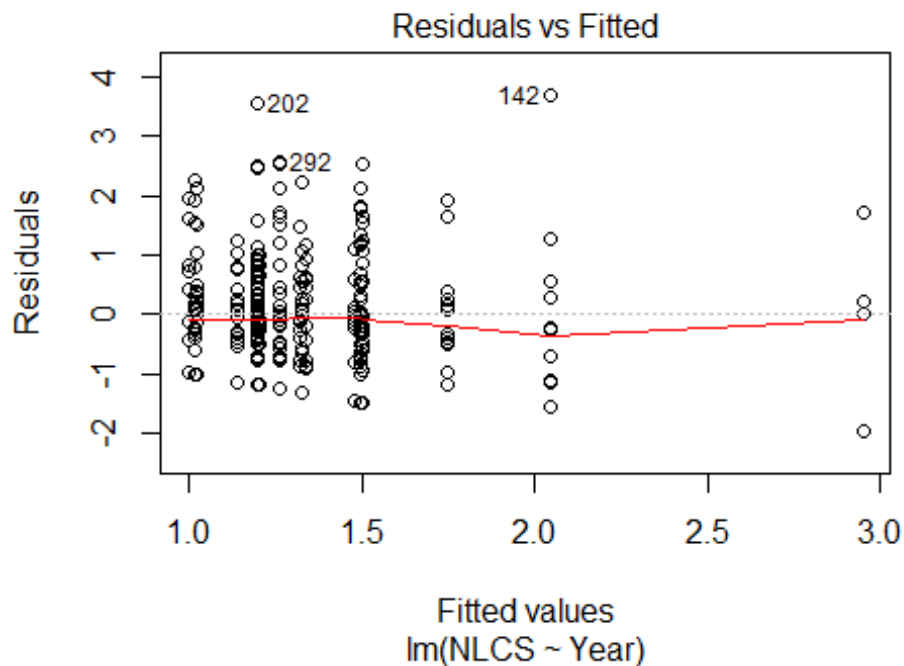
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.42e+00 -4.47e-01 1.53e-16 3.68e-01 2.25e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.474 0.367 1.29 0.2012
## LastAuthorFemale1 -0.355 0.213 -1.67 0.1002
## Year1997 0.500 0.484 1.03 0.3054
## Year1999 0.945 0.420 2.25 0.0278 *
## Year2000 0.272 0.367 0.74 0.4601
## Year2001 0.948 0.408 2.32 0.0233 *
## Year2002 0.798 0.815 0.98 0.3308
## Year2003 0.214 0.367 0.58 0.5607
## Year2004 0.871 0.885 0.98 0.3289
## Year2005 0.191 0.402 0.47 0.6368
## Year2006 1.032 0.457 2.26 0.0272 *
## Year2007 0.714 0.416 1.72 0.0907 .
```

```

## Year2008          0.805      0.429      1.87      0.0653 .
## Year2009          0.728      0.521      1.40      0.1672
## Year2010          0.177      0.447      0.40      0.6932
## Year2011          0.515      0.461      1.12      0.2677
## Year2012          2.105      0.696      3.02      0.0035 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.692
## Multiple R-squared:  0.277, Adjusted R-squared:  0.104
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 75 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.270  0.913  0.960  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      1.19e-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: 84"
## [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
##   15   10   11    5   13   12   15   22   18   11   24   14   25   31   25
## 2011 2012
##   38   45
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   13    9   10    4   12   12   12   20   18   11   24   13   22   30   23
## 2011 2012
##   36   39

```

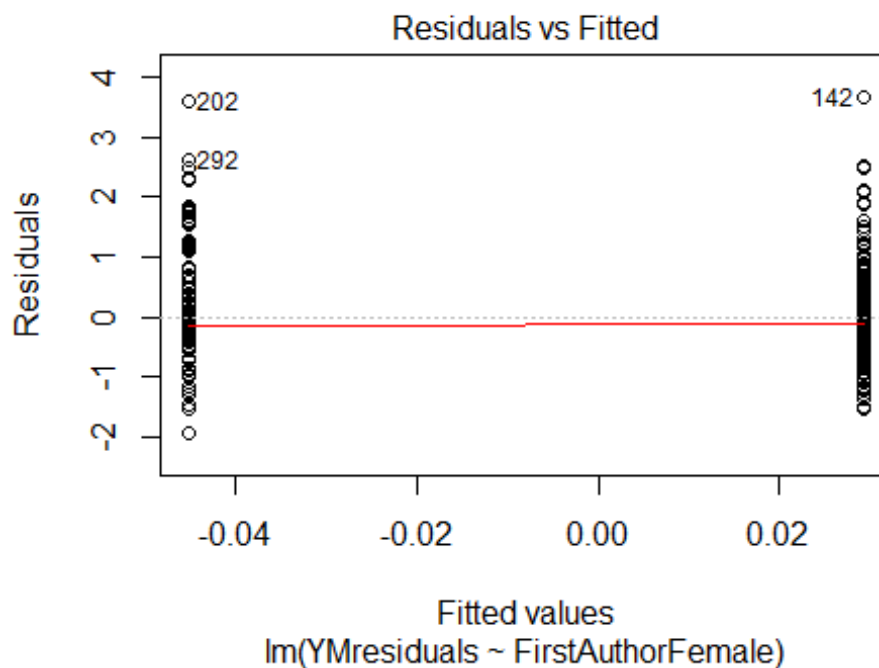
```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 12 9 10 4 12 12 10 18 18 11 24 13 22 29 23
## 2011 2012
## 36 38
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 18, df = 16, p-value = 0.3
```



```
##
## 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: 1.19603316910413"
## [1] "Male first author team size 2018 geometric mean: 1.42466262315448"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
##
## Wilcoxon rank sum test with continuity correction
```

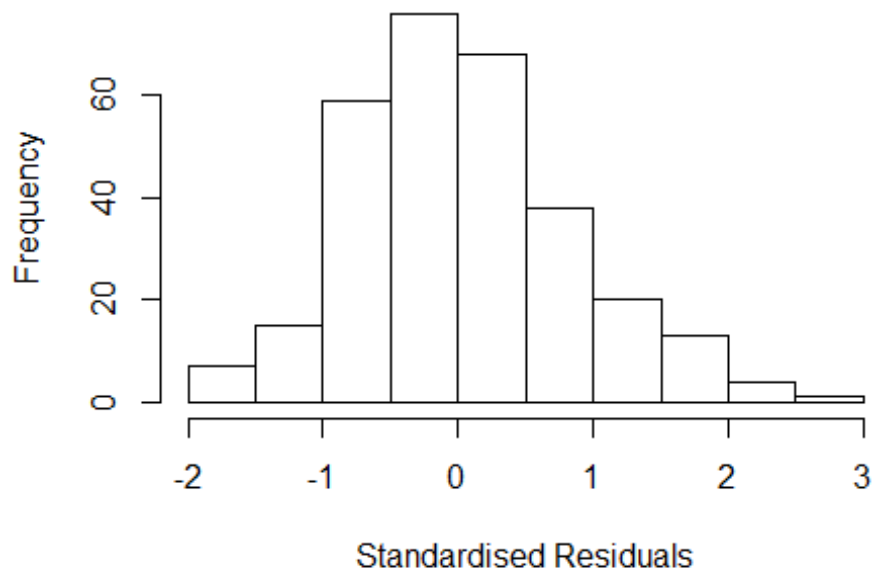
```
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.19623119885132"
## [1] "Male last author team size 2018 geometric mean: 1.44105563982116"

## Warning in wilcox.test.default(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 6.294 1      2.509
## LastAuthorFemale  7.086 1      2.662
## UniqueAuthors    7.175 4      1.279
## Year             11.643 16      1.080
```

Residuals from first and last author and team size



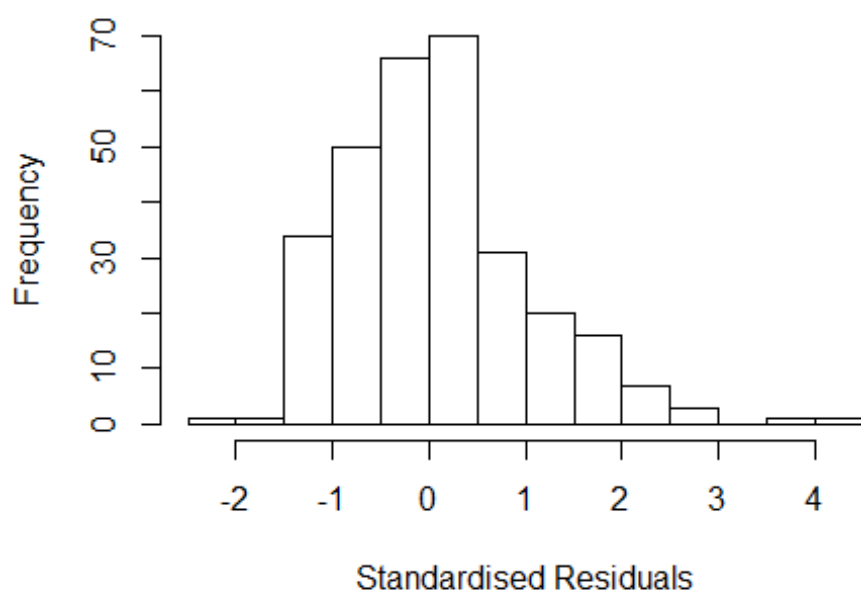
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 142 33646201756 5.737 2005      1210      2      2.645
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9221 -0.5218 -0.0197  0.5061  2.6452
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.6086     0.2191   7.34 2.3e-12 ***
## FirstAuthorFemale1 -0.2126     0.2250  -0.94  0.3456
## LastAuthorFemale1  0.0395     0.2377   0.17  0.8682
## UniqueAuthors2     0.8746     0.2004   4.36 1.8e-05 ***
## UniqueAuthors3     1.3368     0.2083   6.42 5.9e-10 ***
## UniqueAuthors4     1.9297     2.5991   0.74  0.4584
## UniqueAuthors5     1.5340     0.5315   2.89  0.0042 **
## Year1997          -0.5736     0.2697  -2.13  0.0343 *
## Year1998          -0.7513     0.3098  -2.43  0.0159 *
## Year1999           1.4529     0.6022   2.41  0.0165 *
```

```

## Year2000      -0.2982      0.2777      -1.07      0.2839
## Year2001      -0.5648      0.3400      -1.66      0.0978 .
## Year2002      -0.9137      0.3513      -2.60      0.0098 **
## Year2003      -0.5766      0.2943      -1.96      0.0511 .
## Year2004      -0.3841      0.2702      -1.42      0.1563
## Year2005      -0.0508      0.3549      -0.14      0.8863
## Year2006      -0.1557      0.2960      -0.53      0.5994
## Year2007      -0.7691      0.3440      -2.24      0.0262 *
## Year2008      -0.7107      0.2977      -2.39      0.0176 *
## Year2009      -0.3484      0.2731      -1.28      0.2031
## Year2010      -0.7327      0.2630      -2.79      0.0057 **
## Year2011      -0.7290      0.2641      -2.76      0.0062 **
## Year2012      -0.8407      0.2550      -3.30      0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.779
## Multiple R-squared:  0.341, Adjusted R-squared:  0.289
## Convergence in 49 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 279 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.225  0.883  0.948  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      3.32e-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 8.803 1      2.967
## LastAuthorFemale 8.844 1      2.974
## Year      1.677 16      1.016

```


Residuals from first and last author



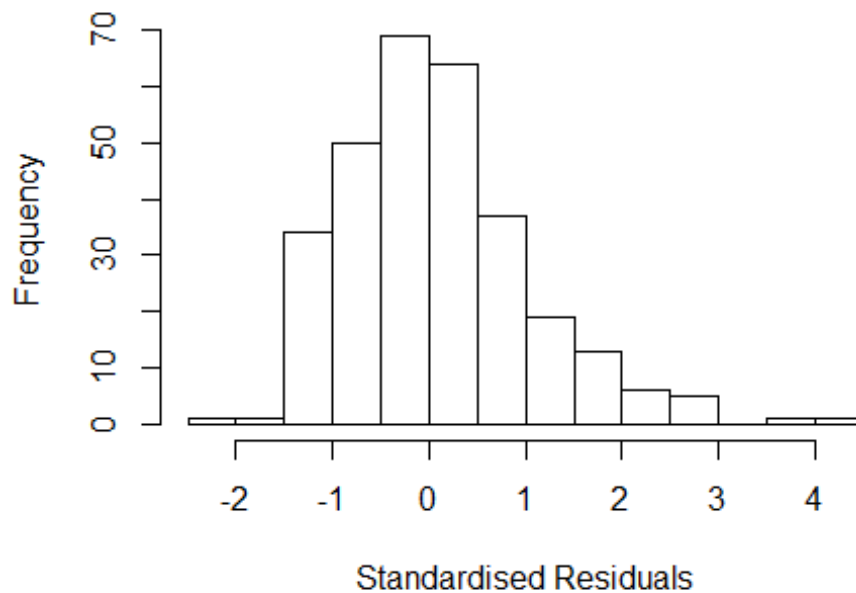
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 142 33646201756 5.737 2005     1210      2      4.065
## 162 33745104091 4.031 2006     1210      2      2.553
## 202 50449104696 4.759 2008     1210      1      3.783
## 292 81855206264 3.843 2011     1210      1      2.815
## 358 84861940126 3.651 2012     1210      1      2.687
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0254 -0.5842 -0.0157  0.5451  4.0651
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.7386     0.2639   6.59 2.2e-10 ***
## FirstAuthorFemale1 -0.3563     0.3027  -1.18  0.2402
## LastAuthorFemale1  0.2403     0.3024   0.79  0.4274
## Year1997          -0.4497     0.3421  -1.31  0.1898
## Year1998          -0.5410     0.3212  -1.68  0.0933 .
## Year1999           1.3907     0.9701   1.43  0.1528
## Year2000          -0.2132     0.3258  -0.65  0.5133
## Year2001          -0.3959     0.3372  -1.17  0.2414
```

```

## Year2002          -1.0466      0.3801    -2.75    0.0063 **
## Year2003          -0.6956      0.3398    -2.05    0.0415 *
## Year2004          -0.5430      0.3040    -1.79    0.0752 .
## Year2005          -0.0667      0.3783    -0.18    0.8601
## Year2006          -0.2610      0.3331    -0.78    0.4339
## Year2007          -0.3929      0.4216    -0.93    0.3522
## Year2008          -0.6464      0.3172    -2.04    0.0425 *
## Year2009          -0.3054      0.3022    -1.01    0.3131
## Year2010          -0.7625      0.3092    -2.47    0.0143 *
## Year2011          -0.5948      0.3186    -1.87    0.0629 .
## Year2012          -0.6587      0.2966    -2.22    0.0271 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.875
## Multiple R-squared:  0.111, Adjusted R-squared:  0.0544
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 113 is an outlier with |weight| <= 0.00026 ( < 0.00033);
## 26 weights are ~= 1. The remaining 274 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0219 0.8660 0.9520 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          3.32e-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 1.243 1          1.115
## Year              1.243 16          1.007

```

Residuals from first author



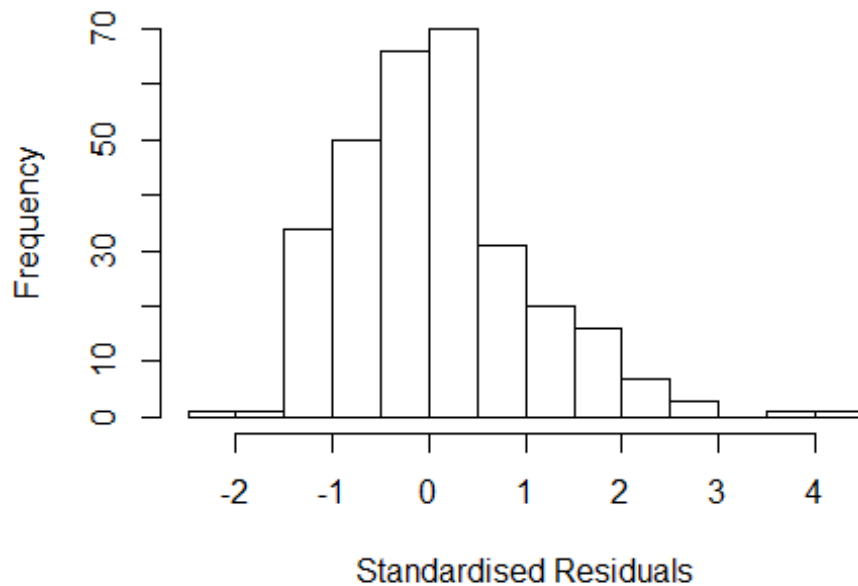
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 142 33646201756 5.737 2005    1210     2    4.065
## 162 33745104091 4.031 2006    1210     2    2.553
## 202 50449104696 4.759 2008    1210     1    3.783
## 292 81855206264 3.843 2011    1210     1    2.815
## 358 84861940126 3.651 2012    1210     1    2.687
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.004 -0.579 -0.038  0.566  4.055
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.7515     0.2629   6.66 1.4e-10 ***
## FirstAuthorFemale1 -0.1406     0.1140  -1.23  0.2184
## Year1997        -0.4263     0.3473  -1.23  0.2206
## Year1998        -0.5247     0.3247  -1.62  0.1072
## Year1999         1.3810     0.8621   1.60  0.1103
## Year2000        -0.2178     0.3258  -0.67  0.5044
## Year2001        -0.3788     0.3384  -1.12  0.2640
## Year2002       -1.0448     0.3784  -2.76  0.0061 **
```

```

## Year2003          -0.6990      0.3388   -2.06   0.0400 *
## Year2004          -0.5430      0.3039   -1.79   0.0751 .
## Year2005          -0.0693      0.3810   -0.18   0.8558
## Year2006          -0.2567      0.3323   -0.77   0.4405
## Year2007          -0.4450      0.3991   -1.11   0.2659
## Year2008          -0.6632      0.3155   -2.10   0.0364 *
## Year2009          -0.3080      0.3026   -1.02   0.3096
## Year2010          -0.7685      0.3066   -2.51   0.0127 *
## Year2011          -0.6017      0.3166   -1.90   0.0584 .
## Year2012          -0.6382      0.2971   -2.15   0.0326 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.889
## Multiple R-squared:  0.11,   Adjusted R-squared:  0.0562
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
##  25 weights are ~ = 1. The remaining 276 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0028  0.8700  0.9530  0.8980  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      3.32e-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 1.329 1          1.153
## Year            1.329 16          1.009

```

Residuals from last author



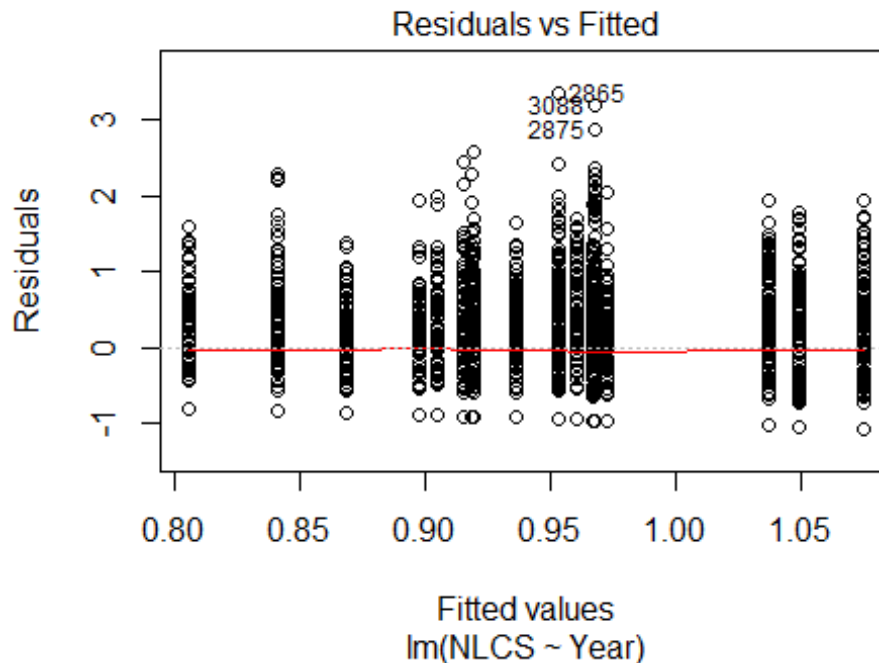
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 142 33646201756 5.737 2005      1210      2      4.065
## 162 33745104091 4.031 2006      1210      2      2.553
## 202 50449104696 4.759 2008      1210      1      3.783
## 292 81855206264 3.843 2011      1210      1      2.815
## 358 84861940126 3.651 2012      1210      1      2.687
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0257 -0.5830 -0.0281  0.5906  4.0732
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.7265     0.2653   6.51 3.5e-10 ***
## LastAuthorFemale1 -0.0712     0.1199  -0.59  0.5533
## Year1997          -0.4105     0.3513  -1.17  0.2435
## Year1998          -0.5054     0.3294  -1.53  0.1260
## Year1999           1.3584     0.8286   1.64  0.1022
## Year2000          -0.2176     0.3266  -0.67  0.5058
## Year2001          -0.3672     0.3427  -1.07  0.2849
## Year2002          -1.0495     0.3836  -2.74  0.0066 **
```

```

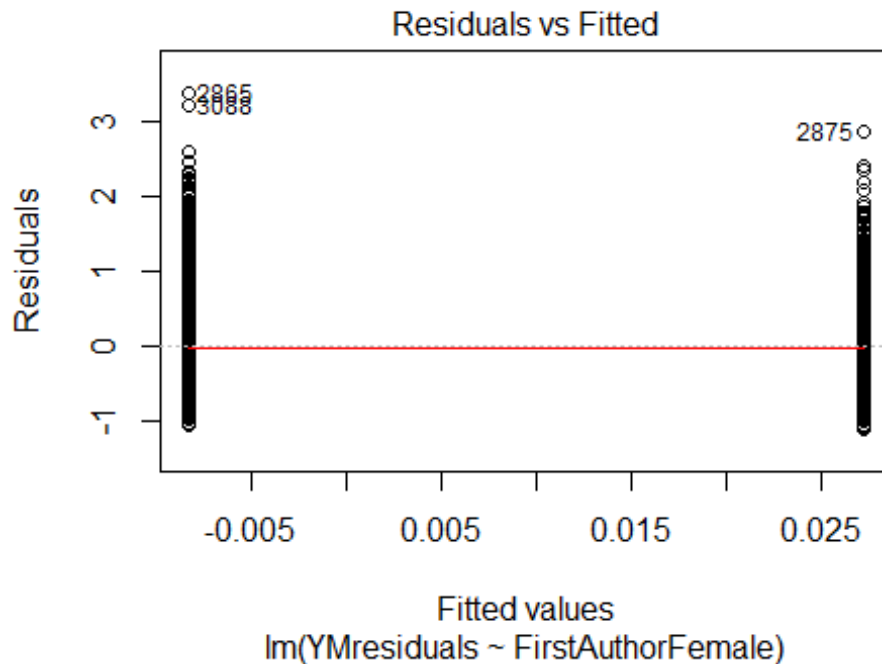
## Year2003      -0.6948      0.3411      -2.04      0.0426 *
## Year2004      -0.5537      0.3048      -1.82      0.0704 .
## Year2005      -0.0627      0.3853      -0.16      0.8708
## Year2006      -0.2685      0.3347      -0.80      0.4231
## Year2007      -0.4517      0.4011      -1.13      0.2610
## Year2008      -0.6652      0.3168      -2.10      0.0366 *
## Year2009      -0.3201      0.3075      -1.04      0.2987
## Year2010      -0.7767      0.3085      -2.52      0.0124 *
## Year2011      -0.5976      0.3192      -1.87      0.0622 .
## Year2012      -0.6234      0.2989      -2.09      0.0379 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.904
## Multiple R-squared:  0.106, Adjusted R-squared:  0.0521
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
##  21 weights are ~ = 1. The remaining 280 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0058 0.8780 0.9570 0.9020 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      3.32e-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: 301"
## [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
## 121 132 117 114 123 139 154 144 163 190 195 202 220 204 214
## 2011 2012
## 296 294

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 105 113 94 102 105 117 132 127 146 167 172 169 190 177 189
## 2011 2012
## 268 257
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 104 113 93 102 105 116 129 127 144 166 171 167 187 176 187
## 2011 2012
## 265 254
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 77, df = 16, p-value = 5e-10
```

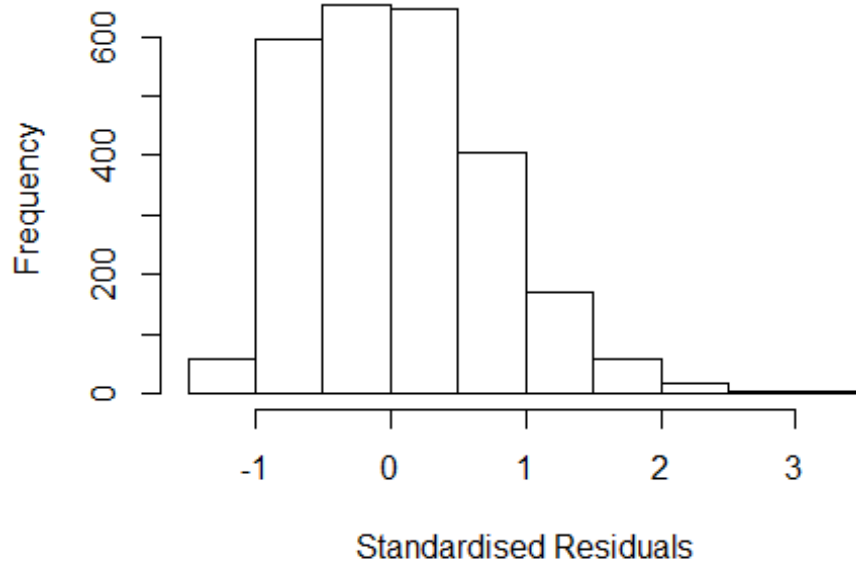


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.3, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 1.20741176943837"
## [1] "Male first author team size 2018 geometric mean: 1.07297392253496"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9700, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.16648845345186"
## [1] "Male last author team size 2018 geometric mean: 1.09414601069819"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9200, 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 6.962 1          2.639
## LastAuthorFemale  7.037 1          2.653
## UniqueAuthors    1.159 4          1.019
## Year              1.201 16         1.006
```


Residuals from first and last author and team size



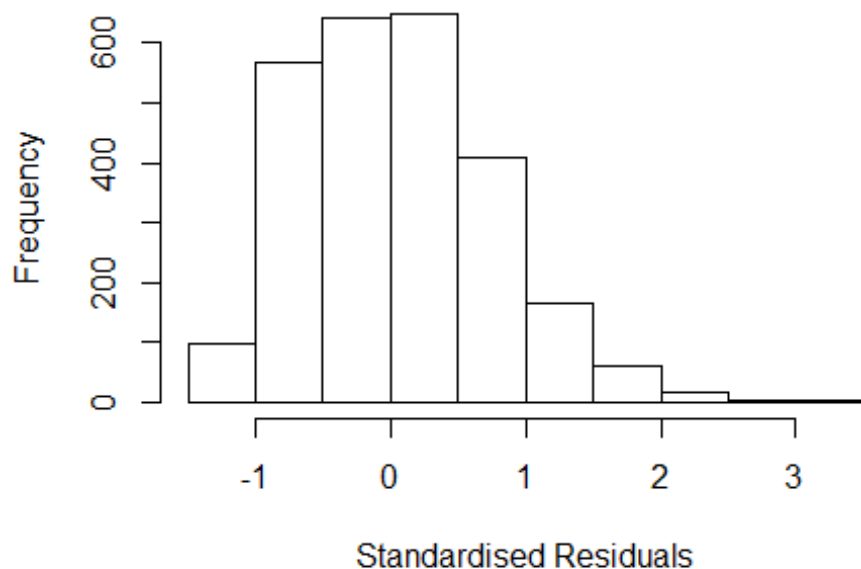
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 901  29644432773 3.510 2002    1202     2    2.634
## 2865 79952786437 4.310 2011    1211     3    3.225
## 2875 84862682664 3.841 2012    1211     1    2.993
## 3088 84863310367 4.162 2012    1211     1    3.353
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33659 -0.50328 -0.00394  0.50091  3.35254
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.92401    0.06432   14.37 < 2e-16 ***
## FirstAuthorFemale1 -0.10851    0.09279   -1.17  0.24235
## LastAuthorFemale1  0.14662    0.09274    1.58  0.11401
## UniqueAuthors2     0.19782    0.05819    3.40  0.00069 ***
## UniqueAuthors3     0.13385    0.11063    1.21  0.22644
## UniqueAuthors4     0.50880    0.14270    3.57  0.00037 ***
## UniqueAuthors5     0.34113    0.21637    1.58  0.11501
```

```

## Year1997      -0.06388    0.09420   -0.68  0.49772
## Year1998      -0.16049    0.09356   -1.72  0.08640 .
## Year1999      -0.05489    0.08664   -0.63  0.52643
## Year2000      -0.08394    0.08975   -0.94  0.34971
## Year2001       0.01637    0.08325    0.20  0.84416
## Year2002      -0.04815    0.09596   -0.50  0.61584
## Year2003      -0.00599    0.09265   -0.06  0.94849
## Year2004      -0.06809    0.08707   -0.78  0.43425
## Year2005      -0.05737    0.08447   -0.68  0.49708
## Year2006      -0.14512    0.08295   -1.75  0.08033 .
## Year2007      -0.01619    0.08138   -0.20  0.84231
## Year2008       0.06828    0.08556    0.80  0.42489
## Year2009       0.07633    0.08398    0.91  0.36354
## Year2010       0.06814    0.08583    0.79  0.42730
## Year2011      -0.03696    0.08084   -0.46  0.64760
## Year2012      -0.11455    0.08839   -1.30  0.19509
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.759
## Multiple R-squared:  0.0183, Adjusted R-squared:  0.00995
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 237 weights are ~= 1. The remaining 2369 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0124 0.8790 0.9480 0.9160 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 7.172 1 2.678
## LastAuthorFemale 7.245 1 2.692
## Year 1.062 16 1.002

```

Residuals from first and last author



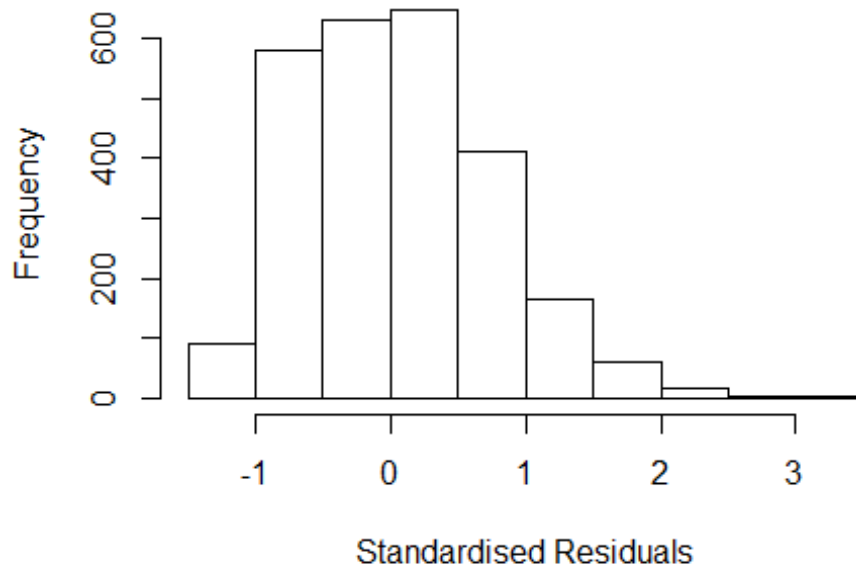
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 901  29644432773 3.510 2002    1202     2    2.624
## 2865 79952786437 4.310 2011    1211     3    3.410
## 2875 84862682664 3.841 2012    1211     1    2.960
## 3088 84863310367 4.162 2012    1211     1    3.331
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.17685 -0.51434 -0.00265  0.49893  3.41033
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.94169    0.06462   14.57  <2e-16 ***
## FirstAuthorFemale1 -0.10525    0.09495   -1.11    0.268
## LastAuthorFemale1  0.15472    0.09494    1.63    0.103
## Year1997        -0.07204    0.09514   -0.76    0.449
## Year1998        -0.17373    0.09370   -1.85    0.064 .
## Year1999        -0.05769    0.08742   -0.66    0.509
## Year2000        -0.08406    0.09020   -0.93    0.351
## Year2001         0.00899    0.08348    0.11    0.914
## Year2002        -0.05536    0.09628   -0.57    0.565
```

```

## Year2003      -0.00886    0.09326   -0.10    0.924
## Year2004      -0.06985    0.08777   -0.80    0.426
## Year2005      -0.06930    0.08486   -0.82    0.414
## Year2006      -0.14986    0.08320   -1.80    0.072 .
## Year2007      -0.02706    0.08153   -0.33    0.740
## Year2008       0.06890    0.08584    0.80    0.422
## Year2009       0.07865    0.08474    0.93    0.353
## Year2010       0.08045    0.08589    0.94    0.349
## Year2011      -0.04201    0.08084   -0.52    0.603
## Year2012      -0.11033    0.08895   -1.24    0.215
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.761
## Multiple R-squared:  0.0115, Adjusted R-squared:  0.00466
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 237 weights are ~= 1. The remaining 2369 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0074 0.8780 0.9490 0.9160 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1          1.012
## Year              1.024 16          1.001

```

Residuals from first author



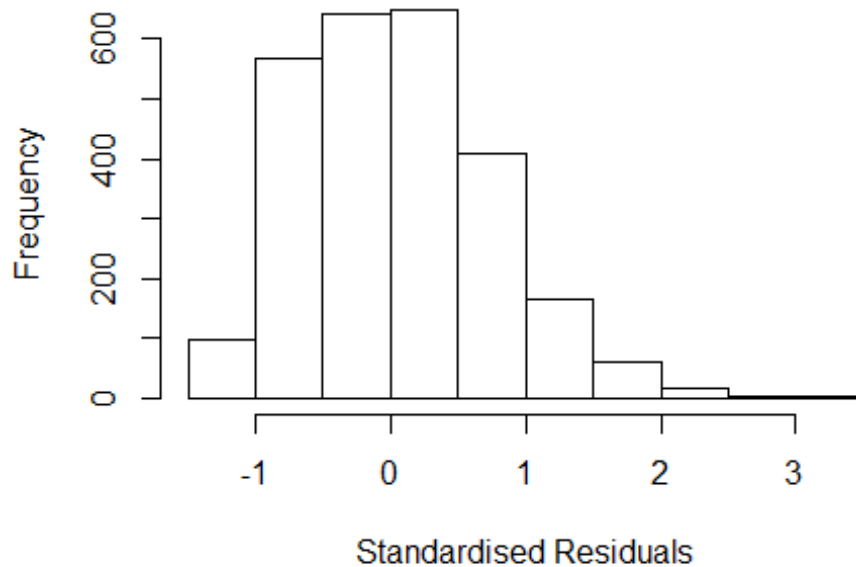
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 901  29644432773 3.510 2002    1202     2    2.624
## 2865 79952786437 4.310 2011    1211     3    3.410
## 2875 84862682664 3.841 2012    1211     1    2.960
## 3088 84863310367 4.162 2012    1211     1    3.331
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q       Max
## -1.064751 -0.513976  0.000544  0.501485  3.407485
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9420    0.0646   14.58  <2e-16 ***
## FirstAuthorFemale1 0.0377    0.0358    1.05   0.293
## Year1997       -0.0691    0.0949   -0.73   0.467
## Year1998       -0.1726    0.0939   -1.84   0.066 .
## Year1999       -0.0577    0.0876   -0.66   0.510
## Year2000       -0.0866    0.0900   -0.96   0.336
## Year2001        0.0115    0.0834    0.14   0.890
## Year2002       -0.0536    0.0962   -0.56   0.577
## Year2003       -0.0101    0.0932   -0.11   0.914
```

```

## Year2004          -0.0675      0.0880    -0.77      0.443
## Year2005          -0.0660      0.0849    -0.78      0.437
## Year2006          -0.1476      0.0833    -1.77      0.076 .
## Year2007          -0.0225      0.0818    -0.27      0.784
## Year2008           0.0739      0.0860      0.86      0.390
## Year2009           0.0851      0.0845      1.01      0.314
## Year2010           0.0811      0.0858      0.95      0.344
## Year2011          -0.0395      0.0809    -0.49      0.625
## Year2012          -0.1028      0.0890    -1.16      0.248
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.763
## Multiple R-squared:  0.0103, Adjusted R-squared:  0.00384
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 241 weights are ~= 1. The remaining 2365 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0082 0.8780 0.9490 0.9160 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.034 1      1.017
## Year      1.034 16      1.001

```

Residuals from last author



```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 901  29644432773 3.510 2002    1202     2    2.624
## 2865 79952786437 4.310 2011    1211     3    3.410
## 2875 84862682664 3.841 2012    1211     1    2.960
## 3088 84863310367 4.162 2012    1211     1    3.331
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.07775 -0.51359 -0.00112  0.49930  3.41230
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.93825    0.06458   14.53  <2e-16 ***
## LastAuthorFemale1 0.05766    0.03594    1.60   0.109
## Year1997       -0.07025    0.09495   -0.74   0.459
## Year1998       -0.17324    0.09380   -1.85   0.065 .
## Year1999       -0.05652    0.08764   -0.64   0.519
## Year2000       -0.08476    0.09000   -0.94   0.346
## Year2001        0.01070    0.08341    0.13   0.898
## Year2002       -0.05322    0.09629   -0.55   0.581
## Year2003       -0.00875    0.09325   -0.09   0.925
```

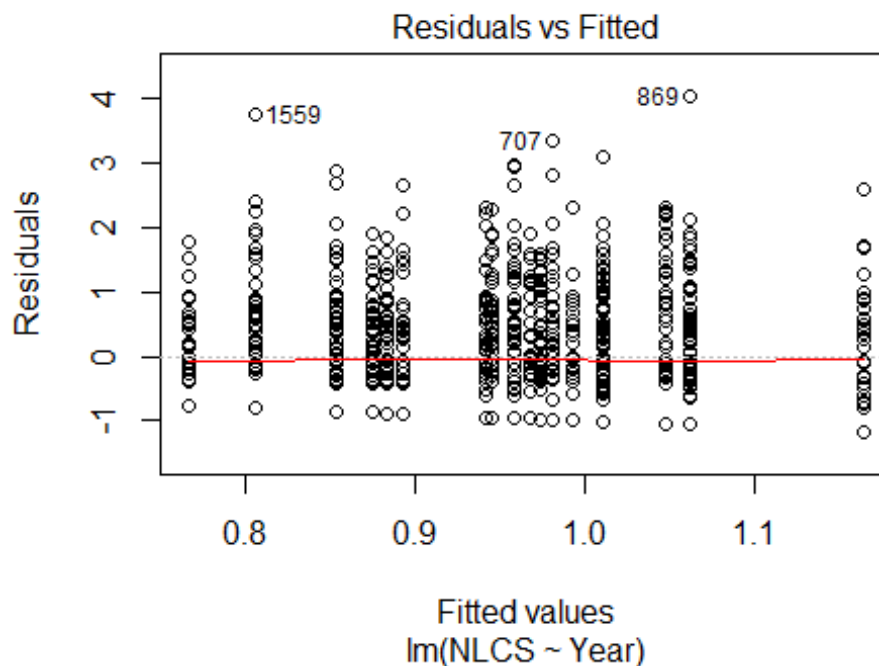
```

## Year2004          -0.06798      0.08791      -0.77      0.439
## Year2005          -0.06737      0.08485      -0.79      0.427
## Year2006          -0.14811      0.08318      -1.78      0.075 .
## Year2007          -0.02435      0.08165      -0.30      0.766
## Year2008           0.07197      0.08588       0.84      0.402
## Year2009           0.08184      0.08458       0.97      0.333
## Year2010           0.08033      0.08575       0.94      0.349
## Year2011          -0.04054      0.08076      -0.50      0.616
## Year2012          -0.10564      0.08892      -1.19      0.235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.762
## Multiple R-squared:  0.011, Adjusted R-squared:  0.00446
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 237 weights are ~= 1. The remaining 2369 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0076 0.8780 0.9490 0.9160 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2606"
## [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
##   61   55   52   70   51   45   89   79  105  102  103  108  125   77   99
## 2011 2012
##  179  134
##

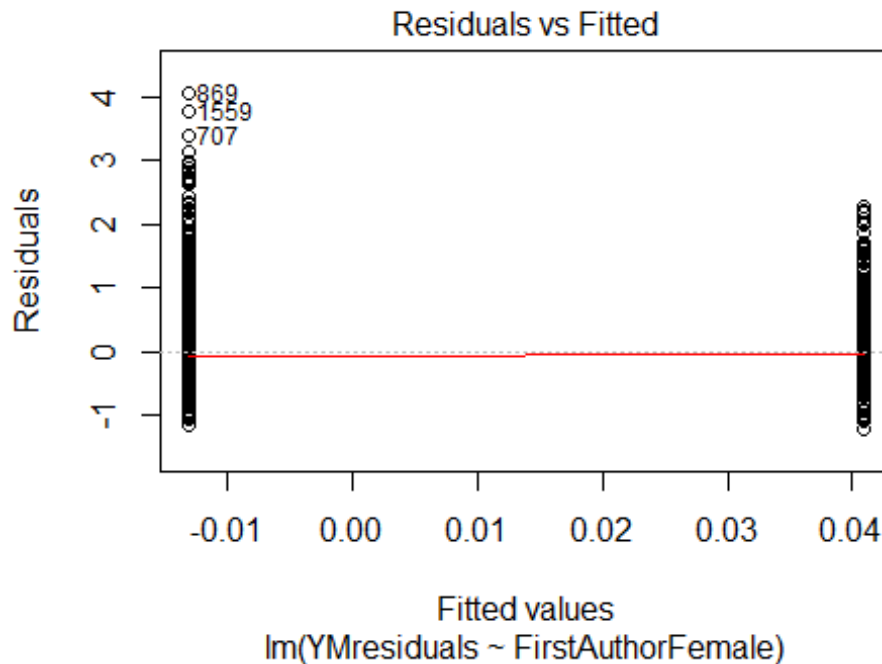
```



```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    54    48    44    63    43    40    81    71    93    86    94    90   116    71    84
## 2011 2012
##   160   115
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    54    47    44    63    42    38    81    71    93    85    93    89   116    70    84
## 2011 2012
##   159   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 = 30, df = 16, p-value = 0.02
```

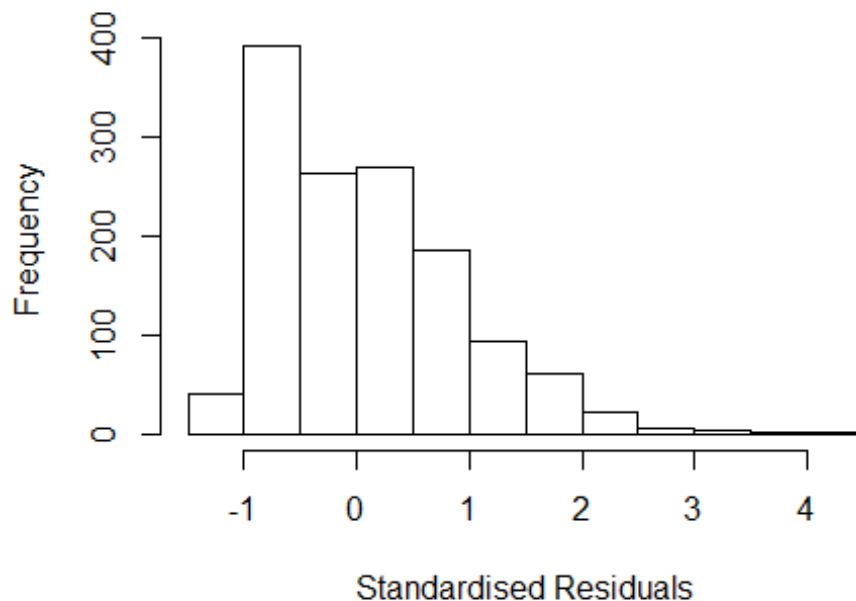


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.54, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 1.10408951367381"
## [1] "Male first author team size 2018 geometric mean: 1.05357292025692"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.07493403181954"
## [1] "Male last author team size 2018 geometric mean: 1.06531889677829"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale  9.954  1          3.155
## LastAuthorFemale 10.289  1          3.208
## UniqueAuthors     1.957  4          1.088
## Year              1.381 16          1.010
```

Residuals from first and last author and team size



```
## [1] "List of 12 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 382   0036748063 3.898 2002    1212      1    3.098
## 391  19144372570 3.928 2002    1212      1    3.128
## 413   85006169108 3.625 2002    1212      1    2.825
## 500   63849171517 3.537 2003    1208      2    2.740
## 704   60950284430 3.788 2005    1208      2    2.960
## 707   61049203364 4.342 2005    1208      2    3.514
## 869   62249126614 5.098 2007    1202      3    4.170
## 938   57749141353 4.118 2008    1212      1    3.195
## 1320  80052368976 3.559 2011    1202      2    2.778
## 1323  80052203106 3.724 2011    1212      1    2.943
## 1511  84864197548 3.215 2012    1212      1    2.569
## 1559  84859974821 4.552 2012    1212      1    3.906
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2144 -0.7814 -0.0424  0.5874  4.1699
##
## Coefficients:
```

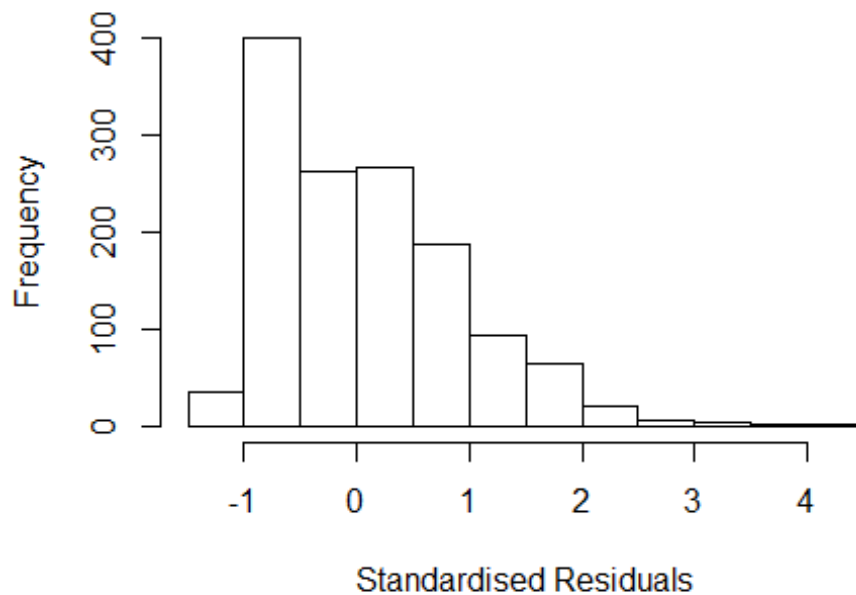
```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.83482    0.13103   6.37 2.6e-10 ***
## FirstAuthorFemale1 -0.01589    0.17475  -0.09  0.928
## LastAuthorFemale1  0.09374    0.17528   0.53  0.593
## UniqueAuthors2     0.16961    0.13952   1.22  0.224
## UniqueAuthors3     0.19196    0.16232   1.18  0.237
## UniqueAuthors4    -0.58181    0.27114  -2.15  0.032 *
## UniqueAuthors5    -0.29466    0.64308  -0.46  0.647
## Year1997           0.30172    0.18816   1.60  0.109
## Year1998          -0.11643    0.16533  -0.70  0.481
## Year1999           0.02522    0.16888   0.15  0.881
## Year2000           0.09414    0.18610   0.51  0.613
## Year2001           0.09090    0.16781   0.54  0.588
## Year2002          -0.03454    0.16933  -0.20  0.838
## Year2003          -0.03806    0.16185  -0.24  0.814
## Year2004           0.09075    0.17033   0.53  0.594
## Year2005          -0.00663    0.16580  -0.04  0.968
## Year2006           0.07401    0.15969   0.46  0.643
## Year2007           0.09328    0.16163   0.58  0.564
## Year2008           0.08825    0.15080   0.59  0.559
## Year2009          -0.01746    0.15932  -0.11  0.913
## Year2010          -0.01885    0.15556  -0.12  0.904
## Year2011          -0.05341    0.14445  -0.37  0.712
## Year2012          -0.18912    0.15373  -1.23  0.219
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.854
## Multiple R-squared:  0.0192, Adjusted R-squared:  0.00281
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 746 is an outlier with |weight| = 0 ( < 7.5e-05);
## 111 weights are ~= 1. The remaining 1230 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0021 0.8960 0.9330 0.9080 0.9840 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           7.45e-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 8.849 1      2.975
## LastAuthorFemale  8.766 1      2.961
## Year              1.076 16      1.002
```

Residuals from first and last author



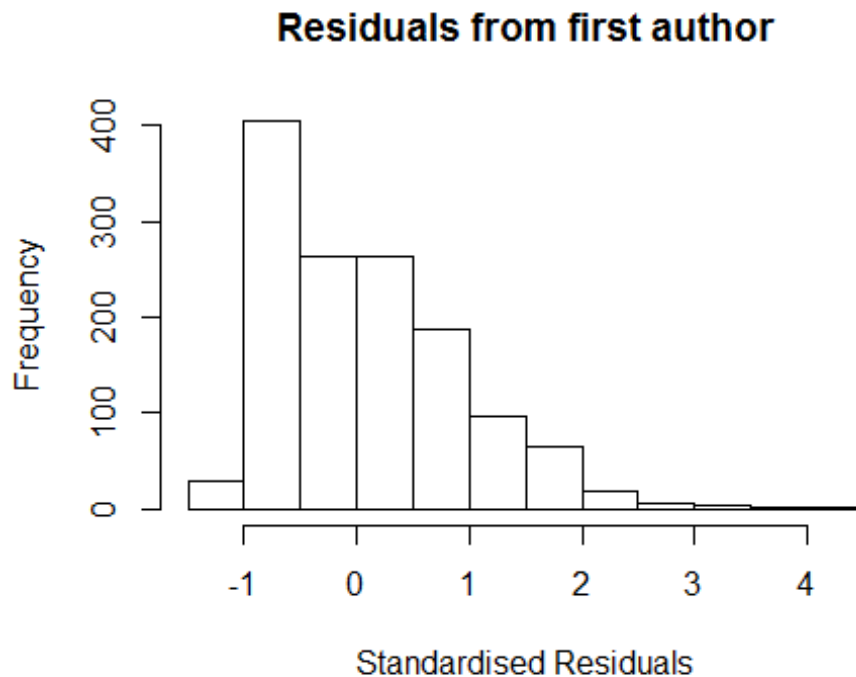
```
## [1] "List of 13 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 82      0000211683 3.761 1997    1212      1    2.640
## 382     0036748063 3.898 2002    1212      1    3.089
## 391    19144372570 3.928 2002    1212      1    3.119
## 413     85006169108 3.625 2002    1212      1    2.816
## 500     63849171517 3.537 2003    1208      2    2.731
## 704     60950284430 3.788 2005    1208      2    2.955
## 707     61049203364 4.342 2005    1208      2    3.509
## 869     62249126614 5.098 2007    1202      3    4.161
## 938     57749141353 4.118 2008    1212      1    3.182
## 1320    80052368976 3.559 2011    1202      2    2.771
## 1323    80052203106 3.724 2011    1212      1    2.936
## 1511    84864197548 3.215 2012    1212      1    2.557
## 1559    84859974821 4.552 2012    1212      1    3.894
##
## 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
## -1.2077 -0.7881 -0.0451  0.5887  4.1611
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.8458      0.1311   6.45 1.6e-10 ***
## FirstAuthorFemale1 0.0376      0.1649   0.23  0.82
## LastAuthorFemale1  0.0489      0.1635   0.30  0.77
## Year1997          0.2754      0.1898   1.45  0.15
## Year1998         -0.1261      0.1647  -0.77  0.44
## Year1999          0.0132      0.1694   0.08  0.94
## Year2000          0.0850      0.1870   0.45  0.65
## Year2001          0.0777      0.1680   0.46  0.64
## Year2002         -0.0365      0.1688  -0.22  0.83
## Year2003         -0.0401      0.1622  -0.25  0.80
## Year2004          0.0942      0.1706   0.55  0.58
## Year2005         -0.0128      0.1633  -0.08  0.94
## Year2006          0.0680      0.1597   0.43  0.67
## Year2007          0.0911      0.1614   0.56  0.57
## Year2008          0.0902      0.1508   0.60  0.55
## Year2009         -0.0247      0.1601  -0.15  0.88
## Year2010         -0.0220      0.1545  -0.14  0.89
## Year2011         -0.0577      0.1441  -0.40  0.69
## Year2012         -0.1881      0.1555  -1.21  0.23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.857
## Multiple R-squared:  0.0146, Adjusted R-squared:  0.00122
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 746 is an outlier with |weight| = 0 ( < 7.5e-05);
## 112 weights are ~ = 1. The remaining 1229 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0035 0.8940 0.9320 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          7.45e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd

```

```
##          0          1000          0
##          psi          subsampling          cov
##          "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.03 1          1.015
## Year              1.03 16          1.001
```



```
## [1] "List of 13 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 82      0000211683 3.761 1997     1212      1     2.640
## 382     0036748063 3.898 2002     1212      1     3.089
## 391    19144372570 3.928 2002     1212      1     3.119
## 413    85006169108 3.625 2002     1212      1     2.816
## 500    63849171517 3.537 2003     1208      2     2.731
## 704    60950284430 3.788 2005     1208      2     2.955
## 707    61049203364 4.342 2005     1208      2     3.509
## 869    62249126614 5.098 2007     1202      3     4.161
## 938    57749141353 4.118 2008     1212      1     3.182
## 1320   80052368976 3.559 2011     1202      2     2.771
## 1323   80052203106 3.724 2011     1212      1     2.936
## 1511   84864197548 3.215 2012     1212      1     2.557
## 1559   84859974821 4.552 2012     1212      1     3.894
##
```

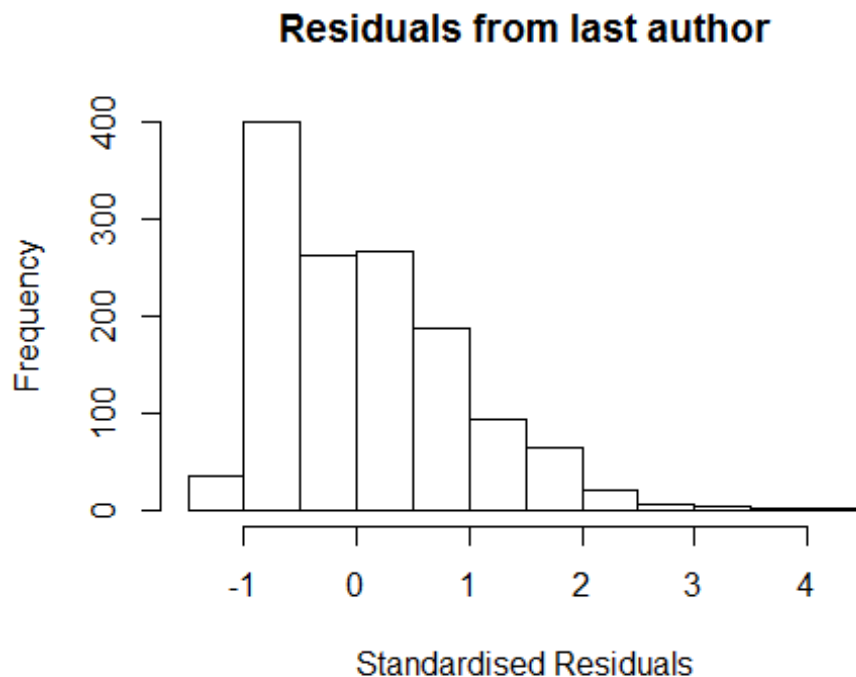
```

## 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
## -1.2054 -0.7896 -0.0466  0.5824  4.1617
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.8467     0.1311   6.46 1.5e-10 ***
## FirstAuthorFemale1 0.0828     0.0563   1.47  0.14
## Year1997          0.2760     0.1896   1.46  0.15
## Year1998         -0.1260     0.1648  -0.76  0.44
## Year1999          0.0135     0.1695   0.08  0.94
## Year2000          0.0858     0.1871   0.46  0.65
## Year2001          0.0778     0.1679   0.46  0.64
## Year2002         -0.0357     0.1688  -0.21  0.83
## Year2003         -0.0406     0.1621  -0.25  0.80
## Year2004          0.0959     0.1706   0.56  0.57
## Year2005         -0.0136     0.1631  -0.08  0.93
## Year2006          0.0677     0.1598   0.42  0.67
## Year2007          0.0896     0.1612   0.56  0.58
## Year2008          0.0911     0.1508   0.60  0.55
## Year2009         -0.0253     0.1601  -0.16  0.87
## Year2010         -0.0213     0.1545  -0.14  0.89
## Year2011         -0.0571     0.1440  -0.40  0.69
## Year2012         -0.1882     0.1554  -1.21  0.23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.857
## Multiple R-squared:  0.0145, Adjusted R-squared:  0.00185
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 746 is an outlier with |weight| = 0 ( < 7.5e-05);
## 117 weights are ~ = 1. The remaining 1224 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0035 0.8940 0.9310 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      7.45e-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 1.019 1      1.009
## Year              1.019 16      1.001
```



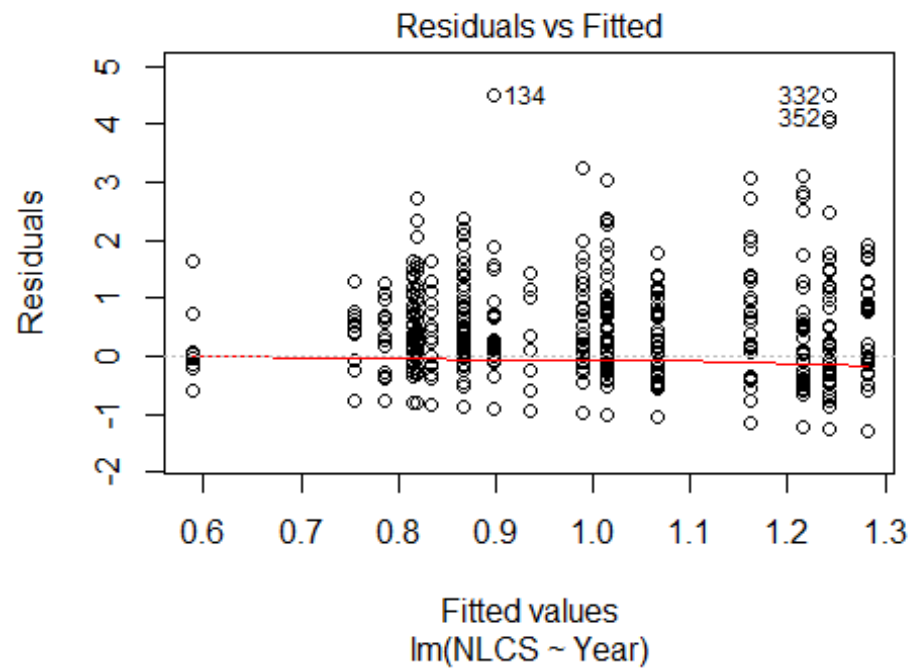
```
## [1] "List of 13 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 82      0000211683 3.761 1997      1212      1      2.640
## 382     0036748063 3.898 2002      1212      1      3.089
## 391     19144372570 3.928 2002      1212      1      3.119
## 413     85006169108 3.625 2002      1212      1      2.816
## 500     63849171517 3.537 2003      1208      2      2.731
## 704     60950284430 3.788 2005      1208      2      2.955
## 707     61049203364 4.342 2005      1208      2      3.509
## 869     62249126614 5.098 2007      1202      3      4.161
## 938     57749141353 4.118 2008      1212      1      3.182
## 1320    80052368976 3.559 2011      1202      2      2.771
## 1323    80052203106 3.724 2011      1212      1      2.936
## 1511    84864197548 3.215 2012      1212      1      2.557
## 1559    84859974821 4.552 2012      1212      1      3.894
```

```
##
## 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
## -1.2045 -0.7884 -0.0454  0.5879  4.1594
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8467    0.1311    6.46 1.5e-10 ***
## LastAuthorFemale1 0.0832    0.0558    1.49   0.14
## Year1997        0.2746    0.1899    1.45   0.15
## Year1998       -0.1261    0.1647   -0.77   0.44
## Year1999        0.0135    0.1691    0.08   0.94
## Year2000        0.0843    0.1868    0.45   0.65
## Year2001        0.0777    0.1679    0.46   0.64
## Year2002       -0.0371    0.1687   -0.22   0.83
## Year2003       -0.0396    0.1623   -0.24   0.81
## Year2004        0.0926    0.1701    0.54   0.59
## Year2005       -0.0117    0.1632   -0.07   0.94
## Year2006        0.0682    0.1596    0.43   0.67
## Year2007        0.0919    0.1613    0.57   0.57
## Year2008        0.0896    0.1508    0.59   0.55
## Year2009       -0.0243    0.1600   -0.15   0.88
## Year2010       -0.0221    0.1545   -0.14   0.89
## Year2011       -0.0583    0.1440   -0.40   0.69
## Year2012       -0.1882    0.1556   -1.21   0.23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.857
## Multiple R-squared:  0.0146, Adjusted R-squared:  0.00194
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 746 is an outlier with |weight| = 0 ( < 7.5e-05);
## 106 weights are ~ 1. The remaining 1235 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0037 0.8940 0.9320 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      7.45e-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
```

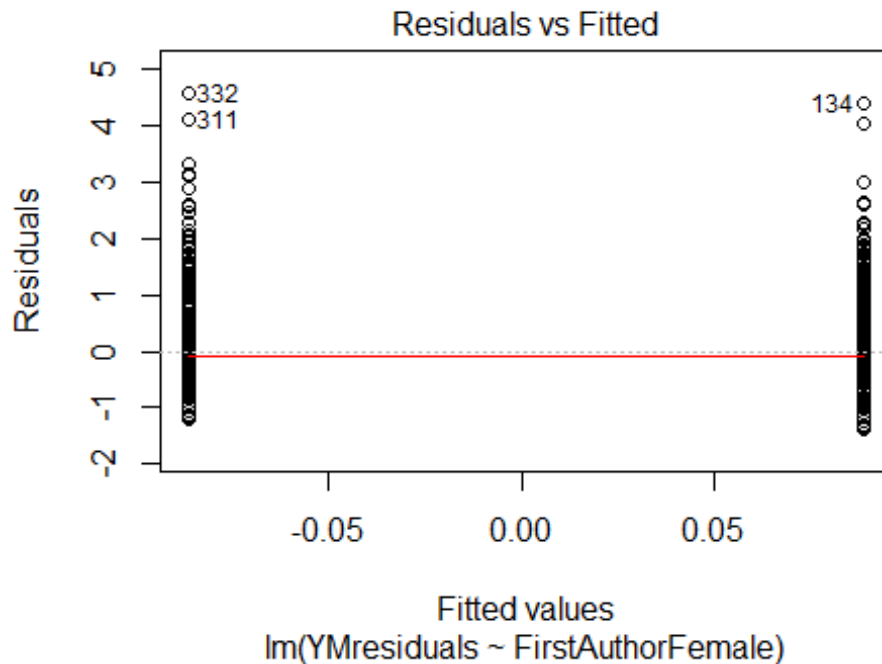
```

##           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: 1342"
## [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
##   20   21   17   20   16   27   52   51   54   76   65   55   73   70   81
## 2011 2012
##   104  125
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18   19   15   18   14   24   41   44   45   68   55   44   62   60   76
## 2011 2012
##    91   109
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18   18   15   17   14   24   40   44   45   68   54   44   61   59   76
## 2011 2012
##    89   106
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 62, df = 16, p-value = 3e-07

```

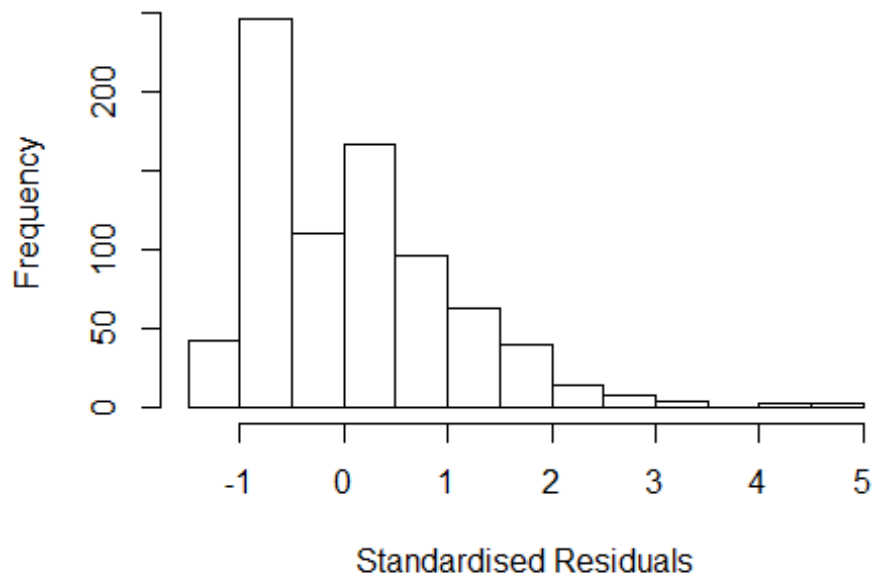


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.43, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 1.16551053761983"
## [1] "Male first author team size 2018 geometric mean: 1.08233606525237"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2600, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.17302707216875"
## [1] "Male last author team size 2018 geometric mean: 1.07031995744419"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2600, 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 5.991 1          2.448
## LastAuthorFemale  6.233 1          2.497
## UniqueAuthors    1.553 4          1.057
## Year              1.655 16         1.016
```

Residuals from first and last author and team size



```
## [1] "List of 16 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 134 60949365420 5.396 2002    1213      1    4.532
## 187 60950331495 4.220 2003    1213      1    3.403
## 240 60950108115 3.874 2004    1213      1    2.808
## 276 60249093788 4.227 2004    1213      1    3.325
## 311 34247665578 5.284 2005    1202      4    4.353
## 332 33646201756 5.737 2005    1210      2    4.815
## 339 84857259490 3.713 2005    1213      1    2.782
## 352 60949545379 5.366 2005    1208      2    4.271
## 378 61049412680 3.958 2006    1213      1    2.872
## 379 61049460890 3.745 2006    1213      1    2.823
## 381 64949189141 4.319 2006    1213      1    3.233
## 425 85014698682 3.958 2006    1213      1    2.872
## 555 77649230762 3.539 2008    1213      2    2.779
## 742 84857704004 4.034 2011    1208      2    3.231
## 820 79953724832 3.382 2011    1210      2    2.722
## 925 84863108187 3.247 2012    1213      1    2.582
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
```

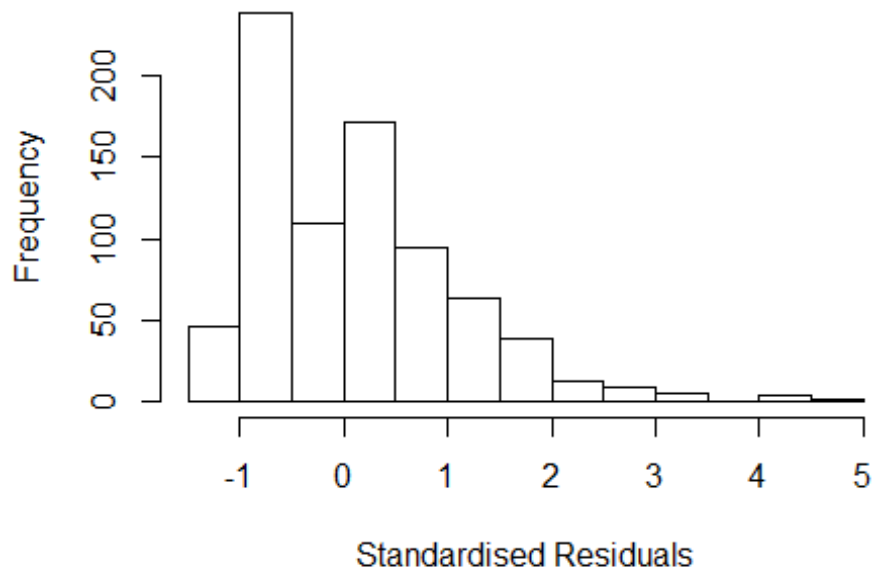
```

##      Min      1Q   Median      3Q      Max
## -1.32960 -0.70190 -0.00611  0.64739  4.81523
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.64030    0.18041   3.55 0.00041 ***
## FirstAuthorFemale1 0.37188    0.15637   2.38 0.01764 *
## LastAuthorFemale1 -0.20768    0.15796  -1.31 0.18898
## UniqueAuthors2    -0.14311    0.12336  -1.16 0.24637
## UniqueAuthors3     0.63089    0.17530   3.60 0.00034 ***
## UniqueAuthors4     0.76686    0.27407   2.80 0.00527 **
## UniqueAuthors5    -0.00929    0.94409  -0.01 0.99215
## Year1997          -0.02507    0.24024  -0.10 0.91691
## Year1998          -0.23635    0.23896  -0.99 0.32294
## Year1999           0.17786    0.21122   0.84 0.40001
## Year2000           0.19292    0.33143   0.58 0.56069
## Year2001           0.06056    0.23210   0.26 0.79421
## Year2002          -0.00521    0.21520  -0.02 0.98068
## Year2003           0.17689    0.23256   0.76 0.44712
## Year2004           0.26162    0.25394   1.03 0.30322
## Year2005           0.29076    0.22734   1.28 0.20128
## Year2006           0.28141    0.23175   1.21 0.22500
## Year2007           0.52511    0.24153   2.17 0.03000 *
## Year2008          -0.04494    0.21291  -0.21 0.83288
## Year2009           0.26749    0.21177   1.26 0.20692
## Year2010           0.06160    0.19693   0.31 0.75451
## Year2011           0.16264    0.20582   0.79 0.42963
## Year2012           0.02503    0.20069   0.12 0.90079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.899
## Multiple R-squared:  0.0524, Adjusted R-squared:  0.0253
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 4 observations c(107,251,268,285)
## are outliers with |weight| = 0 ( < 0.00013);
## 50 weights are ~= 1. The remaining 738 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.120  0.897   0.936   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           1.26e-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 5.641 1      2.375
## LastAuthorFemale 5.725 1      2.393
## Year              1.269 16      1.007
```

Residuals from first and last author



```
## [1] "List of 17 outliers with residuals above 2.5"
##       ScopusId  NLCS Year OneField Fields residuals
## 134 60949365420 5.396 2002    1213     1    4.387
## 187 60950331495 4.220 2003    1213     1    3.401
## 240 60950108115 3.874 2004    1213     1    2.803
## 276 60249093788 4.227 2004    1213     1    3.322
## 311 34247665578 5.284 2005    1202     4    4.374
## 332 33646201756 5.737 2005    1210     2    4.827
## 339 84857259490 3.713 2005    1213     1    2.803
## 352 60949545379 5.366 2005    1208     2    4.290
## 378 61049412680 3.958 2006    1213     1    2.876
## 379 61049460890 3.745 2006    1213     1    2.829
## 381 64949189141 4.319 2006    1213     1    3.237
```



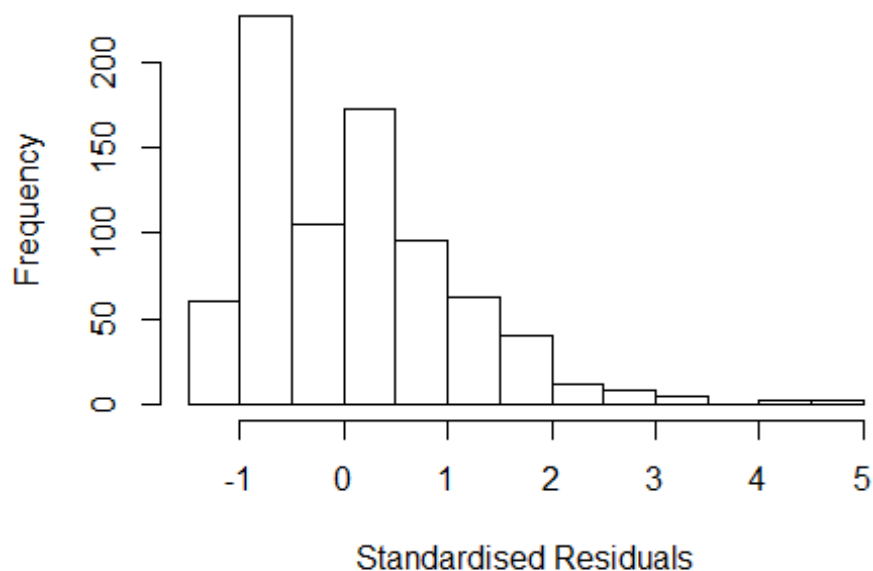
```

## 419 33745104091 4.031 2006      1210      2      3.115
## 425 85014698682 3.958 2006      1213      1      2.876
## 555 77649230762 3.539 2008      1213      2      2.726
## 742 84857704004 4.034 2011      1208      2      3.200
## 820 79953724832 3.382 2011      1210      2      2.548
## 925 84863108187 3.247 2012      1213      1      2.565
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.32490 -0.73312  0.00101  0.65675  4.82681
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.6255     0.1808   3.46 0.00057 ***
## FirstAuthorFemale1 0.3629     0.1541   2.35 0.01878 *
## LastAuthorFemale1 -0.1970     0.1551  -1.27 0.20445
## Year1997          -0.0196     0.2406  -0.08 0.93501
## Year1998          -0.1743     0.2348  -0.74 0.45825
## Year1999           0.1666     0.2112   0.79 0.43029
## Year2000           0.2087     0.3296   0.63 0.52688
## Year2001           0.1076     0.2361   0.46 0.64865
## Year2002           0.0210     0.2163   0.10 0.92266
## Year2003           0.1930     0.2314   0.83 0.40448
## Year2004           0.2795     0.2534   1.10 0.27039
## Year2005           0.2847     0.2271   1.25 0.21036
## Year2006           0.2901     0.2314   1.25 0.21033
## Year2007           0.5334     0.2418   2.21 0.02768 *
## Year2008           0.0213     0.2120   0.10 0.91984
## Year2009           0.3097     0.2138   1.45 0.14783
## Year2010           0.0917     0.1973   0.46 0.64245
## Year2011           0.2082     0.2045   1.02 0.30893
## Year2012           0.0563     0.2014   0.28 0.78001
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.92
## Multiple R-squared:  0.0354, Adjusted R-squared:  0.0129
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 4 observations c(107,251,268,285)
## are outliers with |weight| <= 9.8e-05 ( < 0.00013);
## 53 weights are ~= 1. The remaining 735 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.143  0.895   0.934   0.908   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          1.26e-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 1.139 1          1.067
## Year              1.139 16          1.004
```

Residuals from first author



```
## [1] "List of 17 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 134 60949365420 5.396 2002 1213 1 4.387
## 187 60950331495 4.220 2003 1213 1 3.401
## 240 60950108115 3.874 2004 1213 1 2.803
## 276 60249093788 4.227 2004 1213 1 3.322
```

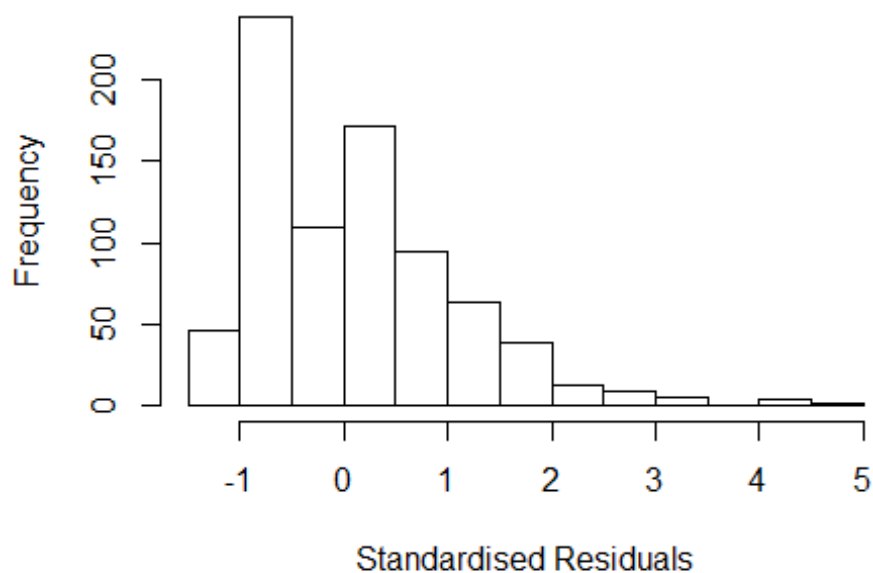
```

## 311 34247665578 5.284 2005      1202      4      4.374
## 332 33646201756 5.737 2005      1210      2      4.827
## 339 84857259490 3.713 2005      1213      1      2.803
## 352 60949545379 5.366 2005      1208      2      4.290
## 378 61049412680 3.958 2006      1213      1      2.876
## 379 61049460890 3.745 2006      1213      1      2.829
## 381 64949189141 4.319 2006      1213      1      3.237
## 419 33745104091 4.031 2006      1210      2      3.115
## 425 85014698682 3.958 2006      1213      1      2.876
## 555 77649230762 3.539 2008      1213      2      2.726
## 742 84857704004 4.034 2011      1208      2      3.200
## 820 79953724832 3.382 2011      1210      2      2.548
## 925 84863108187 3.247 2012      1213      1      2.565
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3402 -0.7171  0.0114  0.6668  4.8400
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.63352    0.17420   3.64 0.00029 ***
## FirstAuthorFemale1 0.18372    0.06951   2.64 0.00838 **
## Year1997       -0.03884    0.23531  -0.17 0.86894
## Year1998       -0.19370    0.23004  -0.84 0.40002
## Year1999        0.17697    0.20515   0.86 0.38862
## Year2000        0.19249    0.32720   0.59 0.55652
## Year2001        0.08359    0.22934   0.36 0.71559
## Year2002       -0.00489    0.20959  -0.02 0.98138
## Year2003        0.18168    0.22687   0.80 0.42350
## Year2004        0.26061    0.24811   1.05 0.29387
## Year2005        0.26346    0.22018   1.20 0.23185
## Year2006        0.27377    0.22636   1.21 0.22687
## Year2007        0.52292    0.23694   2.21 0.02761 *
## Year2008        0.00981    0.20619   0.05 0.96207
## Year2009        0.29142    0.20821   1.40 0.16201
## Year2010        0.07250    0.19178   0.38 0.70550
## Year2011        0.19316    0.19905   0.97 0.33216
## Year2012        0.03810    0.19543   0.19 0.84548
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.919
## Multiple R-squared:  0.0339, Adjusted R-squared:  0.0127
## Convergence in 12 IRWLS iterations
##

```

```
## Robustness weights:
## 4 observations c(107,251,268,285)
## are outliers with |weight| <= 7.5e-05 ( < 0.00013);
## 57 weights are ~= 1. The remaining 731 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.140  0.894  0.930   0.907   0.979   0.999
## Algorithmic parameters:
##       tuning.chi                bb          tuning.psi          refine.tol
##       1.55e+00                5.00e-01          4.69e+00          1.00e-07
##       rel.tol                solve.tol          eps.outlier          eps.x
##       1.00e-07                1.00e-07          1.26e-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 1.162  1          1.078
## Year             1.162 16          1.005
```

Residuals from last author



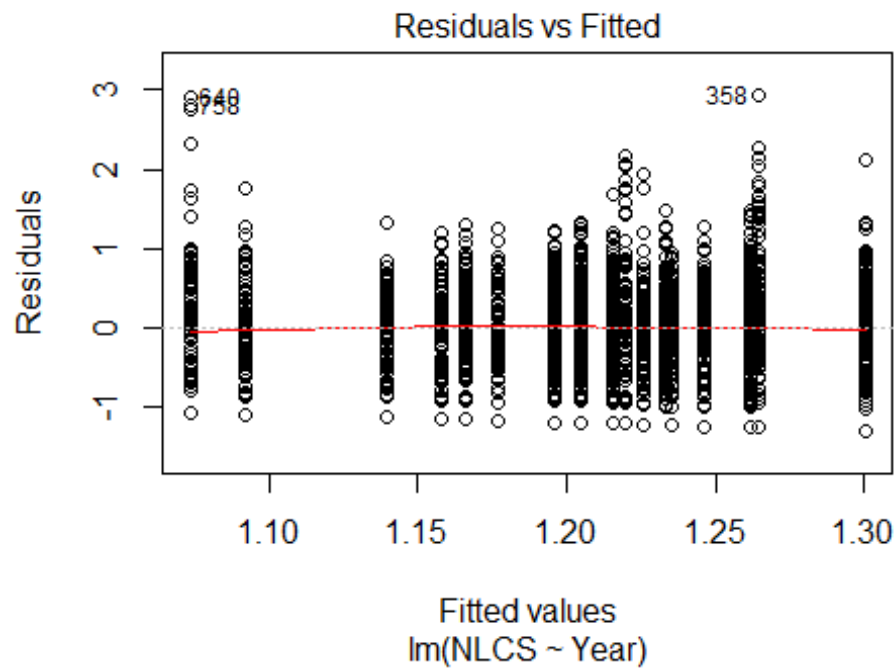
```
## [1] "List of 17 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 134 60949365420 5.396 2002      1213      1      4.387
## 187 60950331495 4.220 2003      1213      1      3.401
## 240 60950108115 3.874 2004      1213      1      2.803
## 276 60249093788 4.227 2004      1213      1      3.322
## 311 34247665578 5.284 2005      1202      4      4.374
## 332 33646201756 5.737 2005      1210      2      4.827
## 339 84857259490 3.713 2005      1213      1      2.803
## 352 60949545379 5.366 2005      1208      2      4.290
## 378 61049412680 3.958 2006      1213      1      2.876
## 379 61049460890 3.745 2006      1213      1      2.829
## 381 64949189141 4.319 2006      1213      1      3.237
## 419 33745104091 4.031 2006      1210      2      3.115
## 425 85014698682 3.958 2006      1213      1      2.876
## 555 77649230762 3.539 2008      1213      2      2.726
## 742 84857704004 4.034 2011      1208      2      3.200
## 820 79953724832 3.382 2011      1210      2      2.548
## 925 84863108187 3.247 2012      1213      1      2.565
##
## Call:
## lmrob(formula = NLCS ~ 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 -0.7606  0.0036  0.6712  4.8250
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6779    0.1668    4.06 5.3e-05 ***
## LastAuthorFemale1 0.1358    0.0701    1.94  0.053 .
## Year1997      -0.0532    0.2316   -0.23  0.819
## Year1998      -0.2076    0.2244   -0.92  0.355
## Year1999       0.1655    0.2020    0.82  0.413
## Year2000       0.1699    0.3208    0.53  0.597
## Year2001       0.0578    0.2237    0.26  0.796
## Year2002      -0.0264    0.2044   -0.13  0.897
## Year2003       0.1575    0.2228    0.71  0.480
## Year2004       0.2444    0.2451    1.00  0.319
## Year2005       0.2340    0.2145    1.09  0.276
## Year2006       0.2503    0.2206    1.13  0.257
## Year2007       0.5061    0.2323    2.18  0.030 *
## Year2008      -0.0113    0.2013   -0.06  0.955
## Year2009       0.2747    0.2029    1.35  0.176
## Year2010       0.0491    0.1865    0.26  0.793
## Year2011       0.1751    0.1940    0.90  0.367
## Year2012       0.0147    0.1899    0.08  0.938
## ---
```

```

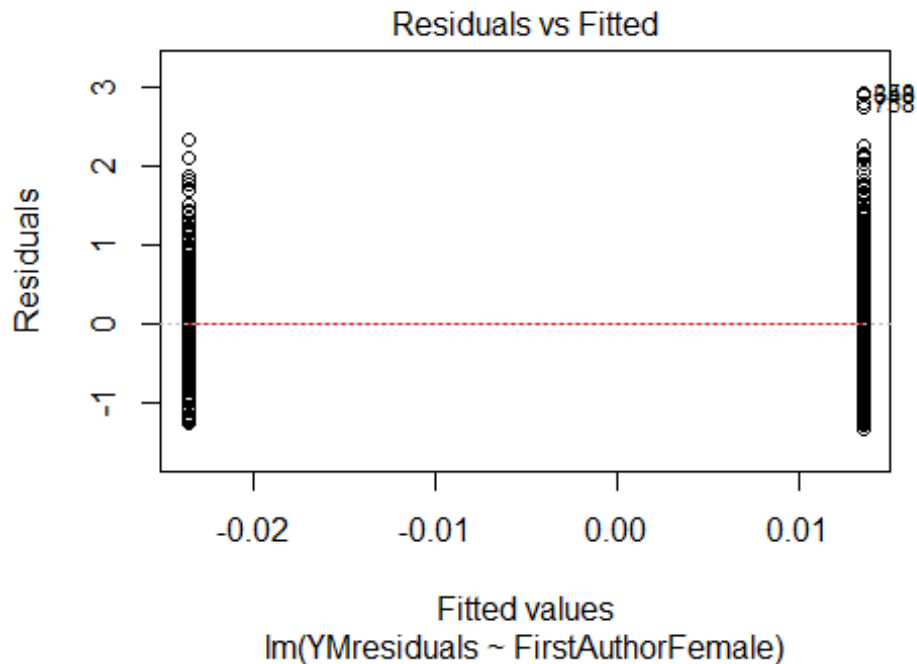
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.92
## Multiple R-squared:  0.0295, Adjusted R-squared:  0.00813
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 4 observations c(107,251,268,285)
## are outliers with |weight| = 0 ( < 0.00013);
## 50 weights are ~= 1. The remaining 738 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.147  0.897  0.934  0.908  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           1.26e-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: 792"
## [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
## 436 279 228 167 169 179 162 159 148 197 234 261 256 354 420
## 2011 2012
## 596 848
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 144 114 107 103 96 69 108 108 105 138 161 186 187 276 320
## 2011 2012
## 486 654
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 118 100 88 83 75 60 90 80 94 122 136 152 158 246 275
## 2011 2012

```

```
## 412 568
## [1] "Heteroscedasticity checks, confirming that there are problems with
## these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 350, df = 16, p-value <2e-16
```

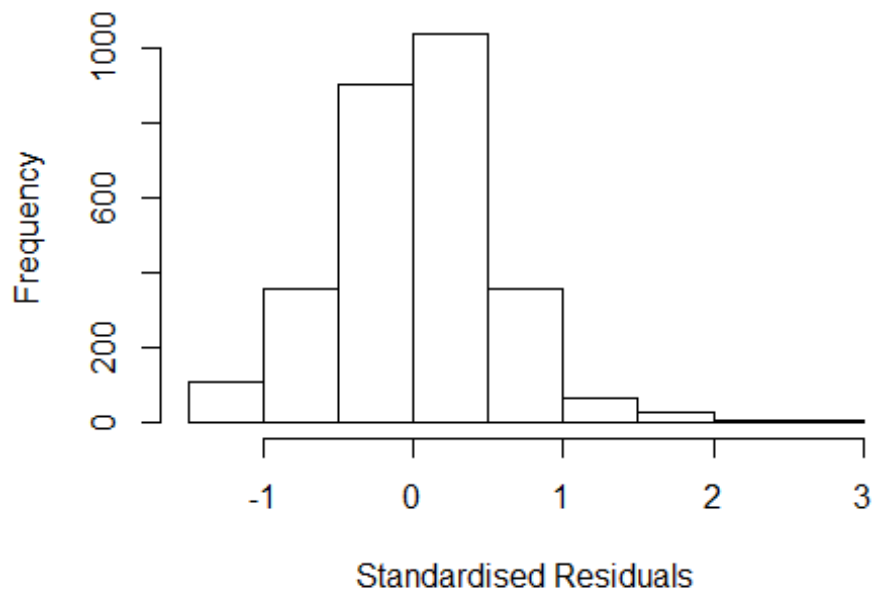


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 12, df = 1, p-value = 5e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.82135994036745"
## [1] "Male first author team size 2018 geometric mean: 4.19313851325457"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.54103067931436"
## [1] "Male last author team size 2018 geometric mean: 4.41405488876659"
##
## 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 1.103 1      1.050
## LastAuthorFemale  1.070 1      1.034
## UniqueAuthors    1.305 4      1.034
## Year             1.379 16      1.010
```


Residuals from first and last author and team size



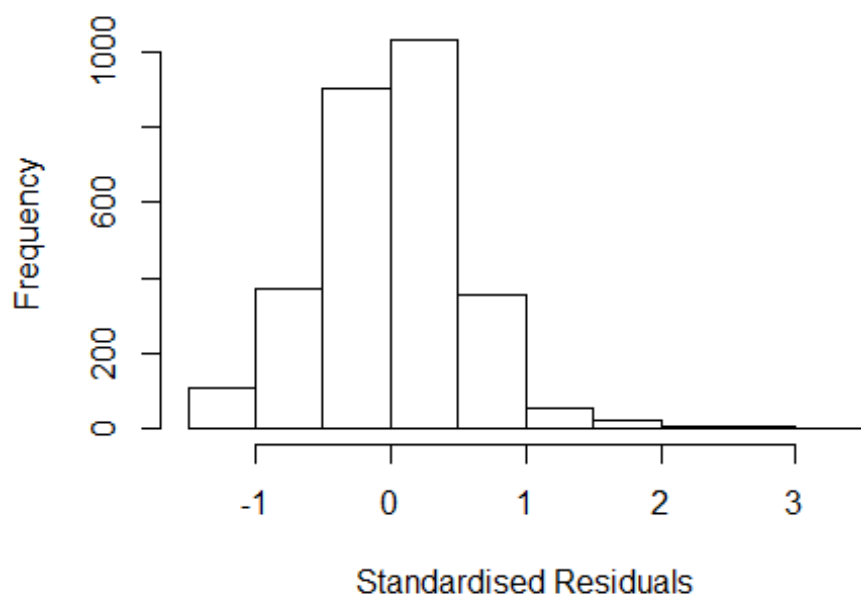
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 358  0029990028 4.195 1996    1300     1    2.868
## 705  0030831104 3.821 1997    1300     1    2.741
## 758 16944363001 3.887 1997    1300     1    2.807
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4225 -0.3147  0.0286  0.3293  2.8676
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10094    0.13045   8.44  <2e-16 ***
## FirstAuthorFemale1 -0.05131    0.02066  -2.48   0.013 *
## LastAuthorFemale1 -0.01186    0.02426  -0.49   0.625
## UniqueAuthors2     0.08180    0.04569   1.79   0.074 .
## UniqueAuthors3     0.01098    0.04591   0.24   0.811
## UniqueAuthors4     0.13512    0.04696   2.88   0.004 **
## UniqueAuthors5     0.22648    0.04408   5.14  3e-07 ***
## Year1997        -0.23542    0.15311  -1.54   0.124
```

```

## Year1998      -0.06955    0.16104   -0.43    0.666
## Year1999      -0.01025    0.13802   -0.07    0.941
## Year2000       0.03484    0.14635    0.24    0.812
## Year2001      -0.02981    0.15273   -0.20    0.845
## Year2002      -0.11344    0.14085   -0.81    0.421
## Year2003       0.03578    0.14097    0.25    0.800
## Year2004       0.09505    0.13830    0.69    0.492
## Year2005       0.08608    0.13515    0.64    0.524
## Year2006       0.00796    0.13514    0.06    0.953
## Year2007       0.01547    0.13275    0.12    0.907
## Year2008       0.12801    0.13470    0.95    0.342
## Year2009       0.03978    0.13261    0.30    0.764
## Year2010       0.02706    0.13183    0.21    0.837
## Year2011       0.05277    0.12998    0.41    0.685
## Year2012      -0.01202    0.12964   -0.09    0.926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.048, Adjusted R-squared:  0.0406
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 4 observations c(75,169,179,198)
## are outliers with |weight| = 0 ( < 3.5e-05);
## 249 weights are ~= 1. The remaining 2604 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0019 0.8630 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      3.50e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.078 1          1.038
## LastAuthorFemale 1.061 1          1.030
## Year              1.104 16          1.003

```

Residuals from first and last author



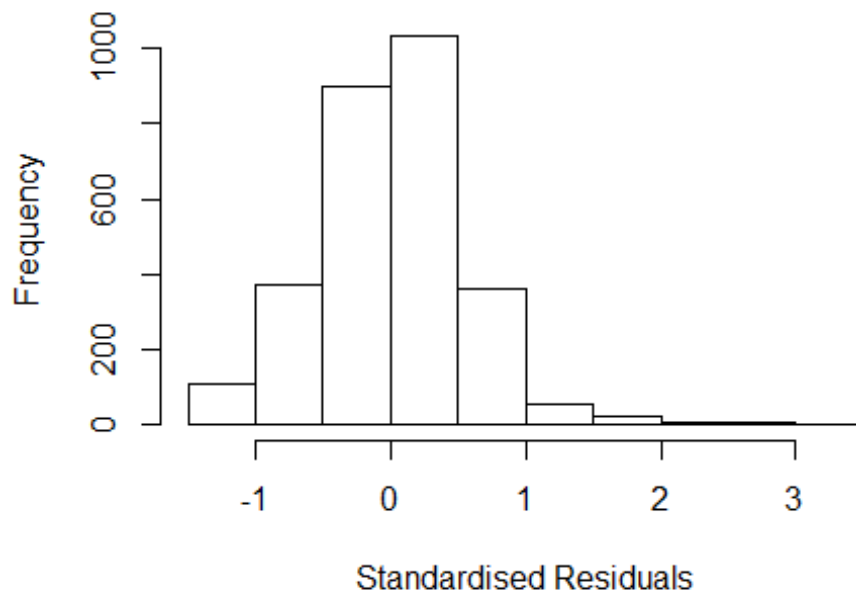
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 358  0029990028 4.195 1996    1300     1    3.030
## 682  0030941260 3.393 1997    1300     2    2.503
## 705  0030831104 3.821 1997    1300     1    2.898
## 758 16944363001 3.887 1997    1300     1    2.964
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3102 -0.3165  0.0191  0.3342  3.0299
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.16506    0.12861     9.06  <2e-16 ***
## FirstAuthorFemale1 -0.04588    0.02065    -2.22   0.026 *
## LastAuthorFemale1 -0.01307    0.02450    -0.53   0.594
## Year1997        -0.22934    0.15276    -1.50   0.133
## Year1998        -0.06426    0.16111    -0.40   0.690
## Year1999         0.00242    0.13894     0.02   0.986
## Year2000         0.06049    0.14778     0.41   0.682
## Year2001         0.01211    0.15368     0.08   0.937
## Year2002        -0.08091    0.14156    -0.57   0.568
```

```

## Year2003          0.05870      0.14273      0.41      0.681
## Year2004          0.11914      0.13883      0.86      0.391
## Year2005          0.11828      0.13632      0.87      0.386
## Year2006          0.03238      0.13623      0.24      0.812
## Year2007          0.05553      0.13397      0.41      0.679
## Year2008          0.15825      0.13584      1.16      0.244
## Year2009          0.08784      0.13364      0.66      0.511
## Year2010          0.08657      0.13258      0.65      0.514
## Year2011          0.11686      0.13079      0.89      0.372
## Year2012          0.05937      0.13030      0.46      0.649
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.494
## Multiple R-squared:  0.0209, Adjusted R-squared:  0.0147
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 4 observations c(75,169,179,198)
## are outliers with |weight| = 0 ( < 3.5e-05);
## 230 weights are ~ = 1. The remaining 2623 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0012 0.8680 0.9530 0.8950 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.50e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1          1.029
## Year              1.058 16          1.002

```

Residuals from first author



```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 358  0029990028 4.195 1996    1300     1    3.030
## 682  0030941260 3.393 1997    1300     2    2.503
## 705  0030831104 3.821 1997    1300     1    2.898
## 758 16944363001 3.887 1997    1300     1    2.964
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3217 -0.3154  0.0208  0.3363  3.0334
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.16158    0.12948   8.97  <2e-16 ***
## FirstAuthorFemale1 -0.04734    0.02047  -2.31   0.021 *
## Year1997        -0.22827    0.15437  -1.48   0.139
## Year1998        -0.06527    0.16124  -0.40   0.686
## Year1999         0.00531    0.13975   0.04   0.970
## Year2000         0.06115    0.14844   0.41   0.680
## Year2001         0.01454    0.15449   0.09   0.925
## Year2002        -0.07799    0.14237  -0.55   0.584
## Year2003         0.06071    0.14353   0.42   0.672
```

```

## Year2004          0.12086    0.13960    0.87    0.387
## Year2005          0.11925    0.13704    0.87    0.384
## Year2006          0.03407    0.13695    0.25    0.804
## Year2007          0.05690    0.13474    0.42    0.673
## Year2008          0.16010    0.13665    1.17    0.241
## Year2009          0.08949    0.13436    0.67    0.505
## Year2010          0.08785    0.13333    0.66    0.510
## Year2011          0.11772    0.13153    0.89    0.371
## Year2012          0.06058    0.13109    0.46    0.644
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.489
## Multiple R-squared:  0.0211, Adjusted R-squared:  0.0152
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 4 observations c(75,169,179,198)
## are outliers with |weight| = 0 ( < 3.5e-05);
## 227 weights are ~= 1. The remaining 2626 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0002 0.8640 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          3.50e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.042 1          1.021
## Year          1.042 16          1.001
##
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 358 0029990028 4.195 1996 1300 1 3.030
## 682 0030941260 3.393 1997 1300 2 2.503
## 705 0030831104 3.821 1997 1300 1 2.898
## 758 16944363001 3.887 1997 1300 1 2.964
##

```

```

## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2863 -0.3135  0.0204  0.3362  3.0379
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.15715     0.12836    9.01  <2e-16 ***
## LastAuthorFemale1 -0.02019     0.02435   -0.83    0.41
## Year1997         -0.22389     0.15250   -1.47    0.14
## Year1998         -0.06540     0.16103   -0.41    0.68
## Year1999         -0.00206     0.13873   -0.01    0.99
## Year2000          0.06024     0.14752    0.41    0.68
## Year2001          0.00823     0.15333    0.05    0.96
## Year2002         -0.08541     0.14139   -0.60    0.55
## Year2003          0.04693     0.14186    0.33    0.74
## Year2004          0.11129     0.13860    0.80    0.42
## Year2005          0.11075     0.13607    0.81    0.42
## Year2006          0.02646     0.13580    0.19    0.85
## Year2007          0.05233     0.13381    0.39    0.70
## Year2008          0.14936     0.13535    1.10    0.27
## Year2009          0.07830     0.13336    0.59    0.56
## Year2010          0.07918     0.13227    0.60    0.55
## Year2011          0.10771     0.13046    0.83    0.41
## Year2012          0.04664     0.12982    0.36    0.72
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.495
## Multiple R-squared:  0.019, Adjusted R-squared:  0.0131
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
##  4 observations c(75,169,179,198)
##  are outliers with |weight| = 0 ( < 3.5e-05);
##  228 weights are ~= 1. The remaining 2625 ones are summarized as
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0009  0.8650  0.9530  0.8950  0.9870  0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.50e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale

```

```

##           500           50           2           1           1000           200
## trace.lev      mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2857"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1301"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 2007 2008 2009 2010 2011 2012
##    1    1    5    2    5    6
##
## 2007 2008 2009 2010 2011 2012
##    1    1    5    1    3    5
##
## 2007 2008 2009 2010 2011 2012
##    1    1    5    0    2    4
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3.16227766016838"
## [1] "Male first author team size 2018 geometric mean: 3.51276529679714"

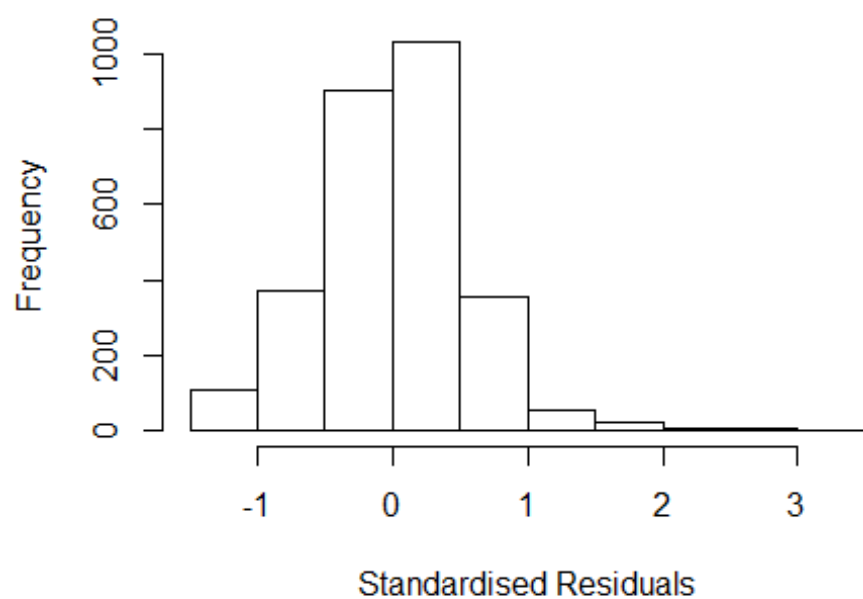
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6.5, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: NaN"
## [1] "Male last author team size 2018 geometric mean: 3.4316649716921"
## [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"

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

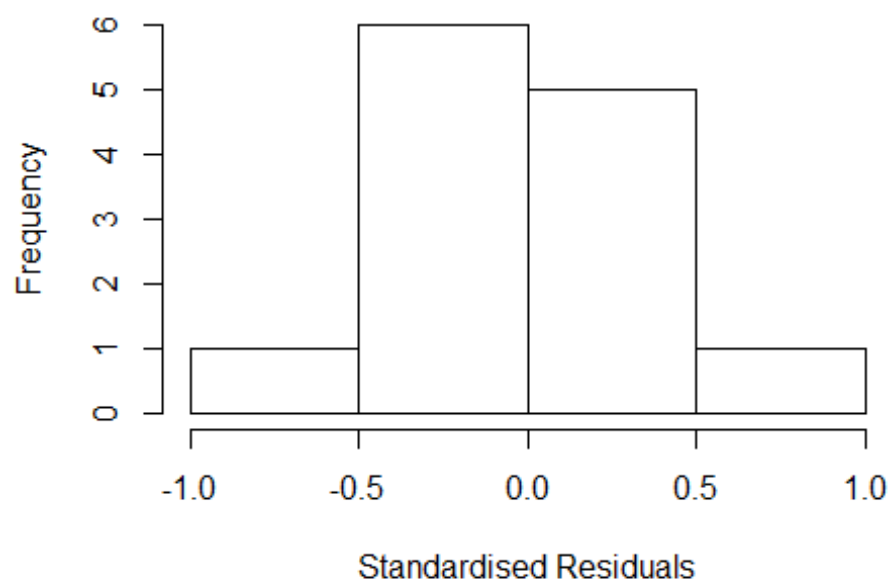
```


Residuals from last author



##		GVIF	Df	$GVIF^{(1/(2*Df))}$
##	FirstAuthorFemale	3.002e+15	1	5.479e+07
##	Year	3.002e+15	4	8.604e+01

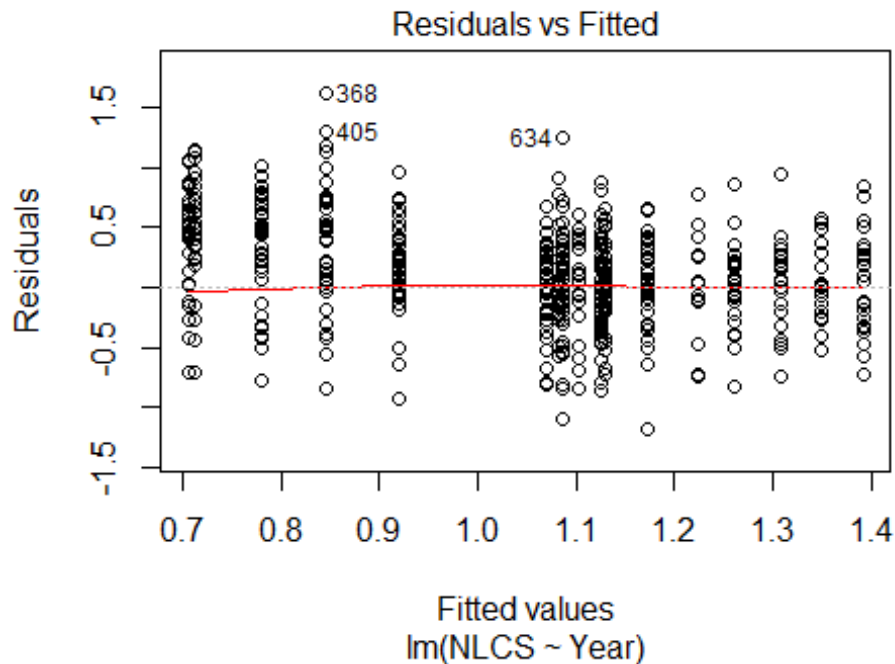
Residuals from first author



```

## [1] "Regression 4: Last author gender, Year as factors"
## [1] "Sample size for the above analysis: 13"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1302"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    32    20    24    29    30    36    34    65    68    64    68    58    51    73    48
## 2011 2012
##    50    63
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    24    14    18    23    21    22    26    50    61    45    59    47    43    62    39
## 2011 2012
##    45    54
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    22    12    17    20    20    19    25    44    59    43    53    43    40    57    34
## 2011 2012
##    38    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 = 87, df = 16, p-value = 1e-11

```



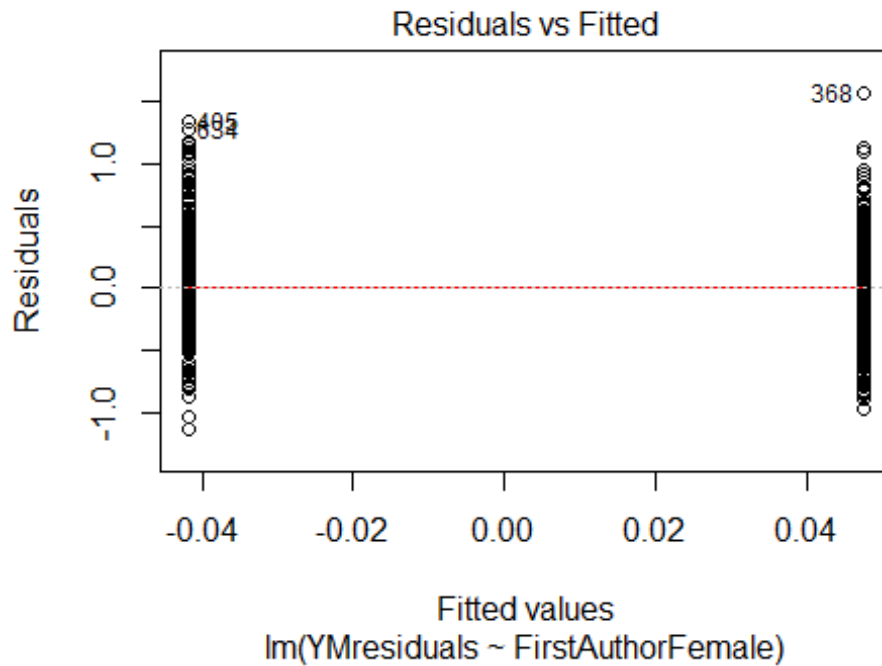
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.3, df = 1, p-value = 0.04

## [1] "Female first author team size 2018 geometric mean: 4.33827180537495"
## [1] "Male first author team size 2018 geometric mean: 4.48832706692029"

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

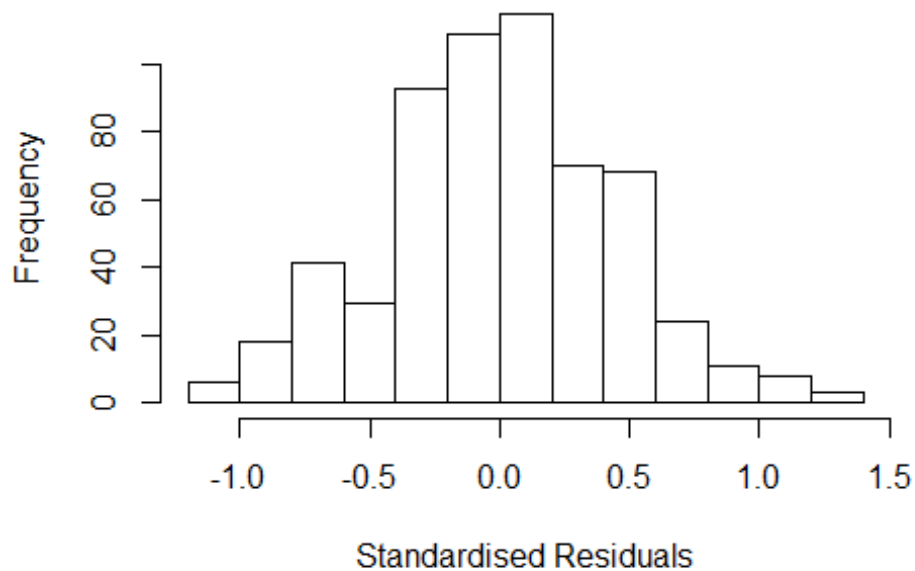
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.60013763781019"
## [1] "Male last author team size 2018 geometric mean: 6.17514291400849"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 45, 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 1.326 1          1.152
## LastAuthorFemale  1.266 1          1.125
## UniqueAuthors     3.264 4          1.159
## Year               3.649 16         1.041
```

Residuals from first and last author and team size



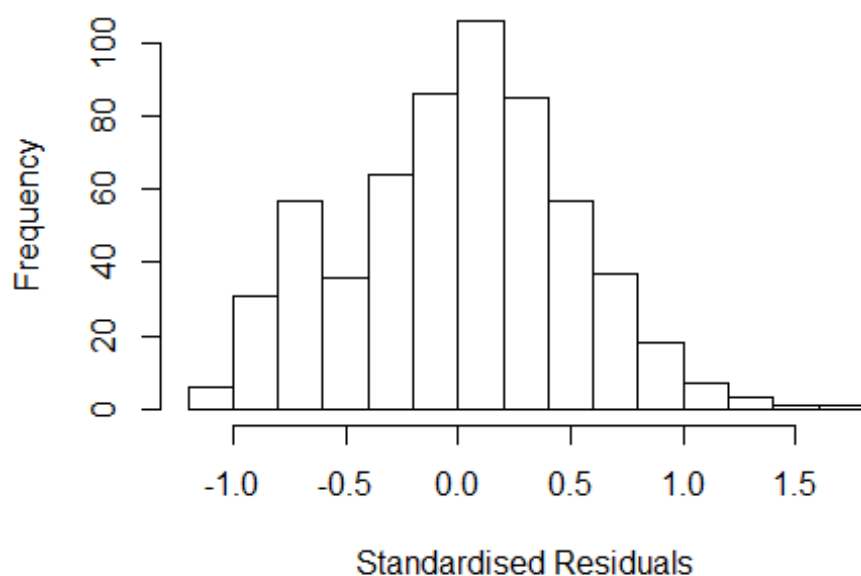
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.16198 -0.27665  0.00239  0.27854  1.35726
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7924    0.1286    6.16 1.4e-09 ***
## FirstAuthorFemale1  0.0439    0.0412    1.06 0.28782
## LastAuthorFemale1 -0.0439    0.0417   -1.06 0.29184
## UniqueAuthors2    0.4938    0.0902    5.47 6.6e-08 ***
## UniqueAuthors3    0.6648    0.0853    7.79 3.1e-14 ***
## UniqueAuthors4    0.7480    0.0797    9.38 < 2e-16 ***
## UniqueAuthors5    0.8049    0.0773   10.41 < 2e-16 ***
## Year1997         -0.3669    0.2206   -1.66 0.09680 .
## Year1998         -0.1539    0.1340   -1.15 0.25128
## Year1999         -0.1177    0.1418   -0.83 0.40680
```

```

## Year2000          -0.0989      0.1507    -0.66    0.51173
## Year2001          -0.4179      0.1397    -2.99    0.00290 **
## Year2002          -0.1244      0.1251    -0.99    0.32052
## Year2003          -0.5303      0.1440    -3.68    0.00025 ***
## Year2004          -0.4386      0.1427    -3.07    0.00222 **
## Year2005          -0.5584      0.1579    -3.54    0.00044 ***
## Year2006          -0.5157      0.1362    -3.79    0.00017 ***
## Year2007          -0.2522      0.1273    -1.98    0.04803 *
## Year2008          -0.2952      0.1272    -2.32    0.02071 *
## Year2009          -0.4544      0.1302    -3.49    0.00052 ***
## Year2010          -0.3368      0.1207    -2.79    0.00543 **
## Year2011          -0.3325      0.1219    -2.73    0.00657 **
## Year2012          -0.3758      0.1233    -3.05    0.00242 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared:  0.367, Adjusted R-squared:  0.343
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 545 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.186  0.833  0.946  0.880  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 1.213 1          1.101
## LastAuthorFemale  1.147 1          1.071
## Year              1.246 16          1.007

```

Residuals from first and last author



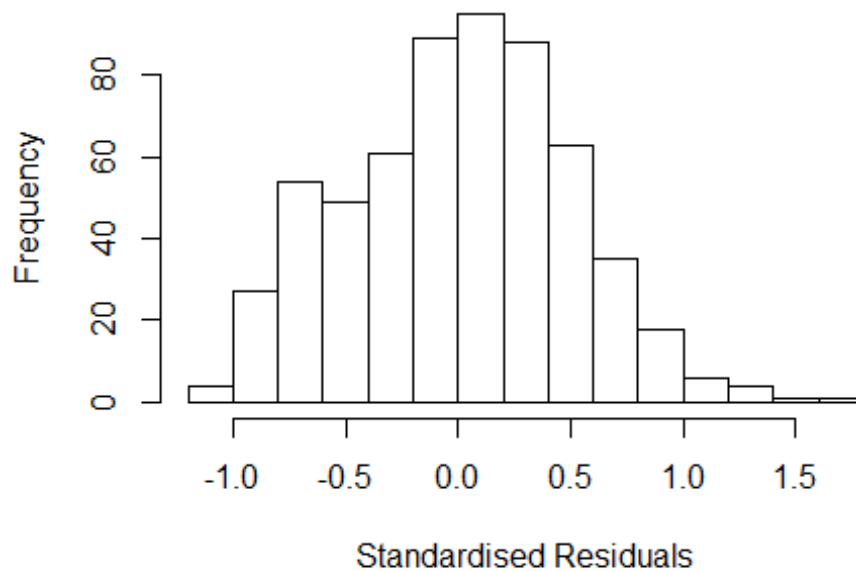
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.163 -0.352 0.024 0.327 1.665
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3380 0.0981 13.64 < 2e-16 ***
## FirstAuthorFemale1 0.1061 0.0462 2.30 0.02206 *
## LastAuthorFemale1 -0.0398 0.0466 -0.85 0.39401
## Year1997 -0.1937 0.1701 -1.14 0.25531
## Year1998 -0.1924 0.1364 -1.41 0.15896
## Year1999 -0.0929 0.1208 -0.77 0.44231
## Year2000 -0.0387 0.1289 -0.30 0.76394
## Year2001 -0.3278 0.1268 -2.59 0.00998 **
## Year2002 -0.1101 0.1202 -0.92 0.36019
## Year2003 -0.7115 0.1777 -4.00 7.1e-05 ***
## Year2004 -0.6104 0.1726 -3.54 0.00044 ***
## Year2005 -0.7371 0.1580 -4.67 3.8e-06 ***
```

```

## Year2006          -0.5252      0.1516   -3.46  0.00057 ***
## Year2007          -0.2673      0.1177   -2.27  0.02346 *
## Year2008          -0.1753      0.1149   -1.53  0.12779
## Year2009          -0.4215      0.1256   -3.36  0.00084 ***
## Year2010          -0.2693      0.1090   -2.47  0.01373 *
## Year2011          -0.2485      0.1102   -2.25  0.02457 *
## Year2012          -0.2589      0.1131   -2.29  0.02245 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.176, Adjusted R-squared:  0.15
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 557 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0992 0.8150 0.9350 0.8810 0.9820 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.145 1      1.070
## Year      1.145 16      1.004

```


Residuals from first author



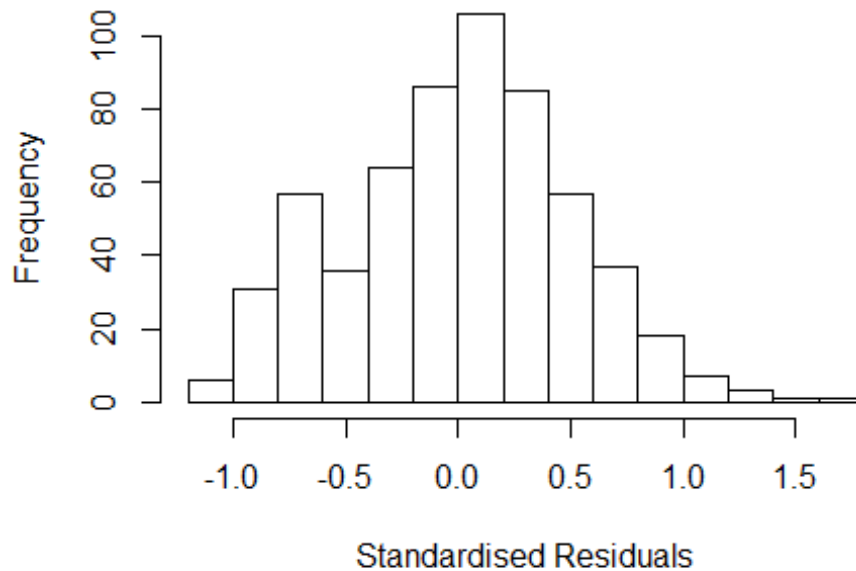
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1517 -0.3451 0.0236 0.3329 1.6442
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3311 0.0976 13.63 < 2e-16 ***
## FirstAuthorFemale1 0.0935 0.0451 2.07 0.03859 *
## Year1997 -0.1923 0.1726 -1.11 0.26557
## Year1998 -0.1949 0.1368 -1.42 0.15492
## Year1999 -0.0947 0.1206 -0.79 0.43228
## Year2000 -0.0383 0.1291 -0.30 0.76657
## Year2001 -0.3216 0.1265 -2.54 0.01126 *
## Year2002 -0.1084 0.1200 -0.90 0.36679
## Year2003 -0.7100 0.1735 -4.09 4.9e-05 ***
## Year2004 -0.6099 0.1653 -3.69 0.00025 ***
## Year2005 -0.7395 0.1556 -4.75 2.5e-06 ***
## Year2006 -0.5314 0.1500 -3.54 0.00043 ***
```

```

## Year2007          -0.2717      0.1179   -2.31  0.02152 *
## Year2008          -0.1794      0.1148   -1.56  0.11854
## Year2009          -0.4255      0.1255   -3.39  0.00075 ***
## Year2010          -0.2675      0.1095   -2.44  0.01485 *
## Year2011          -0.2523      0.1103   -2.29  0.02251 *
## Year2012          -0.2586      0.1132   -2.28  0.02270 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.172, Adjusted R-squared:  0.148
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 554 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.128  0.825  0.937  0.885  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      1.68e-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 1.12 1      1.058
## Year              1.12 16      1.004

```

Residuals from last author



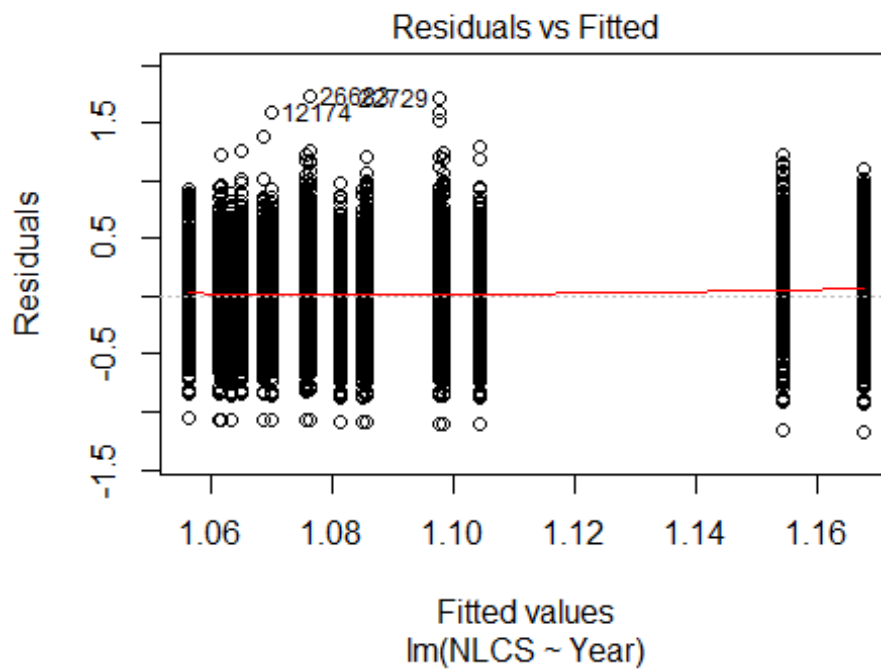
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2022 -0.3501  0.0308  0.3423  1.7012
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.36797    0.09638   14.19  < 2e-16 ***
## LastAuthorFemale1 -0.00366    0.04590   -0.08  0.93649
## Year1997        -0.16669    0.17082   -0.98  0.32956
## Year1998        -0.20520    0.13892   -1.48  0.14020
## Year1999        -0.09568    0.12024   -0.80  0.42649
## Year2000        -0.03184    0.12649   -0.25  0.80136
## Year2001        -0.32722    0.12438   -2.63  0.00875 **
## Year2002        -0.08140    0.11701   -0.70  0.48695
## Year2003        -0.71464    0.18526   -3.86  0.00013 ***
## Year2004        -0.60655    0.17224   -3.52  0.00046 ***
## Year2005        -0.72976    0.15941   -4.58  5.8e-06 ***
## Year2006        -0.52569    0.15368   -3.42  0.00067 ***
```

```

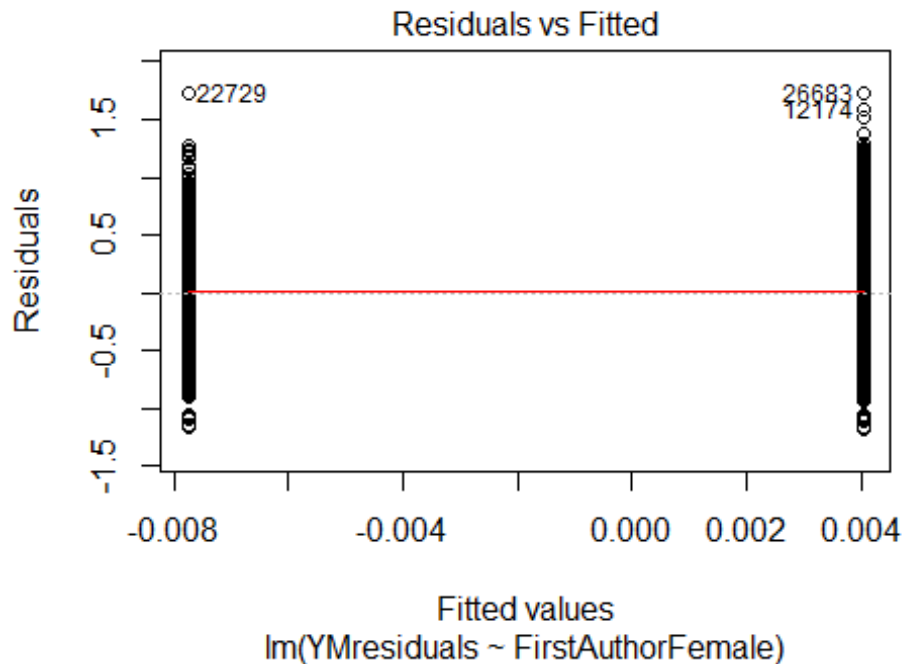
## Year2007          -0.26562      0.11783    -2.25  0.02456 *
## Year2008          -0.16579      0.11285    -1.47  0.14237
## Year2009          -0.40387      0.12494    -3.23  0.00130 **
## Year2010          -0.25539      0.10848    -2.35  0.01889 *
## Year2011          -0.23703      0.10786    -2.20  0.02837 *
## Year2012          -0.23406      0.11096    -2.11  0.03533 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.428
## Multiple R-squared:  0.167, Adjusted R-squared:  0.143
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 558 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.079  0.808  0.935   0.880  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] "Sample size for the above analysis: 595"
## [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
## 1601 1355 1275 1081 1200 1061 1188 1066 1118 1177 1148 1228 1204 1292 1340
## 2011 2012
## 1390 1250
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 831 739 774 711 676 532 801 751 753 777 763 814 797 901 944
## 2011 2012

```

```
## 994 907
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 695 609 658 594 565 449 651 619 636 665 648 698 693 780 802
## 2011 2012
## 826 754
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 500, df = 16, p-value <2e-16
```

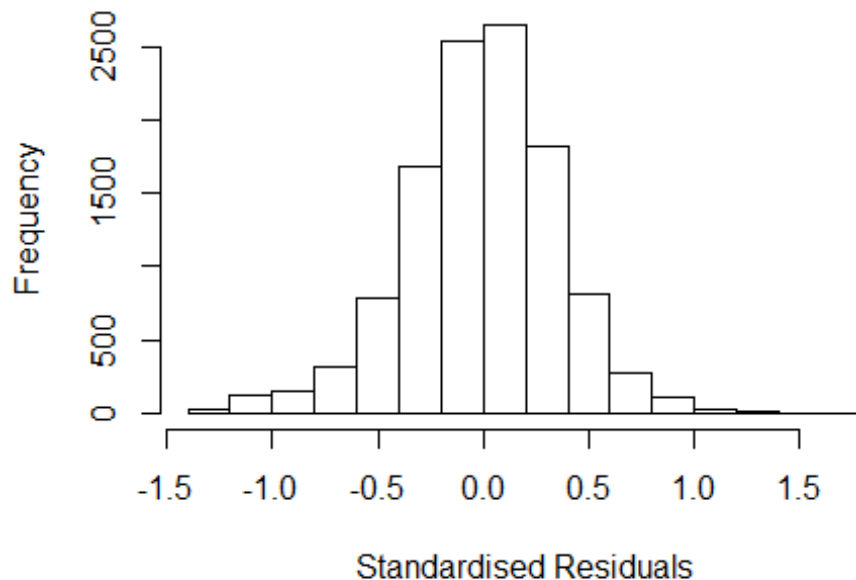


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.2, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.33080877221511"
## [1] "Male first author team size 2018 geometric mean: 3.83321883826937"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.80802534678473"
## [1] "Male last author team size 2018 geometric mean: 4.03353684698492"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, 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 1.024 1          1.012
## LastAuthorFemale  1.014 1          1.007
## UniqueAuthors    1.064 4          1.008
## Year             1.070 16          1.002
```

Residuals from first and last author and team size



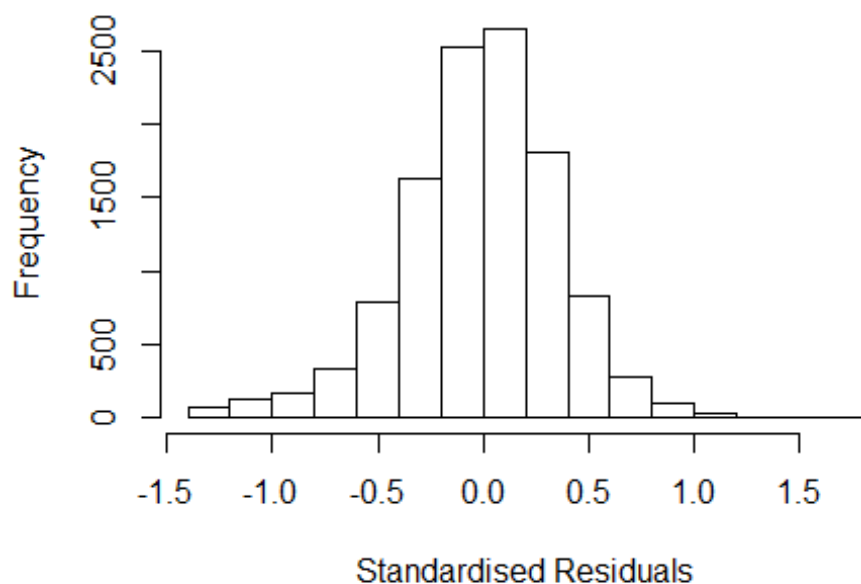
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28872 -0.22141  0.00246  0.21955  1.67946
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.00418    0.03179   31.59 < 2e-16 ***
## FirstAuthorFemale1 -0.01481    0.00689   -2.15  0.03169 *
## LastAuthorFemale1 -0.01054    0.00869   -1.21  0.22488
## UniqueAuthors2     0.18899    0.02842    6.65  3.1e-11 ***
## UniqueAuthors3     0.20961    0.02841    7.38  1.7e-13 ***
## UniqueAuthors4     0.23589    0.02858    8.26 < 2e-16 ***
## UniqueAuthors5     0.27724    0.02820    9.83 < 2e-16 ***
## Year1997           0.00730    0.02574    0.28  0.77682
## Year1998          -0.08826    0.02366   -3.73  0.00019 ***
## Year1999          -0.13035    0.02261   -5.77  8.4e-09 ***
```

```

## Year2000      -0.10752    0.02392   -4.49  7.1e-06 ***
## Year2001      -0.12947    0.02351   -5.51  3.7e-08 ***
## Year2002      -0.12391    0.02120   -5.85  5.2e-09 ***
## Year2003      -0.14812    0.02169   -6.83  9.0e-12 ***
## Year2004      -0.15201    0.02159   -7.04  2.0e-12 ***
## Year2005      -0.15054    0.02169   -6.94  4.2e-12 ***
## Year2006      -0.15548    0.02167   -7.17  7.7e-13 ***
## Year2007      -0.14989    0.02152   -6.97  3.5e-12 ***
## Year2008      -0.14134    0.02202   -6.42  1.4e-10 ***
## Year2009      -0.11886    0.02152   -5.52  3.4e-08 ***
## Year2010      -0.12353    0.02159   -5.72  1.1e-08 ***
## Year2011      -0.14079    0.02187   -6.44  1.3e-10 ***
## Year2012      -0.15164    0.02235   -6.79  1.2e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.327
## Multiple R-squared:  0.0381, Adjusted R-squared:  0.0363
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 4 observations c(4768,9408,9609,11213)
## are outliers with |weight| = 0 ( < 8.8e-06);
## 966 weights are ~ = 1. The remaining 10372 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0078 0.8650 0.9500 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.82e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.017 1 1.008
## LastAuthorFemale 1.009 1 1.004
## Year 1.019 16 1.001

```


Residuals from first and last author



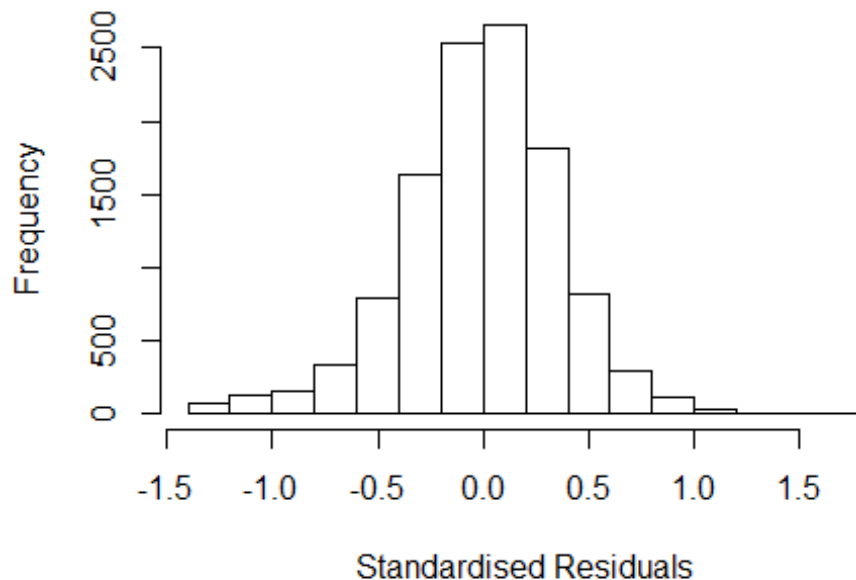
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21998 -0.22450  0.00276  0.21993  1.72354
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.21005    0.01828   66.20 < 2e-16 ***
## FirstAuthorFemale1 -0.00710    0.00691   -1.03  0.30412
## LastAuthorFemale1 -0.00634    0.00871   -0.73  0.46643
## Year1997         0.00993    0.02611    0.38  0.70366
## Year1998        -0.07890    0.02382   -3.31  0.00093 ***
## Year1999        -0.11970    0.02285   -5.24  1.6e-07 ***
## Year2000        -0.09840    0.02428   -4.05  5.1e-05 ***
## Year2001        -0.11626    0.02373   -4.90  9.7e-07 ***
## Year2002        -0.10390    0.02138   -4.86  1.2e-06 ***
## Year2003        -0.13127    0.02186   -6.00  2.0e-09 ***
## Year2004        -0.13481    0.02179   -6.19  6.3e-10 ***
## Year2005        -0.12877    0.02186   -5.89  4.0e-09 ***
```

```

## Year2006          -0.14021      0.02185      -6.42      1.4e-10 ***
## Year2007          -0.13180      0.02170      -6.07      1.3e-09 ***
## Year2008          -0.11754      0.02219      -5.30      1.2e-07 ***
## Year2009          -0.09556      0.02160      -4.42      9.8e-06 ***
## Year2010          -0.09870      0.02179      -4.53      6.0e-06 ***
## Year2011          -0.11855      0.02208      -5.37      8.1e-08 ***
## Year2012          -0.12959      0.02267      -5.72      1.1e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.329
## Multiple R-squared:  0.0145, Adjusted R-squared:  0.013
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 4 observations c(4768,9408,9609,11213)
## are outliers with |weight| = 0 ( < 8.8e-06);
## 981 weights are ~ = 1. The remaining 10357 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0017 0.8640 0.9500 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.82e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.013 1          1.007
## Year              1.013 16          1.000

```

Residuals from first author



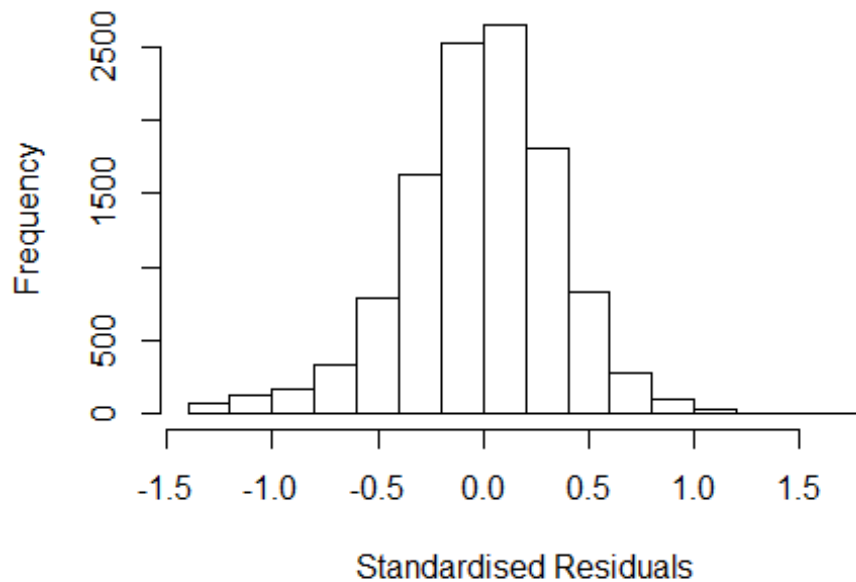
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21910 -0.22445 0.00316 0.21917 1.72448
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20923 0.01824 66.29 < 2e-16 ***
## FirstAuthorFemale1 -0.00747 0.00690 -1.08 0.27903
## Year1997 0.00987 0.02613 0.38 0.70575
## Year1998 -0.07897 0.02383 -3.31 0.00092 ***
## Year1999 -0.11984 0.02286 -5.24 1.6e-07 ***
## Year2000 -0.09864 0.02428 -4.06 4.9e-05 ***
## Year2001 -0.11635 0.02373 -4.90 9.6e-07 ***
## Year2002 -0.10413 0.02140 -4.87 1.2e-06 ***
## Year2003 -0.13127 0.02187 -6.00 2.0e-09 ***
## Year2004 -0.13475 0.02180 -6.18 6.5e-10 ***
## Year2005 -0.12897 0.02187 -5.90 3.8e-09 ***
## Year2006 -0.14029 0.02186 -6.42 1.4e-10 ***
```

```

## Year2007      -0.13182    0.02171   -6.07  1.3e-09 ***
## Year2008      -0.11773    0.02220   -5.30  1.2e-07 ***
## Year2009      -0.09571    0.02161   -4.43  9.6e-06 ***
## Year2010      -0.09887    0.02180   -4.54  5.8e-06 ***
## Year2011      -0.11867    0.02209   -5.37  7.9e-08 ***
## Year2012      -0.12972    0.02267   -5.72  1.1e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.329
## Multiple R-squared:  0.0145, Adjusted R-squared:  0.013
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 4 observations c(4768,9408,9609,11213)
## are outliers with |weight| = 0 ( < 8.8e-06);
## 989 weights are ~ = 1. The remaining 10349 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0015 0.8640 0.9500 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.82e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.006 1      1.003
## Year      1.006 16      1.000

```

Residuals from last author



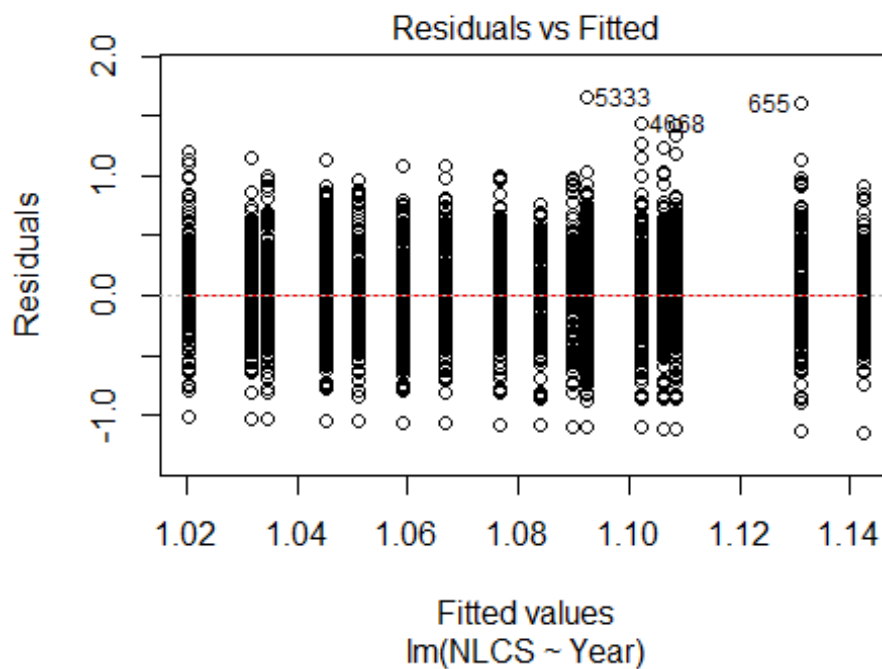
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21818 -0.22373  0.00355  0.22074  1.72604
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20794    0.01813   66.63  < 2e-16 ***
## LastAuthorFemale1 -0.00701    0.00870   -0.81  0.41989
## Year1997         0.01024    0.02610    0.39  0.69472
## Year1998        -0.07873    0.02382   -3.31  0.00095 ***
## Year1999        -0.11977    0.02286   -5.24  1.6e-07 ***
## Year2000        -0.09840    0.02429   -4.05  5.1e-05 ***
## Year2001        -0.11646    0.02372   -4.91  9.3e-07 ***
## Year2002        -0.10406    0.02139   -4.87  1.2e-06 ***
## Year2003        -0.13143    0.02186   -6.01  1.9e-09 ***
## Year2004        -0.13505    0.02180   -6.20  6.0e-10 ***
## Year2005        -0.12897    0.02188   -5.89  3.9e-09 ***
## Year2006        -0.14050    0.02185   -6.43  1.3e-10 ***
```

```

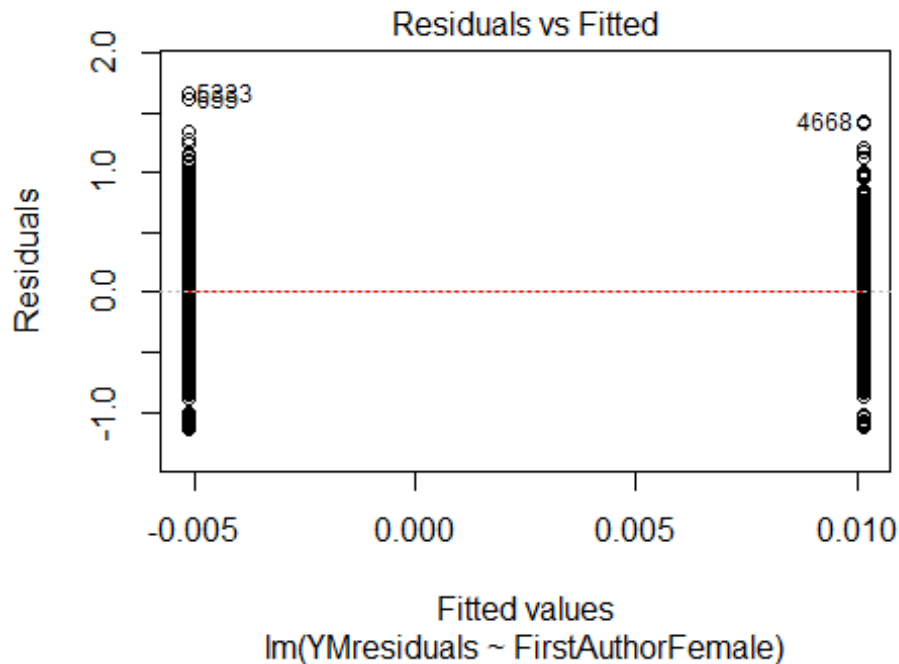
## Year2007          -0.13237      0.02170      -6.10  1.1e-09 ***
## Year2008          -0.11779      0.02219      -5.31  1.1e-07 ***
## Year2009          -0.09596      0.02161      -4.44  9.1e-06 ***
## Year2010          -0.09913      0.02180      -4.55  5.5e-06 ***
## Year2011          -0.11896      0.02208      -5.39  7.3e-08 ***
## Year2012          -0.12997      0.02267      -5.73  1.0e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.329
## Multiple R-squared:  0.0144, Adjusted R-squared:  0.013
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 4 observations c(4768,9408,9609,11213)
## are outliers with |weight| = 0 ( < 8.8e-06);
## 968 weights are ~ = 1. The remaining 10370 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0014 0.8640 0.9500 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.82e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 11342"
## [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
## 309 288 354 287 305 266 308 301 289 517 350 464 373 444 443
## 2011 2012
## 416 518
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 163 155 201 165 164 130 178 171 186 206 224 254 235 279 317
## 2011 2012
## 289 296
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 129 122 166 137 138 108 137 138 156 173 189 199 205 241 259
## 2011 2012
## 243 241
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 31, df = 16, p-value = 0.01
```

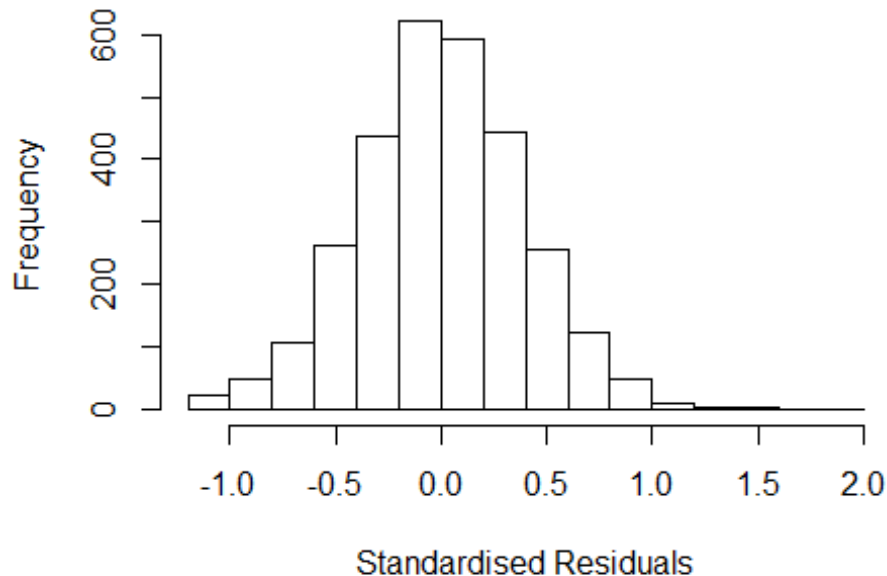


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 3.85209764961214"
## [1] "Male first author team size 2018 geometric mean: 3.47951860262355"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6400, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.30036703679005"
## [1] "Male last author team size 2018 geometric mean: 3.73532277746221"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3400, 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 1.025 1      1.013
## LastAuthorFemale  1.033 1      1.016
## UniqueAuthors    1.106 4      1.013
## Year             1.115 16      1.003
```


Residuals from first and last author and team size



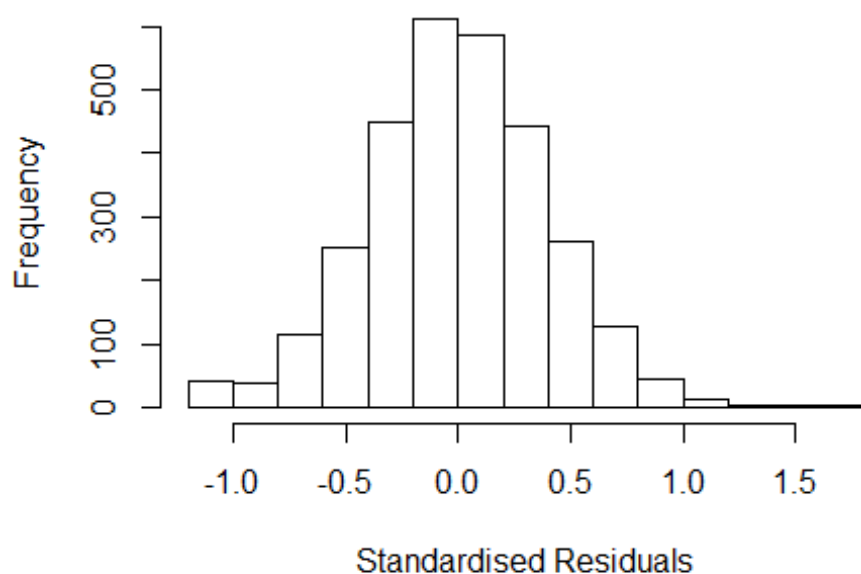
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.15200 -0.25567 -0.00439  0.25629  1.86745
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.913408   0.061496  14.85  < 2e-16 ***
## FirstAuthorFemale1 0.000883   0.015104   0.06  0.95338
## LastAuthorFemale1 0.022358   0.019057   1.17  0.24081
## UniqueAuthors2    0.174434   0.058445   2.98  0.00286 **
## UniqueAuthors3    0.223722   0.058392   3.83  0.00013 ***
## UniqueAuthors4    0.213977   0.059128   3.62  0.00030 ***
## UniqueAuthors5    0.238959   0.058594   4.08  4.7e-05 ***
## Year1997          -0.000369   0.054987  -0.01  0.99464
## Year1998          -0.111701   0.043574  -2.56  0.01041 *
## Year1999          -0.024012   0.048279  -0.50  0.61897
```

```

## Year2000      -0.040058    0.046257   -0.87  0.38656
## Year2001      -0.006598    0.049004   -0.13  0.89291
## Year2002      -0.085851    0.046278   -1.86  0.06368 .
## Year2003      -0.062293    0.046653   -1.34  0.18190
## Year2004      -0.079618    0.041553   -1.92  0.05545 .
## Year2005      -0.009740    0.045343   -0.21  0.82994
## Year2006      -0.067769    0.043197   -1.57  0.11680
## Year2007      -0.022469    0.041162   -0.55  0.58519
## Year2008      -0.033856    0.040406   -0.84  0.40216
## Year2009      -0.035751    0.040475   -0.88  0.37714
## Year2010      -0.016771    0.038173   -0.44  0.66045
## Year2011      -0.076483    0.039860   -1.92  0.05510 .
## Year2012      -0.097957    0.039802   -2.46  0.01391 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.379
## Multiple R-squared:  0.0248, Adjusted R-squared:  0.0175
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 1950 is an outlier with |weight| = 0 ( < 3.4e-05);
## 265 weights are ~= 1. The remaining 2715 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0215 0.8690 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          3.35e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.018 1 1.009
## LastAuthorFemale 1.021 1 1.010
## Year 1.024 16 1.001

```

Residuals from first and last author



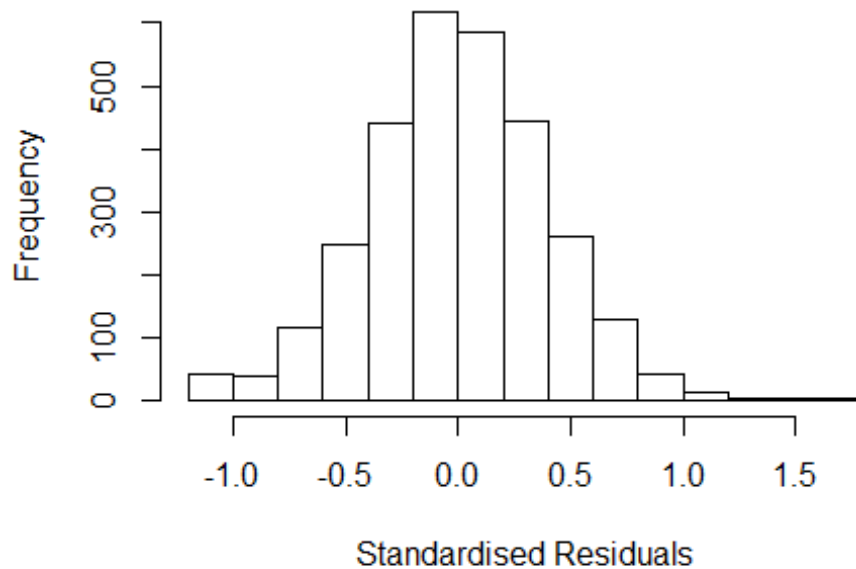
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.14299 -0.25899 -0.00379 0.25780 1.65980
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.105066 0.032278 34.24 <2e-16 ***
## FirstAuthorFemale1 0.008717 0.015105 0.58 0.564
## LastAuthorFemale1 0.026446 0.018940 1.40 0.163
## Year1997 -0.000655 0.055601 -0.01 0.991
## Year1998 -0.110247 0.044927 -2.45 0.014 *
## Year1999 -0.019493 0.048793 -0.40 0.690
## Year2000 -0.035042 0.047811 -0.73 0.464
## Year2001 0.000349 0.049187 0.01 0.994
## Year2002 -0.072274 0.046579 -1.55 0.121
## Year2003 -0.058518 0.046986 -1.25 0.213
## Year2004 -0.068193 0.041897 -1.63 0.104
## Year2005 0.002765 0.046181 0.06 0.952
```

```

## Year2006      -0.057303    0.043841   -1.31    0.191
## Year2007      -0.008943    0.041634   -0.21    0.830
## Year2008      -0.017868    0.041226   -0.43    0.665
## Year2009      -0.025395    0.040974   -0.62    0.535
## Year2010       0.002365    0.038640    0.06    0.951
## Year2011      -0.059046    0.040507   -1.46    0.145
## Year2012      -0.084734    0.040743   -2.08    0.038 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.00864,    Adjusted R-squared:  0.00261
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 255 weights are ~= 1. The remaining 2726 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0186 0.8690 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      3.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.01 1      1.005
## Year      1.01 16      1.000

```

Residuals from first author



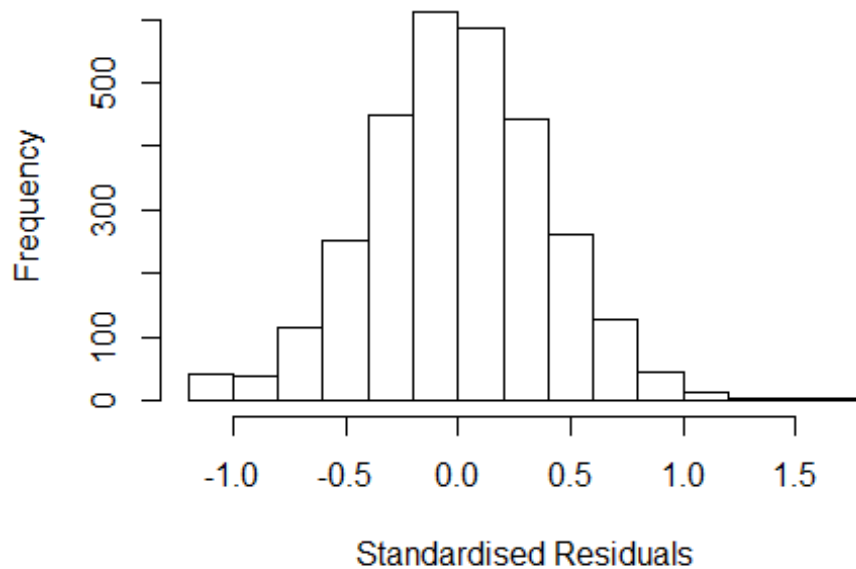
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.12199 -0.26076 -0.00488  0.25781  1.65580
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.109271   0.031960   34.71  <2e-16 ***
## FirstAuthorFemale1 0.010782   0.015068    0.72   0.474
## Year1997        -0.001499   0.055454   -0.03   0.978
## Year1998        -0.110866   0.044719   -2.48   0.013 *
## Year1999        -0.020037   0.048636   -0.41   0.680
## Year2000        -0.035647   0.047659   -0.75   0.455
## Year2001         0.000714   0.049247    0.01   0.988
## Year2002        -0.072349   0.046280   -1.56   0.118
## Year2003        -0.059295   0.046858   -1.27   0.206
## Year2004        -0.068364   0.041750   -1.64   0.102
## Year2005         0.001935   0.046055    0.04   0.966
## Year2006        -0.057655   0.043711   -1.32   0.187
```

```

## Year2007          -0.010395    0.041471    -0.25    0.802
## Year2008          -0.018075    0.041094    -0.44    0.660
## Year2009          -0.024786    0.040891    -0.61    0.544
## Year2010           0.003076    0.038462     0.08    0.936
## Year2011          -0.058813    0.040360    -1.46    0.145
## Year2012          -0.084889    0.040632    -2.09    0.037 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.00795,    Adjusted R-squared:  0.00225
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 254 weights are ~= 1. The remaining 2727 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0195 0.8690 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      3.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.013 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
## -1.13825 -0.25887 -0.00349 0.25763 1.65662
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.107304 0.031928 34.68 <2e-16 ***
## LastAuthorFemale1 0.027503 0.018885 1.46 0.145
## Year1997 -0.000698 0.055651 -0.01 0.990
## Year1998 -0.109658 0.044952 -2.44 0.015 *
## Year1999 -0.019258 0.048827 -0.39 0.693
## Year2000 -0.034549 0.047768 -0.72 0.470
## Year2001 0.000890 0.049217 0.02 0.986
## Year2002 -0.071789 0.046615 -1.54 0.124
## Year2003 -0.057875 0.047012 -1.23 0.218
## Year2004 -0.067436 0.041922 -1.61 0.108
## Year2005 0.003446 0.046200 0.07 0.941
## Year2006 -0.056743 0.043874 -1.29 0.196
```

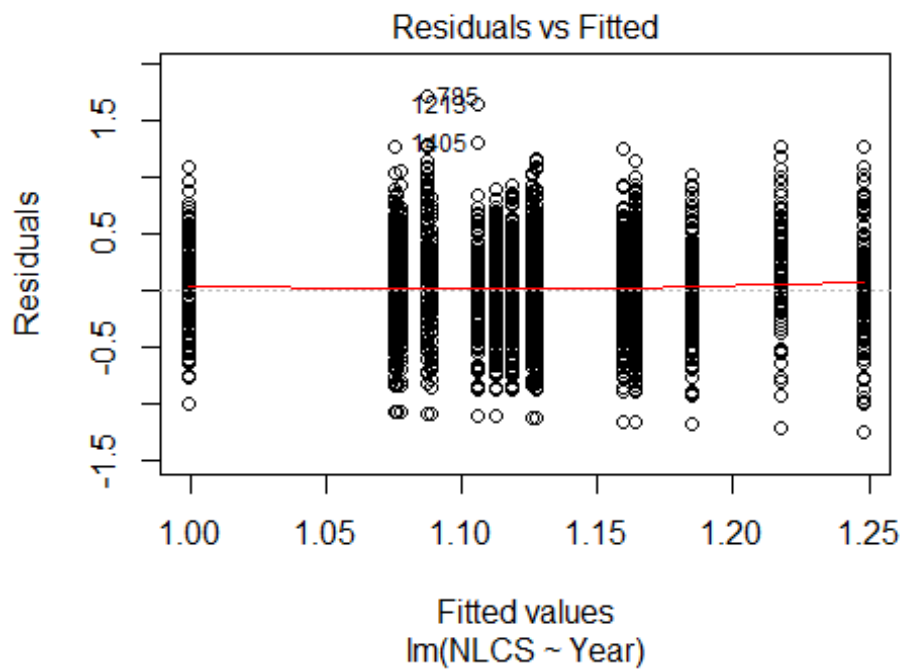
```

## Year2007          -0.008228    0.041673    -0.20    0.843
## Year2008          -0.016920    0.041190    -0.41    0.681
## Year2009          -0.024814    0.040955    -0.61    0.545
## Year2010           0.003069    0.038682     0.08    0.937
## Year2011          -0.058236    0.040524    -1.44    0.151
## Year2012          -0.084198    0.040752    -2.07    0.039 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.00852,    Adjusted R-squared:  0.00283
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 252 weights are ~= 1. The remaining 2729 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0195 0.8680 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      3.35e-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: 2981"
## [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
## 401 348 276 237 210 257 269 273 255 311 352 383 364 369 395
## 2011 2012
## 392 414
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 125 134 95 123 88 84 139 156 151 183 197 236 221 224 261
## 2011 2012

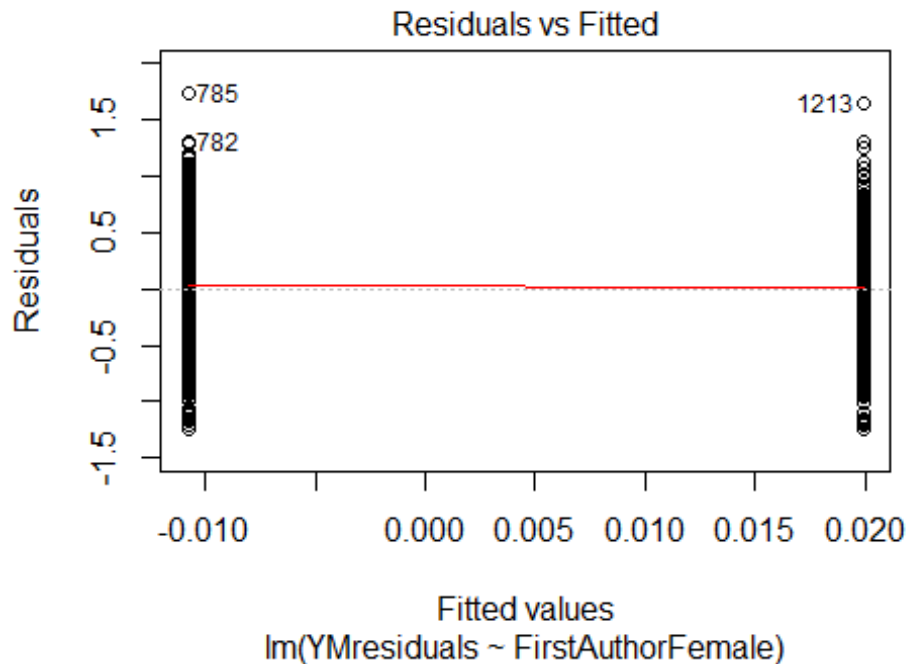
```



```
## 260 264
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 110 113 79 104 72 71 112 128 128 151 162 196 187 188 223
## 2011 2012
## 207 225
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 140, df = 16, p-value <2e-16
```

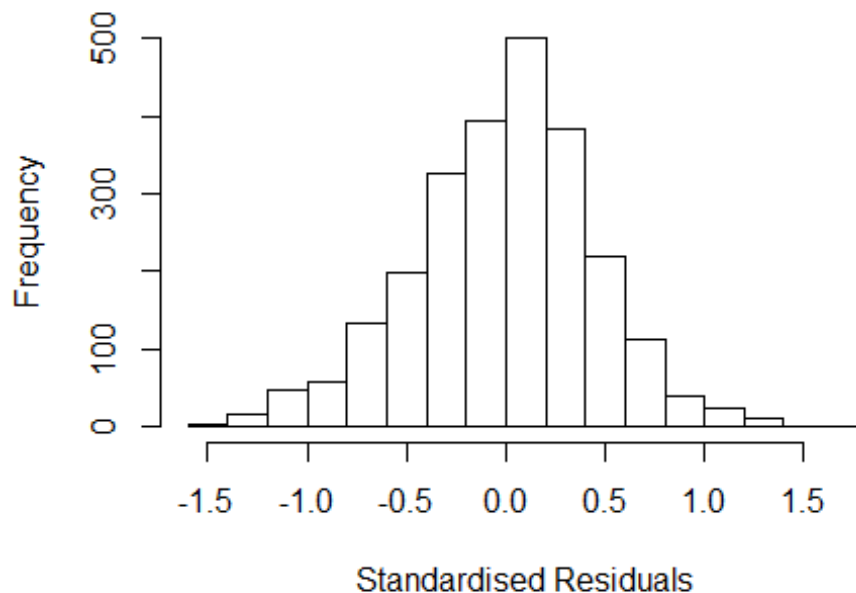


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 4.29371097595593"
## [1] "Male first author team size 2018 geometric mean: 3.58595342376098"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4200, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.37204863652859"
## [1] "Male last author team size 2018 geometric mean: 3.74179399870506"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.060 1          1.030
## LastAuthorFemale  1.035 1          1.018
## UniqueAuthors    1.122 4          1.015
## Year             1.187 16          1.005
```

Residuals from first and last author and team size



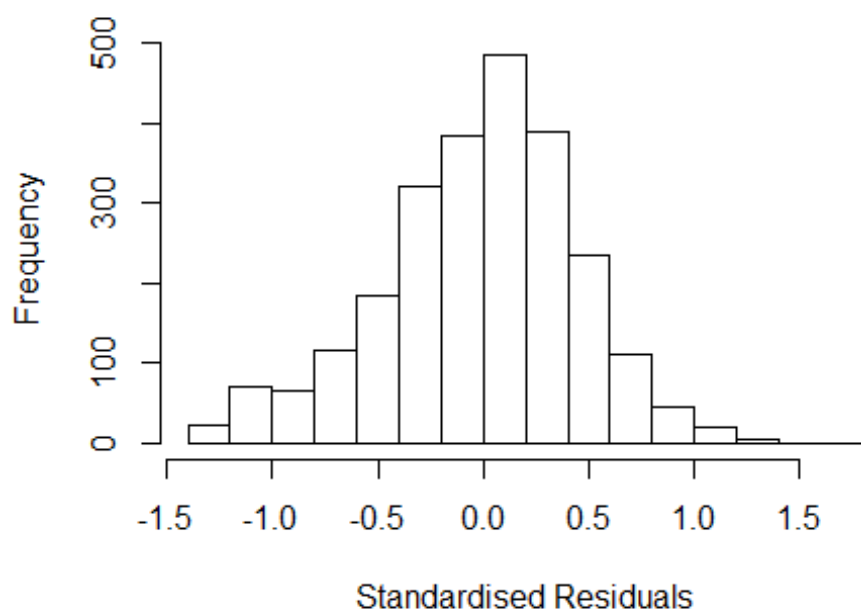
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4616 -0.2991 0.0278 0.2835 1.6766
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97903 0.07997 12.24 < 2e-16 ***
## FirstAuthorFemale1 0.00653 0.01857 0.35 0.72528
## LastAuthorFemale1 0.05961 0.02313 2.58 0.01003 *
## UniqueAuthors2 0.33857 0.05610 6.04 1.8e-09 ***
## UniqueAuthors3 0.35529 0.05566 6.38 2.1e-10 ***
## UniqueAuthors4 0.40548 0.05643 7.19 8.8e-13 ***
## UniqueAuthors5 0.47606 0.05460 8.72 < 2e-16 ***
## Year1997 -0.23542 0.10410 -2.26 0.02382 *
## Year1998 -0.08856 0.09005 -0.98 0.32545
## Year1999 -0.24072 0.07842 -3.07 0.00217 **
```

```

## Year2000      -0.26650    0.08054   -3.31  0.00095 ***
## Year2001      -0.25980    0.07693   -3.38  0.00074 ***
## Year2002      -0.25770    0.07288   -3.54  0.00041 ***
## Year2003      -0.26295    0.07079   -3.71  0.00021 ***
## Year2004      -0.35008    0.07422   -4.72  2.5e-06 ***
## Year2005      -0.27778    0.06782   -4.10  4.3e-05 ***
## Year2006      -0.21446    0.06776   -3.16  0.00157 **
## Year2007      -0.24857    0.06685   -3.72  0.00021 ***
## Year2008      -0.17699    0.06818   -2.60  0.00949 **
## Year2009      -0.18024    0.06807   -2.65  0.00815 **
## Year2010      -0.25203    0.06753   -3.73  0.00019 ***
## Year2011      -0.19182    0.06841   -2.80  0.00509 **
## Year2012      -0.24669    0.06858   -3.60  0.00033 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.423
## Multiple R-squared:  0.0844, Adjusted R-squared:  0.0761
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 198 weights are ~= 1. The remaining 2258 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0809 0.8660 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      4.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1      1.022
## LastAuthorFemale 1.025 1      1.013
## Year 1.067 16      1.002

```

Residuals from first and last author



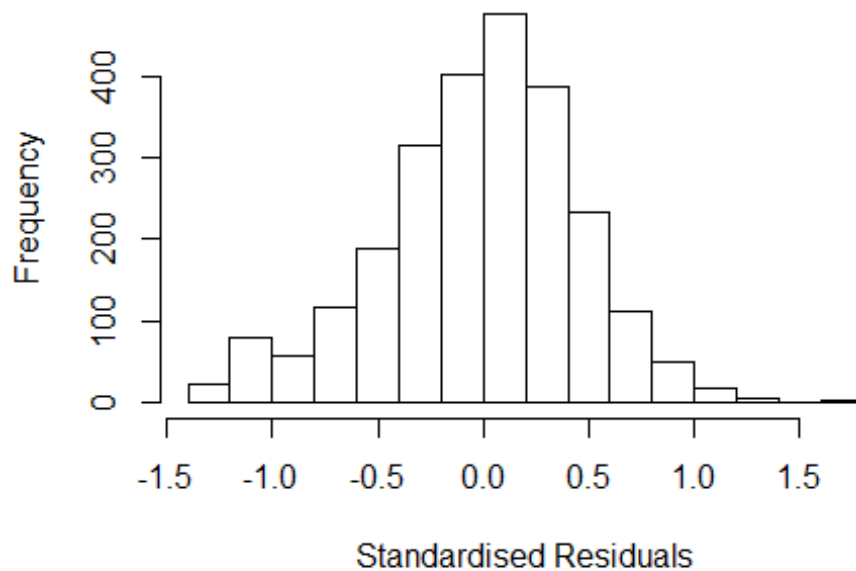
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3414 -0.2921  0.0234  0.2821  1.6588
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3170     0.0616   21.38 < 2e-16 ***
## FirstAuthorFemale1  0.0244     0.0190    1.29  0.19820
## LastAuthorFemale1  0.0550     0.0236    2.33  0.02013 *
## Year1997          -0.2573     0.1083   -2.38  0.01760 *
## Year1998          -0.0886     0.0919   -0.96  0.33523
## Year1999          -0.2401     0.0782   -3.07  0.00215 **
## Year2000          -0.2379     0.0823   -2.89  0.00388 **
## Year2001          -0.2510     0.0803   -3.13  0.00179 **
## Year2002          -0.2454     0.0759   -3.23  0.00124 **
## Year2003          -0.2350     0.0708   -3.32  0.00091 ***
## Year2004          -0.3232     0.0765   -4.22  2.5e-05 ***
## Year2005          -0.2403     0.0681   -3.53  0.00042 ***
```

```

## Year2006          -0.1887      0.0682   -2.77  0.00570 **
## Year2007          -0.1966      0.0676   -2.91  0.00364 **
## Year2008          -0.1400      0.0691   -2.03  0.04265 *
## Year2009          -0.1303      0.0684   -1.91  0.05681 .
## Year2010          -0.2050      0.0683   -3.00  0.00273 **
## Year2011          -0.1479      0.0697   -2.12  0.03392 *
## Year2012          -0.1952      0.0690   -2.83  0.00469 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.43
## Multiple R-squared:  0.0241, Adjusted R-squared:  0.0169
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 195 weights are ~= 1. The remaining 2261 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.105  0.865  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      4.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1      1.021
## Year              1.042 16      1.001

```

Residuals from first author

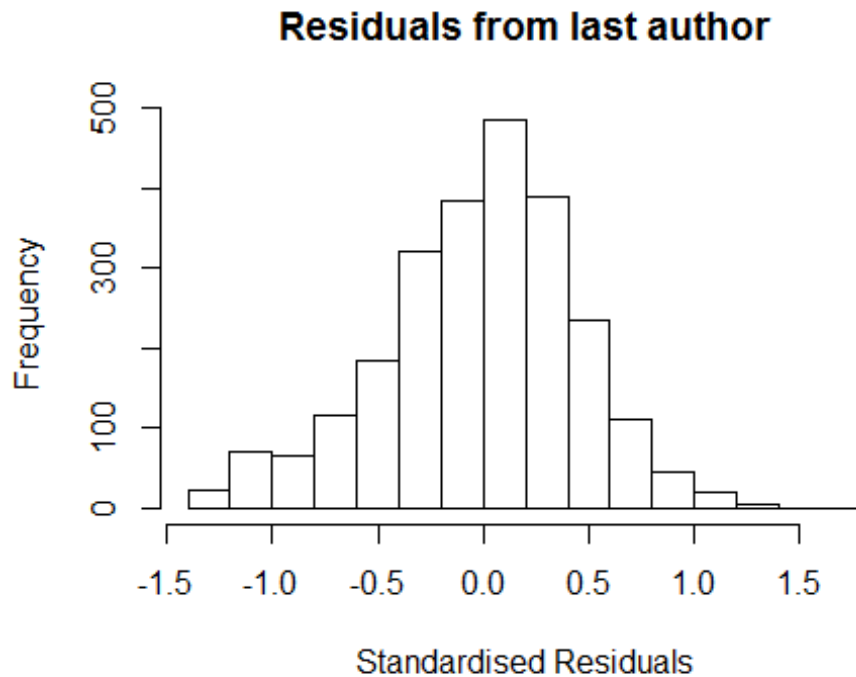


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3561 -0.2895 0.0218 0.2894 1.6491
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3273 0.0617 21.52 < 2e-16 ***
## FirstAuthorFemale1 0.0288 0.0190 1.52 0.12974
## Year1997 -0.2643 0.1078 -2.45 0.01425 *
## Year1998 -0.0928 0.0915 -1.01 0.31075
## Year1999 -0.2453 0.0784 -3.13 0.00178 **
## Year2000 -0.2397 0.0829 -2.89 0.00385 **
## Year2001 -0.2500 0.0807 -3.10 0.00197 **
## Year2002 -0.2499 0.0763 -3.28 0.00107 **
## Year2003 -0.2365 0.0711 -3.32 0.00090 ***
## Year2004 -0.3259 0.0769 -4.24 2.3e-05 ***
## Year2005 -0.2420 0.0685 -3.53 0.00042 ***
## Year2006 -0.1918 0.0685 -2.80 0.00514 **
```

```

## Year2007          -0.1987      0.0679   -2.92  0.00348 **
## Year2008          -0.1403      0.0695   -2.02  0.04365 *
## Year2009          -0.1305      0.0688   -1.90  0.05793 .
## Year2010          -0.2058      0.0687   -3.00  0.00277 **
## Year2011          -0.1478      0.0701   -2.11  0.03497 *
## Year2012          -0.1926      0.0694   -2.78  0.00553 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.431
## Multiple R-squared:  0.022, Adjusted R-squared:  0.0151
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 203 weights are ~= 1. The remaining 2253 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110  0.864  0.950  0.895  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.024 1      1.012
## Year      1.024 16      1.001

```

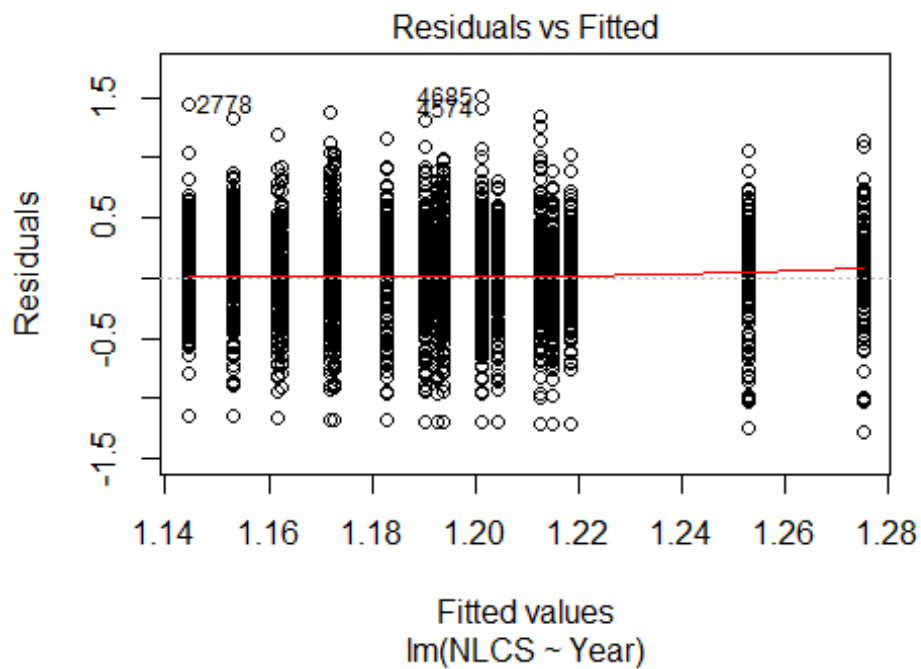
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3214 -0.2922  0.0237  0.2832  1.6752
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3214     0.0617   21.43 < 2e-16 ***
## LastAuthorFemale1  0.0579     0.0237    2.45  0.01449 *
## Year1997        -0.2552     0.1083   -2.36  0.01851 *
## Year1998        -0.0842     0.0916   -0.92  0.35791
## Year1999        -0.2366     0.0780   -3.03  0.00246 **
## Year2000        -0.2374     0.0823   -2.89  0.00394 **
## Year2001        -0.2480     0.0801   -3.10  0.00197 **
## Year2002        -0.2393     0.0755   -3.17  0.00156 **
## Year2003        -0.2311     0.0707   -3.27  0.00111 **
## Year2004        -0.3184     0.0765   -4.16  3.3e-05 ***
## Year2005        -0.2365     0.0680   -3.48  0.00052 ***
## Year2006        -0.1837     0.0680   -2.70  0.00699 **
```

```

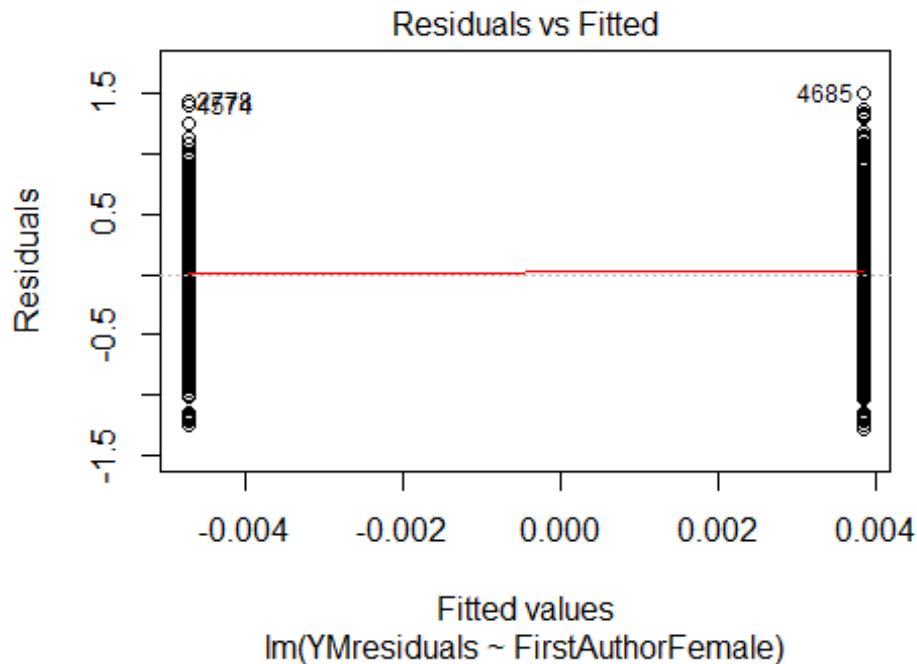
## Year2007          -0.1908      0.0674   -2.83   0.00467 **
## Year2008          -0.1360      0.0690   -1.97   0.04880 *
## Year2009          -0.1244      0.0682   -1.83   0.06807 .
## Year2010          -0.2016      0.0683   -2.95   0.00319 **
## Year2011          -0.1432      0.0696   -2.06   0.03971 *
## Year2012          -0.1909      0.0688   -2.77   0.00560 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.43
## Multiple R-squared:  0.0233, Adjusted R-squared:  0.0165
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 191 weights are ~= 1. The remaining 2265 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0956 0.8640 0.9510 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2456"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
## 287 289 339 285 290 289 299 277 294 337 349 385 363 363 409
## 2011 2012
## 390 425
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 138 152 190 164 156 143 199 184 197 233 229 268 266 259 282
## 2011 2012

```

```
## 274 301
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 118 131 166 129 130 118 171 154 158 211 197 243 237 224 245
## 2011 2012
## 246 275
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 72, df = 16, p-value = 5e-09
```

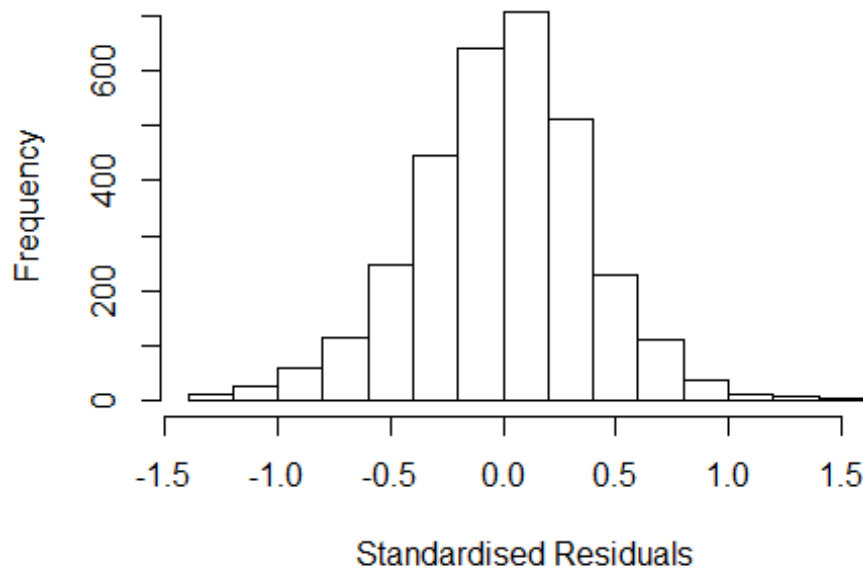


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.7, df = 1, p-value = 0.009
```



```
## [1] "Female first author team size 2018 geometric mean: 6.31880719513006"
## [1] "Male first author team size 2018 geometric mean: 6.29056770526283"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5600, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.3277701579845"
## [1] "Male last author team size 2018 geometric mean: 6.29318536181739"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4800, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.051 1          1.025
## LastAuthorFemale  1.038 1          1.019
## UniqueAuthors    1.145 4          1.017
## Year             1.195 16          1.006
```

Residuals from first and last author and team size



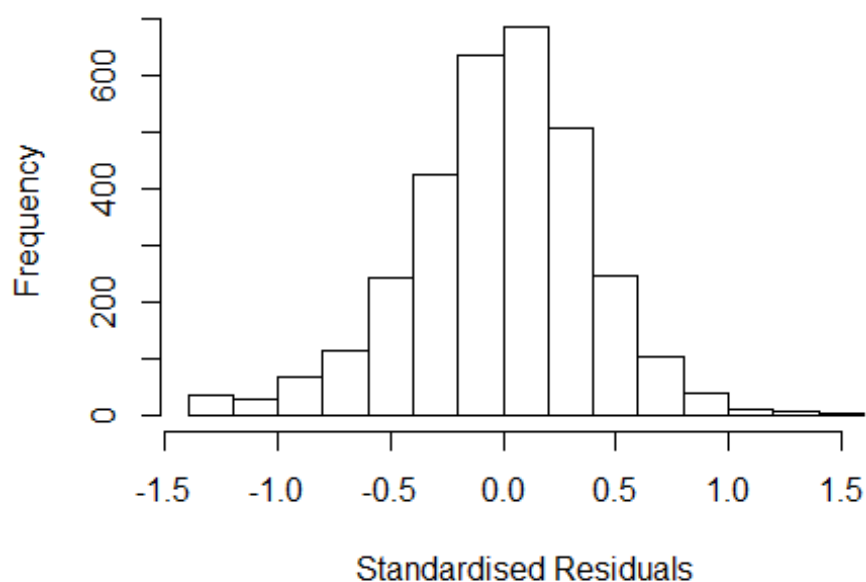
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36138 -0.23843  0.00957  0.23588  1.56311
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.93606    0.07398   12.65 < 2e-16 ***
## FirstAuthorFemale1 -0.02225    0.01365   -1.63  0.10316
## LastAuthorFemale1 -0.00369    0.01529   -0.24  0.80919
## UniqueAuthors2     0.34737    0.06066    5.73  1.1e-08 ***
## UniqueAuthors3     0.36781    0.05951    6.18  7.2e-10 ***
## UniqueAuthors4     0.42630    0.05920    7.20  7.4e-13 ***
## UniqueAuthors5     0.53676    0.05719    9.39 < 2e-16 ***
## Year1997          -0.05819    0.05889   -0.99  0.32321
## Year1998          -0.11144    0.05489   -2.03  0.04241 *
## Year1999          -0.14535    0.05560   -2.61  0.00898 **
```

```

## Year2000      -0.18979    0.05235   -3.63  0.00029 ***
## Year2001      -0.15503    0.05558   -2.79  0.00531 **
## Year2002      -0.17488    0.05162   -3.39  0.00071 ***
## Year2003      -0.23573    0.05095   -4.63  3.9e-06 ***
## Year2004      -0.23279    0.04989   -4.67  3.2e-06 ***
## Year2005      -0.18405    0.04882   -3.77  0.00017 ***
## Year2006      -0.15126    0.05009   -3.02  0.00255 **
## Year2007      -0.19745    0.04941   -4.00  6.6e-05 ***
## Year2008      -0.17298    0.04867   -3.55  0.00038 ***
## Year2009      -0.18450    0.04966   -3.71  0.00021 ***
## Year2010      -0.19722    0.04847   -4.07  4.8e-05 ***
## Year2011      -0.22002    0.05016   -4.39  1.2e-05 ***
## Year2012      -0.21695    0.04786   -4.53  6.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.354
## Multiple R-squared:  0.11,   Adjusted R-squared:  0.103
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 273 weights are ~= 1. The remaining 2880 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0129 0.8640 0.9510 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.17e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.030 1 1.015
## LastAuthorFemale 1.033 1 1.016
## Year 1.063 16 1.002

```

Residuals from first and last author



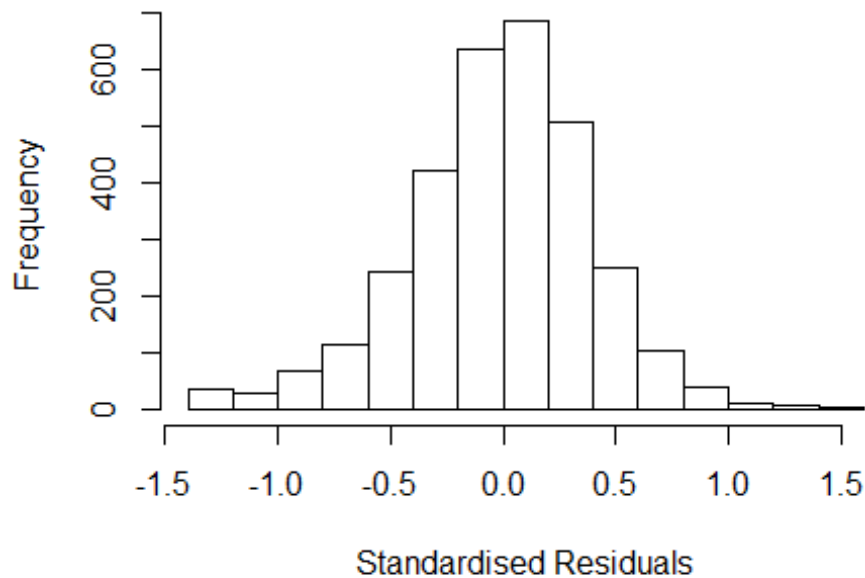
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35416 -0.24256  0.00941  0.24696  1.50361
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.354156   0.042312  32.00  < 2e-16 ***
## FirstAuthorFemale1 -0.017371   0.013996  -1.24  0.21464
## LastAuthorFemale1 -0.000532   0.015735  -0.03  0.97303
## Year1997        -0.036585   0.060080  -0.61  0.54261
## Year1998        -0.119810   0.055990  -2.14  0.03245 *
## Year1999        -0.129842   0.054826  -2.37  0.01793 *
## Year2000        -0.159913   0.051644  -3.10  0.00198 **
## Year2001        -0.126405   0.054688  -2.31  0.02088 *
## Year2002        -0.150890   0.052163  -2.89  0.00385 **
## Year2003        -0.195198   0.050813  -3.84  0.00012 ***
## Year2004        -0.190163   0.049447  -3.85  0.00012 ***
## Year2005        -0.131503   0.049091  -2.68  0.00743 **
```

```

## Year2006      -0.102087    0.050070    -2.04    0.04154 *
## Year2007      -0.146762    0.049268    -2.98    0.00292 **
## Year2008      -0.116831    0.048166    -2.43    0.01534 *
## Year2009      -0.139587    0.049118    -2.84    0.00451 **
## Year2010      -0.135744    0.048726    -2.79    0.00537 **
## Year2011      -0.183374    0.049619    -3.70    0.00022 ***
## Year2012      -0.158601    0.047645    -3.33    0.00088 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.364
## Multiple R-squared:  0.013, Adjusted R-squared:  0.00731
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 265 weights are ~= 1. The remaining 2888 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0501 0.8640 0.9510 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      3.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.03 1      1.015
## Year      1.03 16      1.001

```


Residuals from first author



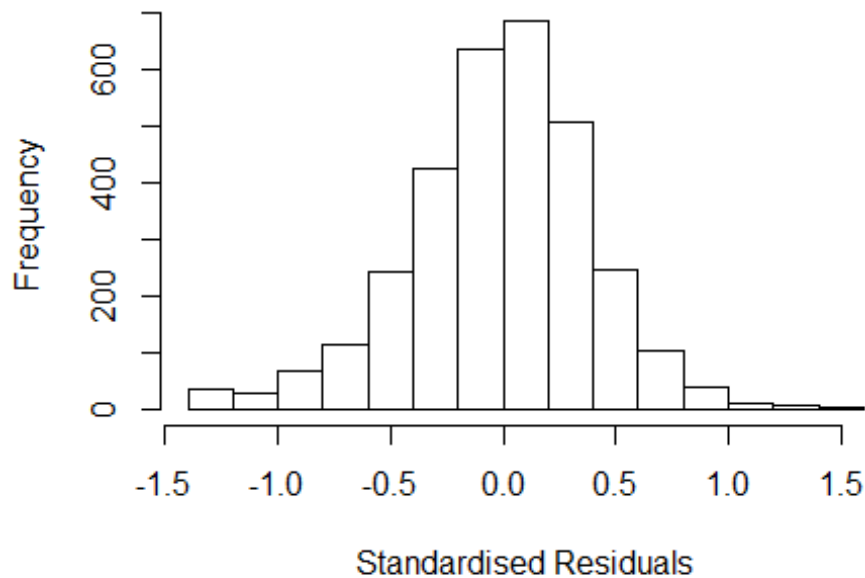
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35399 -0.24283  0.00957  0.24717  1.50374
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3540     0.0422   32.11 < 2e-16 ***
## FirstAuthorFemale1 -0.0174     0.0140   -1.24  0.21485
## Year1997         -0.0365     0.0601   -0.61  0.54334
## Year1998         -0.1198     0.0560   -2.14  0.03253 *
## Year1999         -0.1298     0.0548   -2.37  0.01791 *
## Year2000         -0.1599     0.0516   -3.10  0.00198 **
## Year2001         -0.1264     0.0547   -2.31  0.02096 *
## Year2002         -0.1509     0.0522   -2.89  0.00386 **
## Year2003         -0.1951     0.0508   -3.84  0.00013 ***
## Year2004         -0.1901     0.0495   -3.84  0.00012 ***
## Year2005         -0.1315     0.0491   -2.68  0.00744 **
## Year2006         -0.1021     0.0500   -2.04  0.04115 *
```

```

## Year2007          -0.1467      0.0493   -2.98  0.00293 **
## Year2008          -0.1168      0.0482   -2.42  0.01539 *
## Year2009          -0.1395      0.0491   -2.84  0.00454 **
## Year2010          -0.1357      0.0487   -2.79  0.00535 **
## Year2011          -0.1833      0.0496   -3.69  0.00022 ***
## Year2012          -0.1586      0.0477   -3.33  0.00089 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.364
## Multiple R-squared:  0.013, Adjusted R-squared:  0.00762
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 265 weights are ~= 1. The remaining 2888 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0502 0.8640 0.9510 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      3.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.032 1          1.016
## Year            1.032 16          1.001

```

Residuals from last author



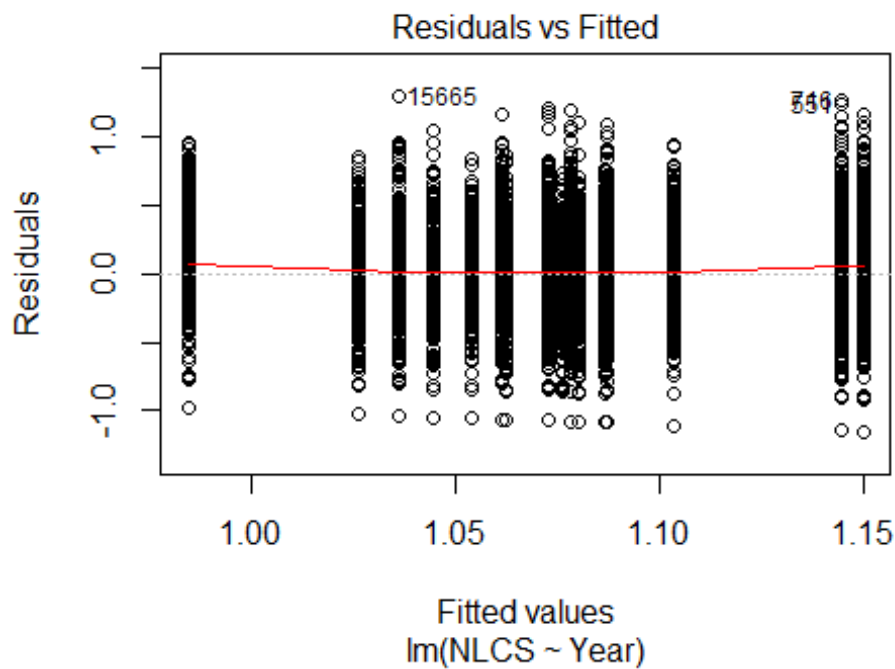
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34875 -0.24349 0.00786 0.24812 1.51064
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.34875 0.04225 31.92 < 2e-16 ***
## LastAuthorFemale1 -0.00217 0.01579 -0.14 0.89048
## Year1997 -0.03833 0.06006 -0.64 0.52345
## Year1998 -0.12110 0.05605 -2.16 0.03081 *
## Year1999 -0.13057 0.05503 -2.37 0.01772 *
## Year2000 -0.16061 0.05181 -3.10 0.00195 **
## Year2001 -0.12911 0.05479 -2.36 0.01851 *
## Year2002 -0.15487 0.05209 -2.97 0.00297 **
## Year2003 -0.19748 0.05091 -3.88 0.00011 ***
## Year2004 -0.19326 0.04952 -3.90 9.7e-05 ***
## Year2005 -0.13406 0.04915 -2.73 0.00642 **
## Year2006 -0.10553 0.05003 -2.11 0.03500 *
```

```

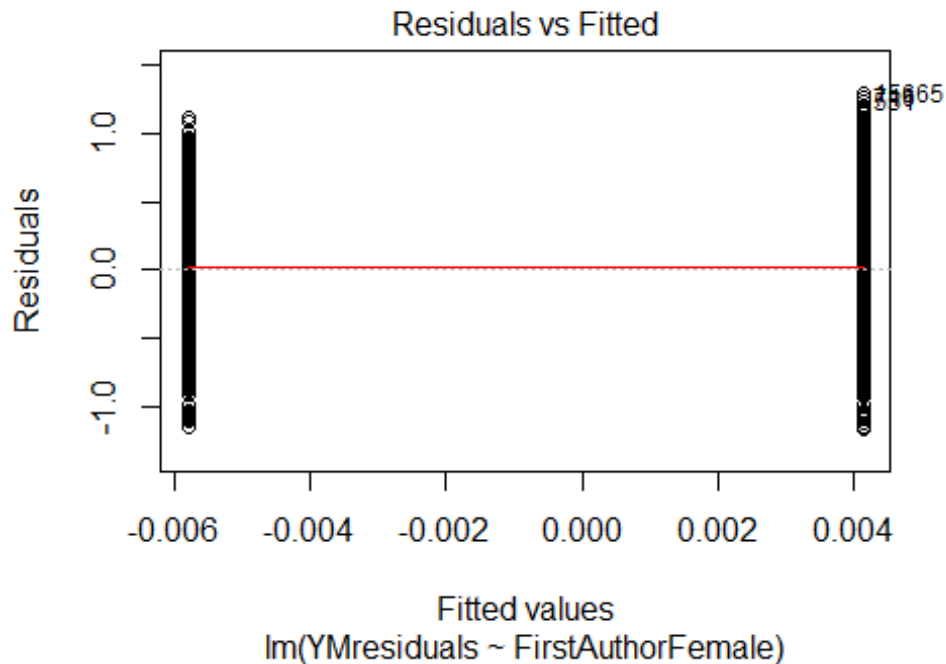
## Year2007          -0.14839      0.04932    -3.01  0.00264 **
## Year2008          -0.11872      0.04825    -2.46  0.01394 *
## Year2009          -0.14191      0.04924    -2.88  0.00398 **
## Year2010          -0.13900      0.04877    -2.85  0.00440 **
## Year2011          -0.18566      0.04975    -3.73  0.00019 ***
## Year2012          -0.16097      0.04776    -3.37  0.00076 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.365
## Multiple R-squared:  0.0124, Adjusted R-squared:  0.00703
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 258 weights are ~ = 1. The remaining 2895 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0481 0.8650 0.9520 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3153"
## [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
## 988 876 800 650 783 710 782 650 656 684 605 759 715 773 792
## 2011 2012
## 715 776
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 500 501 434 414 423 300 546 454 433 464 393 495 490 534 548
## 2011 2012

```

```
## 531 562
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 436 429 369 348 350 255 463 391 364 411 333 426 436 462 460
## 2011 2012
## 446 484
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 420, df = 16, p-value <2e-16
```

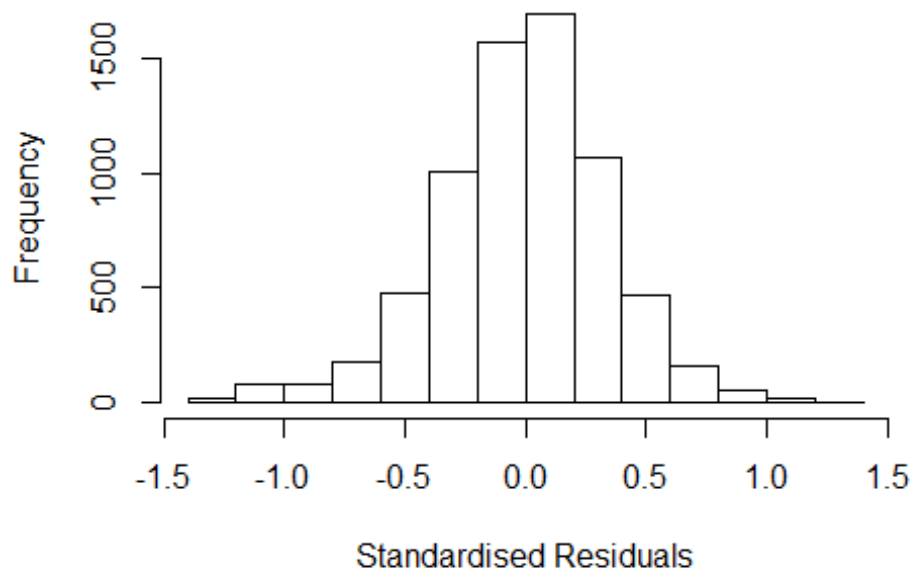


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 17, df = 1, p-value = 5e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.33756640648298"
## [1] "Male first author team size 2018 geometric mean: 4.11580304797044"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.96097670472507"
## [1] "Male last author team size 2018 geometric mean: 4.27787643547703"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6600, 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 1.024 1          1.012
## LastAuthorFemale  1.013 1          1.007
## UniqueAuthors    1.092 4          1.011
## Year             1.105 16          1.003
```

Residuals from first and last author and team size



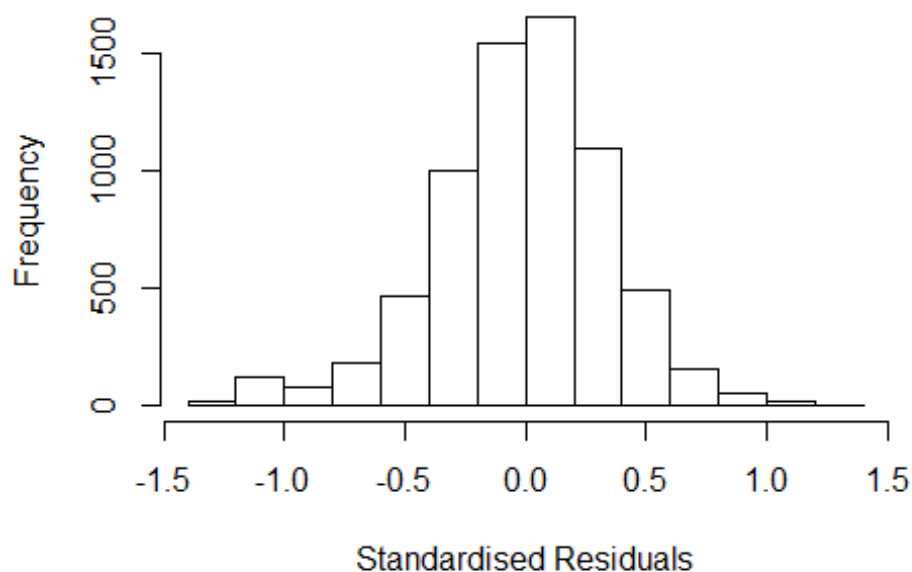
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32718 -0.21655  0.00385  0.20535  1.35967
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.97762    0.03938   24.82 < 2e-16 ***
## FirstAuthorFemale1 -0.01261    0.00825   -1.53  0.12631
## LastAuthorFemale1 -0.00720    0.01013   -0.71  0.47720
## UniqueAuthors2     0.21327    0.03818    5.59 2.4e-08 ***
## UniqueAuthors3     0.24349    0.03795    6.42 1.5e-10 ***
## UniqueAuthors4     0.27107    0.03802    7.13 1.1e-12 ***
## UniqueAuthors5     0.34956    0.03750    9.32 < 2e-16 ***
## Year1997         -0.02669    0.03008   -0.89  0.37506
## Year1998         -0.10514    0.02951   -3.56 0.00037 ***
## Year1999         -0.16069    0.02727   -5.89 4.0e-09 ***
```

```

## Year2000      -0.18515      0.03058      -6.05      1.5e-09 ***
## Year2001      -0.19103      0.03040      -6.28      3.5e-10 ***
## Year2002      -0.15649      0.02528      -6.19      6.4e-10 ***
## Year2003      -0.20578      0.02602      -7.91      3.0e-15 ***
## Year2004      -0.19271      0.02575      -7.48      8.2e-14 ***
## Year2005      -0.18917      0.02489      -7.60      3.4e-14 ***
## Year2006      -0.17198      0.02548      -6.75      1.6e-11 ***
## Year2007      -0.17824      0.02482      -7.18      7.6e-13 ***
## Year2008      -0.18429      0.02540      -7.25      4.5e-13 ***
## Year2009      -0.14994      0.02561      -5.85      5.0e-09 ***
## Year2010      -0.16351      0.02557      -6.40      1.7e-10 ***
## Year2011      -0.21809      0.02644      -8.25      < 2e-16 ***
## Year2012      -0.17520      0.02609      -6.72      2.0e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.316
## Multiple R-squared:  0.0671, Adjusted R-squared:  0.0641
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 575 weights are ~= 1. The remaining 6288 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0253 0.8650 0.9510 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.46e-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 1.016 1          1.008
## LastAuthorFemale 1.009 1          1.005
## Year 1.023 16          1.001

```


Residuals from first and last author



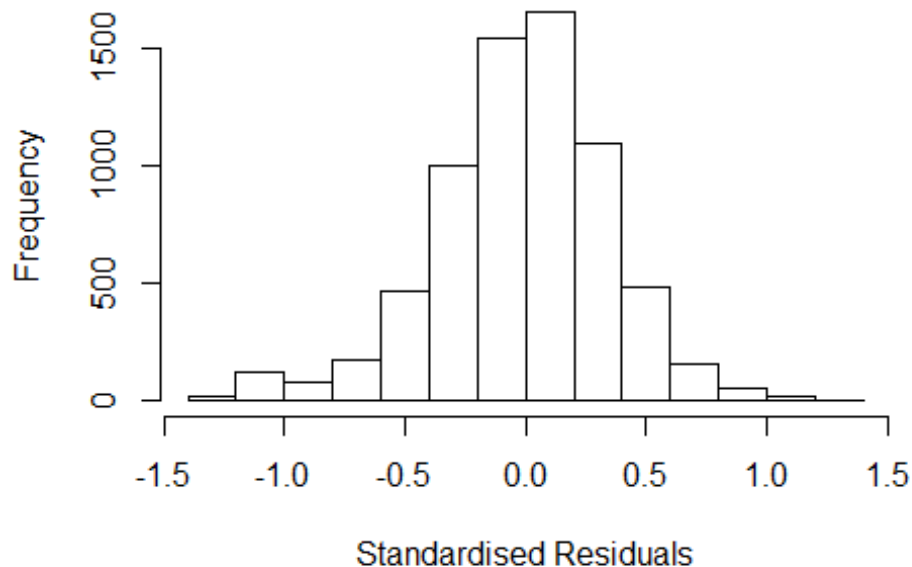
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21049 -0.22222  0.00442  0.21162  1.28031
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.21049    0.02144   56.47 < 2e-16 ***
## FirstAuthorFemale1 -0.00721    0.00834   -0.86  0.3879
## LastAuthorFemale1 -0.00508    0.01023   -0.50  0.6191
## Year1997          -0.01322    0.03046   -0.43  0.6643
## Year1998          -0.06724    0.03016   -2.23  0.0258 *
## Year1999          -0.12291    0.02791   -4.40 1.1e-05 ***
## Year2000          -0.15444    0.03116   -4.96 7.3e-07 ***
## Year2001          -0.15104    0.03105   -4.87 1.2e-06 ***
## Year2002          -0.10580    0.02549   -4.15 3.4e-05 ***
## Year2003          -0.15932    0.02638   -6.04 1.6e-09 ***
## Year2004          -0.14332    0.02588   -5.54 3.2e-08 ***
## Year2005          -0.14228    0.02533   -5.62 2.0e-08 ***
```

```

## Year2006          -0.11867      0.02586      -4.59      4.5e-06 ***
## Year2007          -0.12665      0.02515      -5.04      4.9e-07 ***
## Year2008          -0.12780      0.02587      -4.94      8.0e-07 ***
## Year2009          -0.09615      0.02581      -3.73      0.0002 ***
## Year2010          -0.11263      0.02596      -4.34      1.5e-05 ***
## Year2011          -0.16180      0.02688      -6.02      1.8e-09 ***
## Year2012          -0.12294      0.02663      -4.62      4.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.322
## Multiple R-squared:  0.0185, Adjusted R-squared:  0.0159
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 577 weights are ~= 1. The remaining 6286 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0773 0.8650 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      1.46e-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 1.015 1      1.007
## Year      1.015 16      1.000

```

Residuals from first author



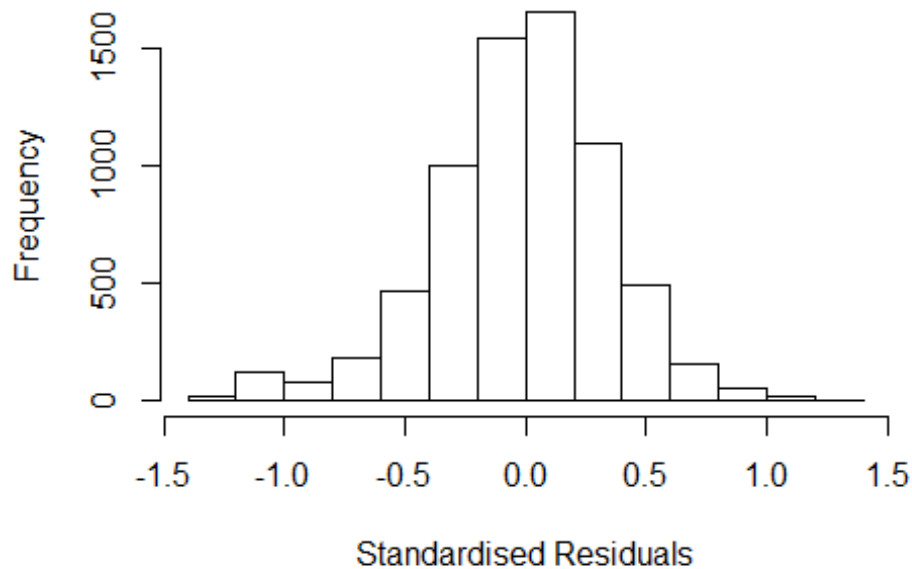
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20960 -0.22232 0.00379 0.21073 1.28130
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20960 0.02134 56.67 < 2e-16 ***
## FirstAuthorFemale1 -0.00749 0.00834 -0.90 0.36931
## Year1997 -0.01303 0.03045 -0.43 0.66873
## Year1998 -0.06715 0.03017 -2.23 0.02605 *
## Year1999 -0.12281 0.02791 -4.40 1.1e-05 ***
## Year2000 -0.15442 0.03116 -4.96 7.4e-07 ***
## Year2001 -0.15119 0.03104 -4.87 1.1e-06 ***
## Year2002 -0.10585 0.02551 -4.15 3.4e-05 ***
## Year2003 -0.15922 0.02638 -6.04 1.7e-09 ***
## Year2004 -0.14309 0.02588 -5.53 3.3e-08 ***
## Year2005 -0.14230 0.02533 -5.62 2.0e-08 ***
## Year2006 -0.11875 0.02586 -4.59 4.5e-06 ***
```

```

## Year2007          -0.12657    0.02516   -5.03  5.0e-07 ***
## Year2008          -0.12799    0.02588   -4.95  7.8e-07 ***
## Year2009          -0.09625    0.02582   -3.73  0.00019 ***
## Year2010          -0.11274    0.02596   -4.34  1.4e-05 ***
## Year2011          -0.16190    0.02688   -6.02  1.8e-09 ***
## Year2012          -0.12301    0.02664   -4.62  3.9e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.322
## Multiple R-squared:  0.0185, Adjusted R-squared:  0.0161
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 579 weights are ~= 1. The remaining 6284 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0765 0.8650 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      1.46e-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 1.008 1      1.004
## Year              1.008 16      1.000

```

Residuals from last author



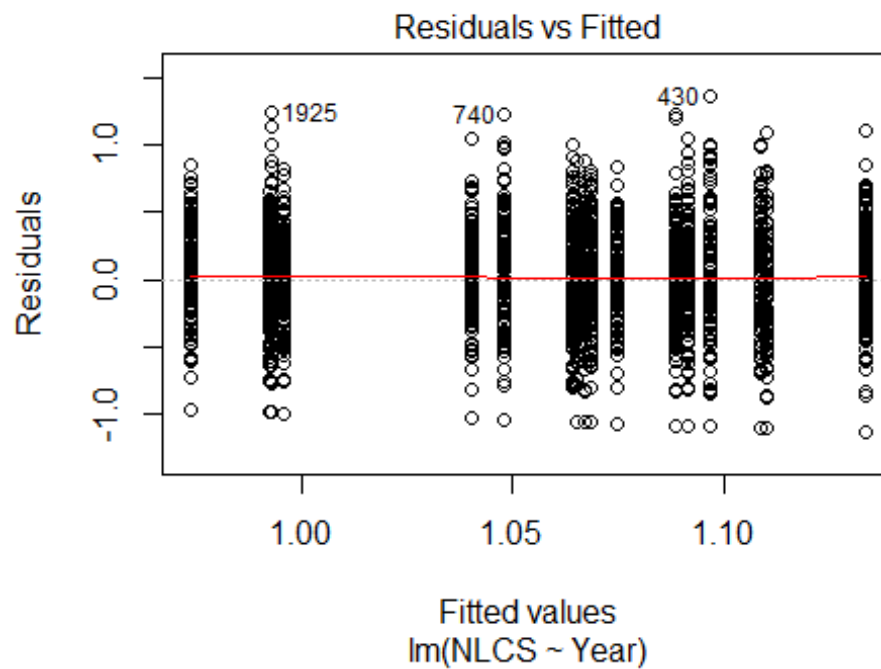
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2077 -0.2222  0.0042  0.2127  1.2836
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20775    0.02112   57.18 < 2e-16 ***
## LastAuthorFemale1 -0.00569    0.01022   -0.56  0.57794
## Year1997        -0.01294    0.03047   -0.42  0.67111
## Year1998        -0.06696    0.03014   -2.22  0.02637 *
## Year1999        -0.12294    0.02792   -4.40  1.1e-05 ***
## Year2000        -0.15384    0.03114   -4.94  8.0e-07 ***
## Year2001        -0.15118    0.03104   -4.87  1.1e-06 ***
## Year2002        -0.10608    0.02551   -4.16  3.2e-05 ***
## Year2003        -0.15949    0.02639   -6.04  1.6e-09 ***
## Year2004        -0.14369    0.02590   -5.55  3.0e-08 ***
## Year2005        -0.14261    0.02534   -5.63  1.9e-08 ***
## Year2006        -0.11888    0.02588   -4.59  4.4e-06 ***
```

```

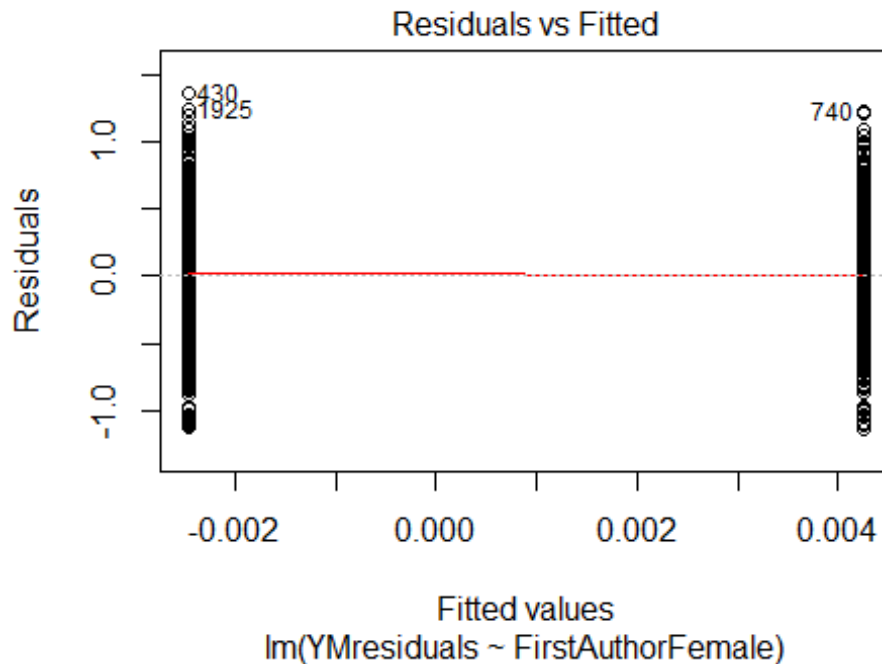
## Year2007          -0.12713      0.02515      -5.06  4.4e-07 ***
## Year2008          -0.12795      0.02587      -4.95  7.7e-07 ***
## Year2009          -0.09656      0.02582      -3.74  0.00019 ***
## Year2010          -0.11317      0.02597      -4.36  1.3e-05 ***
## Year2011          -0.16231      0.02689      -6.04  1.7e-09 ***
## Year2012          -0.12343      0.02664      -4.63  3.7e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.322
## Multiple R-squared:  0.0184, Adjusted R-squared:  0.0159
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 587 weights are ~= 1. The remaining 6276 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0758 0.8650 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      1.46e-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: 6863"
## [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
## 209 205 192 200 206 200 214 205 220 245 195 249 227 243 241
## 2011 2012
## 195 230
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 128 124 109 119 113 84 146 141 153 151 124 161 155 166 173
## 2011 2012

```

```
## 146 162
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 106 98 95 95 98 67 120 124 127 129 108 142 134 141 143
## 2011 2012
## 124 133
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 39, df = 16, p-value = 0.001
```

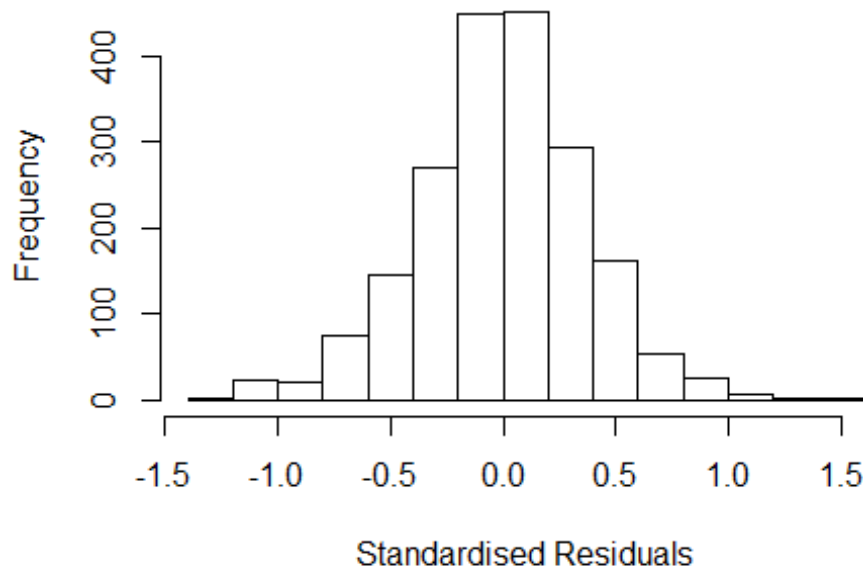


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.009, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 4.76021270084781"
## [1] "Male first author team size 2018 geometric mean: 5.49595663759849"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.6830087289756"
## [1] "Male last author team size 2018 geometric mean: 5.22100343198592"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 820, 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      1.025
## LastAuthorFemale  1.032 1      1.016
## UniqueAuthors     1.236 4      1.027
## Year              1.293 16     1.008
```


Residuals from first and last author and team size



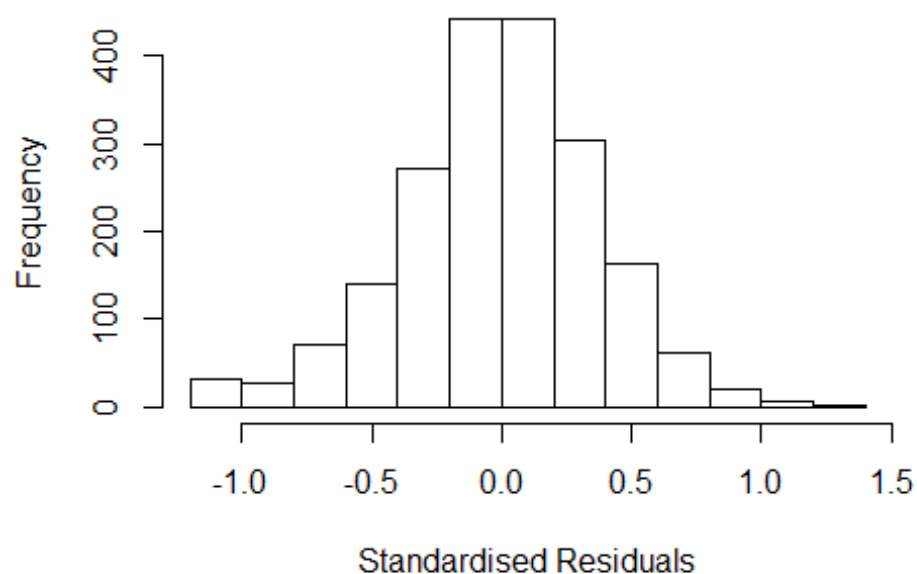
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21243 -0.22730  0.00156  0.22426  1.41367
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8817    0.0847   10.41 < 2e-16 ***
## FirstAuthorFemale1 0.0074    0.0174    0.43  0.67027
## LastAuthorFemale1 0.0473    0.0201    2.35  0.01895 *
## UniqueAuthors2    0.2466    0.0770    3.20  0.00138 **
## UniqueAuthors3    0.2444    0.0766    3.19  0.00145 **
## UniqueAuthors4    0.2598    0.0766    3.39  0.00071 ***
## UniqueAuthors5    0.3308    0.0758    4.36  1.3e-05 ***
## Year1997         -0.0787    0.0588   -1.34  0.18058
## Year1998         -0.0822    0.0631   -1.30  0.19293
## Year1999         -0.0795    0.0547   -1.45  0.14611
```

```

## Year2000          -0.0578      0.0579   -1.00   0.31824
## Year2001          -0.0910      0.0715   -1.27   0.20328
## Year2002          -0.1616      0.0532   -3.04   0.00240 **
## Year2003          -0.2075      0.0540   -3.85   0.00012 ***
## Year2004          -0.1795      0.0542   -3.31   0.00094 ***
## Year2005          -0.1077      0.0496   -2.17   0.02990 *
## Year2006          -0.1394      0.0532   -2.62   0.00890 **
## Year2007          -0.0766      0.0512   -1.50   0.13485
## Year2008          -0.1477      0.0545   -2.71   0.00678 **
## Year2009          -0.0819      0.0525   -1.56   0.11902
## Year2010          -0.1062      0.0500   -2.12   0.03393 *
## Year2011          -0.0303      0.0537   -0.56   0.57260
## Year2012          -0.0749      0.0499   -1.50   0.13370
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.347
## Multiple R-squared:  0.0533, Adjusted R-squared:  0.0427
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 163 weights are ~= 1. The remaining 1821 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0597 0.8640 0.9530 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.04e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.040 1          1.020
## LastAuthorFemale 1.024 1          1.012
## Year 1.057 16          1.002

```

Residuals from first and last author



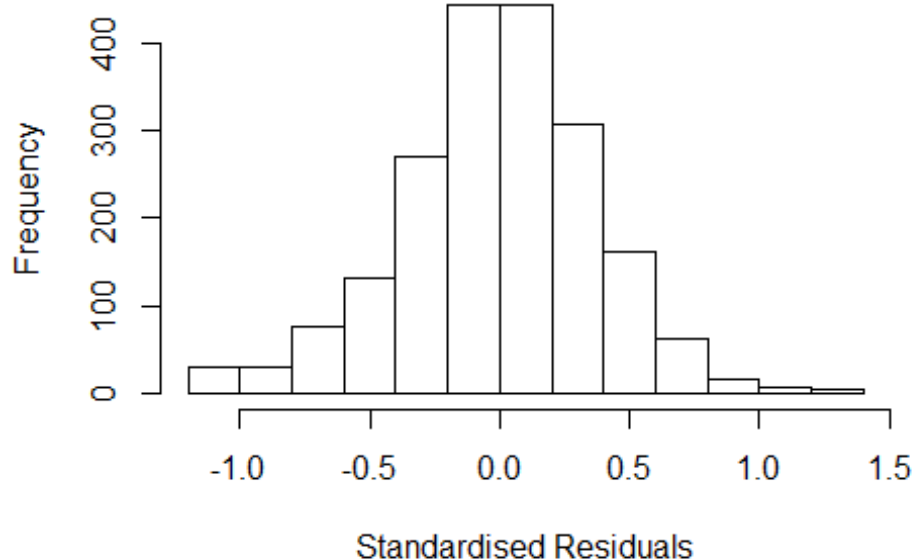
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.14914 -0.23577 0.00471 0.22918 1.39331
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13027 0.04112 27.49 <2e-16 ***
## FirstAuthorFemale1 0.01085 0.01755 0.62 0.5365
## LastAuthorFemale1 0.04503 0.02039 2.21 0.0273 *
## Year1997 -0.06258 0.05759 -1.09 0.2774
## Year1998 -0.07347 0.06253 -1.17 0.2402
## Year1999 -0.06835 0.05296 -1.29 0.1970
## Year2000 -0.05265 0.05570 -0.95 0.3447
## Year2001 -0.06521 0.07067 -0.92 0.3562
## Year2002 -0.13729 0.05216 -2.63 0.0085 **
## Year2003 -0.17483 0.05323 -3.28 0.0010 **
## Year2004 -0.15324 0.05285 -2.90 0.0038 **
## Year2005 -0.08565 0.04837 -1.77 0.0768 .
```

```

## Year2006          -0.10625    0.05288   -2.01    0.0446 *
## Year2007          -0.05936    0.04972   -1.19    0.2327
## Year2008          -0.10680    0.05318   -2.01    0.0447 *
## Year2009          -0.04391    0.05092   -0.86    0.3886
## Year2010          -0.07228    0.04878   -1.48    0.1386
## Year2011           0.00803    0.05325    0.15    0.8802
## Year2012          -0.03406    0.04815   -0.71    0.4795
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.0202, Adjusted R-squared:  0.0112
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 163 weights are ~= 1. The remaining 1821 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0765 0.8670 0.9510 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.035 1      1.018
## Year              1.035 16      1.001

```

Residuals from first author



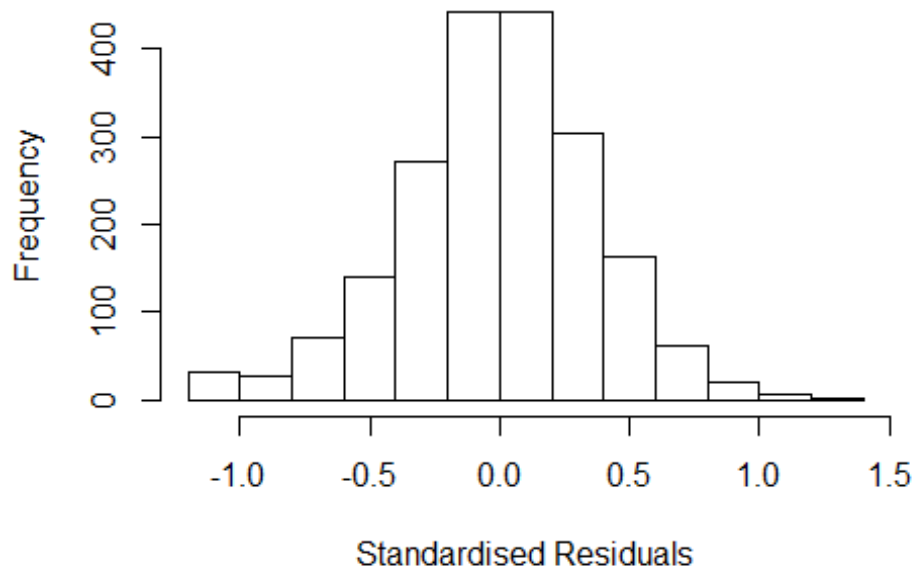
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.16096 -0.23798 0.00539 0.22614 1.38460
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1358 0.0412 27.54 <2e-16 ***
## FirstAuthorFemale1 0.0141 0.0175 0.80 0.4212
## Year1997 -0.0594 0.0577 -1.03 0.3036
## Year1998 -0.0723 0.0625 -1.16 0.2474
## Year1999 -0.0667 0.0533 -1.25 0.2105
## Year2000 -0.0479 0.0561 -0.85 0.3928
## Year2001 -0.0601 0.0710 -0.85 0.3975
## Year2002 -0.1364 0.0525 -2.60 0.0094 **
## Year2003 -0.1729 0.0536 -3.23 0.0013 **
## Year2004 -0.1551 0.0531 -2.92 0.0035 **
## Year2005 -0.0838 0.0487 -1.72 0.0855 .
## Year2006 -0.1033 0.0532 -1.94 0.0522 .
```

```

## Year2007          -0.0567      0.0501   -1.13   0.2580
## Year2008          -0.1062      0.0533   -1.99   0.0463 *
## Year2009          -0.0419      0.0513   -0.82   0.4144
## Year2010          -0.0702      0.0491   -1.43   0.1531
## Year2011           0.0110      0.0537    0.21   0.8370
## Year2012          -0.0330      0.0484   -0.68   0.4950
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.0182, Adjusted R-squared:  0.00969
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 163 weights are ~= 1. The remaining 1821 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0822 0.8680 0.9500 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      5.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.019 1          1.009
## Year              1.019 16          1.001

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1421 -0.2351 0.0045 0.2301 1.3897
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13426 0.04049 28.01 <2e-16 ***
## LastAuthorFemale1 0.04621 0.02035 2.27 0.0233 *
## Year1997 -0.06299 0.05766 -1.09 0.2748
## Year1998 -0.07423 0.06244 -1.19 0.2347
## Year1999 -0.06932 0.05299 -1.31 0.1909
## Year2000 -0.05414 0.05570 -0.97 0.3312
## Year2001 -0.06562 0.07090 -0.93 0.3548
## Year2002 -0.13740 0.05222 -2.63 0.0086 **
## Year2003 -0.17471 0.05339 -3.27 0.0011 **
## Year2004 -0.15371 0.05294 -2.90 0.0037 **
## Year2005 -0.08615 0.04838 -1.78 0.0751 .
## Year2006 -0.10597 0.05287 -2.00 0.0452 *
```

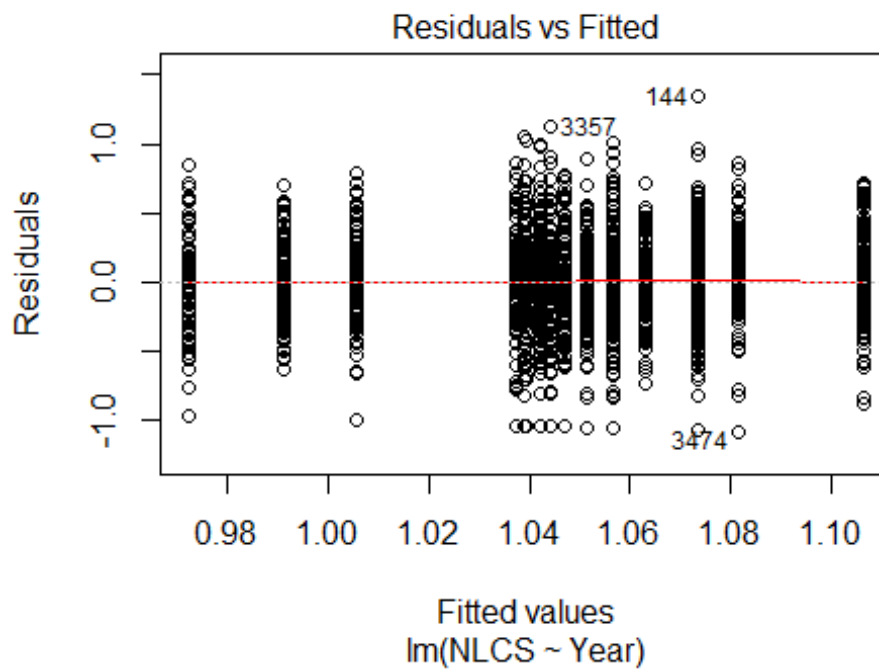
```

## Year2007          -0.05901      0.04975    -1.19    0.2357
## Year2008          -0.10633      0.05319    -2.00    0.0457 *
## Year2009          -0.04474      0.05089    -0.88    0.3794
## Year2010          -0.07214      0.04886    -1.48    0.1400
## Year2011           0.00779      0.05337     0.15    0.8839
## Year2012          -0.03411      0.04821    -0.71    0.4794
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.349
## Multiple R-squared:  0.02,   Adjusted R-squared:  0.0115
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 166 weights are ~= 1. The remaining 1818 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0782 0.8680 0.9510 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1984"
## [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
## 209 189 157 187 141 141 134 161 131 159 190 173 191 217 204
## 2011 2012
## 174 192
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 126 116 95 114 75 79 96 116 100 113 135 123 134 158 147
## 2011 2012

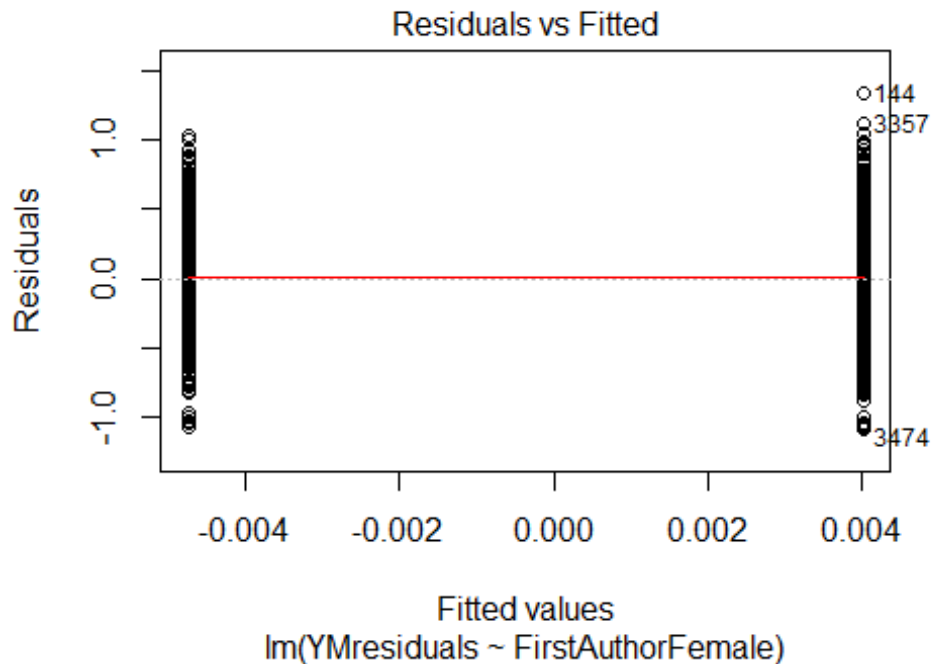
```



```
## 129 131
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 109 99 78 92 63 67 88 100 81 97 119 104 121 136 125
## 2011 2012
## 116 121
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 17, df = 16, p-value = 0.4
```

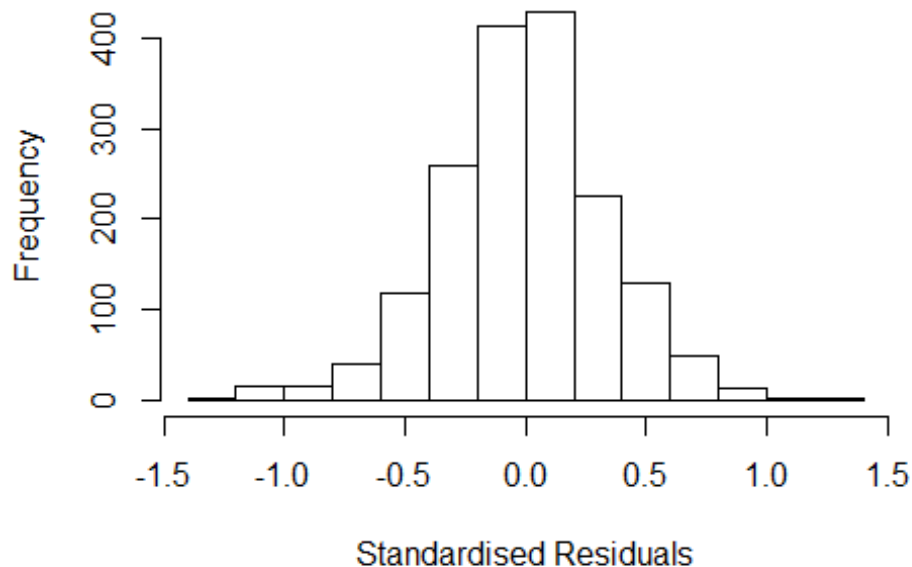


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.1, df = 1, p-value = 0.08
```



```
## [1] "Female first author team size 2018 geometric mean: 3.86893235005681"
## [1] "Male first author team size 2018 geometric mean: 3.89177504132819"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.66514232484556"
## [1] "Male last author team size 2018 geometric mean: 3.96880312081117"
##
## 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 1.089 1      1.043
## LastAuthorFemale  1.069 1      1.034
## UniqueAuthors    1.229 4      1.026
## Year              1.294 16     1.008
```

Residuals from first and last author and team size



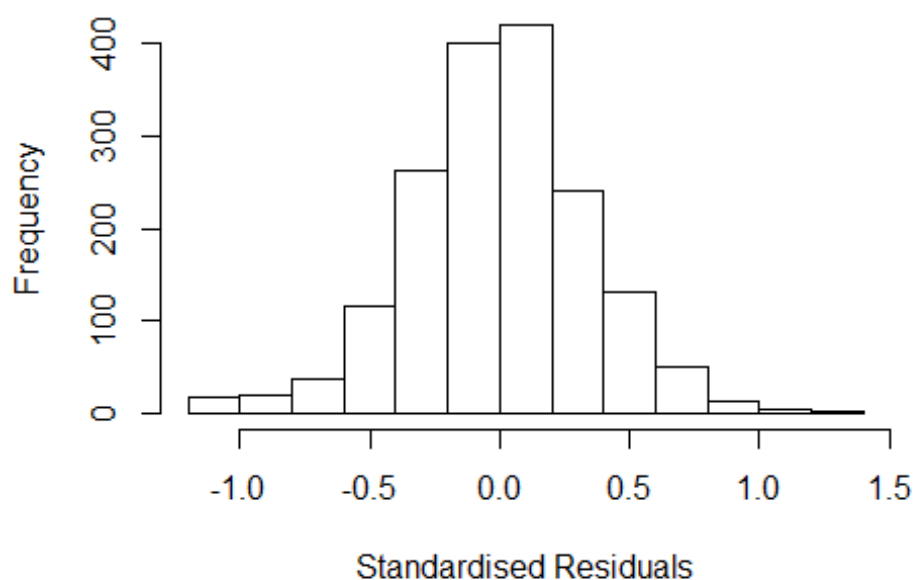
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20215 -0.21119 -0.00523  0.19879  1.32752
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.00253    0.04960   20.21 < 2e-16 ***
## FirstAuthorFemale1 -0.01132    0.01668   -0.68  0.49734
## LastAuthorFemale1  0.02638    0.01930    1.37  0.17198
## UniqueAuthors2     0.08595    0.04302    2.00  0.04590 *
## UniqueAuthors3     0.08183    0.04422    1.85  0.06444 .
## UniqueAuthors4     0.09547    0.04376    2.18  0.02926 *
## UniqueAuthors5     0.16536    0.04334    3.82  0.00014 ***
## Year1997          -0.00334    0.04484   -0.07  0.94072
## Year1998          -0.02520    0.05694   -0.44  0.65814
## Year1999          -0.03319    0.04407   -0.75  0.45151
```

```

## Year2000      -0.04000    0.04986   -0.80  0.42250
## Year2001      -0.06073    0.05467   -1.11  0.26682
## Year2002      -0.12458    0.04703   -2.65  0.00814 **
## Year2003      -0.10021    0.04493   -2.23  0.02586 *
## Year2004      -0.12557    0.04679   -2.68  0.00735 **
## Year2005      -0.06408    0.04551   -1.41  0.15929
## Year2006      -0.05489    0.04263   -1.29  0.19799
## Year2007      -0.04281    0.04419   -0.97  0.33281
## Year2008      -0.01612    0.04183   -0.39  0.70000
## Year2009      -0.08826    0.04100   -2.15  0.03150 *
## Year2010      -0.06297    0.04414   -1.43  0.15391
## Year2011       0.00789    0.04230    0.19  0.85209
## Year2012      -0.06442    0.04513   -1.43  0.15367
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.313
## Multiple R-squared:  0.0316, Adjusted R-squared:  0.019
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 165 weights are ~= 1. The remaining 1551 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0328 0.8570 0.9510 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          5.83e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "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 1.046 1 1.023
## Year 1.098 16 1.003

```

Residuals from first and last author



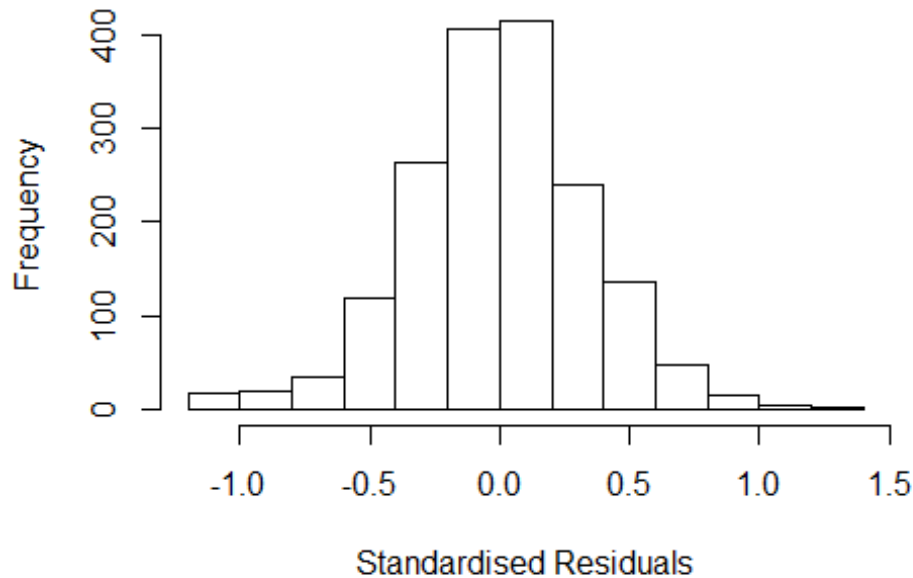
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.14315 -0.21373  0.00193  0.20832  1.32682
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.08918    0.03197   34.07  <2e-16 ***
## FirstAuthorFemale1 -0.00842    0.01670   -0.50   0.614
## LastAuthorFemale1  0.02784    0.01921    1.45   0.147
## Year1997         0.00461    0.04605    0.10   0.920
## Year1998        -0.02022    0.05745   -0.35   0.725
## Year1999        -0.02542    0.04519   -0.56   0.574
## Year2000        -0.03162    0.05005   -0.63   0.528
## Year2001        -0.03536    0.05485   -0.64   0.519
## Year2002        -0.11011    0.04814   -2.29   0.022 *
## Year2003        -0.09208    0.04640   -1.98   0.047 *
## Year2004        -0.11111    0.04651   -2.39   0.017 *
## Year2005        -0.05532    0.04617   -1.20   0.231
```

```

## Year2006      -0.04335    0.04321   -1.00    0.316
## Year2007      -0.02329    0.04414   -0.53    0.598
## Year2008       0.00390    0.04286    0.09    0.928
## Year2009      -0.06096    0.04162   -1.46    0.143
## Year2010      -0.03721    0.04453   -0.84    0.403
## Year2011       0.02613    0.04263    0.61    0.540
## Year2012      -0.04065    0.04541   -0.90    0.371
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.315
## Multiple R-squared:  0.0145, Adjusted R-squared:  0.00402
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 152 weights are ~= 1. The remaining 1564 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0368 0.8590 0.9510 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1      1.029
## Year      1.058 16      1.002

```

Residuals from first author



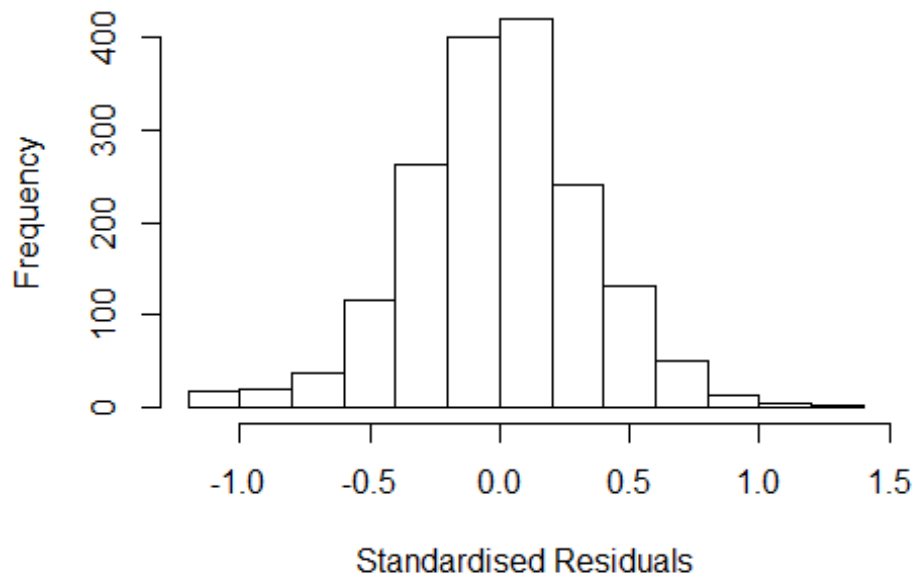
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.120519 -0.214790 -0.000643  0.206119  1.321554
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.09445    0.03179   34.43  <2e-16 ***
## FirstAuthorFemale1 -0.00569    0.01667   -0.34    0.733
## Year1997         0.00501    0.04622    0.11    0.914
## Year1998        -0.02295    0.05762   -0.40    0.690
## Year1999        -0.02433    0.04522   -0.54    0.591
## Year2000        -0.03262    0.05009   -0.65    0.515
## Year2001        -0.03603    0.05530   -0.65    0.515
## Year2002        -0.10968    0.04824   -2.27    0.023 *
## Year2003        -0.09167    0.04657   -1.97    0.049 *
## Year2004        -0.11326    0.04643   -2.44    0.015 *
## Year2005        -0.05382    0.04608   -1.17    0.243
## Year2006        -0.04293    0.04328   -0.99    0.321
```

```

## Year2007          -0.02261    0.04410   -0.51    0.608
## Year2008           0.00583    0.04289    0.14    0.892
## Year2009          -0.05988    0.04162   -1.44    0.150
## Year2010          -0.03489    0.04457   -0.78    0.434
## Year2011           0.02607    0.04278    0.61    0.542
## Year2012          -0.04086    0.04548   -0.90    0.369
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.315
## Multiple R-squared:  0.0131, Adjusted R-squared:  0.00325
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1567 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0394 0.8590 0.9520 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      5.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



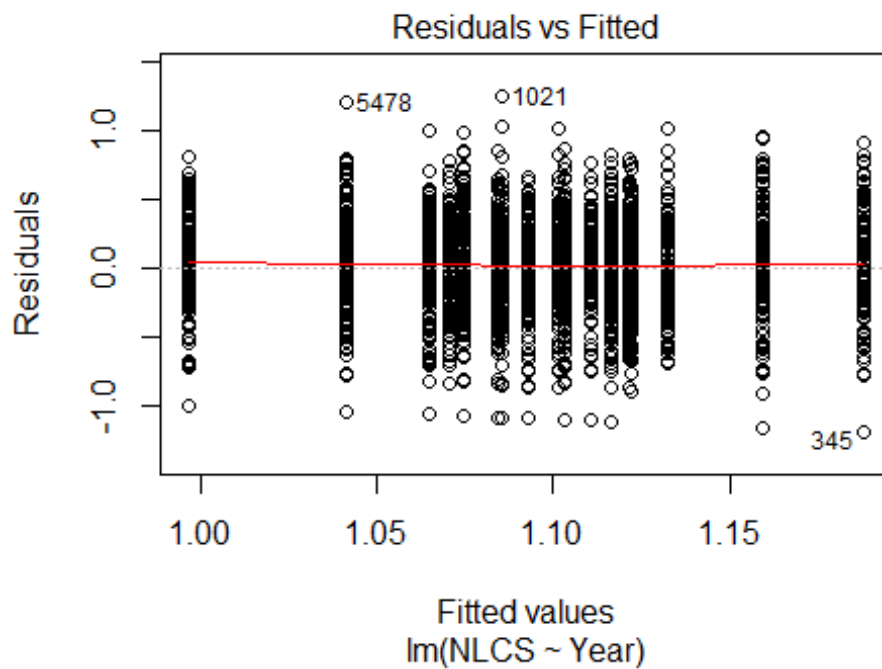
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.13832 -0.21566 0.00256 0.20732 1.32981
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08619 0.03121 34.80 <2e-16 ***
## LastAuthorFemale1 0.02675 0.01914 1.40 0.162
## Year1997 0.00504 0.04607 0.11 0.913
## Year1998 -0.02004 0.05750 -0.35 0.727
## Year1999 -0.02594 0.04521 -0.57 0.566
## Year2000 -0.03115 0.05002 -0.62 0.534
## Year2001 -0.03635 0.05483 -0.66 0.507
## Year2002 -0.11068 0.04820 -2.30 0.022 *
## Year2003 -0.09185 0.04632 -1.98 0.048 *
## Year2004 -0.11228 0.04642 -2.42 0.016 *
## Year2005 -0.05615 0.04613 -1.22 0.224
## Year2006 -0.04389 0.04325 -1.01 0.310
```

```

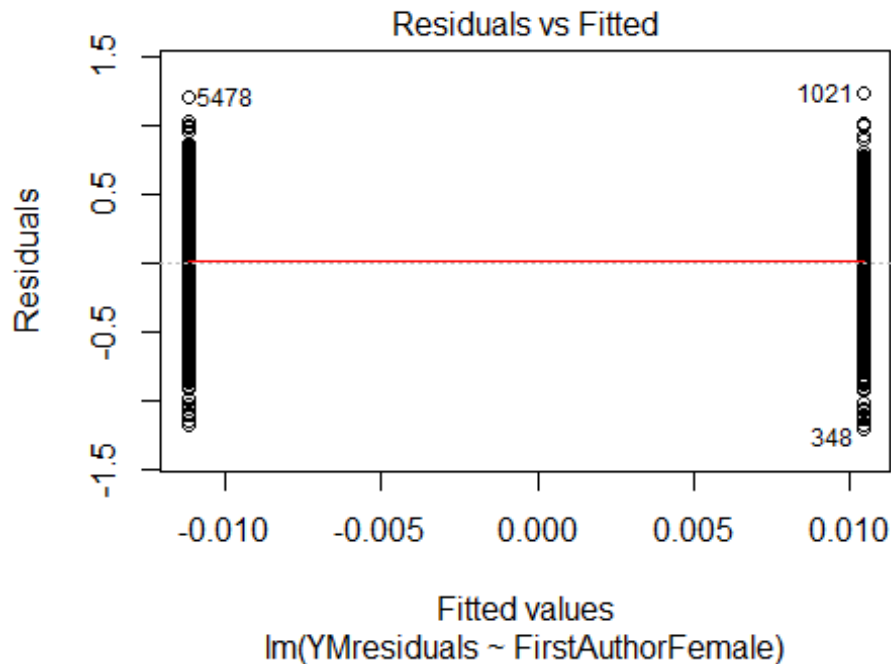
## Year2007          -0.02428      0.04410    -0.55      0.582
## Year2008           0.00310      0.04282      0.07      0.942
## Year2009          -0.06268      0.04150    -1.51      0.131
## Year2010          -0.03864      0.04454    -0.87      0.386
## Year2011           0.02538      0.04264      0.60      0.552
## Year2012          -0.04244      0.04529    -0.94      0.349
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.315
## Multiple R-squared:  0.0143, Adjusted R-squared:  0.00445
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1559 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0356 0.8590 0.9510 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1716"
## [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
## 259 290 279 213 262 294 220 228 235 241 248 300 264 272 286
## 2011 2012
## 291 316
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 134 130 145 100 132 119 135 150 139 155 154 191 183 203 209
## 2011 2012

```

```
## 219 218
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 112 113 125 86 115 107 119 134 122 133 143 172 162 183 179
## 2011 2012
## 192 191
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 66, df = 16, p-value = 5e-08
```

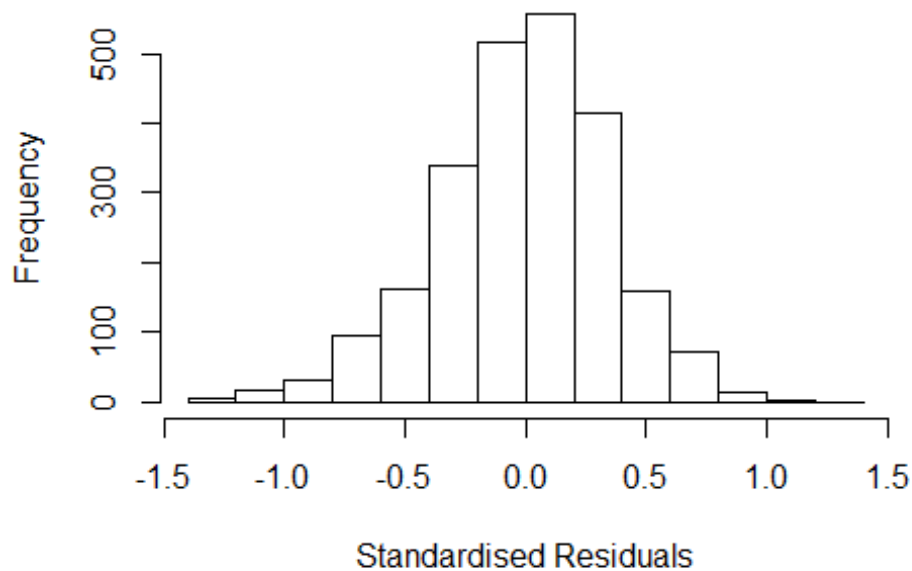


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6, df = 1, p-value = 0.01
```



```
## [1] "Female first author team size 2018 geometric mean: 4.47587123179287"
## [1] "Male first author team size 2018 geometric mean: 4.51049265334221"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3300, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.25398809278863"
## [1] "Male last author team size 2018 geometric mean: 4.60404493333804"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2800, 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 1.063 1 1.031
## LastAuthorFemale 1.048 1 1.024
## UniqueAuthors 1.204 4 1.023
## Year 1.223 16 1.006
```

Residuals from first and last author and team size



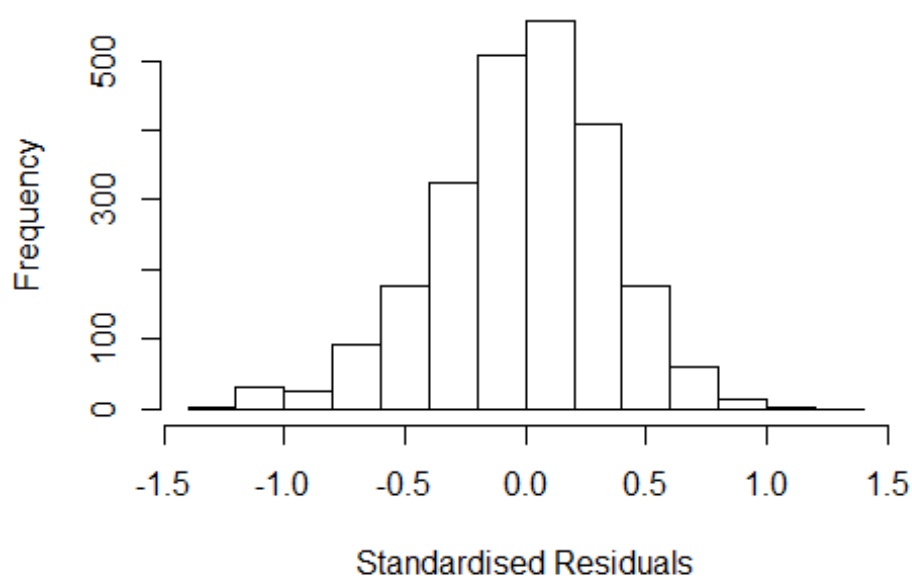
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29871 -0.23188 0.00787 0.22464 1.20281
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.93017 0.06590 14.12 < 2e-16 ***
## FirstAuthorFemale1 -0.01361 0.01488 -0.91 0.36054
## LastAuthorFemale1 -0.02237 0.01692 -1.32 0.18611
## UniqueAuthors2 0.21822 0.05802 3.76 0.00017 ***
## UniqueAuthors3 0.22340 0.05770 3.87 0.00011 ***
## UniqueAuthors4 0.22768 0.05773 3.94 8.3e-05 ***
## UniqueAuthors5 0.28623 0.05708 5.01 5.7e-07 ***
## Year1997 0.08231 0.05655 1.46 0.14562
## Year1998 -0.06383 0.05138 -1.24 0.21423
## Year1999 0.00939 0.05520 0.17 0.86501
```

```

## Year2000      -0.01141    0.05136   -0.22  0.82427
## Year2001      -0.02739    0.05250   -0.52  0.60194
## Year2002      -0.02448    0.05013   -0.49  0.62531
## Year2003      -0.07819    0.04881   -1.60  0.10928
## Year2004      -0.05626    0.05094   -1.10  0.26952
## Year2005      -0.03443    0.05094   -0.68  0.49918
## Year2006      -0.06028    0.05008   -1.20  0.22879
## Year2007      -0.06315    0.04704   -1.34  0.17962
## Year2008      -0.03346    0.04827   -0.69  0.48831
## Year2009      -0.03119    0.04729   -0.66  0.50967
## Year2010      -0.05347    0.05134   -1.04  0.29770
## Year2011      -0.06559    0.04969   -1.32  0.18695
## Year2012      -0.12328    0.04999   -2.47  0.01374 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.334
## Multiple R-squared:  0.0433, Adjusted R-squared:  0.0344
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 180 weights are ~= 1. The remaining 2208 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0961 0.8710 0.9510 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          4.19e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.046 1          1.023
## LastAuthorFemale 1.037 1          1.018
## Year 1.059 16          1.002

```

Residuals from first and last author



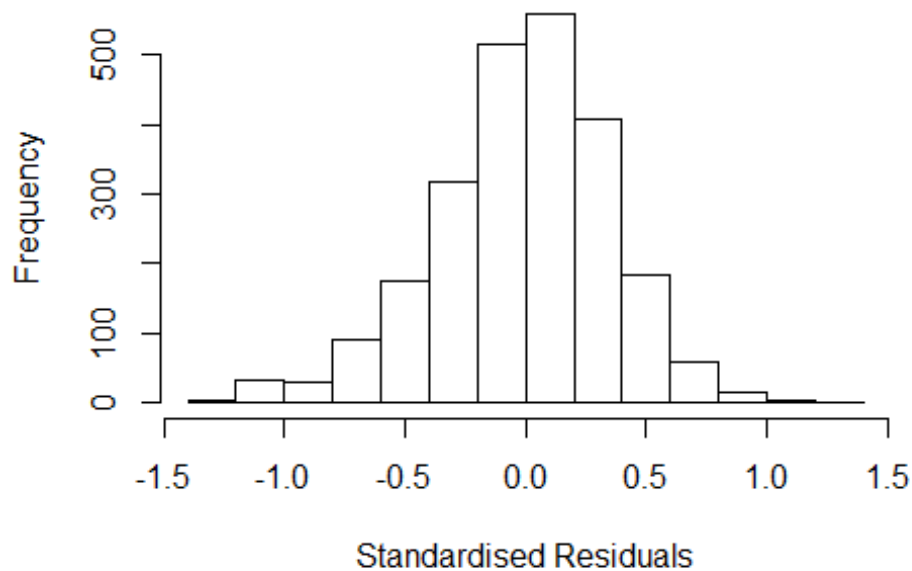
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2326 -0.2274 0.0112 0.2272 1.2602
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14242 0.04162 27.45 <2e-16 ***
## FirstAuthorFemale1 -0.00613 0.01488 -0.41 0.681
## LastAuthorFemale1 -0.02782 0.01699 -1.64 0.102
## Year1997 0.09015 0.05707 1.58 0.114
## Year1998 -0.04184 0.05227 -0.80 0.424
## Year1999 0.02062 0.05611 0.37 0.713
## Year2000 0.00762 0.05144 0.15 0.882
## Year2001 -0.01131 0.05332 -0.21 0.832
## Year2002 -0.00740 0.05104 -0.14 0.885
## Year2003 -0.04808 0.04957 -0.97 0.332
## Year2004 -0.03144 0.05107 -0.62 0.538
## Year2005 -0.01018 0.05109 -0.20 0.842
```

```

## Year2006      -0.03453    0.05032   -0.69    0.493
## Year2007      -0.04050    0.04751   -0.85    0.394
## Year2008      -0.00625    0.04872   -0.13    0.898
## Year2009      -0.00434    0.04776   -0.09    0.928
## Year2010      -0.01779    0.05148   -0.35    0.730
## Year2011      -0.04006    0.05035   -0.80    0.426
## Year2012      -0.09709    0.05095   -1.91    0.057 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.336
## Multiple R-squared:  0.013, Adjusted R-squared:  0.00548
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 180 weights are ~= 1. The remaining 2208 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.129  0.869  0.951  0.897  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.19e-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 1.033 1      1.016
## Year      1.033 16      1.001

```


Residuals from first author



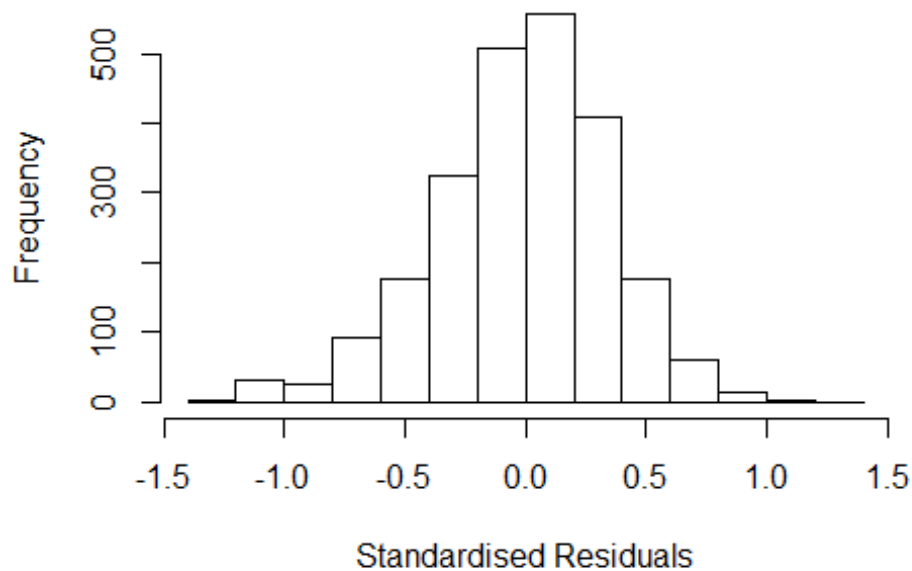
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.227 -0.229 0.011 0.226 1.235
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13645 0.04137 27.47 <2e-16 ***
## FirstAuthorFemale1 -0.00954 0.01480 -0.64 0.519
## Year1997 0.09050 0.05710 1.59 0.113
## Year1998 -0.03871 0.05209 -0.74 0.457
## Year1999 0.02116 0.05594 0.38 0.705
## Year2000 0.00666 0.05131 0.13 0.897
## Year2001 -0.01085 0.05320 -0.20 0.838
## Year2002 -0.00767 0.05096 -0.15 0.880
## Year2003 -0.04699 0.04953 -0.95 0.343
## Year2004 -0.02986 0.05106 -0.58 0.559
## Year2005 -0.01225 0.05108 -0.24 0.811
## Year2006 -0.03681 0.05025 -0.73 0.464
```

```

## Year2007          -0.04016    0.04740   -0.85    0.397
## Year2008          -0.00736    0.04870   -0.15    0.880
## Year2009          -0.00463    0.04768   -0.10    0.923
## Year2010          -0.01804    0.05137   -0.35    0.726
## Year2011          -0.04014    0.05020   -0.80    0.424
## Year2012          -0.09924    0.05101   -1.95    0.052 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.336
## Multiple R-squared:  0.0116, Adjusted R-squared:  0.00455
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 2203 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.148  0.868  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      4.19e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.025 1          1.013
## Year            1.025 16          1.001

```

Residuals from last author



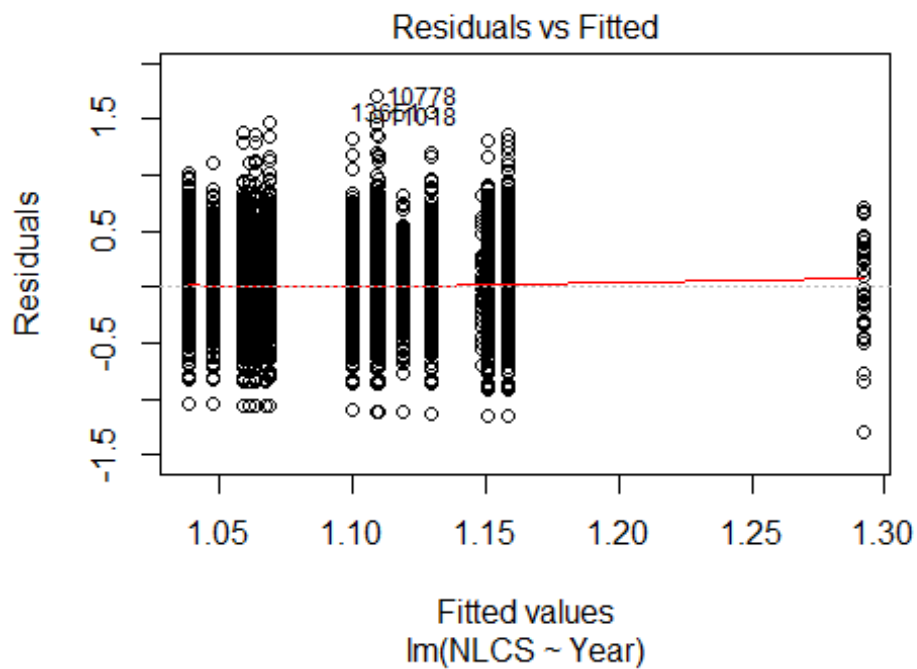
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2303 -0.2262 0.0105 0.2260 1.2635
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14069 0.04150 27.48 <2e-16 ***
## LastAuthorFemale1 -0.02873 0.01691 -1.70 0.089 .
## Year1997 0.08958 0.05702 1.57 0.116
## Year1998 -0.04249 0.05230 -0.81 0.417
## Year1999 0.01996 0.05609 0.36 0.722
## Year2000 0.00739 0.05149 0.14 0.886
## Year2001 -0.01177 0.05332 -0.22 0.825
## Year2002 -0.00753 0.05106 -0.15 0.883
## Year2003 -0.04919 0.04957 -0.99 0.321
## Year2004 -0.03192 0.05111 -0.62 0.532
## Year2005 -0.01092 0.05104 -0.21 0.831
## Year2006 -0.03586 0.05015 -0.71 0.475
```

```

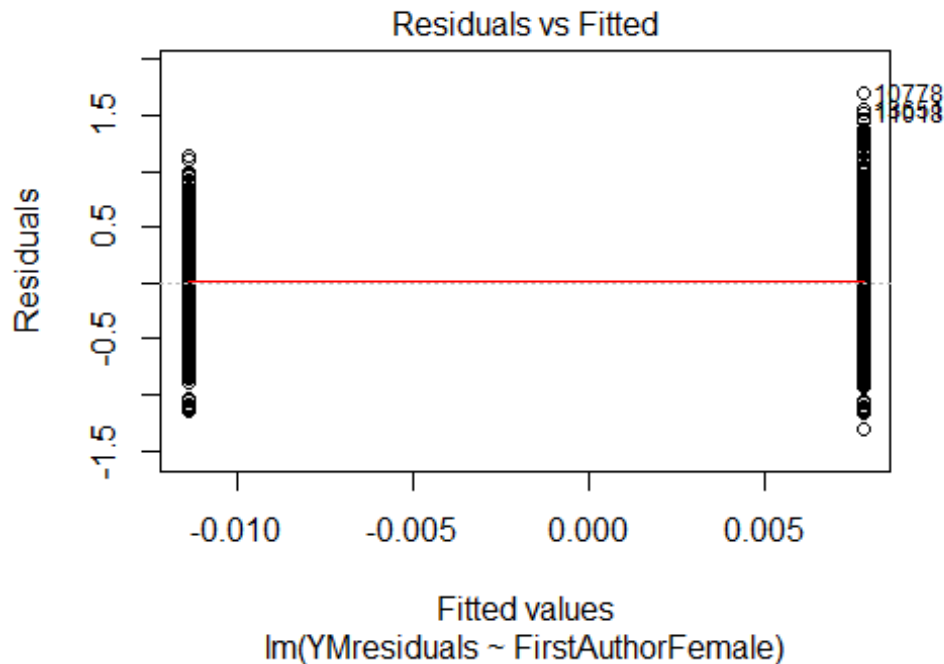
## Year2007          -0.04174      0.04739   -0.88      0.378
## Year2008          -0.00740      0.04868   -0.15      0.879
## Year2009          -0.00521      0.04770   -0.11      0.913
## Year2010          -0.01873      0.05147   -0.36      0.716
## Year2011          -0.04154      0.05025   -0.83      0.408
## Year2012          -0.09836      0.05090   -1.93      0.053 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.336
## Multiple R-squared:  0.0129, Adjusted R-squared:  0.00579
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 179 weights are ~= 1. The remaining 2209 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.126  0.869  0.951  0.897  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.19e-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: 2388"
## [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
## 898 759 91 576 63 694 704 596 632 692 747 751 712 795 754
## 2011 2012
## 708 823
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 389 363 42 300 30 225 429 364 371 405 420 449 419 458 458
## 2011 2012

```

```
## 449 508
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 331 310 39 263 27 190 339 298 320 336 355 380 349 397 399
## 2011 2012
## 394 434
## [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 = 5e-16
```

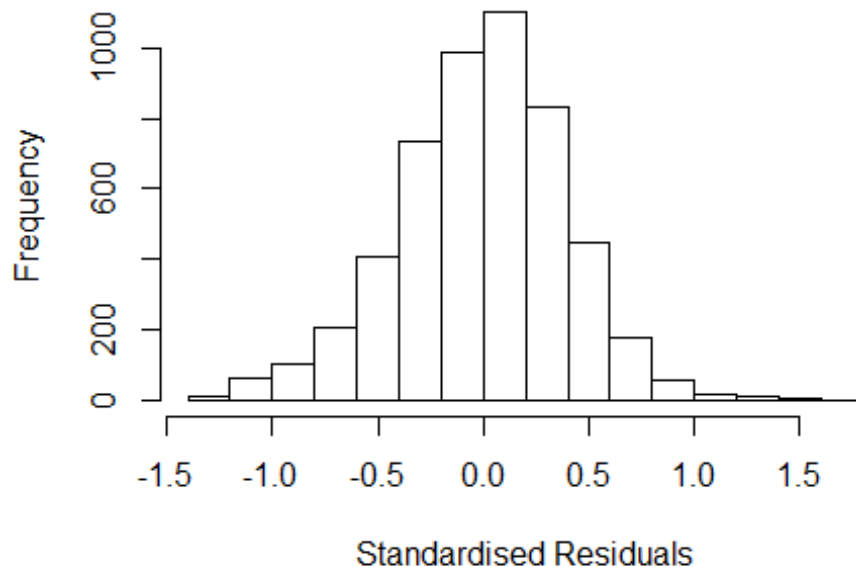


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 33, df = 1, p-value = 1e-08
```



```
## [1] "Female first author team size 2018 geometric mean: 4.47427870771528"
## [1] "Male first author team size 2018 geometric mean: 3.92479108239544"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 23000, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.05702737675677"
## [1] "Male last author team size 2018 geometric mean: 4.19713749622097"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "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  1.026 1      1.013
## UniqueAuthors    1.082 4      1.010
## Year              1.105 16      1.003
```

Residuals from first and last author and team size



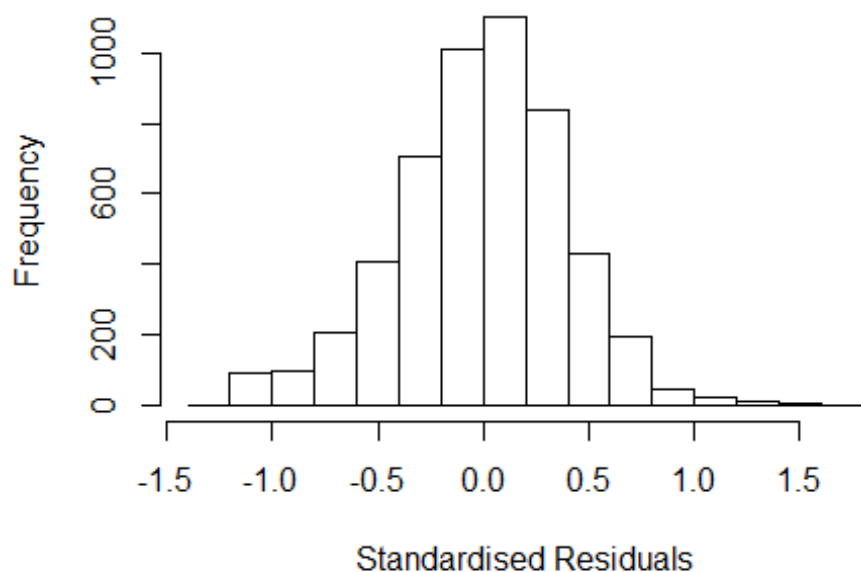
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2719 -0.2540 0.0105 0.2544 1.6287
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0914 0.0363 30.10 < 2e-16 ***
## FirstAuthorFemale1 -0.0138 0.0112 -1.24 0.21659
## LastAuthorFemale1 -0.0158 0.0129 -1.22 0.22141
## UniqueAuthors2 0.0885 0.0293 3.02 0.00252 **
## UniqueAuthors3 0.1188 0.0294 4.03 5.5e-05 ***
## UniqueAuthors4 0.1101 0.0302 3.64 0.00027 ***
## UniqueAuthors5 0.1804 0.0286 6.31 3.1e-10 ***
## Year1997 -0.0304 0.0382 -0.79 0.42685
## Year1998 0.1697 0.0749 2.26 0.02357 *
## Year1999 -0.1187 0.0361 -3.29 0.00102 **
```

```

## Year2000          -0.0352      0.0831   -0.42  0.67220
## Year2001          -0.0756      0.0362   -2.09  0.03661 *
## Year2002          -0.1473      0.0353   -4.17  3.1e-05 ***
## Year2003          -0.1292      0.0345   -3.75  0.00018 ***
## Year2004          -0.1406      0.0331   -4.25  2.2e-05 ***
## Year2005          -0.1341      0.0330   -4.06  5.0e-05 ***
## Year2006          -0.1213      0.0340   -3.57  0.00036 ***
## Year2007          -0.1317      0.0323   -4.08  4.7e-05 ***
## Year2008          -0.1127      0.0339   -3.33  0.00088 ***
## Year2009          -0.0866      0.0339   -2.55  0.01077 *
## Year2010          -0.0832      0.0325   -2.56  0.01049 *
## Year2011          -0.0596      0.0339   -1.76  0.07916 .
## Year2012          -0.1256      0.0343   -3.66  0.00026 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0276, Adjusted R-squared:  0.0235
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 442 weights are ~= 1. The remaining 4719 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0208 0.8670 0.9500 0.8980 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.94e-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 1.034 1 1.017
## LastAuthorFemale 1.026 1 1.013
## Year 1.038 16 1.001

```


Residuals from first and last author



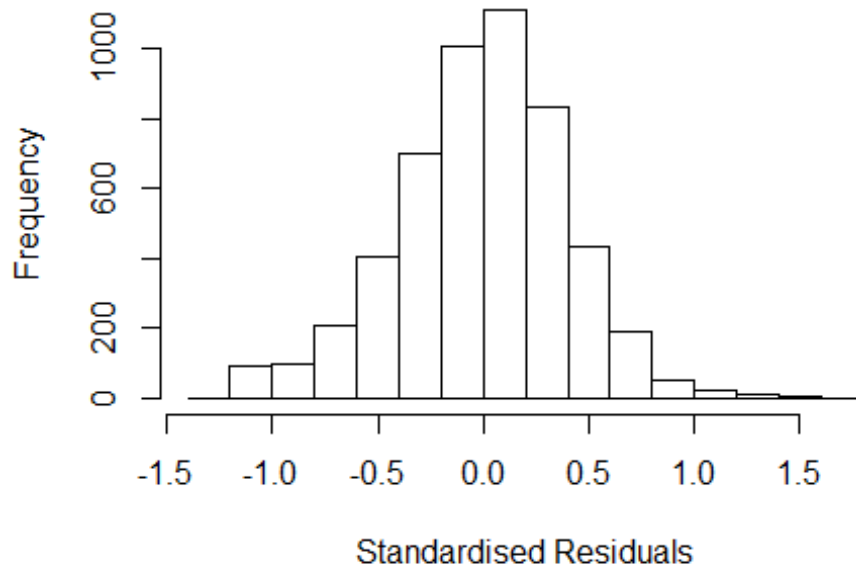
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3321 -0.2546 0.0117 0.2552 1.6887
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19187 0.02743 43.45 < 2e-16 ***
## FirstAuthorFemale1 -0.00735 0.01116 -0.66 0.51029
## LastAuthorFemale1 -0.01582 0.01294 -1.22 0.22149
## Year1997 -0.02158 0.03803 -0.57 0.57043
## Year1998 0.14019 0.07840 1.79 0.07383 .
## Year1999 -0.10747 0.03612 -2.98 0.00294 **
## Year2000 -0.04275 0.08635 -0.50 0.62060
## Year2001 -0.05414 0.03585 -1.51 0.13103
## Year2002 -0.13323 0.03553 -3.75 0.00018 ***
## Year2003 -0.11095 0.03456 -3.21 0.00134 **
## Year2004 -0.12001 0.03306 -3.63 0.00029 ***
## Year2005 -0.11168 0.03285 -3.40 0.00068 ***
```

```

## Year2006      -0.10559    0.03392   -3.11  0.00186 **
## Year2007      -0.10583    0.03233   -3.27  0.00107 **
## Year2008      -0.08758    0.03392   -2.58  0.00986 **
## Year2009      -0.06657    0.03374   -1.97  0.04856 *
## Year2010      -0.05709    0.03261   -1.75  0.08013 .
## Year2011      -0.03617    0.03398   -1.06  0.28713
## Year2012      -0.09821    0.03435   -2.86  0.00426 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0121, Adjusted R-squared:  0.00866
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 466 weights are ~= 1. The remaining 4695 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0067 0.8670 0.9490 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.94e-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 1.022 1      1.011
## Year      1.022 16      1.001

```

Residuals from first author



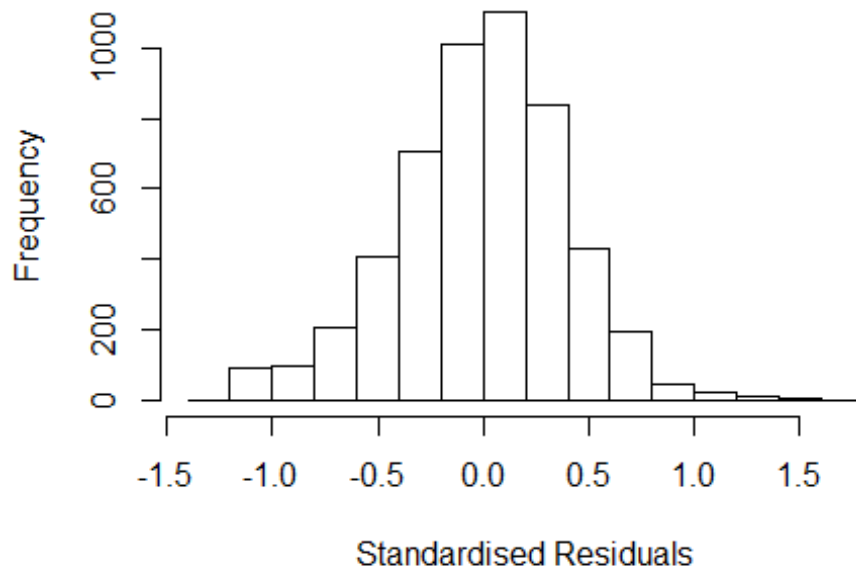
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3287 -0.2537 0.0111 0.2556 1.6917
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18836 0.02714 43.79 < 2e-16 ***
## FirstAuthorFemale1 -0.00927 0.01111 -0.83 0.40442
## Year1997 -0.02017 0.03798 -0.53 0.59551
## Year1998 0.14034 0.07842 1.79 0.07360 .
## Year1999 -0.10666 0.03603 -2.96 0.00308 **
## Year2000 -0.04260 0.08614 -0.49 0.62092
## Year2001 -0.05412 0.03586 -1.51 0.13127
## Year2002 -0.13198 0.03547 -3.72 0.00020 ***
## Year2003 -0.11040 0.03456 -3.19 0.00141 **
## Year2004 -0.11938 0.03306 -3.61 0.00031 ***
## Year2005 -0.11142 0.03282 -3.39 0.00069 ***
## Year2006 -0.10487 0.03390 -3.09 0.00199 **
```

```

## Year2007          -0.10520    0.03232   -3.25  0.00114 **
## Year2008          -0.08652    0.03389   -2.55  0.01072 *
## Year2009          -0.06605    0.03374   -1.96  0.05030 .
## Year2010          -0.05674    0.03262   -1.74  0.08195 .
## Year2011          -0.03603    0.03398   -1.06  0.28903
## Year2012          -0.09764    0.03434   -2.84  0.00448 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0118, Adjusted R-squared:  0.00851
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 467 weights are ~= 1. The remaining 4694 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0063 0.8660 0.9490 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.94e-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 1.015 1          1.007
## Year              1.015 16          1.000

```

Residuals from last author



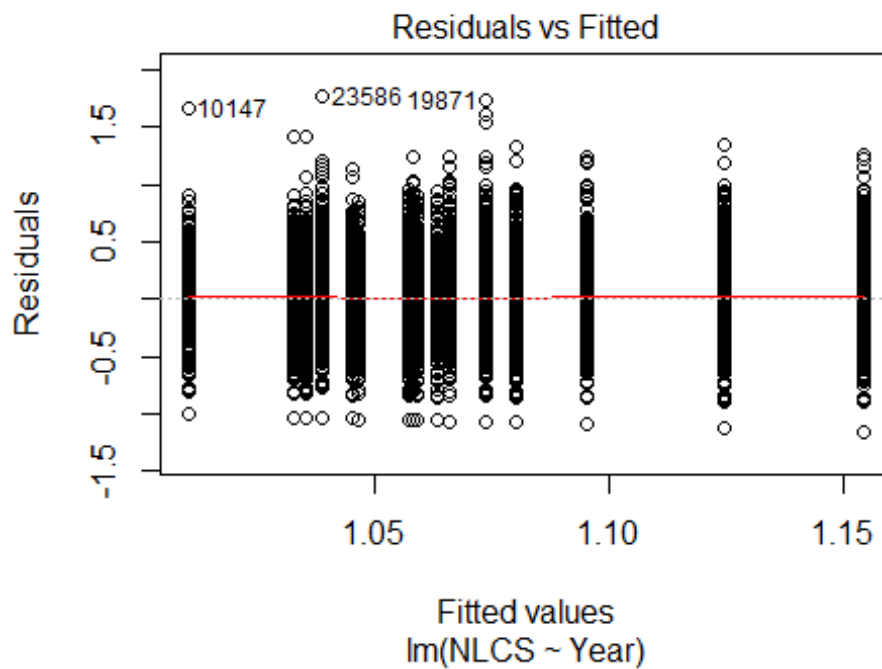
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3289 -0.2541 0.0102 0.2565 1.6915
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1898 0.0272 43.80 < 2e-16 ***
## LastAuthorFemale1 -0.0170 0.0129 -1.32 0.18630
## Year1997 -0.0218 0.0381 -0.57 0.56647
## Year1998 0.1391 0.0783 1.78 0.07550 .
## Year1999 -0.1079 0.0361 -2.99 0.00282 **
## Year2000 -0.0423 0.0866 -0.49 0.62515
## Year2001 -0.0549 0.0358 -1.53 0.12553
## Year2002 -0.1342 0.0355 -3.78 0.00016 ***
## Year2003 -0.1113 0.0345 -3.22 0.00128 **
## Year2004 -0.1207 0.0331 -3.65 0.00026 ***
## Year2005 -0.1125 0.0328 -3.43 0.00062 ***
## Year2006 -0.1066 0.0339 -3.15 0.00167 **
```

```

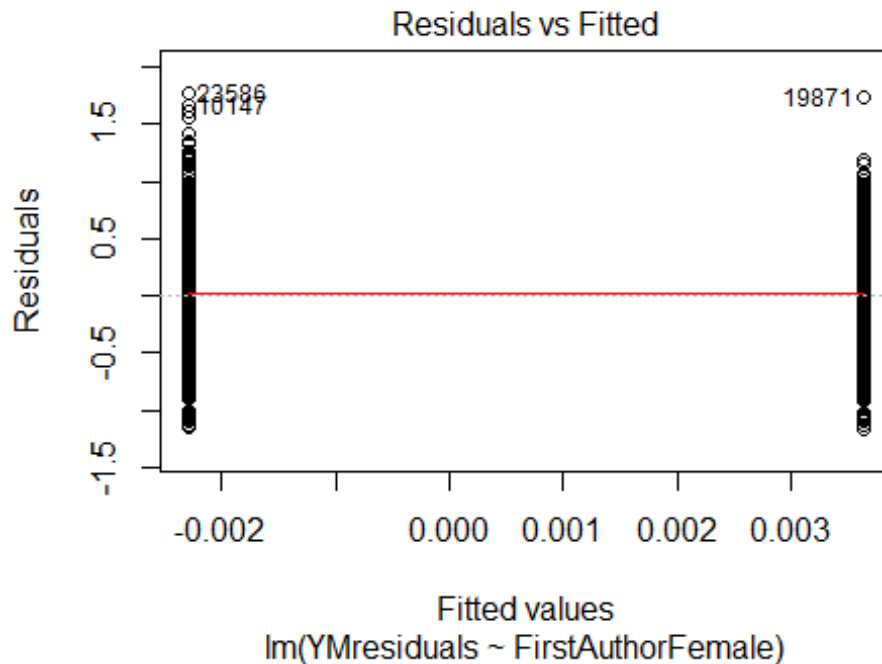
## Year2007          -0.1070      0.0323   -3.31  0.00093 ***
## Year2008          -0.0883      0.0339   -2.60  0.00927 **
## Year2009          -0.0672      0.0337   -1.99  0.04639 *
## Year2010          -0.0579      0.0326   -1.78  0.07547 .
## Year2011          -0.0372      0.0339   -1.10  0.27263
## Year2012          -0.0997      0.0342   -2.91  0.00362 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.012, Adjusted R-squared:  0.00877
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 468 weights are ~= 1. The remaining 4693 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0063 0.8670 0.9490 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.94e-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: 5161"
## [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
## 1215 1170 1059 938 980 920 1011 889 950 970 982 1069 1073 1211 1245
## 2011 2012
## 1156 1146
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 666 664 528 559 488 391 659 607 621 640 630 715 713 838 863
## 2011 2012

```

```
## 829 802
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 565 559 437 476 407 320 536 500 523 553 524 605 627 723 723
## 2011 2012
## 698 695
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 200, df = 16, p-value <2e-16
```

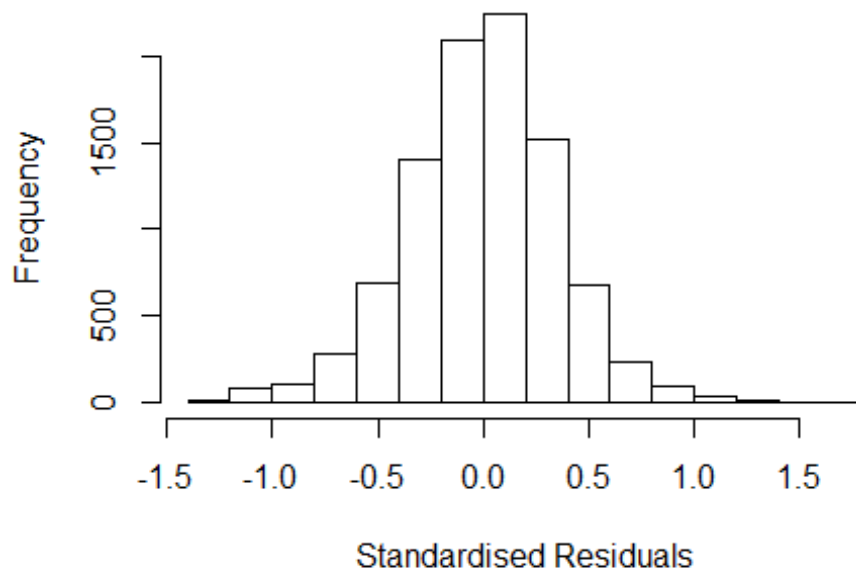


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 36, df = 1, p-value = 2e-09
```



```
## [1] "Female first author team size 2018 geometric mean: 4.15985634502202"
## [1] "Male first author team size 2018 geometric mean: 4.04167527750527"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.80308209220715"
## [1] "Male last author team size 2018 geometric mean: 4.1769369806469"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24000, 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 1.023 1 1.012
## LastAuthorFemale 1.012 1 1.006
## UniqueAuthors 1.101 4 1.012
## Year 1.104 16 1.003
```


Residuals from first and last author and team size



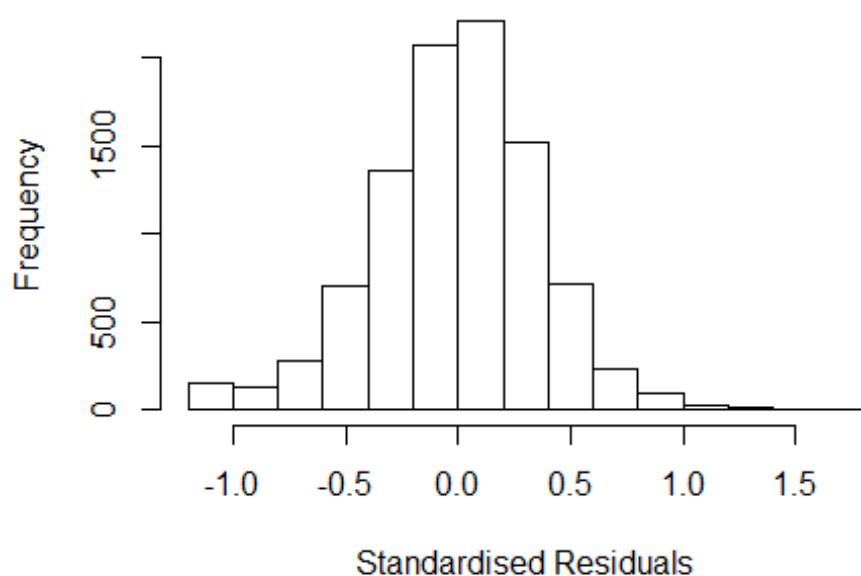
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27397 -0.22228 0.00602 0.21914 1.70346
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9.41e-01 3.13e-02 30.06 < 2e-16 ***
## FirstAuthorFemale1 8.07e-04 7.27e-03 0.11 0.912
## LastAuthorFemale1 -4.11e-05 8.89e-03 0.00 0.996
## UniqueAuthors2 1.96e-01 2.95e-02 6.65 3.0e-11 ***
## UniqueAuthors3 2.30e-01 2.94e-02 7.83 5.3e-15 ***
## UniqueAuthors4 2.47e-01 2.96e-02 8.35 < 2e-16 ***
## UniqueAuthors5 3.32e-01 2.90e-02 11.44 < 2e-16 ***
## Year1997 -9.85e-03 2.50e-02 -0.39 0.694
## Year1998 -6.32e-02 2.56e-02 -2.47 0.014 *
## Year1999 -1.35e-01 2.28e-02 -5.93 3.0e-09 ***
```

```

## Year2000      -1.30e-01  2.38e-02  -5.46  4.8e-08 ***
## Year2001      -1.08e-01  2.49e-02  -4.33  1.5e-05 ***
## Year2002      -1.36e-01  2.21e-02  -6.14  8.6e-10 ***
## Year2003      -1.77e-01  2.22e-02  -7.96  1.9e-15 ***
## Year2004      -1.55e-01  2.17e-02  -7.17  7.9e-13 ***
## Year2005      -1.60e-01  2.11e-02  -7.59  3.6e-14 ***
## Year2006      -1.36e-01  2.15e-02  -6.34  2.4e-10 ***
## Year2007      -1.35e-01  2.07e-02  -6.50  8.5e-11 ***
## Year2008      -1.41e-01  2.11e-02  -6.69  2.3e-11 ***
## Year2009      -1.17e-01  2.10e-02  -5.58  2.5e-08 ***
## Year2010      -1.30e-01  2.08e-02  -6.25  4.2e-10 ***
## Year2011      -1.35e-01  2.18e-02  -6.18  6.7e-10 ***
## Year2012      -1.54e-01  2.22e-02  -6.94  4.2e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.329
## Multiple R-squared:  0.0573, Adjusted R-squared:  0.0551
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 4 observations c(3715,7731,7896,9337)
## are outliers with |weight| = 0 ( < 1.1e-05);
## 796 weights are ~ = 1. The remaining 8671 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8670 0.9510 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "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 1.009 1 1.004
## Year 1.018 16 1.001

```

Residuals from first and last author



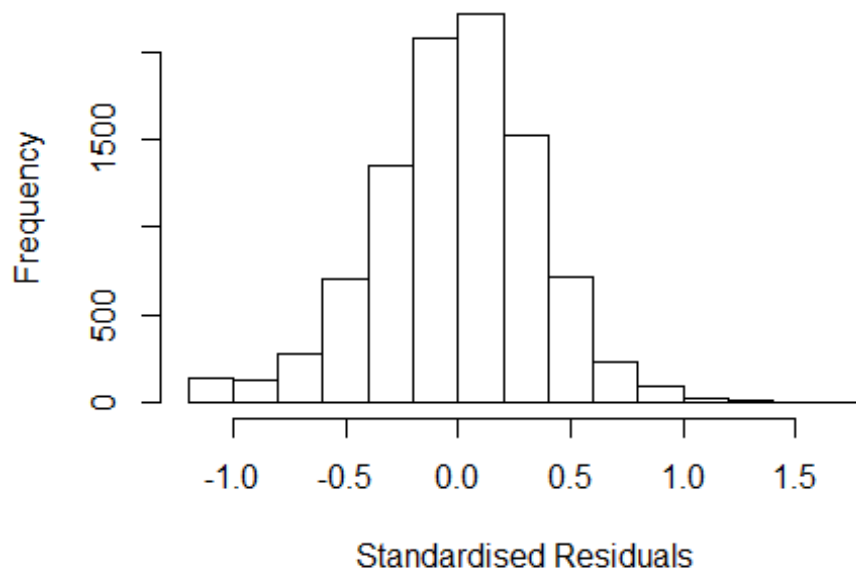
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.16980 -0.22495 0.00667 0.22311 1.75962
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16144 0.01709 67.95 < 2e-16 ***
## FirstAuthorFemale1 0.00836 0.00738 1.13 0.25701
## LastAuthorFemale1 0.00439 0.00902 0.49 0.62686
## Year1997 -0.00837 0.02556 -0.33 0.74316
## Year1998 -0.03901 0.02627 -1.48 0.13760
## Year1999 -0.11440 0.02313 -4.95 7.7e-07 ***
## Year2000 -0.10353 0.02393 -4.33 1.5e-05 ***
## Year2001 -0.07807 0.02522 -3.10 0.00197 **
## Year2002 -0.10406 0.02260 -4.61 4.2e-06 ***
## Year2003 -0.14039 0.02268 -6.19 6.3e-10 ***
## Year2004 -0.12611 0.02188 -5.76 8.5e-09 ***
## Year2005 -0.12389 0.02136 -5.80 6.8e-09 ***
```

```

## Year2006          -0.10421    0.02182   -4.78  1.8e-06 ***
## Year2007          -0.09954    0.02103   -4.73  2.2e-06 ***
## Year2008          -0.09852    0.02127   -4.63  3.7e-06 ***
## Year2009          -0.07987    0.02118   -3.77  0.00016 ***
## Year2010          -0.08882    0.02109   -4.21  2.6e-05 ***
## Year2011          -0.09354    0.02204   -4.24  2.2e-05 ***
## Year2012          -0.11706    0.02263   -5.17  2.3e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.334
## Multiple R-squared:  0.0115, Adjusted R-squared:  0.00965
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 4 observations c(3715,7731,7896,9337)
## are outliers with |weight| = 0 ( < 1.1e-05);
## 803 weights are ~ = 1. The remaining 8664 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8650 0.9510 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.06e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.012 1          1.006
## Year              1.012 16          1.000

```

Residuals from first author



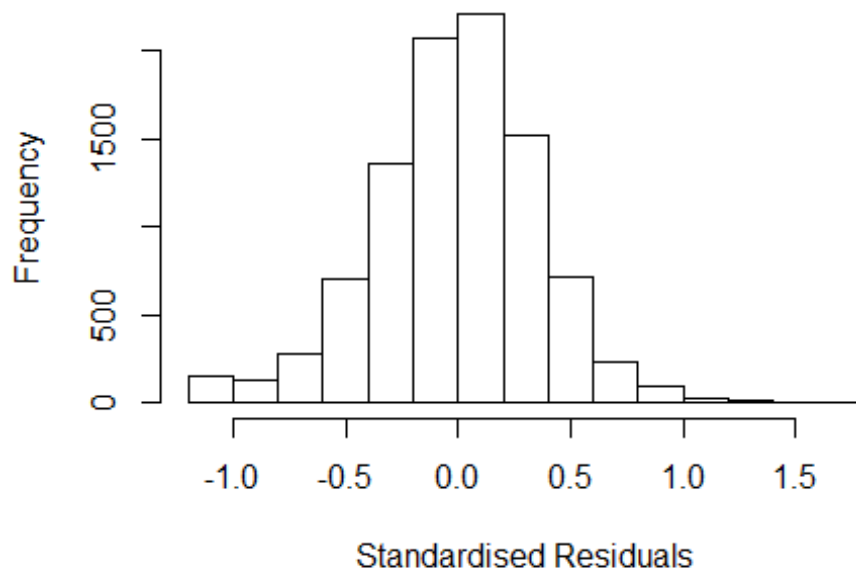
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17078 -0.22533 0.00694 0.22327 1.75872
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16214 0.01703 68.24 < 2e-16 ***
## FirstAuthorFemale1 0.00863 0.00737 1.17 0.24157
## Year1997 -0.00830 0.02555 -0.32 0.74539
## Year1998 -0.03912 0.02627 -1.49 0.13653
## Year1999 -0.11441 0.02313 -4.95 7.7e-07 ***
## Year2000 -0.10347 0.02393 -4.32 1.5e-05 ***
## Year2001 -0.07776 0.02521 -3.08 0.00204 **
## Year2002 -0.10403 0.02259 -4.60 4.2e-06 ***
## Year2003 -0.14031 0.02268 -6.19 6.4e-10 ***
## Year2004 -0.12616 0.02188 -5.76 8.4e-09 ***
## Year2005 -0.12376 0.02136 -5.80 7.0e-09 ***
## Year2006 -0.10408 0.02182 -4.77 1.9e-06 ***
```

```

## Year2007          -0.09954      0.02103      -4.73      2.2e-06 ***
## Year2008          -0.09847      0.02127      -4.63      3.7e-06 ***
## Year2009          -0.07982      0.02118      -3.77      0.00016 ***
## Year2010          -0.08876      0.02109      -4.21      2.6e-05 ***
## Year2011          -0.09343      0.02204      -4.24      2.3e-05 ***
## Year2012          -0.11686      0.02262      -5.17      2.4e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.334
## Multiple R-squared:  0.0115, Adjusted R-squared:  0.00974
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 4 observations c(3715,7731,7896,9337)
## are outliers with |weight| = 0 ( < 1.1e-05);
## 794 weights are ~ = 1. The remaining 8673 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8650 0.9510 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.06e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.007 1          1.003
## Year            1.007 16          1.000

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.16944 -0.22499  0.00655  0.22364  1.75600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.16427    0.01685   69.11 < 2e-16 ***
## LastAuthorFemale1 0.00517    0.00902    0.57  0.56668
## Year1997       -0.00877    0.02556   -0.34  0.73148
## Year1998       -0.03910    0.02626   -1.49  0.13654
## Year1999       -0.11429    0.02313   -4.94  7.9e-07 ***
## Year2000       -0.10351    0.02391   -4.33  1.5e-05 ***
## Year2001       -0.07778    0.02521   -3.09  0.00204 **
## Year2002       -0.10362    0.02258   -4.59  4.5e-06 ***
## Year2003       -0.14020    0.02269   -6.18  6.7e-10 ***
## Year2004       -0.12581    0.02187   -5.75  9.1e-09 ***
## Year2005       -0.12347    0.02136   -5.78  7.7e-09 ***
## Year2006       -0.10377    0.02180   -4.76  2.0e-06 ***
```

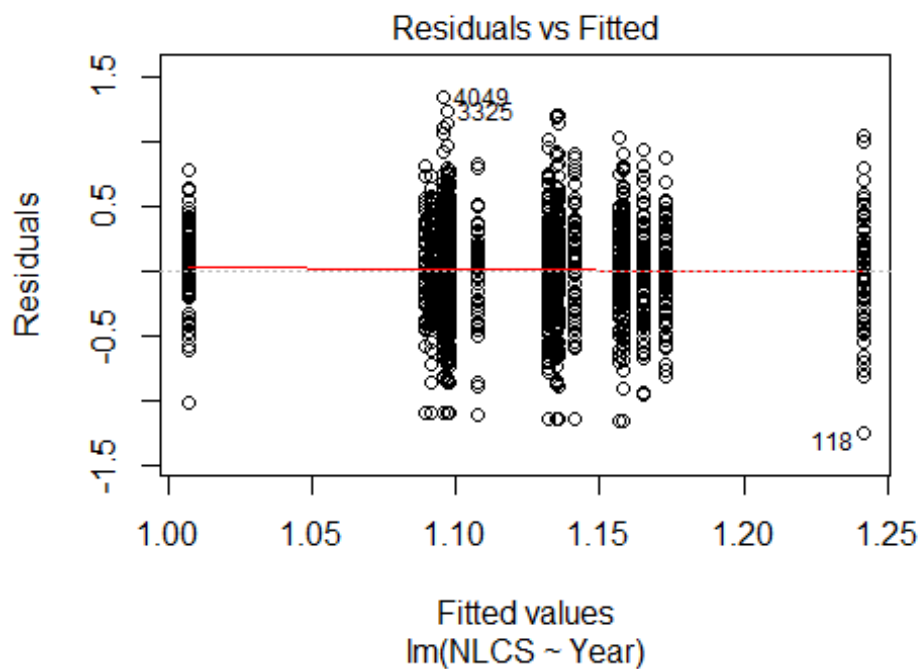
```

## Year2007          -0.09885      0.02101    -4.70  2.6e-06 ***
## Year2008          -0.09801      0.02128    -4.61  4.2e-06 ***
## Year2009          -0.07921      0.02116    -3.74  0.00018 ***
## Year2010          -0.08804      0.02108    -4.18  3.0e-05 ***
## Year2011          -0.09299      0.02205    -4.22  2.5e-05 ***
## Year2012          -0.11627      0.02261    -5.14  2.8e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.334
## Multiple R-squared:  0.0114, Adjusted R-squared:  0.00962
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 4 observations c(3715,7731,7896,9337)
## are outliers with |weight| = 0 ( < 1.1e-05);
## 804 weights are ~ = 1. The remaining 8663 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0008 0.8640 0.9510 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.06e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 9471"
## [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
## 107 135 104 115 123 128 139 153 174 184 153 200 217 244 329
## 2011 2012
## 285 252
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

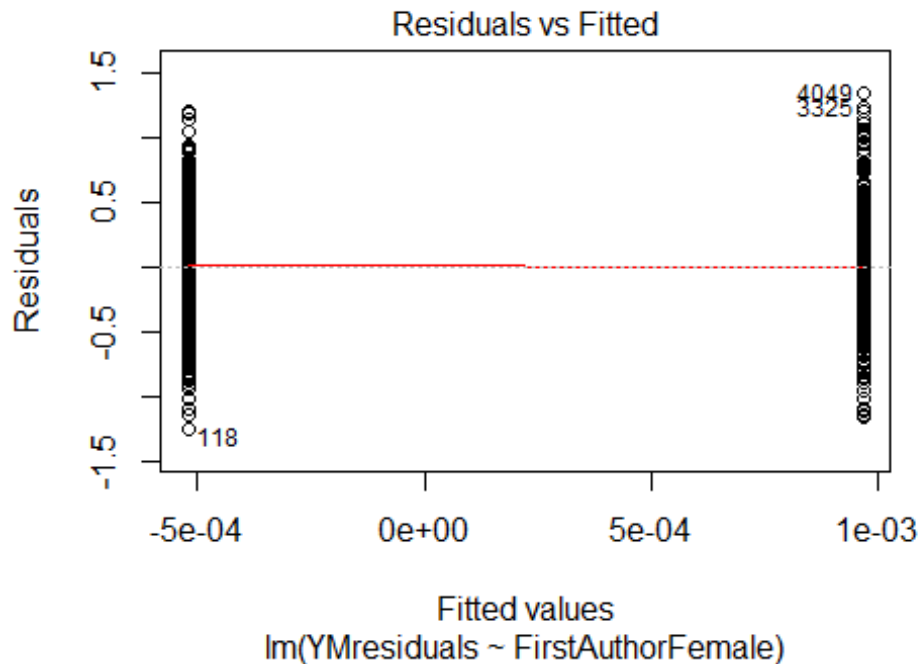
```



```
## 73 70 58 63 68 58 80 92 129 125 98 122 147 161 227
## 2011 2012
## 213 187
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 64 63 52 47 61 51 67 76 102 100 87 101 127 132 196
## 2011 2012
## 182 160
## [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
```

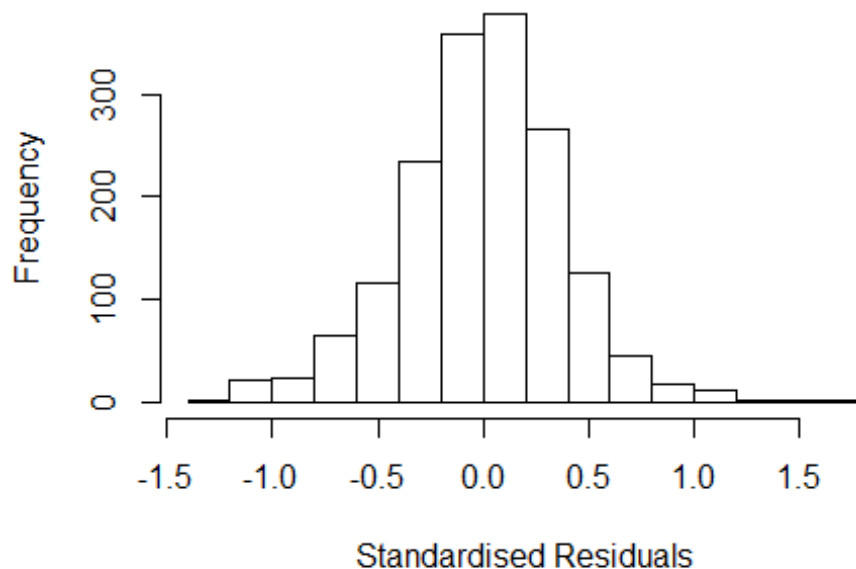


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.033, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 5.48264989767015"
## [1] "Male first author team size 2018 geometric mean: 5.1336398167281"
##
## 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] "Female last author team size 2018 geometric mean: 5.03916252298714"
## [1] "Male last author team size 2018 geometric mean: 5.33910178761729"
##
## 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 1.049 1      1.024
## LastAuthorFemale  1.064 1      1.032
## UniqueAuthors    1.349 4      1.038
## Year             1.428 16      1.011
```

Residuals from first and last author and team size



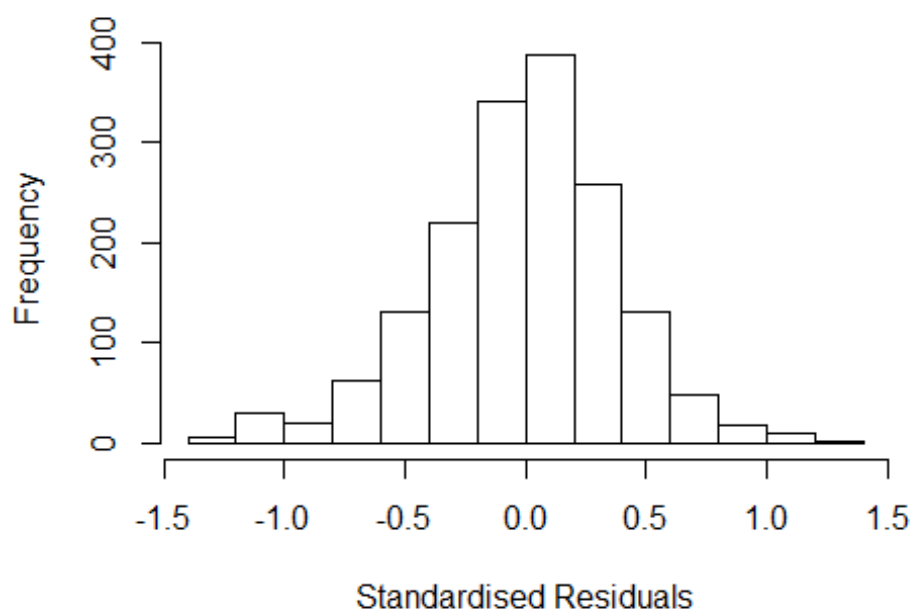
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28087 -0.23525  0.00584  0.23113  1.70519
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9158    0.0880   10.41 < 2e-16 ***
## FirstAuthorFemale1 -0.0146    0.0196   -0.75  0.45572
## LastAuthorFemale1  0.0459    0.0248    1.85  0.06446 .
## UniqueAuthors2    0.3187    0.0828    3.85  0.00012 ***
## UniqueAuthors3    0.3415    0.0793    4.31  1.8e-05 ***
## UniqueAuthors4    0.3945    0.0805    4.90  1.1e-06 ***
## UniqueAuthors5    0.4372    0.0775    5.64  2.0e-08 ***
## Year1997         -0.0844    0.0817   -1.03  0.30207
## Year1998         -0.1181    0.0872   -1.35  0.17590
## Year1999         -0.2139    0.0724   -2.95  0.00318 **
```

```

## Year2000          -0.1930      0.0693    -2.79   0.00540 **
## Year2001          -0.0965      0.0725    -1.33   0.18333
## Year2002          -0.1268      0.0738    -1.72   0.08594 .
## Year2003          -0.2498      0.0641    -3.90   0.00010 ***
## Year2004          -0.2105      0.0634    -3.32   0.00093 ***
## Year2005          -0.1571      0.0616    -2.55   0.01087 *
## Year2006          -0.1329      0.0615    -2.16   0.03097 *
## Year2007          -0.1311      0.0606    -2.17   0.03050 *
## Year2008          -0.1553      0.0639    -2.43   0.01520 *
## Year2009          -0.2135      0.0605    -3.53   0.00043 ***
## Year2010          -0.2012      0.0591    -3.40   0.00068 ***
## Year2011          -0.1693      0.0594    -2.85   0.00442 **
## Year2012          -0.2133      0.0617    -3.46   0.00056 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.349
## Multiple R-squared:  0.0715, Adjusted R-squared:  0.0591
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 1592 is an outlier with |weight| = 0 ( < 6e-05);
## 140 weights are ~= 1. The remaining 1527 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0037 0.8670 0.9510 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           6.00e-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 1.056 1 1.028
## LastAuthorFemale 1.058 1 1.029
## Year 1.106 16 1.003

```

Residuals from first and last author



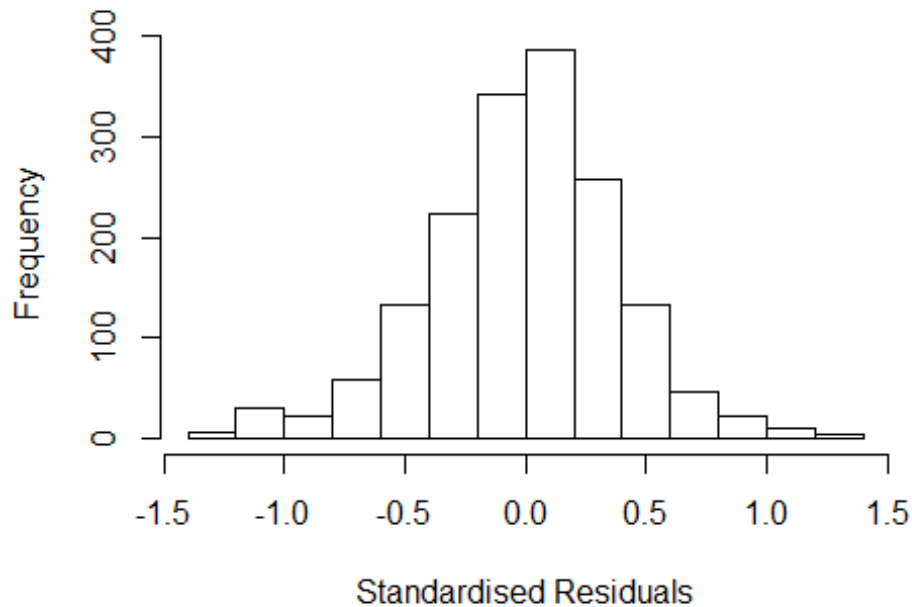
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22054 -0.24027  0.00848  0.23542  1.31015
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.25799    0.05284   23.81  <2e-16 ***
## FirstAuthorFemale1 -0.00954    0.01976   -0.48   0.6294
## LastAuthorFemale1  0.03651    0.02461    1.48   0.1380
## Year1997        -0.03744    0.08146   -0.46   0.6458
## Year1998        -0.10629    0.09174   -1.16   0.2468
## Year1999        -0.18720    0.07009   -2.67   0.0076 **
## Year2000        -0.15888    0.06900   -2.30   0.0214 *
## Year2001        -0.06223    0.07362   -0.85   0.3981
## Year2002        -0.09570    0.07424   -1.29   0.1975
## Year2003        -0.20162    0.06478   -3.11   0.0019 **
## Year2004        -0.15377    0.06322   -2.43   0.0151 *
## Year2005        -0.11744    0.06118   -1.92   0.0551 .
```

```

## Year2006      -0.07333    0.06148   -1.19    0.2331
## Year2007      -0.07858    0.06050   -1.30    0.1942
## Year2008      -0.09493    0.06394   -1.48    0.1378
## Year2009      -0.16170    0.06105   -2.65    0.0082 **
## Year2010      -0.14420    0.05892   -2.45    0.0145 *
## Year2011      -0.12873    0.06057   -2.13    0.0337 *
## Year2012      -0.15611    0.06259   -2.49    0.0127 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.354
## Multiple R-squared:  0.0166, Adjusted R-squared:  0.00589
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 156 weights are ~= 1. The remaining 1512 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.142  0.861  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      6.00e-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 1.051 1      1.025
## Year              1.051 16      1.002

```

Residuals from first author



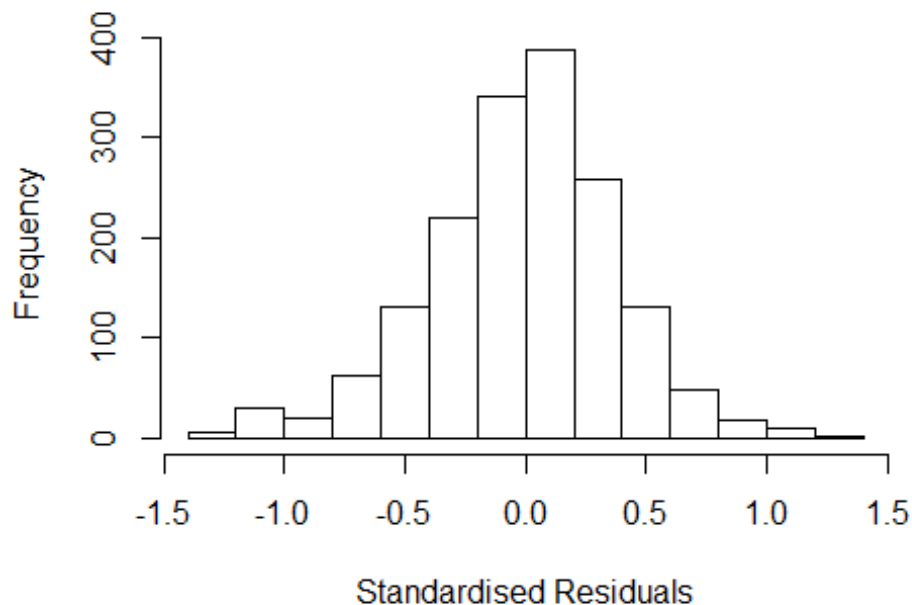
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22368 -0.24507  0.00834  0.23448  1.33937
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.26158    0.05284   23.87  <2e-16 ***
## FirstAuthorFemale1 -0.00636    0.01976   -0.32  0.7476
## Year1997          -0.03790    0.08156   -0.46  0.6422
## Year1998          -0.10394    0.09127   -1.14  0.2549
## Year1999          -0.18569    0.07010   -2.65  0.0082 **
## Year2000          -0.15508    0.06938   -2.24  0.0255 *
## Year2001          -0.05926    0.07384   -0.80  0.4224
## Year2002          -0.09436    0.07416   -1.27  0.2034
## Year2003          -0.19991    0.06497   -3.08  0.0021 **
## Year2004          -0.15122    0.06343   -2.38  0.0172 *
## Year2005          -0.11281    0.06113   -1.85  0.0652 .
## Year2006          -0.07403    0.06158   -1.20  0.2295
```

```

## Year2007          -0.07473    0.06060   -1.23    0.2177
## Year2008          -0.09306    0.06402   -1.45    0.1463
## Year2009          -0.15917    0.06118   -2.60    0.0094 **
## Year2010          -0.14283    0.05907   -2.42    0.0157 *
## Year2011          -0.12646    0.06072   -2.08    0.0374 *
## Year2012          -0.15559    0.06272   -2.48    0.0132 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.354
## Multiple R-squared:  0.0151, Adjusted R-squared:  0.005
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 160 weights are ~= 1. The remaining 1508 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.121  0.861  0.948  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      6.00e-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 1.054 1          1.027
## Year            1.054 16          1.002

```


Residuals from last author



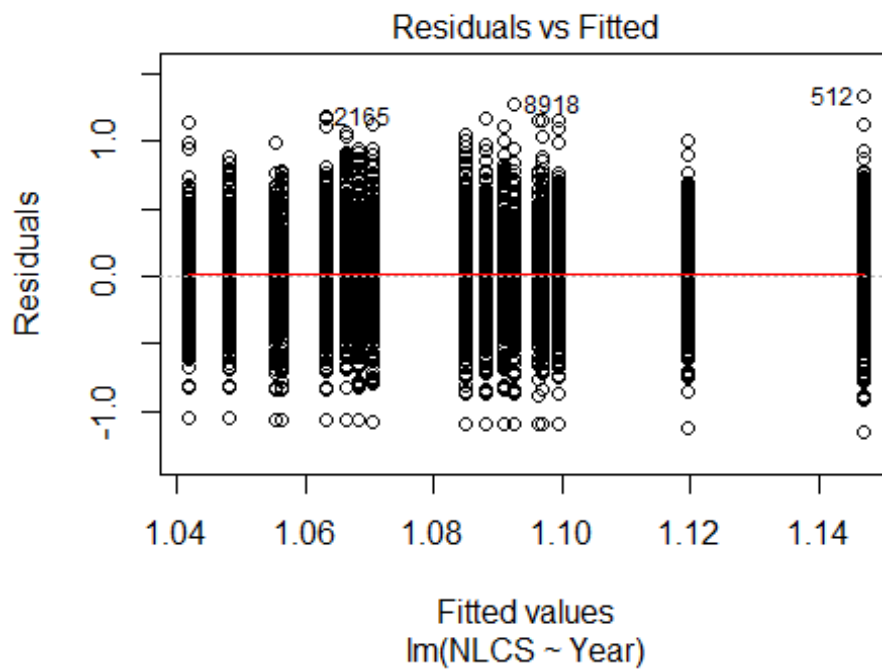
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21771 -0.24048  0.00756  0.23481  1.30531
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2555     0.0526   23.86  <2e-16 ***
## LastAuthorFemale1  0.0353     0.0246    1.43  0.1518
## Year1997         -0.0378     0.0815   -0.46  0.6430
## Year1998         -0.1061     0.0916   -1.16  0.2470
## Year1999         -0.1870     0.0701   -2.67  0.0078 **
## Year2000         -0.1589     0.0691   -2.30  0.0216 *
## Year2001         -0.0629     0.0737   -0.85  0.3934
## Year2002         -0.0951     0.0744   -1.28  0.2013
## Year2003         -0.2024     0.0649   -3.12  0.0018 **
## Year2004         -0.1540     0.0634   -2.43  0.0152 *
## Year2005         -0.1177     0.0613   -1.92  0.0550 .
## Year2006         -0.0742     0.0617   -1.20  0.2293
```

```

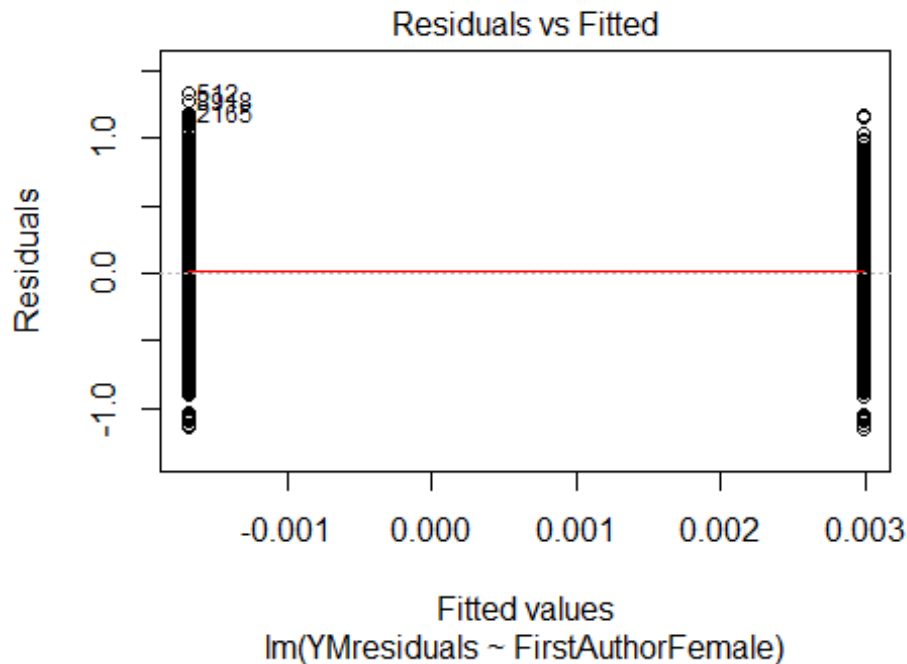
## Year2007          -0.0796      0.0606   -1.31    0.1897
## Year2008          -0.0952      0.0642   -1.48    0.1383
## Year2009          -0.1628      0.0612   -2.66    0.0079 **
## Year2010          -0.1448      0.0591   -2.45    0.0143 *
## Year2011          -0.1298      0.0607   -2.14    0.0326 *
## Year2012          -0.1571      0.0628   -2.50    0.0124 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.354
## Multiple R-squared:  0.0165, Adjusted R-squared:  0.00637
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 158 weights are ~ = 1. The remaining 1510 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.145  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      6.00e-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: 1668"
## [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
## 831 800 756 670 675 587 694 594 596 592 657 803 678 617 609
## 2011 2012
## 659 655
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 422 437 415 376 308 244 450 354 379 403 439 549 472 442 417
## 2011 2012

```

```
## 478 460
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 372 385 337 317 263 214 379 286 324 346 369 473 400 370 373
## 2011 2012
## 407 398
## [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
```

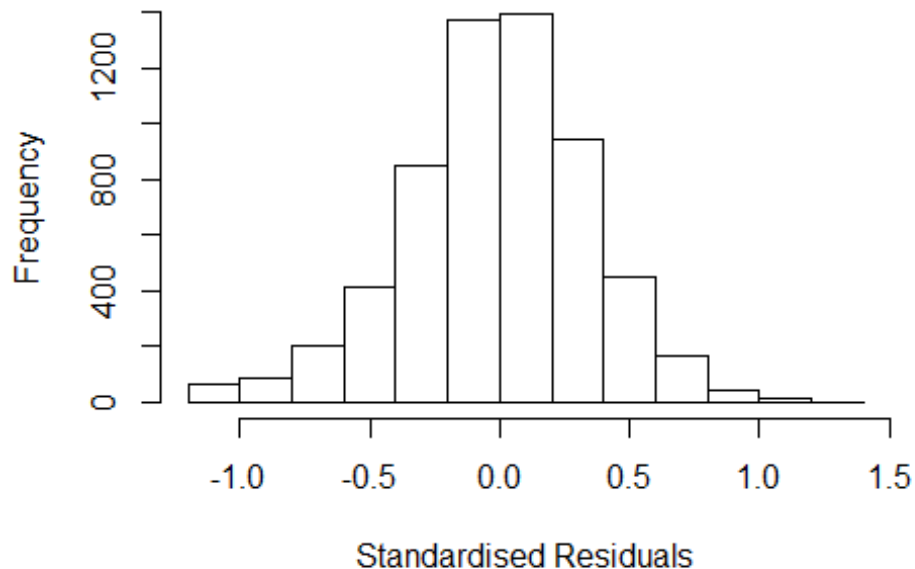


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.8, df = 1, p-value = 0.003
```



```
## [1] "Female first author team size 2018 geometric mean: 3.9787058309294"
## [1] "Male first author team size 2018 geometric mean: 3.61467819038702"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 19000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.61257220774336"
## [1] "Male last author team size 2018 geometric mean: 3.84253407242893"
##
## 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 1.045 1      1.022
## LastAuthorFemale  1.022 1      1.011
## UniqueAuthors    1.110 4      1.013
## Year             1.168 16      1.005
```

Residuals from first and last author and team size



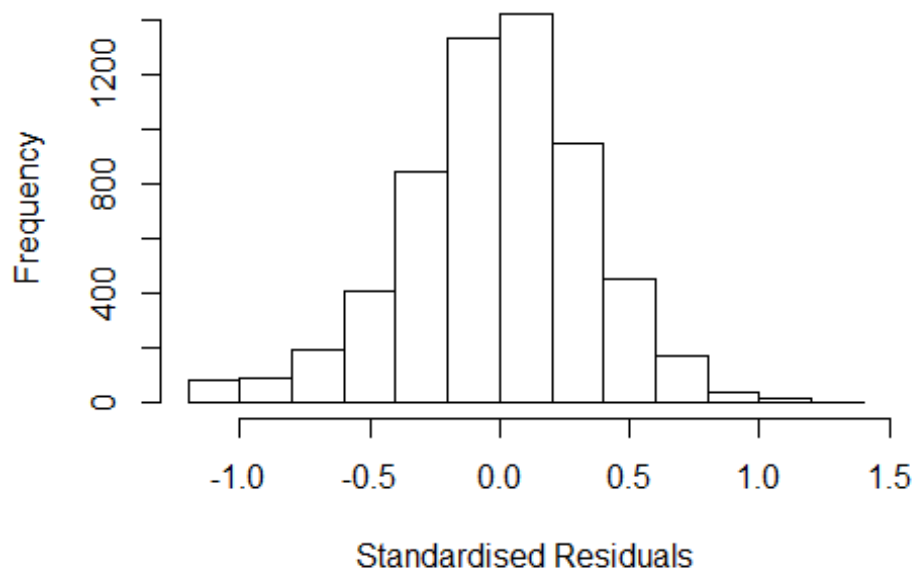
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1763 -0.2207  0.0019  0.2207  1.3470
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.00402    0.03653   27.48 < 2e-16 ***
## FirstAuthorFemale1  0.00223    0.00943    0.24  0.81323
## LastAuthorFemale1 -0.01942    0.01138   -1.71  0.08793 .
## UniqueAuthors2     0.14541    0.03285    4.43  9.7e-06 ***
## UniqueAuthors3     0.17227    0.03266    5.27  1.4e-07 ***
## UniqueAuthors4     0.17365    0.03331    5.21  1.9e-07 ***
## UniqueAuthors5     0.23648    0.03287    7.20  7.0e-13 ***
## Year1997          -0.07843    0.02955   -2.65  0.00797 **
## Year1998          -0.06775    0.03043   -2.23  0.02605 *
## Year1999          -0.12158    0.02878   -4.22  2.4e-05 ***
```

```

## Year2000      -0.11218    0.03081   -3.64  0.00027 ***
## Year2001      -0.08535    0.02997   -2.85  0.00442 **
## Year2002      -0.08216    0.02831   -2.90  0.00372 **
## Year2003      -0.08219    0.02891   -2.84  0.00448 **
## Year2004      -0.10458    0.02735   -3.82  0.00013 ***
## Year2005      -0.13256    0.02739   -4.84  1.3e-06 ***
## Year2006      -0.06602    0.02699   -2.45  0.01449 *
## Year2007      -0.08115    0.02548   -3.18  0.00146 **
## Year2008      -0.06771    0.02716   -2.49  0.01270 *
## Year2009      -0.05925    0.02697   -2.20  0.02803 *
## Year2010      -0.10291    0.02791   -3.69  0.00023 ***
## Year2011      -0.07827    0.02695   -2.90  0.00370 **
## Year2012      -0.10341    0.02893   -3.57  0.00035 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.329
## Multiple R-squared:  0.0291, Adjusted R-squared:  0.0256
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 498 weights are ~= 1. The remaining 5515 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0569 0.8650 0.9510 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.66e-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 1.044 1 1.022
## LastAuthorFemale 1.019 1 1.010
## Year 1.060 16 1.002

```

Residuals from first and last author



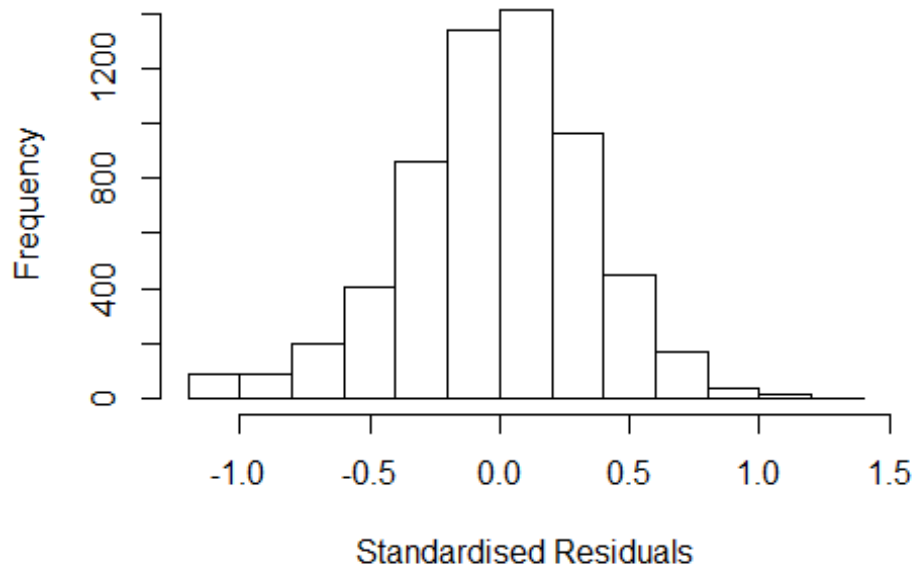
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15818 -0.22459 0.00541 0.21874 1.34131
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15818 0.02112 54.83 < 2e-16 ***
## FirstAuthorFemale1 0.00615 0.00949 0.65 0.51713
## LastAuthorFemale1 -0.02249 0.01145 -1.96 0.04956 *
## Year1997 -0.07533 0.02956 -2.55 0.01084 *
## Year1998 -0.05999 0.03042 -1.97 0.04867 *
## Year1999 -0.10800 0.02877 -3.75 0.00018 ***
## Year2000 -0.09592 0.03065 -3.13 0.00176 **
## Year2001 -0.06174 0.02950 -2.09 0.03643 *
## Year2002 -0.06754 0.02850 -2.37 0.01783 *
## Year2003 -0.05592 0.02885 -1.94 0.05262 .
## Year2004 -0.07567 0.02719 -2.78 0.00540 **
## Year2005 -0.10919 0.02739 -3.99 6.8e-05 ***
```

```

## Year2006      -0.04182    0.02711   -1.54  0.12298
## Year2007      -0.05145    0.02546   -2.02  0.04334 *
## Year2008      -0.03722    0.02735   -1.36  0.17367
## Year2009      -0.03035    0.02694   -1.13  0.26005
## Year2010      -0.07461    0.02802   -2.66  0.00777 **
## Year2011      -0.04732    0.02687   -1.76  0.07826 .
## Year2012      -0.07098    0.02891   -2.46  0.01411 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.332
## Multiple R-squared:  0.00684,    Adjusted R-squared:  0.00386
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 474 weights are ~= 1. The remaining 5539 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.066  0.866  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      1.66e-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 1.042 1      1.021
## Year      1.042 16      1.001

```


Residuals from first author



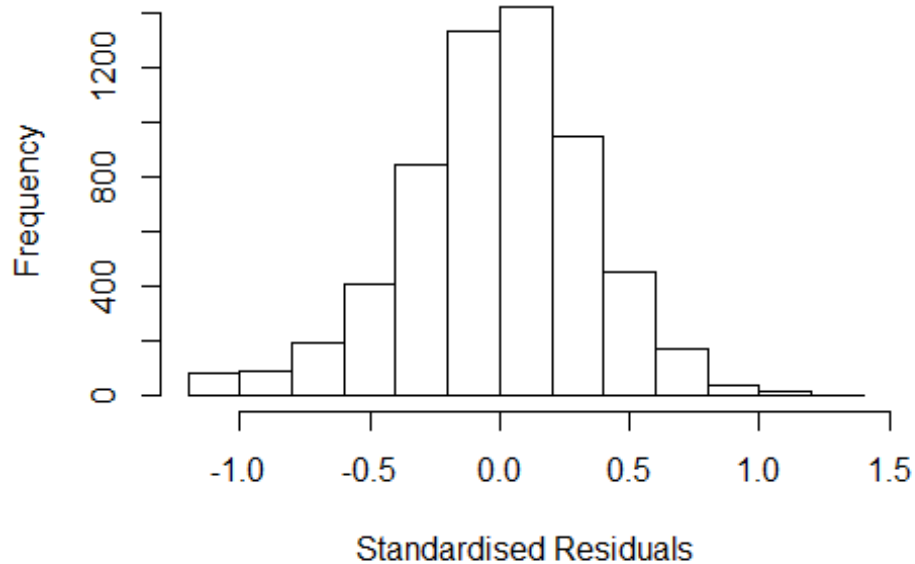
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15563 -0.22224 0.00551 0.21974 1.32137
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15563 0.02109 54.80 < 2e-16 ***
## FirstAuthorFemale1 0.00473 0.00949 0.50 0.61840
## Year1997 -0.07617 0.02956 -2.58 0.01000 *
## Year1998 -0.06052 0.03044 -1.99 0.04682 *
## Year1999 -0.10838 0.02879 -3.76 0.00017 ***
## Year2000 -0.09760 0.03061 -3.19 0.00144 **
## Year2001 -0.06244 0.02949 -2.12 0.03429 *
## Year2002 -0.06887 0.02850 -2.42 0.01570 *
## Year2003 -0.05790 0.02883 -2.01 0.04462 *
## Year2004 -0.07739 0.02720 -2.84 0.00446 **
## Year2005 -0.11133 0.02738 -4.07 4.8e-05 ***
## Year2006 -0.04387 0.02707 -1.62 0.10517
```

```

## Year2007          -0.05288    0.02549   -2.07  0.03809 *
## Year2008          -0.03910    0.02735   -1.43  0.15284
## Year2009          -0.03137    0.02697   -1.16  0.24472
## Year2010          -0.07612    0.02801   -2.72  0.00660 **
## Year2011          -0.04915    0.02686   -1.83  0.06728 .
## Year2012          -0.07275    0.02897   -2.51  0.01206 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.333
## Multiple R-squared:  0.00617,    Adjusted R-squared:  0.00335
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 487 weights are ~= 1. The remaining 5526 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.079  0.866  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      1.66e-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 1.018 1          1.009
## Year              1.018 16          1.001

```

Residuals from last author



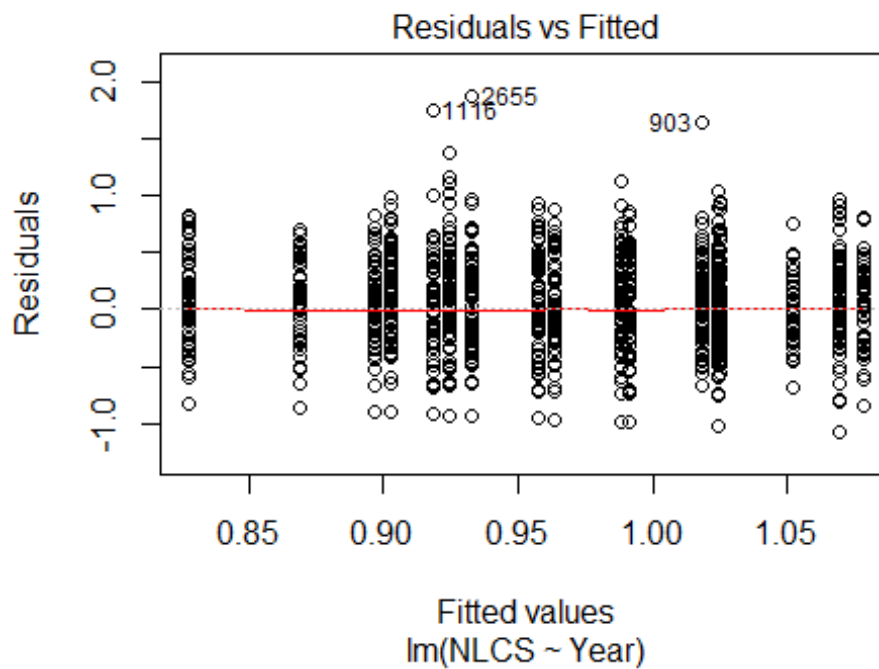
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.15992 -0.22414  0.00477  0.21885  1.33900
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1599     0.0209   55.37 < 2e-16 ***
## LastAuthorFemale1 -0.0219     0.0114   -1.92  0.05552 .
## Year1997         -0.0755     0.0296   -2.55  0.01071 *
## Year1998         -0.0602     0.0304   -1.98  0.04772 *
## Year1999         -0.1079     0.0288   -3.75  0.00018 ***
## Year2000         -0.0957     0.0306   -3.12  0.00180 **
## Year2001         -0.0617     0.0295   -2.09  0.03661 *
## Year2002         -0.0673     0.0285   -2.36  0.01830 *
## Year2003         -0.0560     0.0289   -1.94  0.05264 .
## Year2004         -0.0753     0.0272   -2.77  0.00565 **
## Year2005         -0.1089     0.0274   -3.97  7.2e-05 ***
## Year2006         -0.0413     0.0271   -1.52  0.12789
```

```

## Year2007          -0.0508      0.0254   -2.00   0.04574 *
## Year2008          -0.0365      0.0273   -1.33   0.18201
## Year2009          -0.0296      0.0269   -1.10   0.27188
## Year2010          -0.0736      0.0279   -2.63   0.00844 **
## Year2011          -0.0462      0.0268   -1.72   0.08486 .
## Year2012          -0.0698      0.0289   -2.42   0.01566 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.332
## Multiple R-squared:  0.00677,    Adjusted R-squared:  0.00395
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 484 weights are ~= 1. The remaining 5529 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0672 0.8660 0.9520 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.66e-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: 6013"
## [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
## 120 127 134 112 110 108 121 97 97 122 145 146 139 121 133
## 2011 2012
## 126 134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 60 65 72 59 51 40 78 67 68 80 94 104 96 73 92
## 2011 2012

```

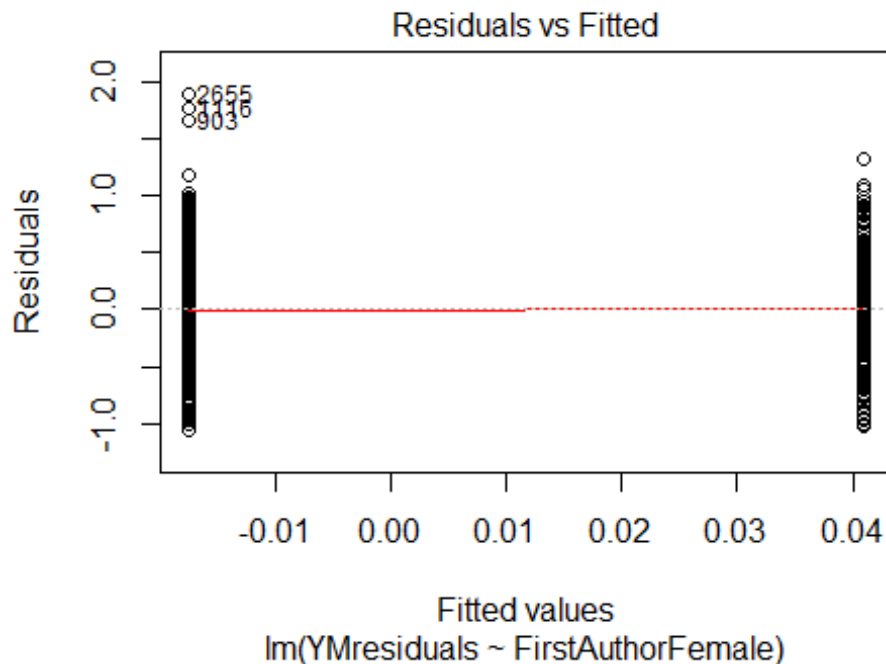
```
## 74 85
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 52 68 49 41 33 56 50 55 70 78 91 76 65 78
## 2011 2012
## 66 72
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 22, df = 16, p-value = 0.1
```



```
##
## 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: 3.46356519907113"
## [1] "Male first author team size 2018 geometric mean: 3.54940456367511"
## Warning in wilcox.test.default(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: 2.81319791511651"
## [1] "Male last author team size 2018 geometric mean: 3.80806902252752"

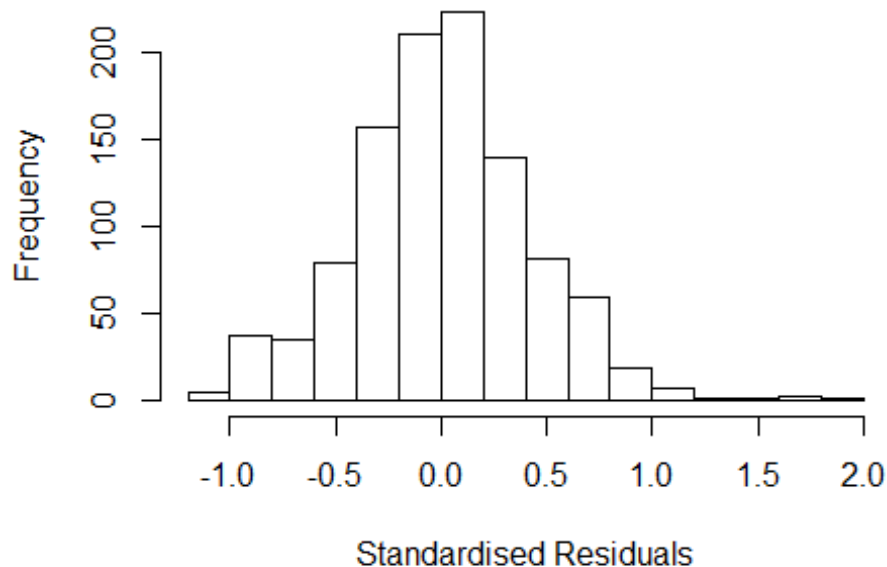
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "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	1.101	1	1.049
UniqueAuthors	1.336	4	1.037
Year	1.449	16	1.012

Residuals from first and last author and team size



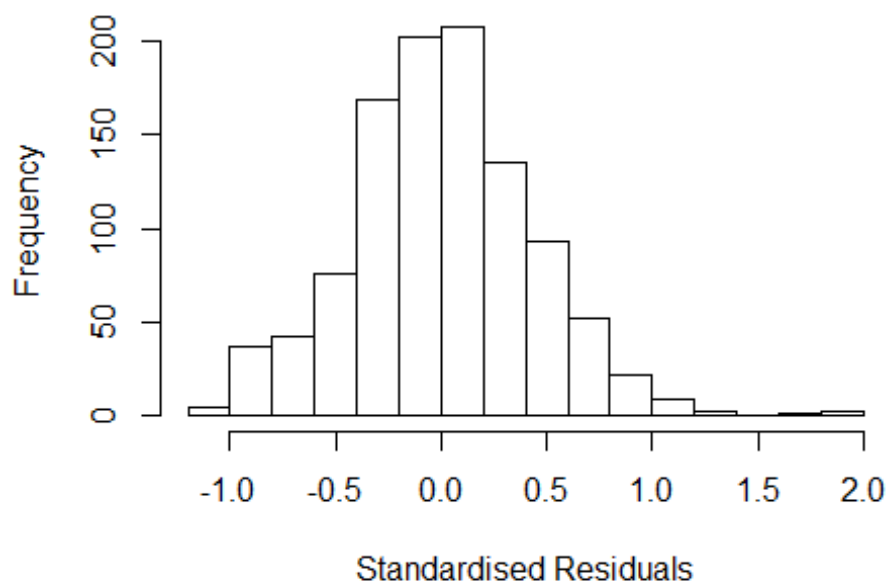
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.13664 -0.25591 0.00466 0.25993 1.83322
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.78862 0.06019 13.10 < 2e-16 ***
## FirstAuthorFemale1 0.05785 0.02829 2.04 0.0411 *
## LastAuthorFemale1 -0.01386 0.03799 -0.36 0.7153
## UniqueAuthors2 0.13099 0.05464 2.40 0.0167 *
## UniqueAuthors3 0.12570 0.05645 2.23 0.0262 *
## UniqueAuthors4 0.19001 0.06009 3.16 0.0016 **
## UniqueAuthors5 0.33552 0.05698 5.89 5.3e-09 ***
## Year1997 0.01375 0.07491 0.18 0.8544
## Year1998 -0.06628 0.06269 -1.06 0.2906
## Year1999 -0.09989 0.06969 -1.43 0.1521
```

```

## Year2000          0.06074      0.06873      0.88      0.3770
## Year2001          0.01776      0.06808      0.26      0.7942
## Year2002          0.00373      0.06577      0.06      0.9548
## Year2003         -0.08882      0.06964     -1.28      0.2025
## Year2004         -0.17408      0.06679     -2.61      0.0093 **
## Year2005          0.01251      0.07257      0.17      0.8632
## Year2006          0.01457      0.06303      0.23      0.8172
## Year2007         -0.04173      0.06632     -0.63      0.5293
## Year2008         -0.08405      0.07130     -1.18      0.2388
## Year2009          0.12233      0.06947      1.76      0.0785 .
## Year2010          0.01192      0.06483      0.18      0.8542
## Year2011          0.01588      0.06942      0.23      0.8191
## Year2012         -0.11645      0.06417     -1.81      0.0699 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.385
## Multiple R-squared:  0.0985, Adjusted R-squared:  0.0792
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(385,1041)
## are outliers with |weight| <= 8.7e-05 ( < 9.5e-05);
## 95 weights are ~= 1. The remaining 957 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.016  0.863   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           9.49e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.075 1           1.037
## LastAuthorFemale  1.088 1           1.043
## Year              1.132 16           1.004

```


Residuals from first and last author



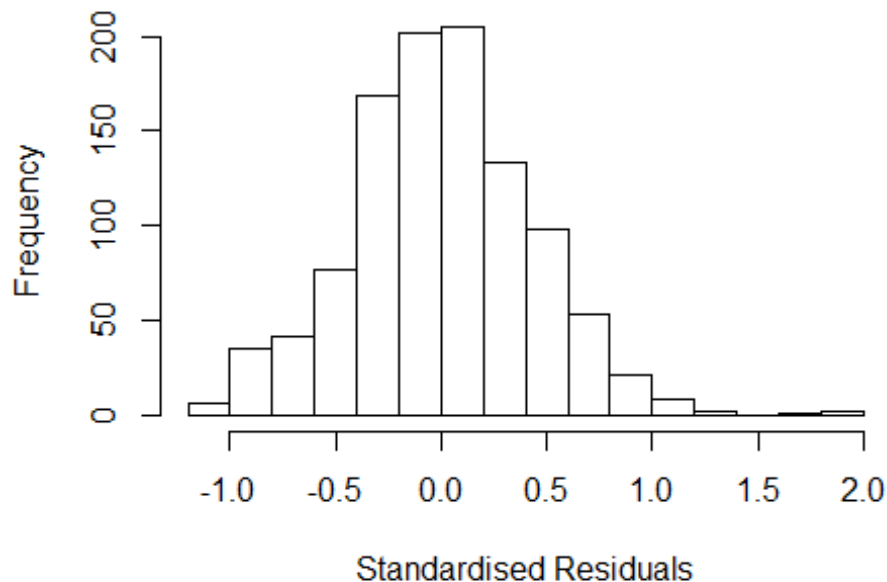
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09438 -0.26281 -0.00599  0.27069  1.94885
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9321    0.0475   19.62  <2e-16 ***
## FirstAuthorFemale1 0.0824    0.0289    2.85  0.0045 **
## LastAuthorFemale1 -0.0217    0.0382   -0.57  0.5699
## Year1997          0.0087    0.0810    0.11  0.9145
## Year1998         -0.0532    0.0673   -0.79  0.4295
## Year1999         -0.0943    0.0740   -1.27  0.2030
## Year2000          0.1125    0.0762    1.48  0.1405
## Year2001          0.0842    0.0736    1.14  0.2532
## Year2002          0.0312    0.0689    0.45  0.6513
## Year2003         -0.0706    0.0751   -0.94  0.3473
## Year2004         -0.1674    0.0682   -2.45  0.0143 *
## Year2005          0.0274    0.0739    0.37  0.7105
```

```

## Year2006          0.0559      0.0679      0.82      0.4109
## Year2007         -0.0188      0.0685     -0.27      0.7840
## Year2008         -0.0586      0.0743     -0.79      0.4307
## Year2009          0.1623      0.0710      2.28      0.0225 *
## Year2010          0.0787      0.0677      1.16      0.2453
## Year2011          0.0428      0.0726      0.59      0.5559
## Year2012         -0.0770      0.0685     -1.12      0.2614
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.404
## Multiple R-squared:  0.0435, Adjusted R-squared:  0.0269
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1041 is an outlier with |weight| = 0 ( < 9.5e-05);
## 92 weights are ~= 1. The remaining 961 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0084 0.8650 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      9.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.053 1          1.026
## Year              1.053 16          1.002

```

Residuals from first author



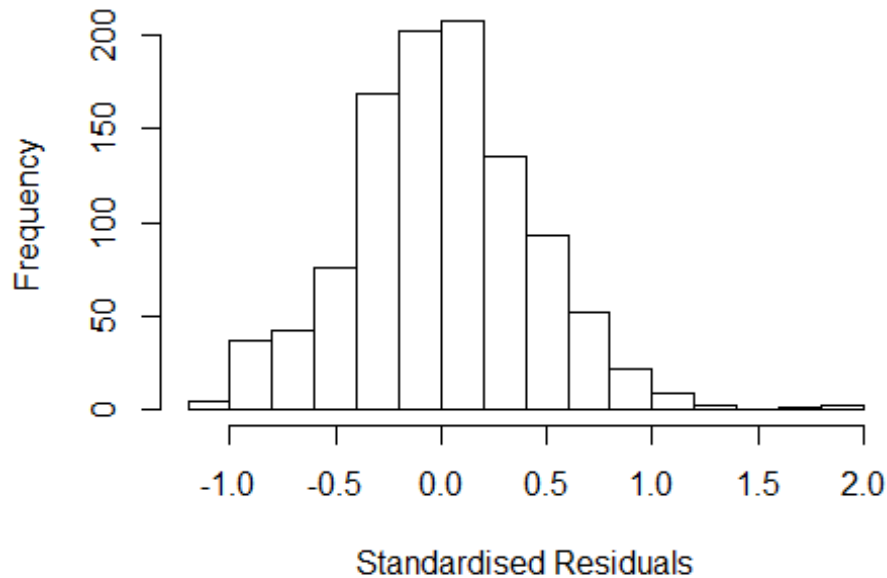
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09144 -0.26360 -0.00378  0.27449  1.95034
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.93224    0.04766   19.56  <2e-16 ***
## FirstAuthorFemale1 0.08004    0.02865    2.79   0.0053 **
## Year1997         0.00515    0.08047    0.06   0.9490
## Year1998        -0.05523    0.06745   -0.82   0.4131
## Year1999        -0.09603    0.07420   -1.29   0.1959
## Year2000         0.10719    0.07572    1.42   0.1572
## Year2001         0.08245    0.07343    1.12   0.2618
## Year2002         0.02993    0.06924    0.43   0.6656
## Year2003        -0.07436    0.07515   -0.99   0.3227
## Year2004        -0.16935    0.06828   -2.48   0.0133 *
## Year2005         0.02449    0.07400    0.33   0.7407
## Year2006         0.05270    0.06779    0.78   0.4371
```

```

## Year2007          -0.02218      0.06846    -0.32    0.7460
## Year2008          -0.06207      0.07412    -0.84    0.4026
## Year2009           0.15921      0.07107     2.24    0.0253 *
## Year2010           0.07545      0.06812     1.11    0.2683
## Year2011           0.03950      0.07238     0.55    0.5854
## Year2012          -0.07858      0.06879    -1.14    0.2536
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.407
## Multiple R-squared:  0.0429, Adjusted R-squared:  0.0272
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1041 is an outlier with |weight| = 0 ( < 9.5e-05);
## 94 weights are ~= 1. The remaining 959 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0109 0.8660 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      9.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.062 1          1.031
## Year            1.062 16          1.002

```

Residuals from last author



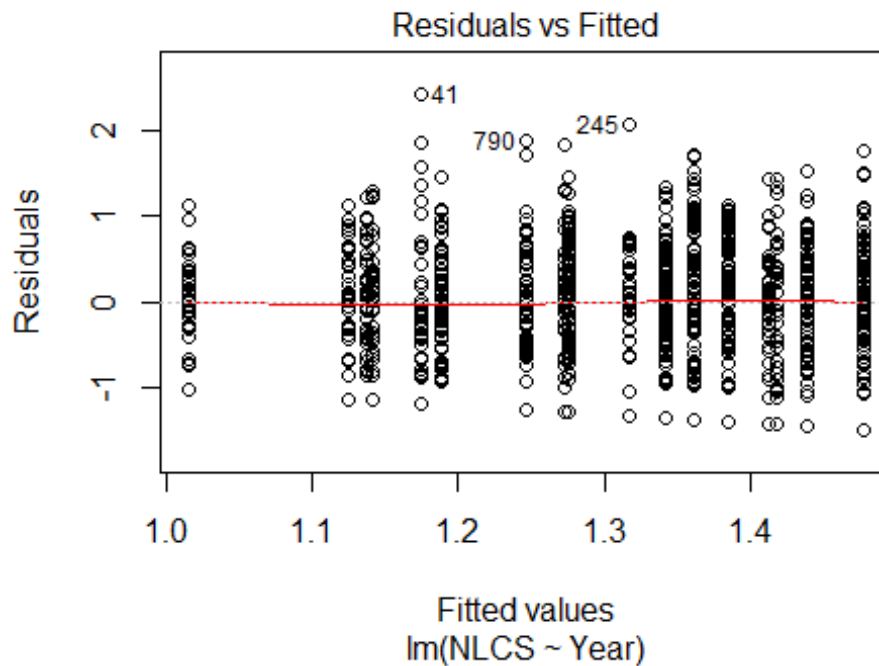
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.11536 -0.27677 0.00177 0.27211 1.91669
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95977 0.04751 20.20 <2e-16 ***
## LastAuthorFemale1 -0.00644 0.03805 -0.17 0.866
## Year1997 -0.00259 0.08126 -0.03 0.975
## Year1998 -0.06260 0.06907 -0.91 0.365
## Year1999 -0.10319 0.07553 -1.37 0.172
## Year2000 0.10829 0.07649 1.42 0.157
## Year2001 0.07886 0.07246 1.09 0.277
## Year2002 0.02604 0.07020 0.37 0.711
## Year2003 -0.08347 0.07692 -1.09 0.278
## Year2004 -0.17373 0.06991 -2.48 0.013 *
## Year2005 0.02068 0.07460 0.28 0.782
## Year2006 0.05283 0.06861 0.77 0.441
```

```

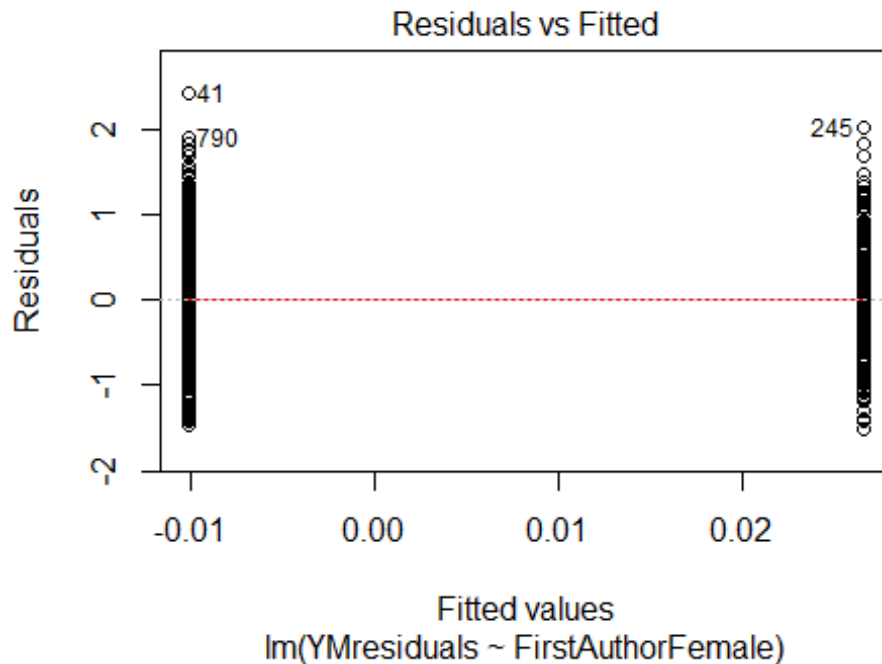
## Year2007          -0.02478      0.06996    -0.35      0.723
## Year2008          -0.06796      0.07403    -0.92      0.359
## Year2009           0.15559      0.07115     2.19      0.029 *
## Year2010           0.07736      0.06972     1.11      0.267
## Year2011           0.03751      0.07345     0.51      0.610
## Year2012          -0.07246      0.06964    -1.04      0.298
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.0353, Adjusted R-squared:  0.0194
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1041 is an outlier with |weight| <= 3e-07 ( < 9.5e-05);
## 98 weights are ~= 1. The remaining 955 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.017  0.865   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      9.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1054"
## [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
##   51   53   53   52   51   73   49   56   44   60   80  105  141  125  124
## 2011 2012
##  120  114
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   43   39   36   34   39   61   38   41   33   43   63   80  112  100   99

```

```
## 2011 2012
## 96 93
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 39 38 33 31 37 55 35 37 31 38 61 73 107 89 85
## 2011 2012
## 81 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 = 27, df = 16, p-value = 0.04
```

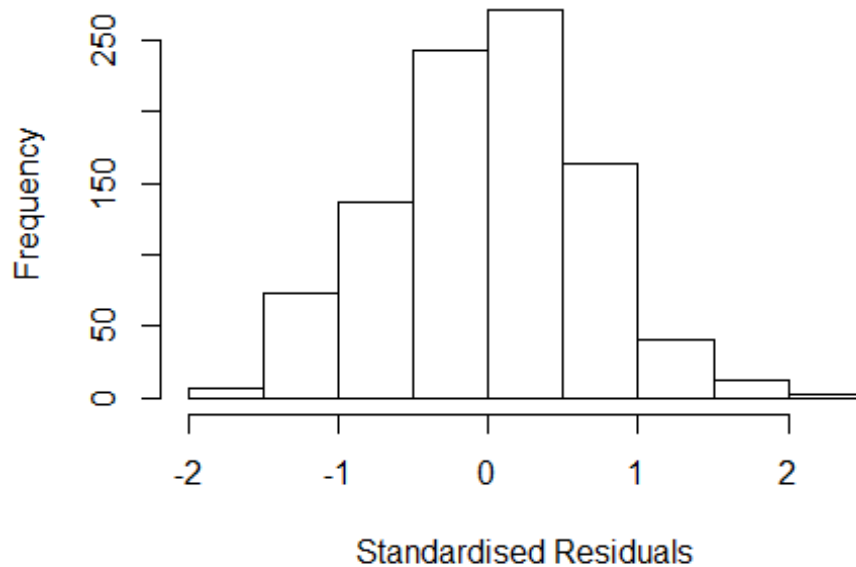


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.34, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 2.06699573931857"
## [1] "Male first author team size 2018 geometric mean: 1.93486479302598"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 890, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.66439353652919"
## [1] "Male last author team size 2018 geometric mean: 2.11409786769712"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 520, 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 1.903 1          1.380
## LastAuthorFemale  1.791 1          1.338
## UniqueAuthors    1.324 4          1.036
## Year              1.440 16         1.011
```


Residuals from first and last author and team size



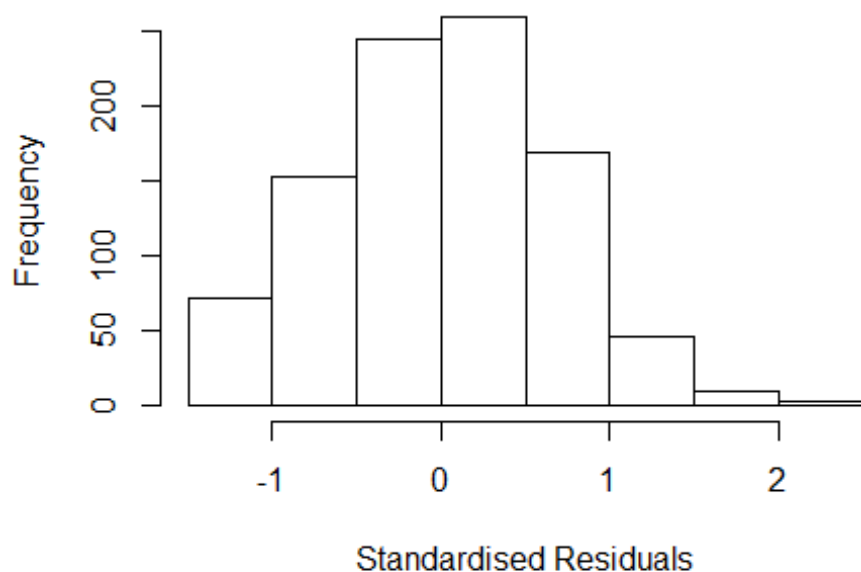
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6569 -0.4478 0.0175 0.4526 2.2150
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8837 0.1050 8.42 < 2e-16 ***
## FirstAuthorFemale1 0.0495 0.0660 0.75 0.4533
## LastAuthorFemale1 -0.0262 0.0656 -0.40 0.6899
## UniqueAuthors2 0.2116 0.0529 4.00 6.9e-05 ***
## UniqueAuthors3 0.3729 0.0655 5.69 1.7e-08 ***
## UniqueAuthors4 0.3310 0.1378 2.40 0.0165 *
## UniqueAuthors5 0.1456 0.1180 1.23 0.2174
## Year1997 0.1755 0.1468 1.20 0.2324
## Year1998 -0.0174 0.1455 -0.12 0.9046
## Year1999 0.2590 0.1523 1.70 0.0894 .
```

```

## Year2000          0.1391      0.1595      0.87      0.3833
## Year2001          0.1718      0.1271      1.35      0.1769
## Year2002          0.2404      0.1475      1.63      0.1034
## Year2003          0.1615      0.1598      1.01      0.3125
## Year2004          0.3930      0.1743      2.25      0.0244 *
## Year2005          0.3546      0.1439      2.46      0.0139 *
## Year2006          0.1550      0.1319      1.18      0.2401
## Year2007          0.3948      0.1328      2.97      0.0030 **
## Year2008          0.3342      0.1192      2.80      0.0052 **
## Year2009          0.1895      0.1290      1.47      0.1424
## Year2010          0.3923      0.1289      3.04      0.0024 **
## Year2011          0.3508      0.1327      2.64      0.0084 **
## Year2012          0.2543      0.1565      1.63      0.1044
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.662
## Multiple R-squared:  0.0797, Adjusted R-squared:  0.0579
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 91 weights are ~= 1. The remaining 860 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.240  0.870  0.946  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          1.05e-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 1.773 1          1.331
## LastAuthorFemale 1.720 1          1.311
## Year          1.161 16          1.005

```

Residuals from first and last author



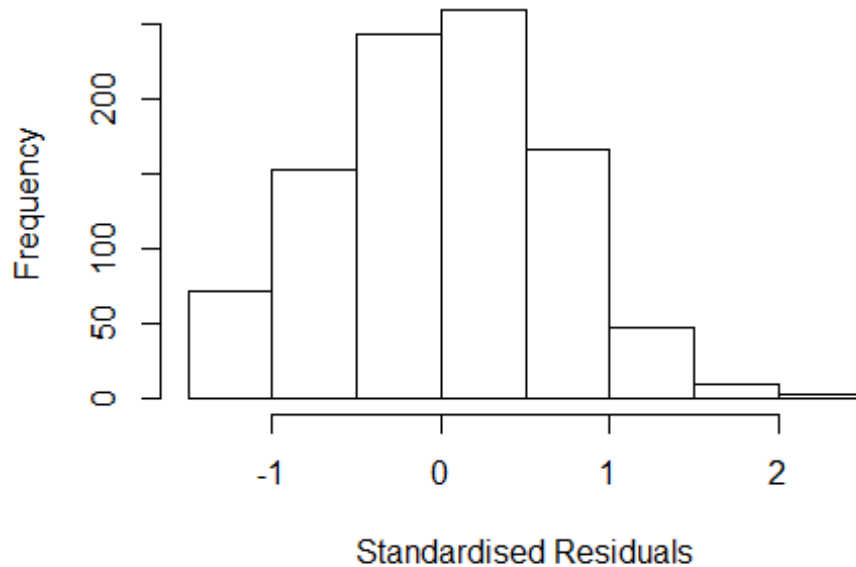
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4631 -0.4655 0.0136 0.4832 2.1153
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9694 0.0999 9.70 < 2e-16 ***
## FirstAuthorFemale1 0.0541 0.0652 0.83 0.40674
## LastAuthorFemale1 -0.0291 0.0657 -0.44 0.65832
## Year1997 0.1621 0.1394 1.16 0.24536
## Year1998 0.0210 0.1391 0.15 0.88020
## Year1999 0.2712 0.1465 1.85 0.06452 .
## Year2000 0.1636 0.1574 1.04 0.29886
## Year2001 0.1861 0.1273 1.46 0.14402
## Year2002 0.2393 0.1490 1.61 0.10874
## Year2003 0.1863 0.1560 1.19 0.23273
## Year2004 0.4229 0.1847 2.29 0.02230 *
## Year2005 0.4633 0.1375 3.37 0.00078 ***
```

```

## Year2006          0.2217      0.1317      1.68  0.09271 .
## Year2007          0.4622      0.1289      3.59  0.00035 ***
## Year2008          0.4057      0.1141      3.56  0.00039 ***
## Year2009          0.2536      0.1244      2.04  0.04182 *
## Year2010          0.4686      0.1252      3.74  0.00019 ***
## Year2011          0.4068      0.1309      3.11  0.00195 **
## Year2012          0.3291      0.1518      2.17  0.03039 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.69
## Multiple R-squared:  0.0391, Adjusted R-squared:  0.0206
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 82 weights are ~= 1. The remaining 869 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.327  0.871  0.951  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      1.05e-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 1.098 1      1.048
## Year              1.098 16      1.003

```

Residuals from first author



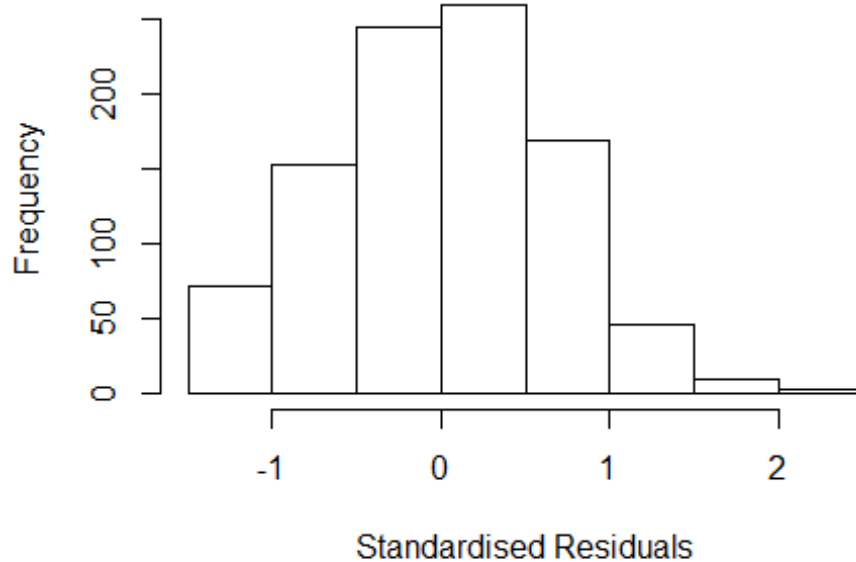
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.472 -0.466 0.011 0.476 2.108
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9656 0.0991 9.74 < 2e-16 ***
## FirstAuthorFemale1 0.0372 0.0514 0.72 0.46914
## Year1997 0.1624 0.1393 1.17 0.24381
## Year1998 0.0218 0.1392 0.16 0.87544
## Year1999 0.2700 0.1467 1.84 0.06593 .
## Year2000 0.1641 0.1570 1.05 0.29609
## Year2001 0.1860 0.1270 1.46 0.14335
## Year2002 0.2404 0.1487 1.62 0.10614
## Year2003 0.1836 0.1561 1.18 0.23982
## Year2004 0.4253 0.1852 2.30 0.02185 *
## Year2005 0.4618 0.1370 3.37 0.00078 ***
## Year2006 0.2226 0.1316 1.69 0.09111 .
```

```

## Year2007          0.4630      0.1288      3.60  0.00034 ***
## Year2008          0.4064      0.1138      3.57  0.00037 ***
## Year2009          0.2526      0.1243      2.03  0.04249 *
## Year2010          0.4694      0.1247      3.76  0.00018 ***
## Year2011          0.4092      0.1305      3.14  0.00177 **
## Year2012          0.3330      0.1506      2.21  0.02731 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.69
## Multiple R-squared:  0.039, Adjusted R-squared:  0.0215
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 867 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.330  0.868  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      1.05e-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 1.066 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
## -1.4494 -0.4625 0.0202 0.4818 2.1351
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97690 0.10023 9.75 < 2e-16 ***
## LastAuthorFemale1 0.00293 0.05175 0.06 0.95482
## Year1997 0.16106 0.13896 1.16 0.24671
## Year1998 0.01843 0.13996 0.13 0.89526
## Year1999 0.26610 0.14746 1.80 0.07146 .
## Year2000 0.16009 0.15723 1.02 0.30885
## Year2001 0.18253 0.12779 1.43 0.15351
## Year2002 0.23353 0.14913 1.57 0.11770
## Year2003 0.18071 0.15715 1.15 0.25045
## Year2004 0.42046 0.18684 2.25 0.02465 *
## Year2005 0.46559 0.13714 3.40 0.00072 ***
## Year2006 0.22411 0.13196 1.70 0.08977 .
```

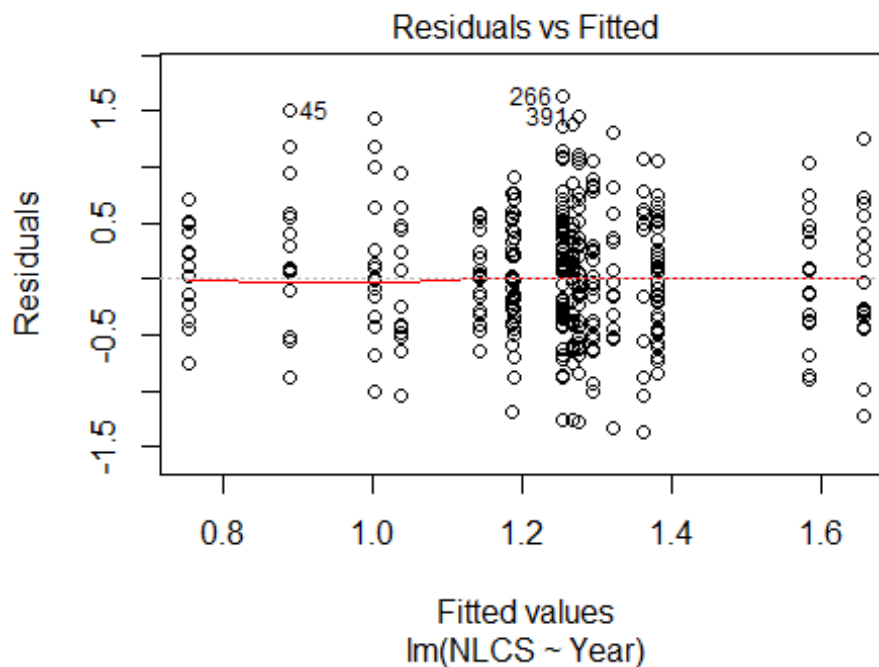
```

## Year2007      0.45905      0.12972      3.54  0.00042 ***
## Year2008      0.40913      0.11414      3.58  0.00036 ***
## Year2009      0.24860      0.12492      1.99  0.04687 *
## Year2010      0.46956      0.12532      3.75  0.00019 ***
## Year2011      0.40734      0.13130      3.10  0.00198 **
## Year2012      0.33258      0.15161      2.19  0.02851 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.689
## Multiple R-squared:  0.0385, Adjusted R-squared:  0.021
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 874 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.317  0.871  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      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 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
##   17   25   34   18   27   25   16   19   21   30   24   22   40   30   45
## 2011 2012
##   42   39
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   16   19   22   14   18   17   11   14   20   27   19   16   34   23   39
## 2011 2012

```



```
## 34 29
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 16 16 19 13 16 15 11 12 18 25 19 12 33 18 30
## 2011 2012
## 29 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 = 0.06
```



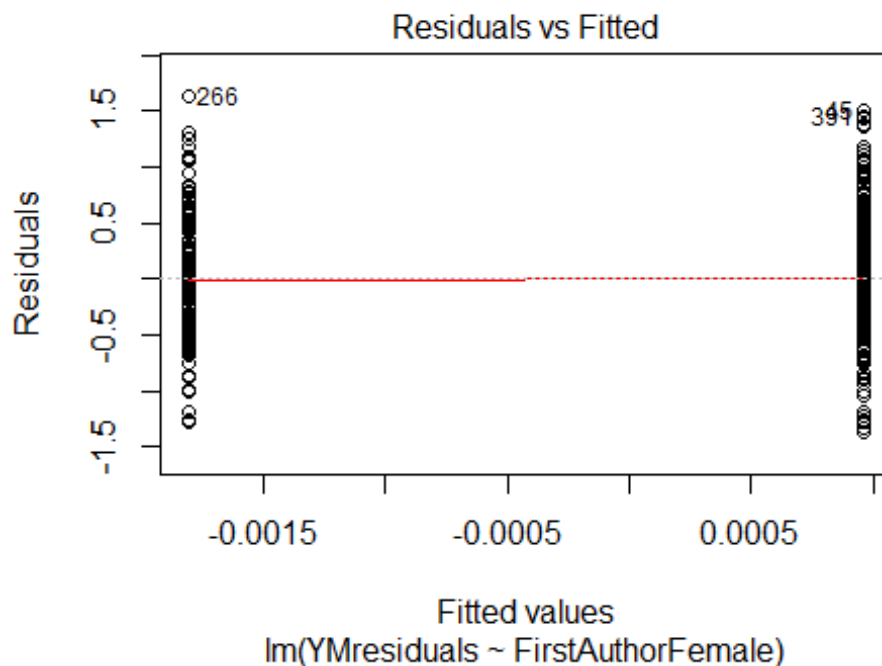
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.043, df = 1, p-value = 0.8

## [1] "Female first author team size 2018 geometric mean: 2.24593836038571"
## [1] "Male first author team size 2018 geometric mean: 2.11856722029779"

## Warning in wilcox.test.default(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 = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.66112628872142"
## [1] "Male last author team size 2018 geometric mean: 1.75213146865393"

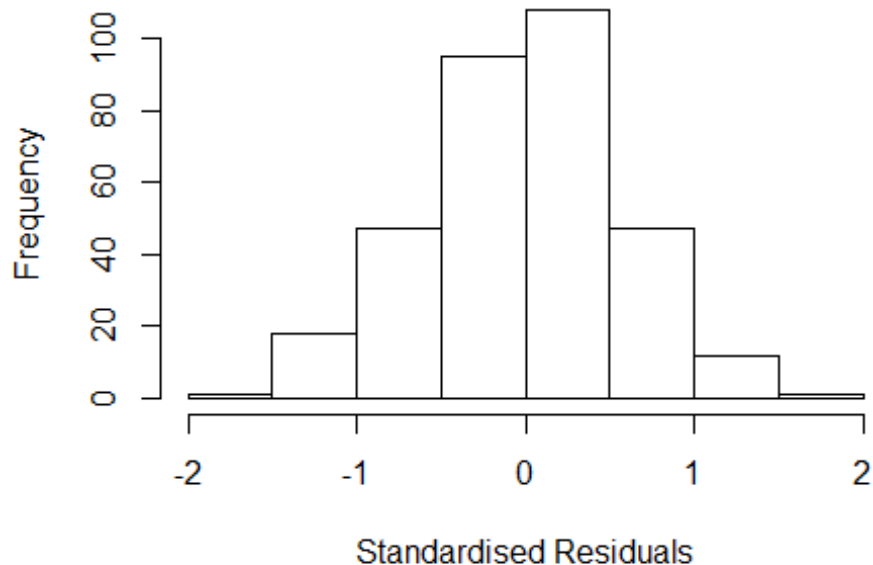
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



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

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	2.280	1	1.510
LastAuthorFemale	2.410	1	1.552
UniqueAuthors	3.175	4	1.155
Year	3.540	16	1.040

Residuals from first and last author and team size



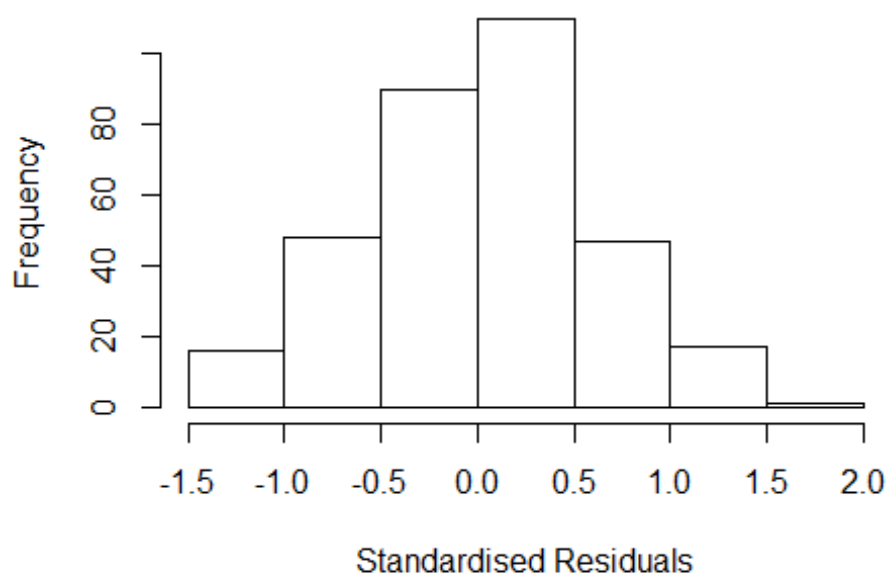
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6504 -0.3675 0.0155 0.4044 1.6189
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.68332 0.14337 4.77 2.9e-06 ***
## FirstAuthorFemale1 0.00514 0.09806 0.05 0.95822
## LastAuthorFemale1 -0.05080 0.09886 -0.51 0.60769
## UniqueAuthors2 0.24507 0.08980 2.73 0.00672 **
## UniqueAuthors3 0.29329 0.09763 3.00 0.00288 **
## UniqueAuthors4 0.30070 0.13102 2.30 0.02240 *
## UniqueAuthors5 -0.22316 0.17687 -1.26 0.20800
## Year1997 0.08309 0.22960 0.36 0.71770
## Year1998 0.40962 0.18299 2.24 0.02591 *
## Year1999 0.20109 0.20730 0.97 0.33279
```

```

## Year2000      0.03501      0.22047      0.16      0.87393
## Year2001      0.86382      0.23179      3.73      0.00023 ***
## Year2002      0.72199      0.39674      1.82      0.06977 .
## Year2003      0.55917      0.21412      2.61      0.00946 **
## Year2004      0.34416      0.19348      1.78      0.07627 .
## Year2005      0.53924      0.20441      2.64      0.00877 **
## Year2006      0.38528      0.16517      2.33      0.02032 *
## Year2007      0.87084      0.22312      3.90      0.00012 ***
## Year2008      0.42477      0.19927      2.13      0.03383 *
## Year2009      0.40440      0.18180      2.22      0.02685 *
## Year2010      0.61181      0.16351      3.74      0.00022 ***
## Year2011      0.50624      0.17742      2.85      0.00462 **
## Year2012      0.43119      0.16754      2.57      0.01054 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.589
## Multiple R-squared:  0.168, Adjusted R-squared:  0.108
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 316 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.412  0.881  0.959  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          3.04e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.177 1      1.475
## LastAuthorFemale  2.092 1      1.446
## Year              1.396 16      1.010

```

Residuals from first and last author



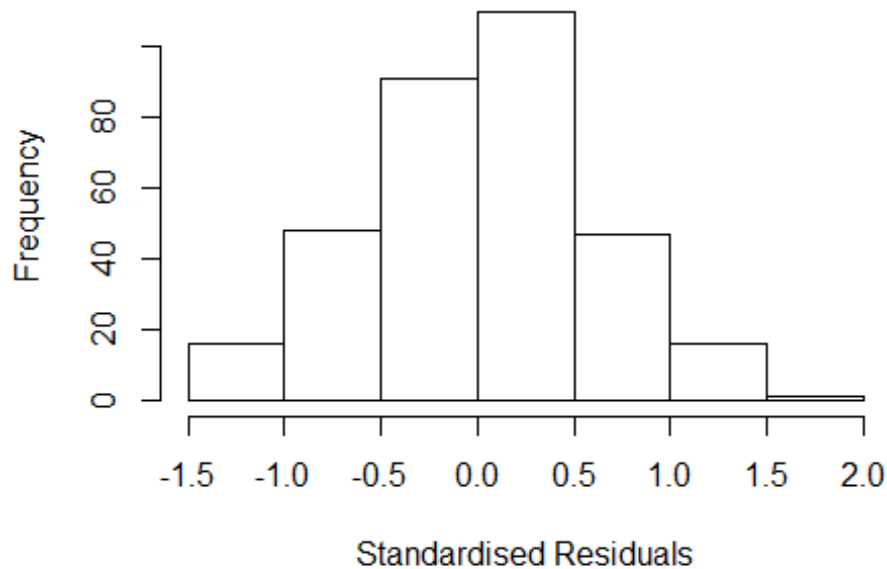
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4451 -0.3841 0.0217 0.4061 1.5656
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7656 0.1375 5.57 5.6e-08 ***
## FirstAuthorFemale1 0.0360 0.1002 0.36 0.71989
## LastAuthorFemale1 -0.0605 0.0981 -0.62 0.53821
## Year1997 0.0579 0.2428 0.24 0.81174
## Year1998 0.4308 0.1920 2.24 0.02557 *
## Year1999 0.1973 0.2218 0.89 0.37439
## Year2000 0.0755 0.2110 0.36 0.72067
## Year2001 0.8807 0.2246 3.92 0.00011 ***
## Year2002 0.6796 0.3437 1.98 0.04894 *
## Year2003 0.5263 0.2062 2.55 0.01119 *
## Year2004 0.3596 0.1802 2.00 0.04686 *
## Year2005 0.5149 0.2116 2.43 0.01552 *
```

```

## Year2006          0.3913      0.1597      2.45  0.01484 *
## Year2007          0.8666      0.2264      3.83  0.00016 ***
## Year2008          0.5329      0.1905      2.80  0.00548 **
## Year2009          0.4555      0.1782      2.56  0.01108 *
## Year2010          0.6810      0.1631      4.18  3.9e-05 ***
## Year2011          0.5647      0.1812      3.12  0.00200 **
## Year2012          0.4735      0.1683      2.81  0.00521 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.126, Adjusted R-squared:  0.0751
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 298 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.470  0.879  0.953  0.908  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.04e-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 1.183 1          1.088
## Year              1.183 16          1.005

```

Residuals from first author



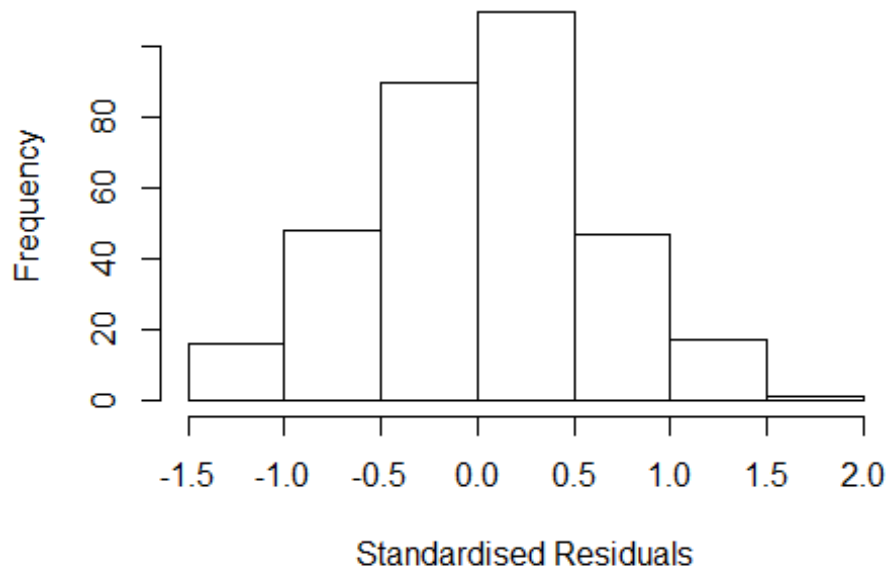
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4427 -0.3868 0.0342 0.4092 1.6113
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.76078 0.13713 5.55 6.2e-08 ***
## FirstAuthorFemale1 -0.00545 0.07411 -0.07 0.94146
## Year1997 0.05349 0.24278 0.22 0.82576
## Year1998 0.43209 0.19252 2.24 0.02551 *
## Year1999 0.19528 0.21982 0.89 0.37503
## Year2000 0.08922 0.20741 0.43 0.66739
## Year2001 0.88038 0.22796 3.86 0.00014 ***
## Year2002 0.68188 0.34735 1.96 0.05052 .
## Year2003 0.52340 0.20705 2.53 0.01197 *
## Year2004 0.36564 0.18098 2.02 0.04421 *
## Year2005 0.51540 0.21221 2.43 0.01572 *
## Year2006 0.38474 0.15912 2.42 0.01619 *
```

```

## Year2007          0.86737      0.22739      3.81  0.00016 ***
## Year2008          0.53568      0.19018      2.82  0.00516 **
## Year2009          0.45166      0.17871      2.53  0.01199 *
## Year2010          0.67507      0.16161      4.18  3.8e-05 ***
## Year2011          0.55704      0.17984      3.10  0.00213 **
## Year2012          0.47446      0.16815      2.82  0.00508 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.593
## Multiple R-squared:  0.125, Adjusted R-squared:  0.077
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 297 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.440  0.874  0.952  0.907  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.04e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.14 1      1.068
## Year              1.14 16      1.004

```


Residuals from last author



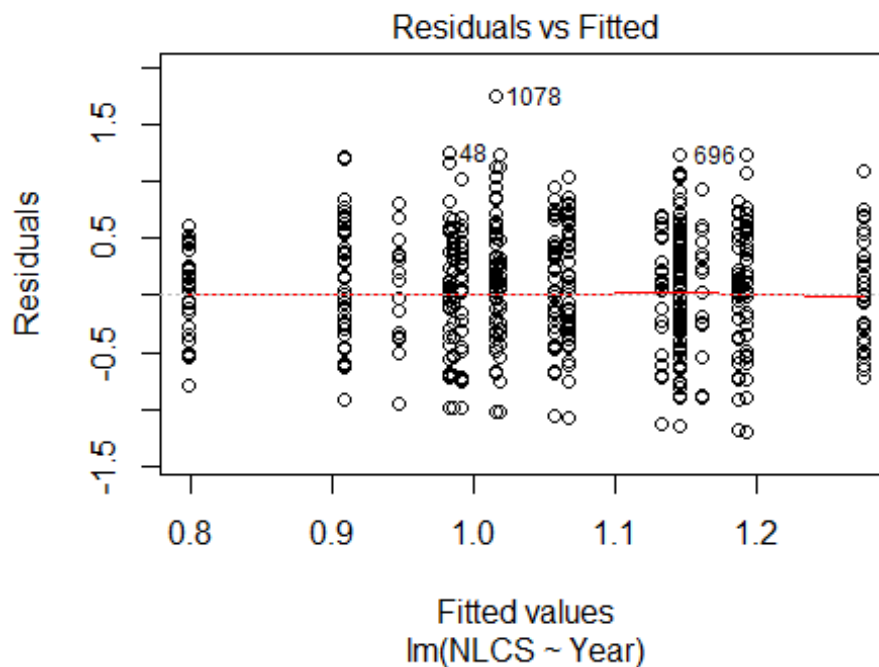
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4562 -0.3872 0.0337 0.4015 1.5977
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7685 0.1374 5.59 4.9e-08 ***
## LastAuthorFemale1 -0.0361 0.0723 -0.50 0.61794
## Year1997 0.0605 0.2441 0.25 0.80455
## Year1998 0.4359 0.1921 2.27 0.02394 *
## Year1999 0.1948 0.2213 0.88 0.37931
## Year2000 0.0881 0.2070 0.43 0.67079
## Year2001 0.8840 0.2269 3.90 0.00012 ***
## Year2002 0.6876 0.3414 2.01 0.04484 *
## Year2003 0.5218 0.2054 2.54 0.01155 *
## Year2004 0.3651 0.1805 2.02 0.04400 *
## Year2005 0.5157 0.2126 2.43 0.01585 *
## Year2006 0.3917 0.1599 2.45 0.01483 *
```

```

## Year2007          0.8662      0.2258      3.84  0.00015 ***
## Year2008          0.5384      0.1880      2.86  0.00448 **
## Year2009          0.4538      0.1778      2.55  0.01117 *
## Year2010          0.6824      0.1630      4.19  3.7e-05 ***
## Year2011          0.5616      0.1811      3.10  0.00210 **
## Year2012          0.4738      0.1683      2.82  0.00518 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.125, Adjusted R-squared:  0.0776
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 302 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.452  0.881  0.954  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      3.04e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 329"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
##   35   27   37   24   31   41   31   26   40   26   63   63   54   66   74
## 2011 2012
##   77   67
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   30   21   30   17   25   31   28   23   35   19   56   51   45   55   49
## 2011 2012

```

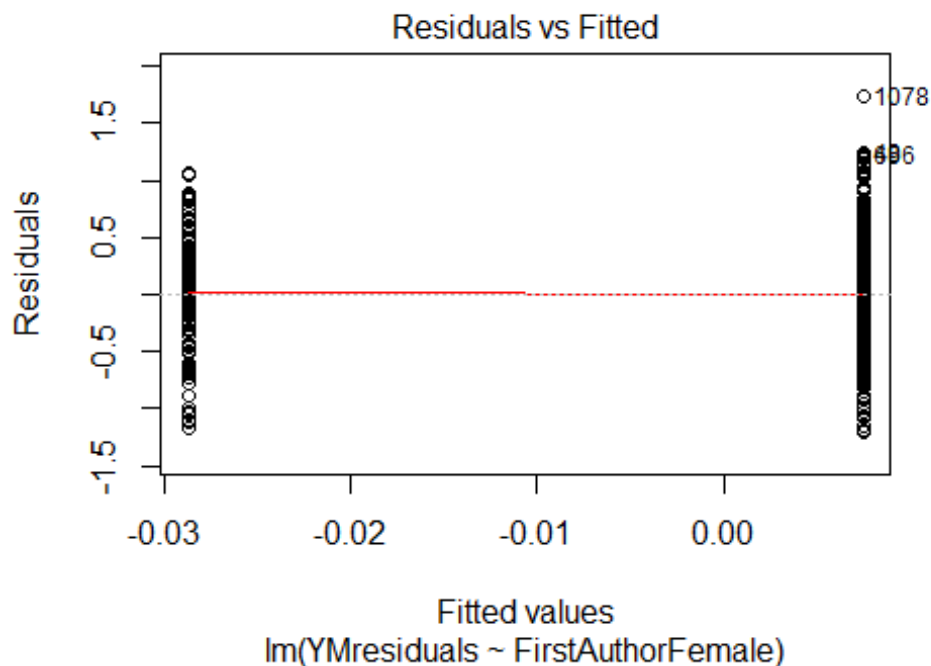
```
##      64      44
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    28    20    30    15    23    29    27    22    32    19    54    44    45    51    42
## 2011 2012
##    51    39
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 13, df = 16, p-value = 0.7
```



```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.34, df = 1, p-value = 0.6
## [1] "Female first author team size 2018 geometric mean: 1.8949771719079"
## [1] "Male first author team size 2018 geometric mean: 1.80319152958882"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

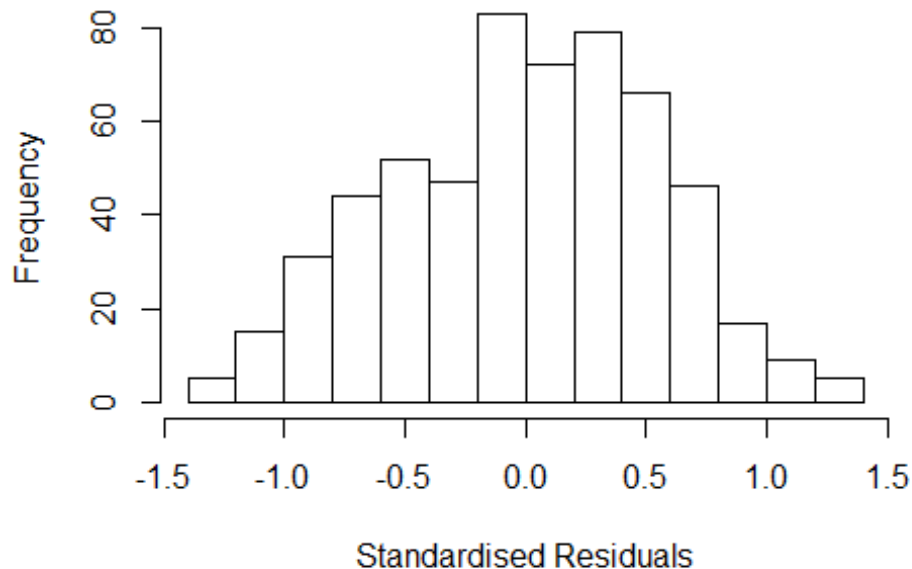
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 320, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.84908039834368"
## [1] "Male last author team size 2018 geometric mean: 1.82454318422679"

## Warning in wilcox.test.default(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 = 0.9
## 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 1.429  1      1.196
## LastAuthorFemale  1.429  1      1.196
## UniqueAuthors    1.430  4      1.046
## Year              1.637 16      1.016
```

Residuals from first and last author and team size



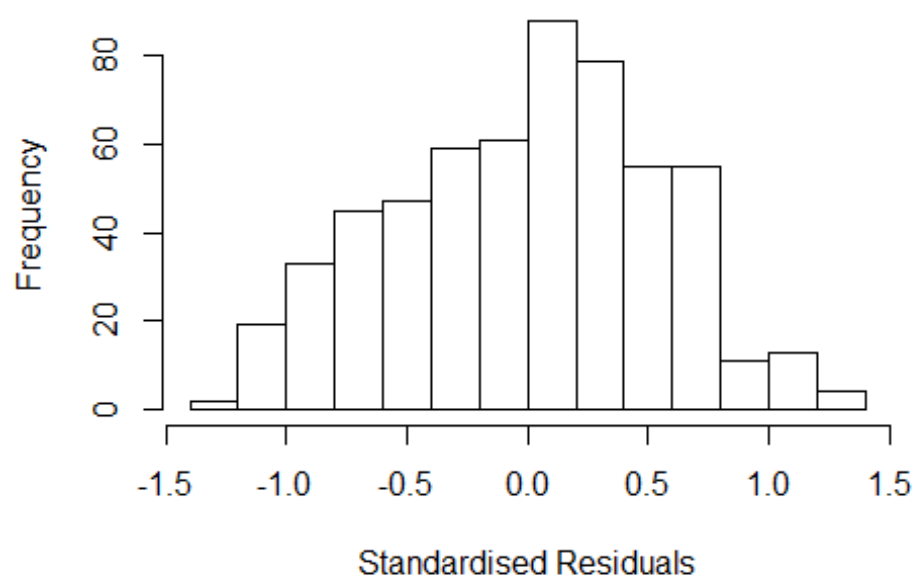
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.331 -0.423 0.045 0.401 1.329
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90448 0.12555 7.20 1.9e-12 ***
## FirstAuthorFemale1 0.01741 0.06788 0.26 0.7977
## LastAuthorFemale1 -0.05593 0.06515 -0.86 0.3910
## UniqueAuthors2 0.17487 0.05377 3.25 0.0012 **
## UniqueAuthors3 0.28185 0.07515 3.75 0.0002 ***
## UniqueAuthors4 0.11426 0.22520 0.51 0.6121
## UniqueAuthors5 0.06675 0.21412 0.31 0.7554
## Year1997 0.00753 0.17662 0.04 0.9660
## Year1998 -0.14102 0.14331 -0.98 0.3255
## Year1999 0.13553 0.16887 0.80 0.4226
```

```

## Year2000      0.18340      0.16683      1.10      0.2721
## Year2001      0.23494      0.15097      1.56      0.1202
## Year2002     -0.00842      0.16072     -0.05      0.9582
## Year2003      0.01134      0.17176      0.07      0.9474
## Year2004      0.26148      0.14838      1.76      0.0786 .
## Year2005     -0.08109      0.17261     -0.47      0.6387
## Year2006      0.14492      0.15196      0.95      0.3407
## Year2007     -0.09152      0.16144     -0.57      0.5710
## Year2008      0.22645      0.15486      1.46      0.1442
## Year2009      0.10966      0.15202      0.72      0.4710
## Year2010      0.12862      0.15307      0.84      0.4011
## Year2011     -0.01095      0.14971     -0.07      0.9417
## Year2012      0.04085      0.16765      0.24      0.8076
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.583
## Multiple R-squared:  0.0763, Adjusted R-squared:  0.0393
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 526 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.581  0.884  0.950  0.919  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.75e-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 1.476 1      1.215
## LastAuthorFemale  1.427 1      1.195
## Year              1.202 16      1.006

```

Residuals from first and last author



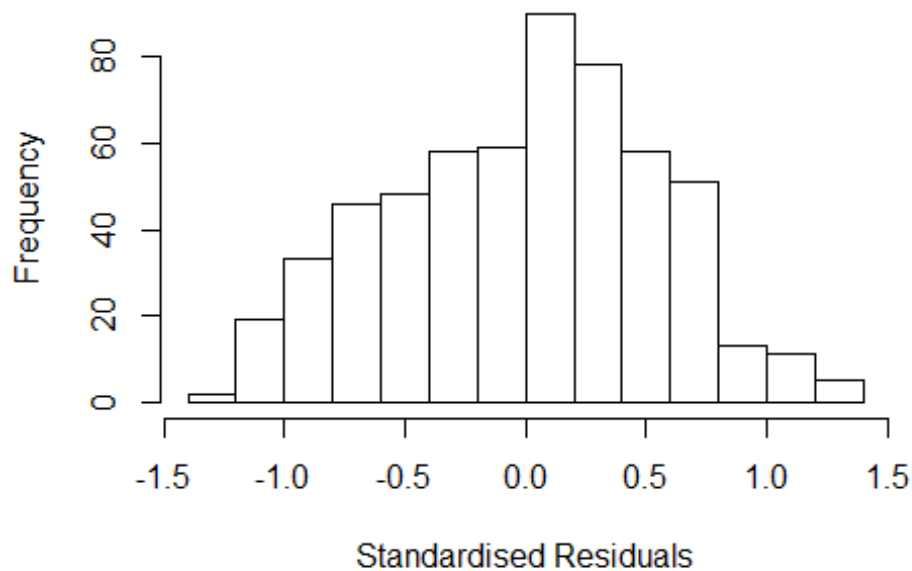
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2418 -0.4128 0.0428 0.3922 1.2736
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95444 0.12157 7.85 2.2e-14 ***
## FirstAuthorFemale1 -0.00141 0.06942 -0.02 0.984
## LastAuthorFemale1 -0.03574 0.06604 -0.54 0.589
## Year1997 0.03776 0.17221 0.22 0.827
## Year1998 -0.14723 0.14011 -1.05 0.294
## Year1999 0.11564 0.17390 0.67 0.506
## Year2000 0.21825 0.16763 1.30 0.193
## Year2001 0.28732 0.15160 1.90 0.059 .
## Year2002 0.03347 0.15878 0.21 0.833
## Year2003 0.04532 0.16833 0.27 0.788
## Year2004 0.31483 0.14935 2.11 0.035 *
## Year2005 0.00782 0.17158 0.05 0.964
```

```

## Year2006          0.20065      0.14659      1.37      0.172
## Year2007         -0.02091      0.15840     -0.13      0.895
## Year2008          0.25390      0.15404      1.65      0.100 .
## Year2009          0.17389      0.14807      1.17      0.241
## Year2010          0.20185      0.15315      1.32      0.188
## Year2011          0.03990      0.14994      0.27      0.790
## Year2012          0.10276      0.16265      0.63      0.528
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.59
## Multiple R-squared:  0.0441, Adjusted R-squared:  0.0129
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 521 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.621  0.885   0.948   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      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 1.122 1      1.059
## Year              1.122 16      1.004

```


Residuals from first author



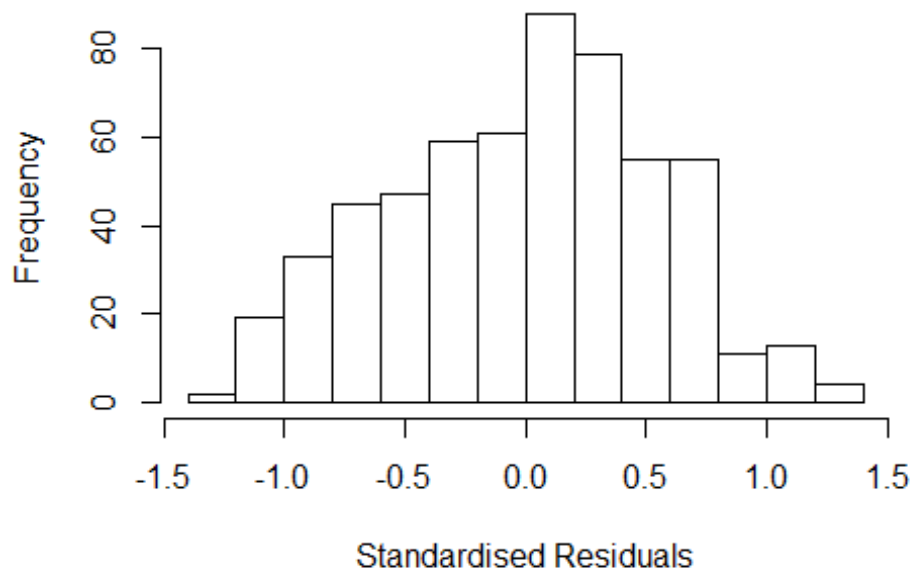
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2388 -0.4160 0.0473 0.3904 1.2795
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9485 0.1198 7.92 1.3e-14 ***
## FirstAuthorFemale1 -0.0209 0.0607 -0.35 0.730
## Year1997 0.0435 0.1708 0.25 0.799
## Year1998 -0.1443 0.1391 -1.04 0.300
## Year1999 0.1208 0.1730 0.70 0.485
## Year2000 0.2163 0.1676 1.29 0.197
## Year2001 0.2903 0.1507 1.93 0.055 .
## Year2002 0.0377 0.1572 0.24 0.811
## Year2003 0.0487 0.1671 0.29 0.771
## Year2004 0.3194 0.1486 2.15 0.032 *
## Year2005 0.0107 0.1711 0.06 0.950
## Year2006 0.2006 0.1459 1.38 0.170
```

```

## Year2007          -0.0231      0.1587   -0.15    0.884
## Year2008          0.2566      0.1532    1.67    0.095 .
## Year2009          0.1747      0.1476    1.18    0.237
## Year2010          0.2052      0.1523    1.35    0.178
## Year2011          0.0374      0.1496    0.25    0.803
## Year2012          0.1033      0.1621    0.64    0.524
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.589
## Multiple R-squared:  0.0437, Adjusted R-squared:  0.0143
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 523 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.616  0.883   0.948   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      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 1.083 1          1.041
## Year            1.083 16          1.002

```

Residuals from last author



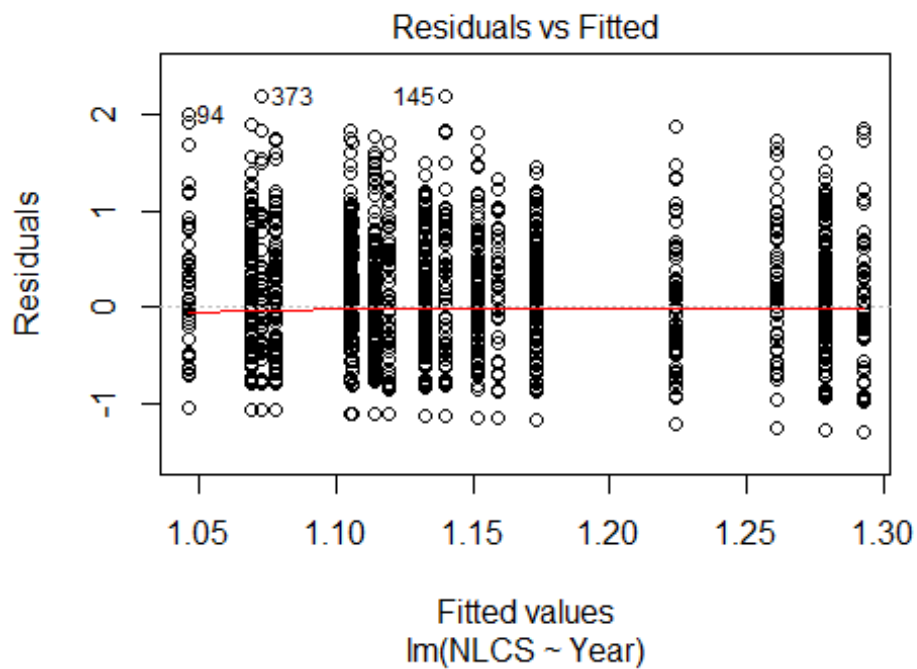
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2415 -0.4128 0.0432 0.3923 1.2735
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95445 0.12154 7.85 2.1e-14 ***
## LastAuthorFemale1 -0.03647 0.05752 -0.63 0.526
## Year1997 0.03760 0.17208 0.22 0.827
## Year1998 -0.14731 0.14000 -1.05 0.293
## Year1999 0.11543 0.17365 0.66 0.506
## Year2000 0.21830 0.16735 1.30 0.193
## Year2001 0.28700 0.15010 1.91 0.056 .
## Year2002 0.03327 0.15840 0.21 0.834
## Year2003 0.04511 0.16765 0.27 0.788
## Year2004 0.31464 0.14917 2.11 0.035 *
## Year2005 0.00768 0.17144 0.04 0.964
## Year2006 0.20048 0.14613 1.37 0.171
```

```

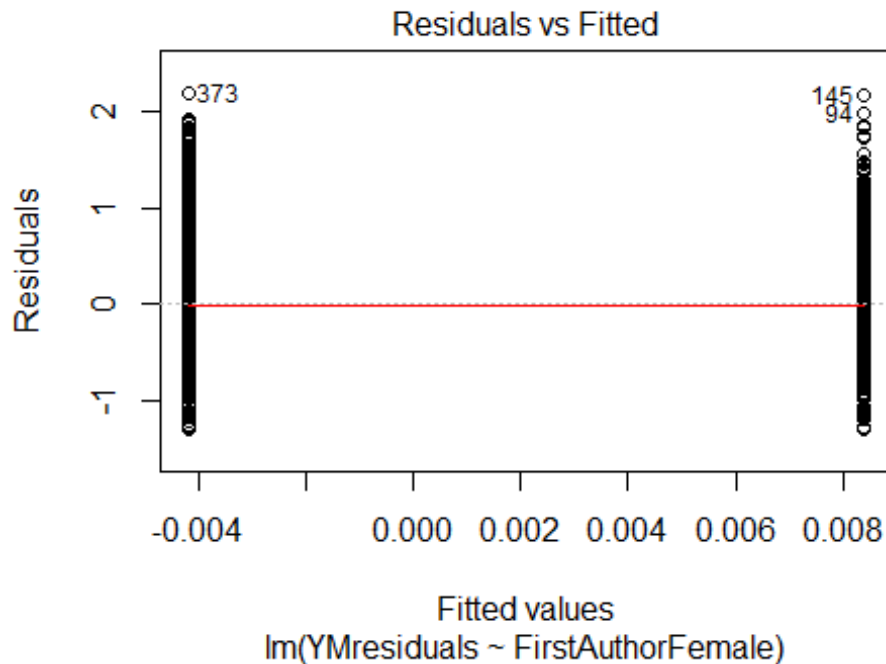
## Year2007          -0.02101      0.15825    -0.13      0.894
## Year2008           0.25382      0.15393      1.65      0.100 .
## Year2009           0.17368      0.14761      1.18      0.240
## Year2010           0.20169      0.15297      1.32      0.188
## Year2011           0.03981      0.14985      0.27      0.791
## Year2012           0.10237      0.16097      0.64      0.525
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.591
## Multiple R-squared:  0.0441, Adjusted R-squared:  0.0147
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 521 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.621  0.885   0.949   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      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: 571"
## [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
##   69   91   78   76   64   82   87   82   93  122  153  170  223  223  243
## 2011 2012
##  189  202
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   62   69   59   51   53   61   66   58   77  102  130  143  179  194  194
## 2011 2012

```

```
## 158 160
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 60 63 57 46 49 58 63 54 71 91 122 129 166 177 181
## 2011 2012
## 136 139
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 18, df = 16, p-value = 0.3
```

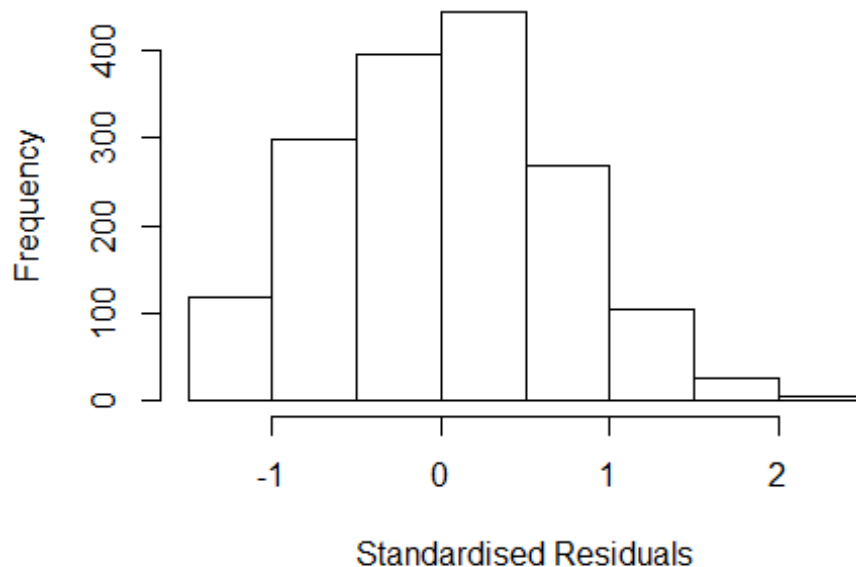


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 2.17248030854999"
## [1] "Male first author team size 2018 geometric mean: 1.9033825029995"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.0240043697194"
## [1] "Male last author team size 2018 geometric mean: 1.99527960058832"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, 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 1.392 1          1.180
## LastAuthorFemale  1.393 1          1.180
## UniqueAuthors    1.201 4          1.023
## Year              1.194 16         1.006
```

Residuals from first and last author and team size



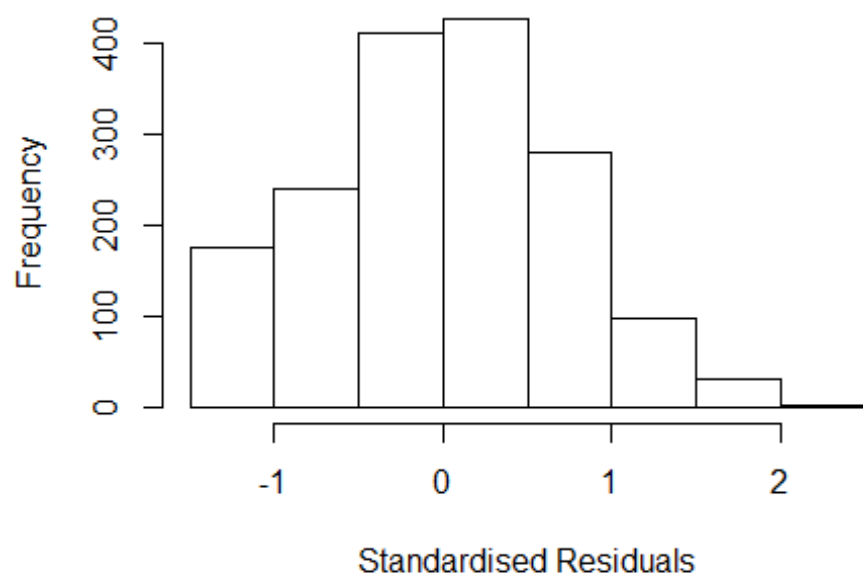
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3659 -0.5014 0.0141 0.4835 2.3767
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.915830 0.107077 8.55 < 2e-16 ***
## FirstAuthorFemale1 -0.031968 0.042791 -0.75 0.4551
## LastAuthorFemale1 0.092001 0.043254 2.13 0.0336 *
## UniqueAuthors2 0.261244 0.041742 6.26 4.9e-10 ***
## UniqueAuthors3 0.220335 0.051174 4.31 1.8e-05 ***
## UniqueAuthors4 0.165635 0.086476 1.92 0.0556 .
## UniqueAuthors5 0.299508 0.094665 3.16 0.0016 **
## Year1997 0.033769 0.154965 0.22 0.8275
## Year1998 0.216540 0.138282 1.57 0.1176
## Year1999 -0.033486 0.147669 -0.23 0.8206
```

```

## Year2000      0.056668    0.143427    0.40    0.6928
## Year2001      0.155200    0.152548    1.02    0.3091
## Year2002      0.086197    0.135567    0.64    0.5250
## Year2003      0.091980    0.148548    0.62    0.5359
## Year2004     -0.033056    0.139297   -0.24    0.8124
## Year2005      0.000616    0.133431    0.00    0.9963
## Year2006      0.001028    0.124100    0.01    0.9934
## Year2007      0.037081    0.121566    0.31    0.7604
## Year2008      0.087109    0.118225    0.74    0.4613
## Year2009      0.060565    0.116343    0.52    0.6027
## Year2010     -0.011446    0.119753   -0.10    0.9239
## Year2011      0.188821    0.120854    1.56    0.1184
## Year2012     -0.011251    0.118357   -0.10    0.9243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.704
## Multiple R-squared:  0.0399, Adjusted R-squared:  0.0271
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 145 weights are ~= 1. The remaining 1517 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.231  0.862  0.948  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      6.02e-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 1.341 1      1.158
## LastAuthorFemale  1.334 1      1.155
## Year              1.029 16      1.001

```


Residuals from first and last author



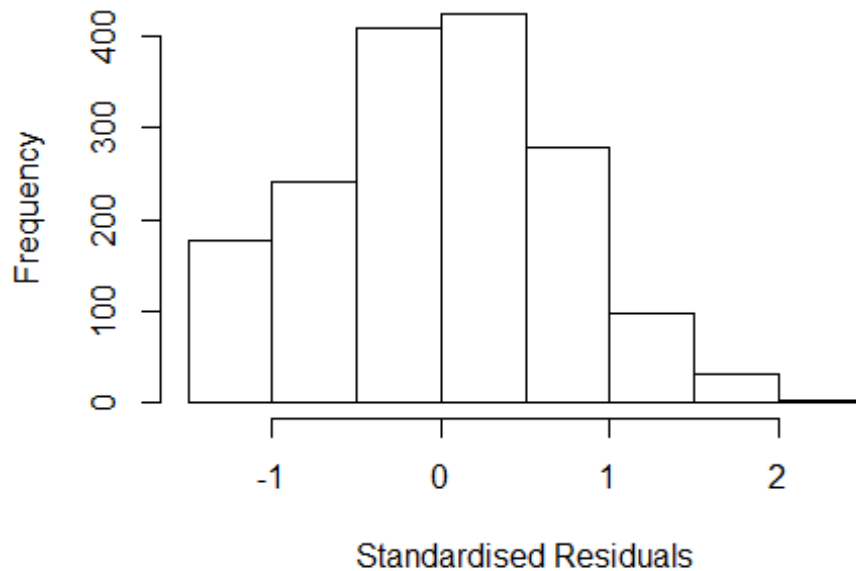
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2933 -0.4982 0.0029 0.4948 2.2861
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0055 0.1102 9.12 <2e-16 ***
## FirstAuthorFemale1 -0.0108 0.0429 -0.25 0.801
## LastAuthorFemale1 0.0832 0.0433 1.92 0.055 .
## Year1997 0.0184 0.1565 0.12 0.906
## Year1998 0.2186 0.1433 1.53 0.127
## Year1999 -0.0325 0.1489 -0.22 0.827
## Year2000 0.0710 0.1470 0.48 0.629
## Year2001 0.1855 0.1560 1.19 0.234
## Year2002 0.1233 0.1385 0.89 0.373
## Year2003 0.1221 0.1541 0.79 0.428
## Year2004 0.0115 0.1439 0.08 0.936
## Year2005 0.0535 0.1389 0.39 0.700
```

```

## Year2006          0.0364      0.1282      0.28      0.777
## Year2007          0.0853      0.1242      0.69      0.492
## Year2008          0.1400      0.1219      1.15      0.251
## Year2009          0.0908      0.1202      0.76      0.450
## Year2010          0.0585      0.1228      0.48      0.634
## Year2011          0.2154      0.1251      1.72      0.085 .
## Year2012          0.0435      0.1229      0.35      0.723
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.729
## Multiple R-squared:  0.0109, Adjusted R-squared:  0.000113
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 130 weights are ~= 1. The remaining 1532 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.304  0.871  0.950  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      6.02e-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 1.016 1      1.008
## Year              1.016 16      1.000

```

Residuals from first author



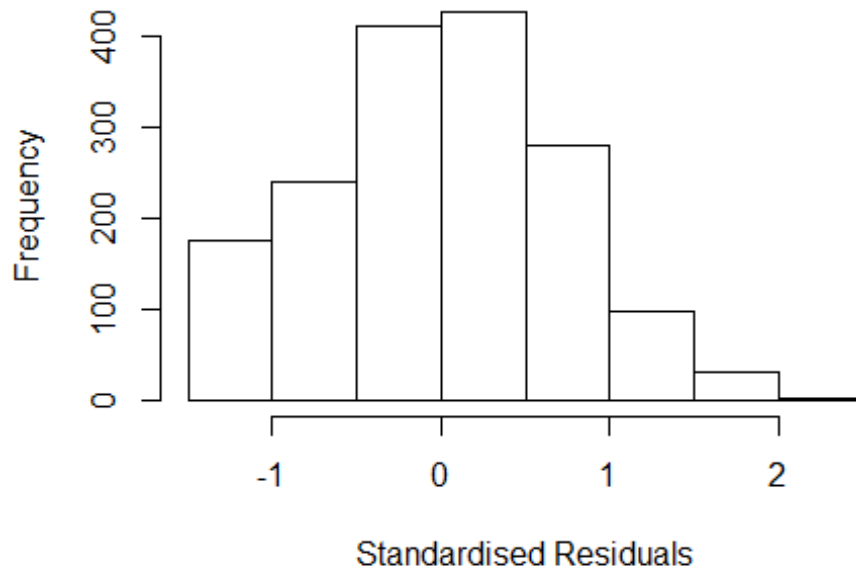
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26522 -0.50198 0.00416 0.49104 2.28122
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01562 0.10988 9.24 <2e-16 ***
## FirstAuthorFemale1 0.02980 0.03735 0.80 0.425
## Year1997 0.02838 0.15626 0.18 0.856
## Year1998 0.21843 0.14341 1.52 0.128
## Year1999 -0.03784 0.14851 -0.25 0.799
## Year2000 0.07715 0.14799 0.52 0.602
## Year2001 0.18026 0.15629 1.15 0.249
## Year2002 0.12754 0.13881 0.92 0.358
## Year2003 0.12127 0.15478 0.78 0.433
## Year2004 0.00985 0.14360 0.07 0.945
## Year2005 0.05791 0.13935 0.42 0.678
## Year2006 0.03234 0.12815 0.25 0.801
```

```

## Year2007          0.08597    0.12398    0.69    0.488
## Year2008          0.14218    0.12176    1.17    0.243
## Year2009          0.09531    0.12005    0.79    0.427
## Year2010          0.06239    0.12290    0.51    0.612
## Year2011          0.21980    0.12509    1.76    0.079 .
## Year2012          0.04605    0.12282    0.37    0.708
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.729
## Multiple R-squared:  0.00877,    Adjusted R-squared:  -0.00148
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1533 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.307  0.871  0.950  0.916  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.02e-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 1.01 1          1.005
## Year              1.01 16          1.000

```

Residuals from last author



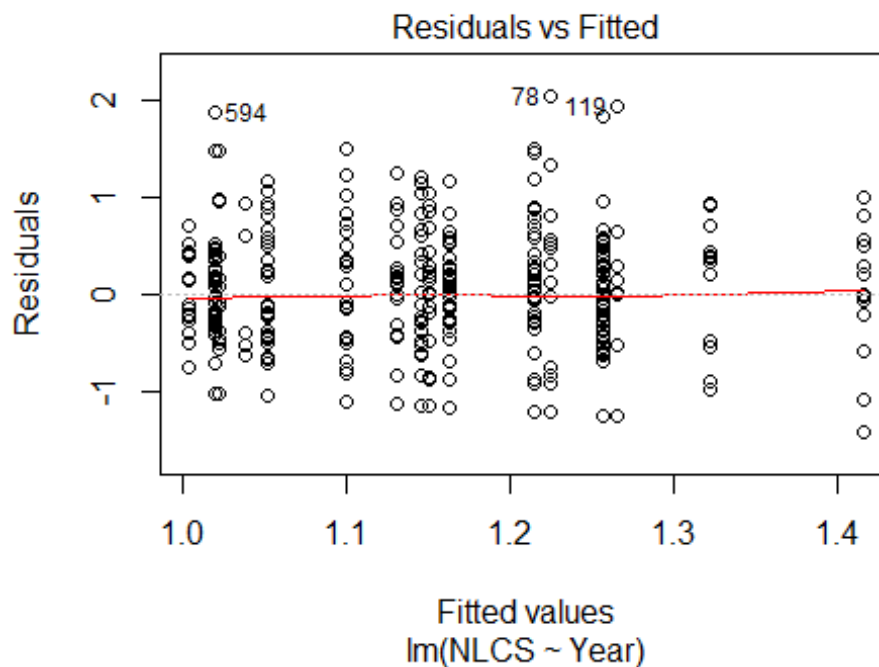
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29683 -0.49886 0.00262 0.49286 2.28837
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0035 0.1101 9.12 <2e-16 ***
## LastAuthorFemale1 0.0777 0.0377 2.06 0.040 *
## Year1997 0.0187 0.1565 0.12 0.905
## Year1998 0.2183 0.1434 1.52 0.128
## Year1999 -0.0328 0.1488 -0.22 0.825
## Year2000 0.0718 0.1472 0.49 0.626
## Year2001 0.1856 0.1560 1.19 0.234
## Year2002 0.1236 0.1386 0.89 0.373
## Year2003 0.1219 0.1543 0.79 0.430
## Year2004 0.0115 0.1439 0.08 0.936
## Year2005 0.0539 0.1390 0.39 0.698
## Year2006 0.0359 0.1282 0.28 0.779
```

```

## Year2007          0.0856      0.1242      0.69      0.491
## Year2008          0.1395      0.1219      1.14      0.253
## Year2009          0.0911      0.1203      0.76      0.449
## Year2010          0.0582      0.1228      0.47      0.636
## Year2011          0.2157      0.1252      1.72      0.085 .
## Year2012          0.0430      0.1229      0.35      0.727
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.728
## Multiple R-squared:  0.0109, Adjusted R-squared:  0.000686
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 126 weights are ~= 1. The remaining 1536 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.302  0.872  0.950  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      6.02e-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: 1662"
## [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
##   31   29   26   16   15   25   23   22   22   34   34   53   38   48   40
## 2011 2012
##   41   44
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   20   15   15    9    5   14   16   15   18   25   28   35   26   37   33
## 2011 2012

```

```
## 21 31
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 19 13 15 7 4 14 15 12 16 23 22 31 23 31 29
## 2011 2012
## 19 26
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 31, df = 16, p-value = 0.01
```



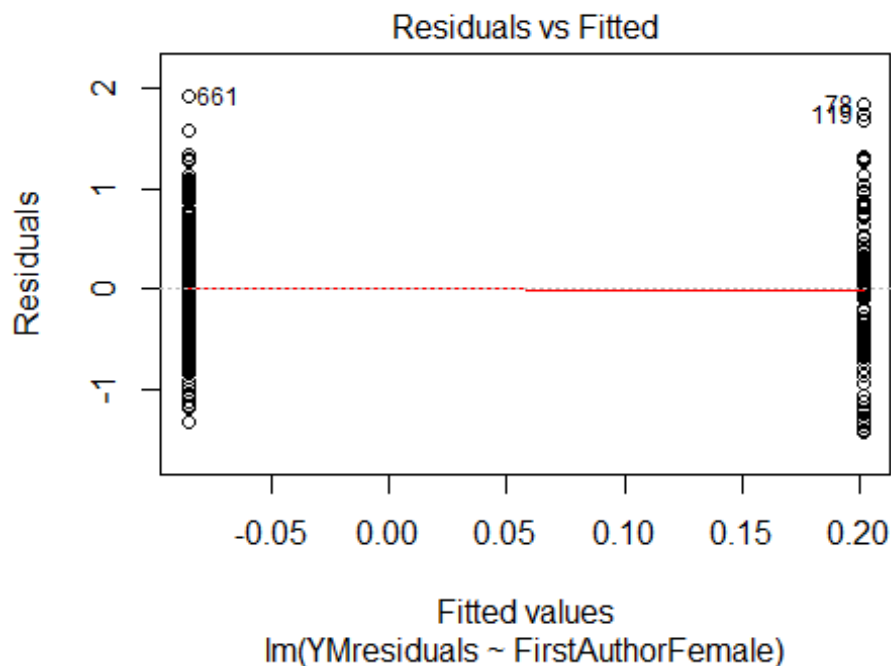
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.1, df = 1, p-value = 0.04

## [1] "Female first author team size 2018 geometric mean: 4.21716332650875"
## [1] "Male first author team size 2018 geometric mean: 2.02153015930144"

## Warning in wilcox.test.default(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 = 0.04
## 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: 2.32634909917539"

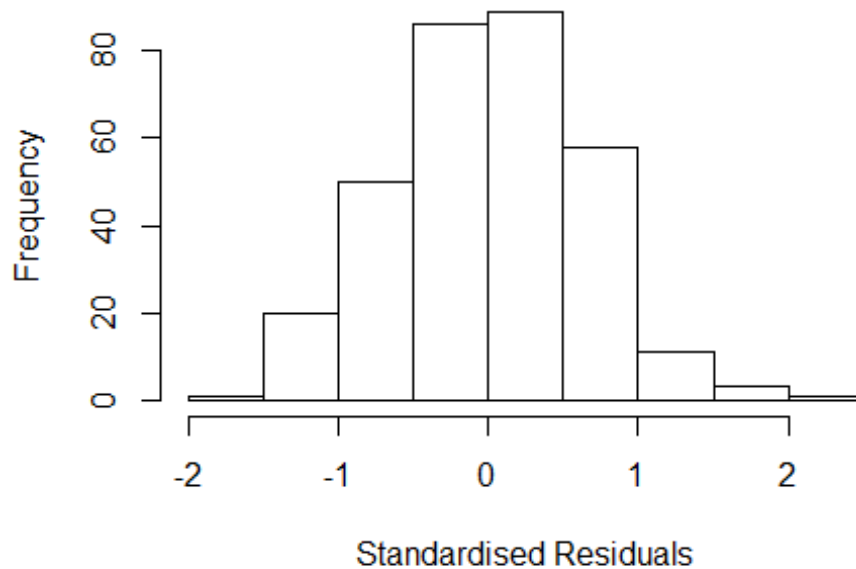
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16, p-value = 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	2.286	1	1.512
LastAuthorFemale	2.334	1	1.528
UniqueAuthors	2.380	4	1.115
Year	3.314	16	1.038

Residuals from first and last author and team size



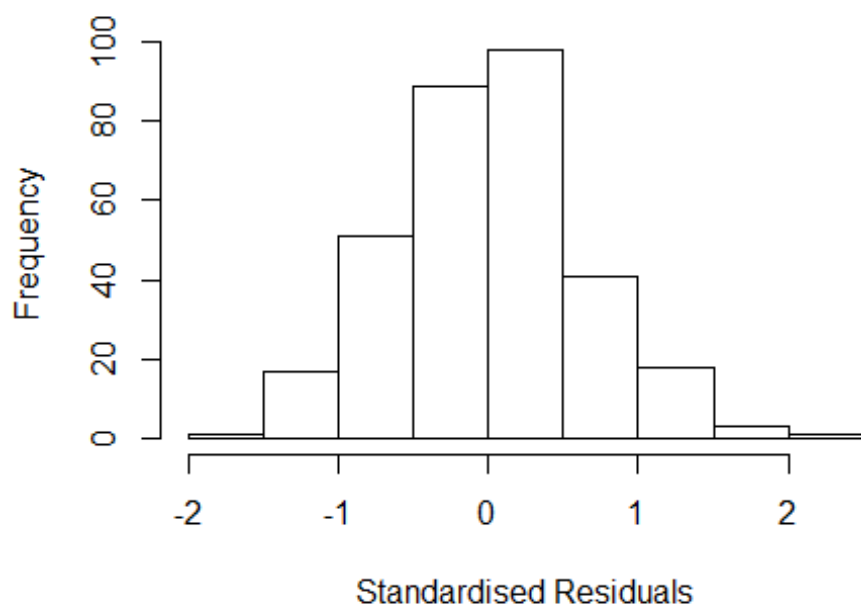
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7052 -0.4447 0.0137 0.4270 2.1626
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.82811 0.16489 5.02 8.8e-07 ***
## FirstAuthorFemale1 0.30984 0.12294 2.52 0.0122 *
## LastAuthorFemale1 -0.03284 0.12029 -0.27 0.7850
## UniqueAuthors2 0.25801 0.08713 2.96 0.0033 **
## UniqueAuthors3 0.36371 0.12118 3.00 0.0029 **
## UniqueAuthors4 0.43477 0.18565 2.34 0.0199 *
## UniqueAuthors5 0.26990 0.29538 0.91 0.3616
## Year1997 -0.01695 0.31872 -0.05 0.9576
## Year1998 0.36270 0.24906 1.46 0.1464
## Year1999 0.06553 0.41877 0.16 0.8758
```

```

## Year2000      0.15275    0.48115    0.32    0.7511
## Year2001      0.29632    0.24999    1.19    0.2368
## Year2002     -0.05942    0.19444   -0.31    0.7601
## Year2003     -0.11596    0.28047   -0.41    0.6796
## Year2004      0.05466    0.23902    0.23    0.8193
## Year2005      0.16891    0.18096    0.93    0.3514
## Year2006     -0.12344    0.22640   -0.55    0.5860
## Year2007      0.04730    0.19173    0.25    0.8053
## Year2008      0.00627    0.22841    0.03    0.9781
## Year2009      0.30920    0.20070    1.54    0.1245
## Year2010     -0.12053    0.19947   -0.60    0.5461
## Year2011      0.09930    0.22136    0.45    0.6541
## Year2012     -0.11207    0.19363   -0.58    0.5632
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.653
## Multiple R-squared:  0.115, Adjusted R-squared:  0.0493
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 299 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.250  0.887  0.955  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      3.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.153  1      1.467
## LastAuthorFemale  2.265  1      1.505
## Year              1.720 16      1.017

```

Residuals from first and last author



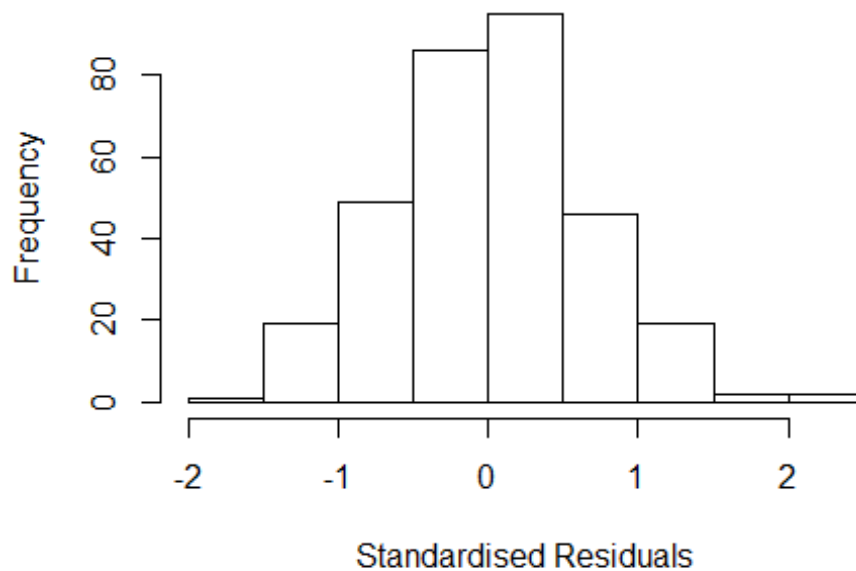
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5768 -0.4378 0.0177 0.4334 2.0408
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03804 0.16243 6.39 6.3e-10 ***
## FirstAuthorFemale1 0.27047 0.12141 2.23 0.027 *
## LastAuthorFemale1 -0.06609 0.12240 -0.54 0.590
## Year1997 -0.07727 0.38440 -0.20 0.841
## Year1998 0.35064 0.24883 1.41 0.160
## Year1999 -0.07350 0.41328 -0.18 0.859
## Year2000 0.04695 0.41512 0.11 0.910
## Year2001 0.23257 0.26217 0.89 0.376
## Year2002 -0.10014 0.20064 -0.50 0.618
## Year2003 -0.05802 0.28885 -0.20 0.841
## Year2004 0.02836 0.25916 0.11 0.913
## Year2005 0.20378 0.17940 1.14 0.257
```

```

## Year2006      -0.08573    0.23657   -0.36    0.717
## Year2007      0.05960    0.18943    0.31    0.753
## Year2008      0.00217    0.22379    0.01    0.992
## Year2009      0.26826    0.20534    1.31    0.192
## Year2010     -0.15512    0.19731   -0.79    0.432
## Year2011      0.18989    0.22010    0.86    0.389
## Year2012     -0.12916    0.20807   -0.62    0.535
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.656
## Multiple R-squared:  0.0702, Adjusted R-squared:  0.0144
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 299 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.313  0.872  0.956  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      3.13e-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 1.277 1      1.130
## Year              1.277 16      1.008

```

Residuals from first author



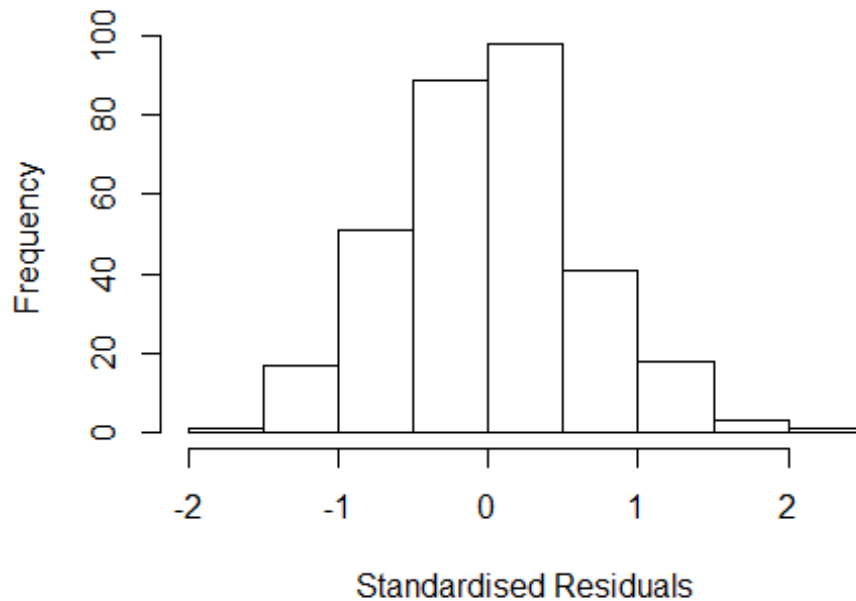
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5383 -0.4328 0.0212 0.4125 2.0872
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03381 0.16535 6.25 1.4e-09 ***
## FirstAuthorFemale1 0.23547 0.09418 2.50 0.013 *
## Year1997 -0.08444 0.38205 -0.22 0.825
## Year1998 0.33865 0.24752 1.37 0.172
## Year1999 -0.08649 0.40038 -0.22 0.829
## Year2000 0.04387 0.40785 0.11 0.914
## Year2001 0.22004 0.26534 0.83 0.408
## Year2002 -0.11142 0.20088 -0.55 0.580
## Year2003 -0.06721 0.28926 -0.23 0.816
## Year2004 0.02288 0.26012 0.09 0.930
## Year2005 0.20890 0.18152 1.15 0.251
## Year2006 -0.09480 0.23527 -0.40 0.687
```

```

## Year2007          0.05457    0.19027    0.29    0.774
## Year2008          0.00799    0.22252    0.04    0.971
## Year2009          0.26901    0.20675    1.30    0.194
## Year2010         -0.16150    0.19776   -0.82    0.415
## Year2011          0.17967    0.22068    0.81    0.416
## Year2012         -0.12900    0.20964   -0.62    0.539
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.663
## Multiple R-squared:  0.0687, Adjusted R-squared:  0.0161
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 296 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.301  0.873  0.955  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      3.13e-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 1.334 1          1.155
## Year            1.334 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
## -1.42225 -0.41150  0.00875  0.46091  2.32870
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0713     0.1607   6.67 1.3e-10 ***
## LastAuthorFemale1  0.0795     0.0954   0.83  0.41
## Year1997        -0.0987     0.3813  -0.26  0.80
## Year1998         0.3510     0.2442   1.44  0.15
## Year1999        -0.1930     0.3762  -0.51  0.61
## Year2000         0.0416     0.4425   0.09  0.93
## Year2001         0.2291     0.2735   0.84  0.40
## Year2002        -0.1314     0.1966  -0.67  0.50
## Year2003        -0.1025     0.2962  -0.35  0.73
## Year2004         0.0372     0.2580   0.14  0.89
## Year2005         0.1960     0.1780   1.10  0.27
## Year2006        -0.0730     0.2519  -0.29  0.77
```

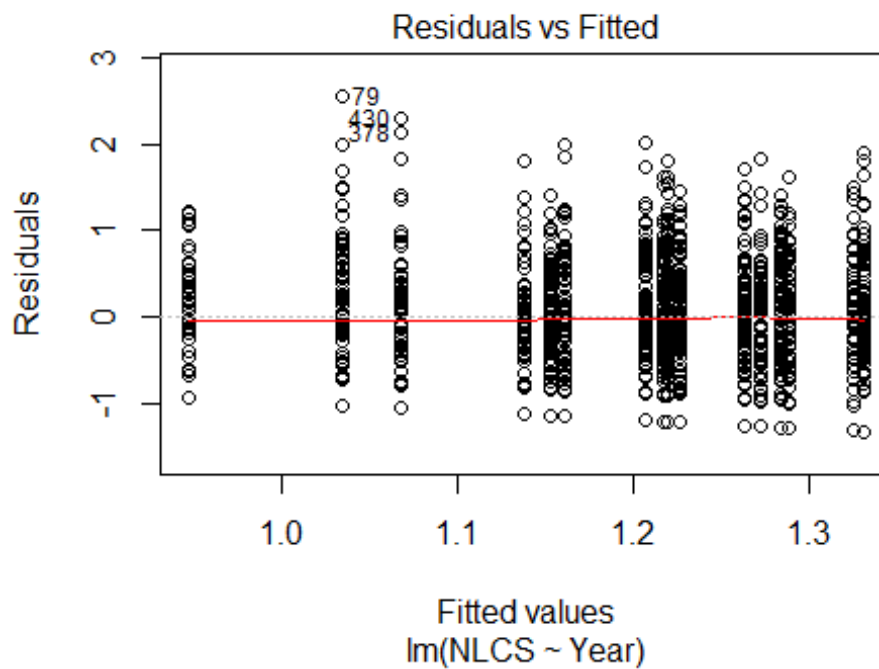
```

## Year2007          0.0887      0.1878      0.47      0.64
## Year2008          0.0413      0.2176      0.19      0.85
## Year2009          0.2610      0.2065      1.26      0.21
## Year2010         -0.1357      0.1985     -0.68      0.49
## Year2011          0.1466      0.2185      0.67      0.50
## Year2012         -0.0807      0.2109     -0.38      0.70
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.653
## Multiple R-squared:  0.0514, Adjusted R-squared:  -0.00216
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 298 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.178  0.871  0.954  0.907  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.13e-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: 319"
## [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
##   93   93   98   80   72  100  116   85   73   93  109  114  134  124  136
## 2011 2012
##  136  121
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   85   70   75   66   61   77   94   70   55   70   89   95  105  104  108
## 2011 2012

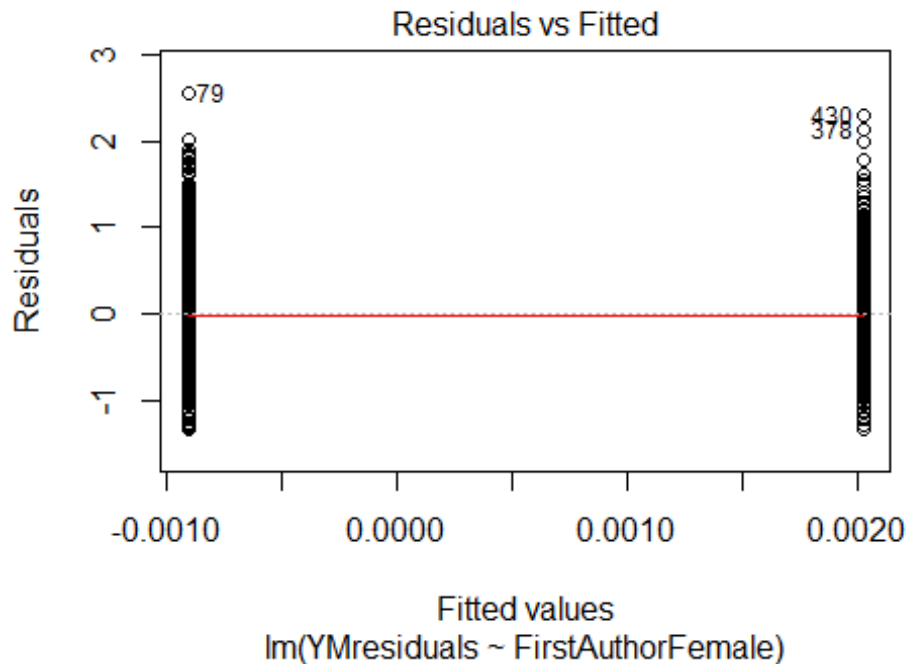
```



```
## 109 102
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 79 65 72 60 60 70 87 62 50 57 81 90 99 96 102
## 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 = 23, df = 16, p-value = 0.1
```

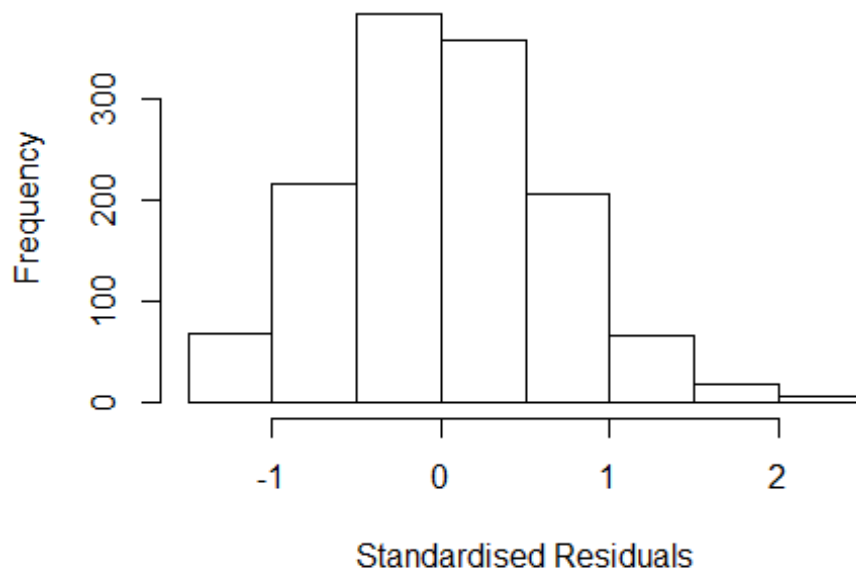


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.24, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 2.0710421447781"
## [1] "Male first author team size 2018 geometric mean: 1.95576768458185"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.91208994663022"
## [1] "Male last author team size 2018 geometric mean: 2.04281069879599"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 880, 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 1.547 1      1.244
## LastAuthorFemale  1.604 1      1.267
## UniqueAuthors    1.305 4      1.034
## Year              1.411 16     1.011
```

Residuals from first and last author and team size



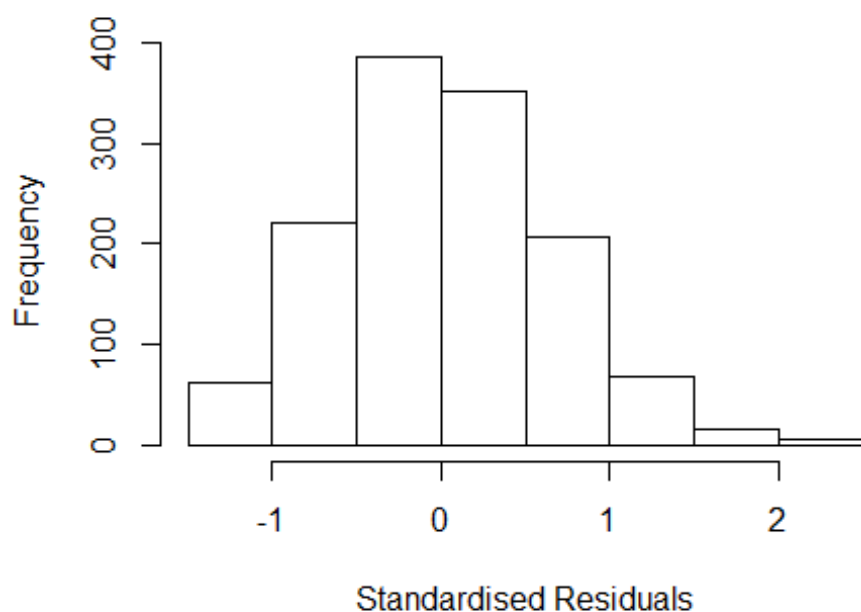
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.36781 -0.41809 -0.00697 0.43792 2.42802
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.88392 0.09226 9.58 < 2e-16 ***
## FirstAuthorFemale1 -0.00267 0.04706 -0.06 0.95477
## LastAuthorFemale1 0.00762 0.04968 0.15 0.87811
## UniqueAuthors2 0.04698 0.04263 1.10 0.27073
## UniqueAuthors3 0.02379 0.05560 0.43 0.66881
## UniqueAuthors4 0.13549 0.10649 1.27 0.20348
## UniqueAuthors5 0.14865 0.11392 1.30 0.19216
## Year1997 0.19990 0.12003 1.67 0.09607 .
## Year1998 0.21962 0.12412 1.77 0.07706 .
## Year1999 0.06411 0.11996 0.53 0.59316
```

```

## Year2000      0.03111      0.12546      0.25      0.80419
## Year2001      0.28325      0.13056      2.17      0.03023 *
## Year2002      0.32003      0.10934      2.93      0.00348 **
## Year2003      0.36622      0.12231      2.99      0.00280 **
## Year2004      0.35327      0.13839      2.55      0.01080 *
## Year2005      0.32762      0.12682      2.58      0.00989 **
## Year2006      0.36018      0.11508      3.13      0.00179 **
## Year2007      0.28308      0.11289      2.51      0.01227 *
## Year2008      0.22424      0.10723      2.09      0.03670 *
## Year2009      0.28058      0.10966      2.56      0.01062 *
## Year2010      0.23504      0.10902      2.16      0.03127 *
## Year2011      0.39071      0.11689      3.34      0.00085 ***
## Year2012      0.28506      0.11568      2.46      0.01386 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.646
## Multiple R-squared:  0.0324, Adjusted R-squared:  0.0159
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 115 weights are ~= 1. The remaining 1201 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.128  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          7.60e-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 1.503 1      1.226
## LastAuthorFemale  1.529 1      1.236
## Year              1.145 16      1.004

```

Residuals from first and last author

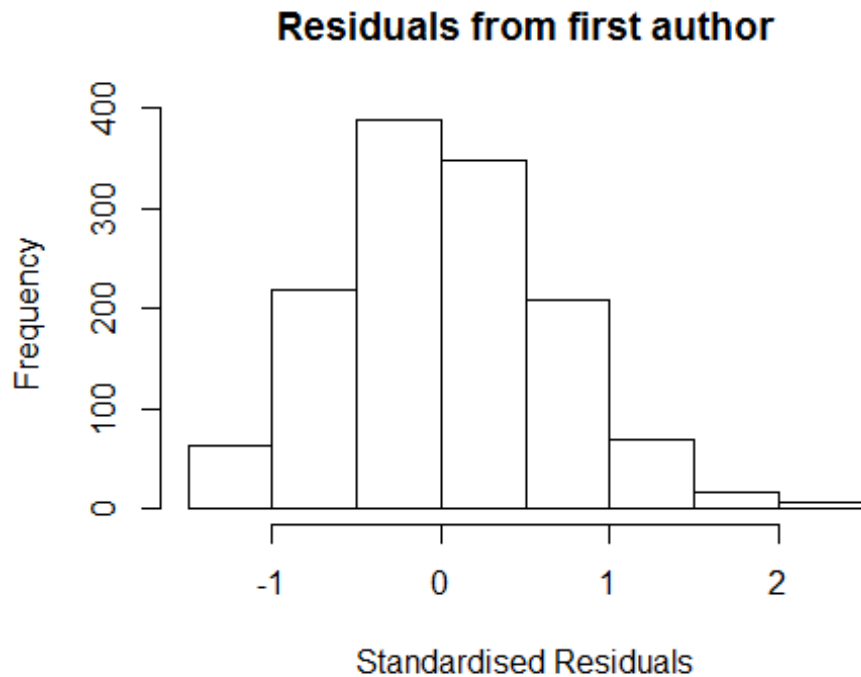


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3248 -0.4211 -0.0112 0.4487 2.4025
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9000 0.0903 9.97 < 2e-16 ***
## FirstAuthorFemale1 -0.0040 0.0467 -0.09 0.93170
## LastAuthorFemale1 0.0141 0.0489 0.29 0.77262
## Year1997 0.2003 0.1194 1.68 0.09374 .
## Year1998 0.2280 0.1230 1.85 0.06415 .
## Year1999 0.0684 0.1185 0.58 0.56399
## Year2000 0.0330 0.1246 0.26 0.79132
## Year2001 0.2875 0.1296 2.22 0.02673 *
## Year2002 0.3262 0.1089 3.00 0.00280 **
## Year2003 0.3749 0.1211 3.10 0.00200 **
## Year2004 0.3643 0.1381 2.64 0.00844 **
## Year2005 0.3422 0.1275 2.68 0.00736 **
```

```

## Year2006          0.3821      0.1136      3.36  0.00079 ***
## Year2007          0.2966      0.1120      2.65  0.00818 **
## Year2008          0.2385      0.1070      2.23  0.02598 *
## Year2009          0.2963      0.1093      2.71  0.00682 **
## Year2010          0.2519      0.1088      2.31  0.02078 *
## Year2011          0.4106      0.1162      3.53  0.00042 ***
## Year2012          0.2995      0.1155      2.59  0.00963 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.649
## Multiple R-squared:  0.0297, Adjusted R-squared:  0.0162
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 112 weights are ~= 1. The remaining 1204 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.141  0.867  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      7.60e-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 1.05 1      1.025
## Year              1.05 16      1.002

```

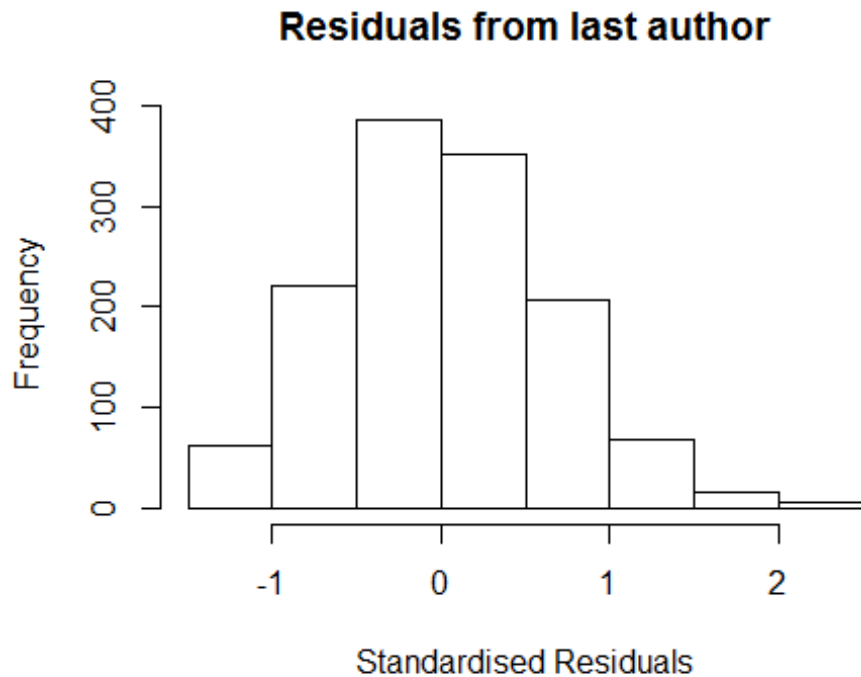


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3169 -0.4240 -0.0146 0.4504 2.4079
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90266 0.08982 10.05 < 2e-16 ***
## FirstAuthorFemale1 0.00288 0.03911 0.07 0.94133
## Year1997 0.20136 0.11947 1.69 0.09213 .
## Year1998 0.22600 0.12276 1.84 0.06585 .
## Year1999 0.06759 0.11849 0.57 0.56848
## Year2000 0.03338 0.12481 0.27 0.78915
## Year2001 0.28739 0.12975 2.21 0.02694 *
## Year2002 0.32669 0.10896 3.00 0.00277 **
## Year2003 0.37484 0.12135 3.09 0.00205 **
## Year2004 0.36320 0.13791 2.63 0.00855 **
## Year2005 0.34000 0.12659 2.69 0.00732 **
## Year2006 0.38082 0.11343 3.36 0.00081 ***
```

```

## Year2007          0.29559    0.11191    2.64  0.00836 **
## Year2008          0.23846    0.10711    2.23  0.02617 *
## Year2009          0.29457    0.10910    2.70  0.00702 **
## Year2010          0.25078    0.10869    2.31  0.02119 *
## Year2011          0.41136    0.11637    3.53  0.00042 ***
## Year2012          0.29843    0.11539    2.59  0.00981 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.648
## Multiple R-squared:  0.0296, Adjusted R-squared:  0.0169
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1207 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.138  0.866  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      7.60e-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 1.07 1      1.034
## Year      1.07 16      1.002

```

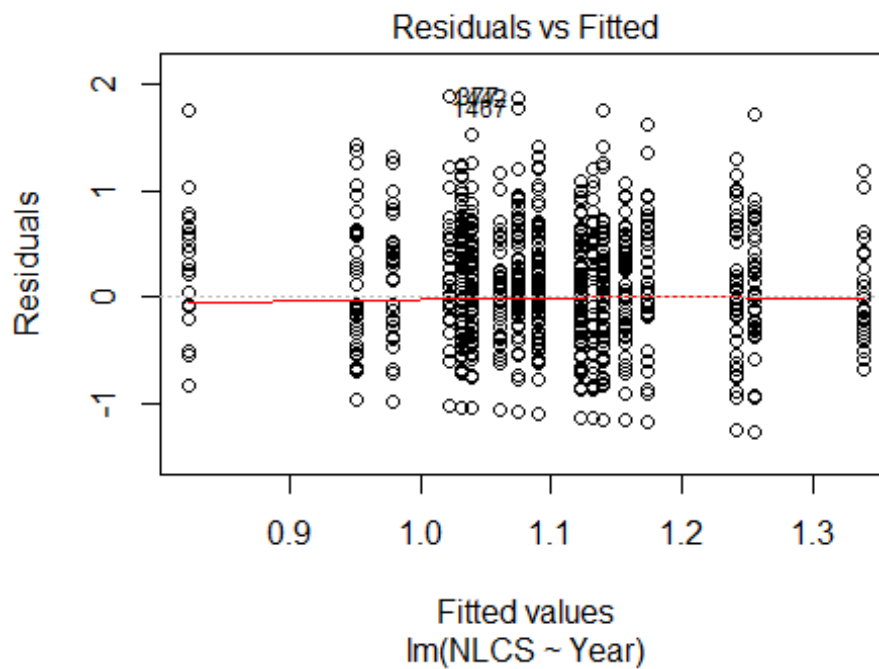
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3224 -0.4216 -0.0125 0.4469 2.4011
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8996 0.0902 9.97 < 2e-16 ***
## LastAuthorFemale1 0.0121 0.0410 0.30 0.76769
## Year1997 0.2004 0.1194 1.68 0.09361 .
## Year1998 0.2276 0.1231 1.85 0.06458 .
## Year1999 0.0681 0.1185 0.58 0.56525
## Year2000 0.0330 0.1247 0.26 0.79110
## Year2001 0.2873 0.1296 2.22 0.02682 *
## Year2002 0.3261 0.1089 2.99 0.00281 **
## Year2003 0.3748 0.1212 3.09 0.00204 **
## Year2004 0.3639 0.1377 2.64 0.00831 **
## Year2005 0.3416 0.1270 2.69 0.00726 **
## Year2006 0.3814 0.1130 3.38 0.00076 ***
```

```

## Year2007          0.2963      0.1118      2.65  0.00816 **
## Year2008          0.2380      0.1070      2.23  0.02621 *
## Year2009          0.2959      0.1093      2.71  0.00688 **
## Year2010          0.2515      0.1085      2.32  0.02062 *
## Year2011          0.4106      0.1163      3.53  0.00043 ***
## Year2012          0.2992      0.1154      2.59  0.00960 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.648
## Multiple R-squared:  0.0297, Adjusted R-squared:  0.017
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 112 weights are ~= 1. The remaining 1204 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.141  0.867  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      7.60e-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: 1316"
## [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
##   53   61   48   47   54   42   52   47   55   75   58   88  101  111  112
## 2011 2012
##  106   84
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   42   40   32   32   39   32   38   36   45   54   44   71   78   88   90
## 2011 2012

```

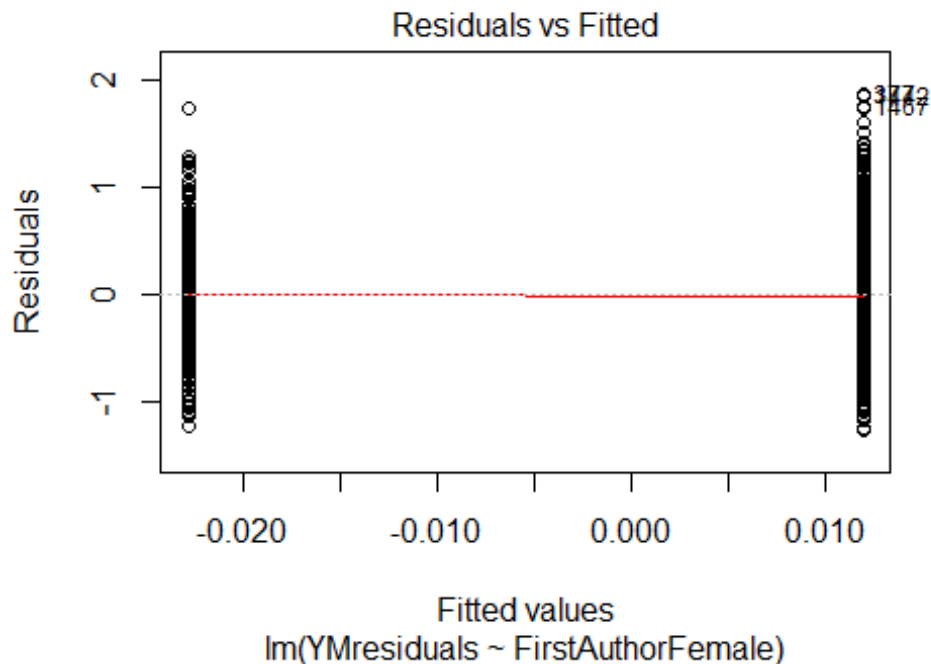
```
##      87      62
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   38   36   32   25   37   28   37   34   41   45   43   63   74   85   83
## 2011 2012
##   77   59
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```



```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.7, df = 1, p-value = 0.1
## [1] "Female first author team size 2018 geometric mean: 2.11525406380944"
## [1] "Male first author team size 2018 geometric mean: 1.96712220439502"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

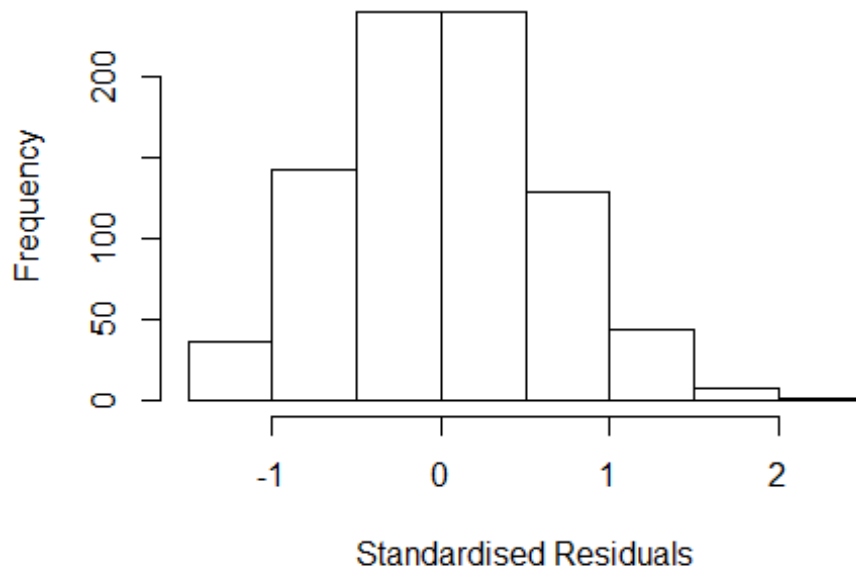
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 720, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.81553029019423"
## [1] "Male last author team size 2018 geometric mean: 2.15274029367827"

## Warning in wilcox.test.default(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 = 0.2
## 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 1.311  1      1.145
## LastAuthorFemale  1.290  1      1.136
## UniqueAuthors    1.385  4      1.042
## Year              1.550 16      1.014
```

Residuals from first and last author and team size



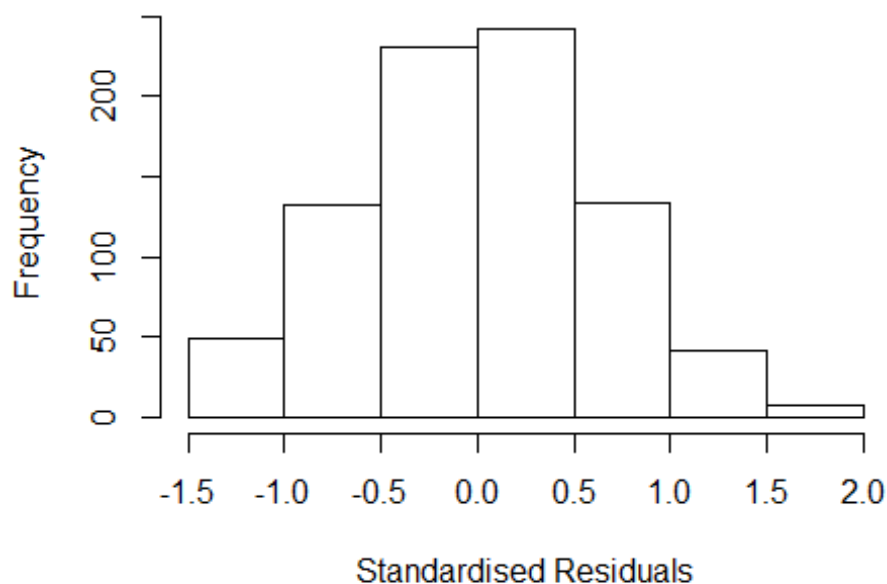
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.358536 -0.406546  0.000436  0.429030  2.034382
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1181    0.1280   8.74 < 2e-16 ***
## FirstAuthorFemale1 -0.0419    0.0499  -0.84  0.4005
## LastAuthorFemale1  0.0683    0.0495   1.38  0.1683
## UniqueAuthors2     0.1444    0.0532   2.72  0.0067 **
## UniqueAuthors3     0.2566    0.0632   4.06 5.4e-05 ***
## UniqueAuthors4     0.2002    0.0999   2.00  0.0453 *
## UniqueAuthors5    -0.0469    0.2417  -0.19  0.8461
## Year1997           0.0960    0.1801   0.53  0.5940
## Year1998           0.1044    0.1454   0.72  0.4730
## Year1999          -0.3724    0.1936  -1.92  0.0548 .
```

```

## Year2000          -0.0763      0.1885   -0.40   0.6857
## Year2001          -0.2577      0.2001   -1.29   0.1981
## Year2002          -0.2941      0.1689   -1.74   0.0820 .
## Year2003          -0.1096      0.1823   -0.60   0.5476
## Year2004          -0.3430      0.1606   -2.14   0.0330 *
## Year2005          -0.0266      0.1529   -0.17   0.8618
## Year2006          -0.1663      0.1512   -1.10   0.2718
## Year2007          -0.1397      0.1429   -0.98   0.3285
## Year2008          -0.2200      0.1430   -1.54   0.1244
## Year2009          -0.1673      0.1413   -1.18   0.2370
## Year2010          -0.2094      0.1470   -1.42   0.1546
## Year2011          -0.2309      0.1451   -1.59   0.1120
## Year2012          -0.1432      0.1483   -0.97   0.3345
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.62
## Multiple R-squared:  0.0554, Adjusted R-squared:  0.0299
## 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.
##  0.260  0.872  0.948  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.19e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.274 1      1.129
## LastAuthorFemale 1.268 1      1.126
## Year 1.139 16      1.004

```

Residuals from first and last author



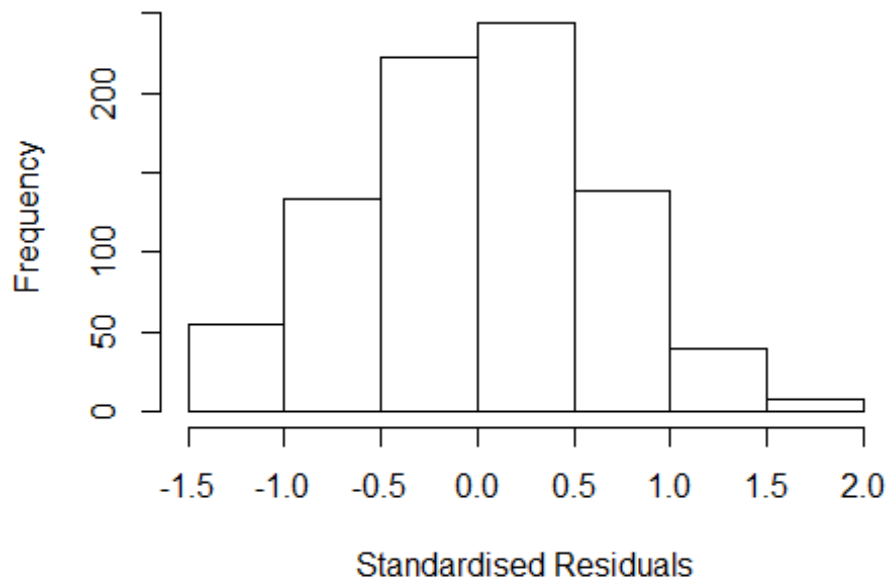
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3054 -0.4246 0.0147 0.4258 1.9654
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20968 0.12550 9.64 <2e-16 ***
## FirstAuthorFemale1 -0.03645 0.05024 -0.73 0.468
## LastAuthorFemale1 0.06862 0.05002 1.37 0.171
## Year1997 0.06356 0.17567 0.36 0.718
## Year1998 0.10350 0.14547 0.71 0.477
## Year1999 -0.39611 0.17898 -2.21 0.027 *
## Year2000 -0.07861 0.18691 -0.42 0.674
## Year2001 -0.27011 0.19972 -1.35 0.177
## Year2002 -0.28767 0.16713 -1.72 0.086 .
## Year2003 -0.10218 0.18602 -0.55 0.583
## Year2004 -0.32308 0.16497 -1.96 0.051 .
## Year2005 -0.00693 0.15571 -0.04 0.965
```

```

## Year2006      -0.15082    0.15270   -0.99    0.324
## Year2007      -0.11897    0.14244   -0.84    0.404
## Year2008      -0.19191    0.14208   -1.35    0.177
## Year2009      -0.14200    0.14180   -1.00    0.317
## Year2010      -0.17152    0.14715   -1.17    0.244
## Year2011      -0.21251    0.14739   -1.44    0.150
## Year2012      -0.12695    0.14783   -0.86    0.391
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.63
## Multiple R-squared:  0.0322, Adjusted R-squared:  0.0109
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 767 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.310  0.877   0.948   0.912   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.19e-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 1.085 1      1.042
## Year              1.085 16      1.003

```


Residuals from first author



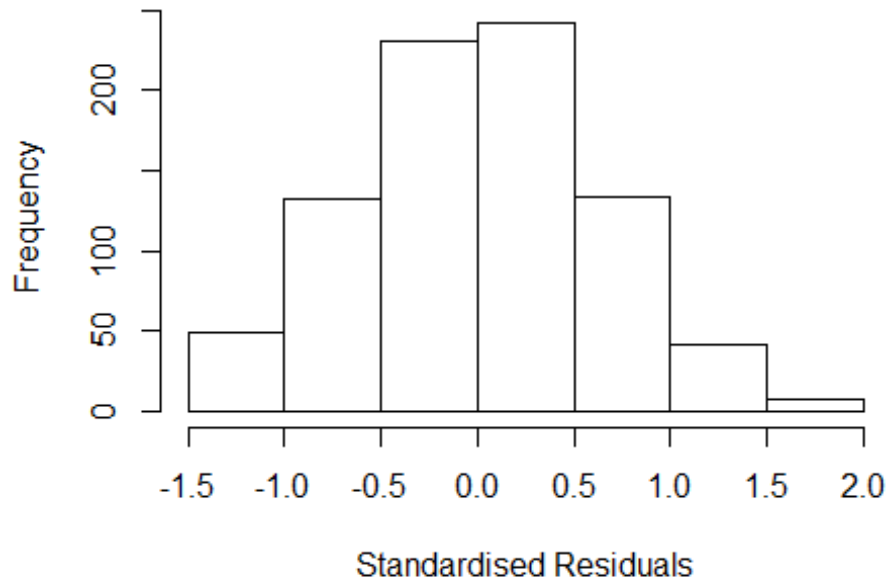
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2805 -0.4208 0.0143 0.4322 1.9554
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21238 0.12513 9.69 <2e-16 ***
## FirstAuthorFemale1 -0.00626 0.04638 -0.13 0.893
## Year1997 0.06812 0.17555 0.39 0.698
## Year1998 0.11164 0.14575 0.77 0.444
## Year1999 -0.39177 0.17894 -2.19 0.029 *
## Year2000 -0.06331 0.18735 -0.34 0.735
## Year2001 -0.26275 0.20219 -1.30 0.194
## Year2002 -0.27862 0.16729 -1.67 0.096 .
## Year2003 -0.08904 0.18472 -0.48 0.630
## Year2004 -0.30596 0.16438 -1.86 0.063 .
## Year2005 0.00790 0.15540 0.05 0.959
## Year2006 -0.14022 0.15159 -0.93 0.355
```

```

## Year2007      -0.10732    0.14163   -0.76    0.449
## Year2008      -0.18239    0.14157   -1.29    0.198
## Year2009      -0.13182    0.14128   -0.93    0.351
## Year2010      -0.16006    0.14665   -1.09    0.275
## Year2011      -0.20398    0.14735   -1.38    0.167
## Year2012      -0.11208    0.14741   -0.76    0.447
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.633
## Multiple R-squared:  0.0299, Adjusted R-squared:  0.00973
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 779 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.320  0.882  0.951  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      1.19e-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 1.077 1          1.038
## Year            1.077 16          1.002

```

Residuals from last author



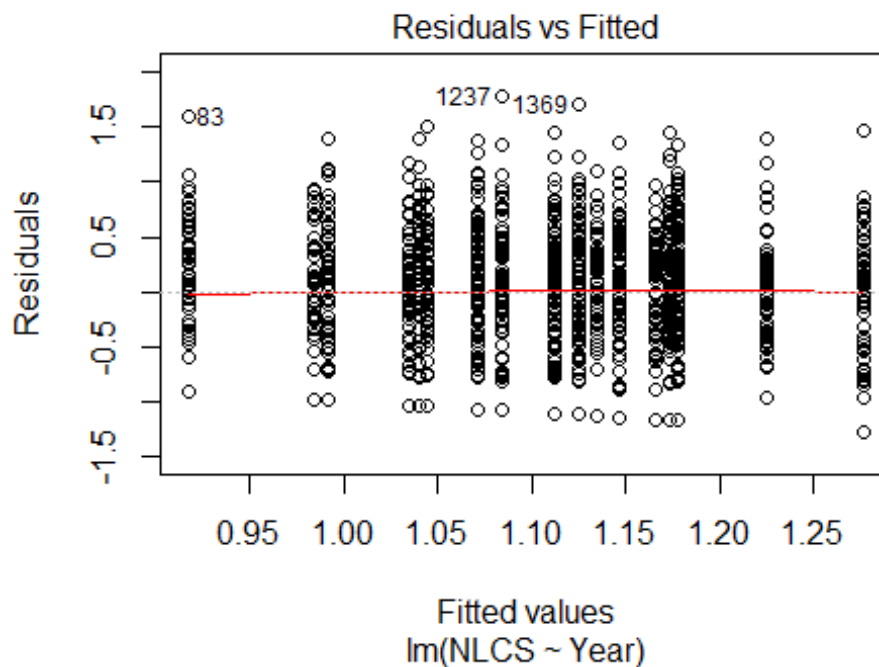
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3174 -0.4249 0.0192 0.4350 1.9757
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20686 0.12604 9.58 <2e-16 ***
## LastAuthorFemale1 0.05227 0.04624 1.13 0.259
## Year1997 0.05823 0.17584 0.33 0.741
## Year1998 0.09714 0.14571 0.67 0.505
## Year1999 -0.40112 0.17864 -2.25 0.025 *
## Year2000 -0.07680 0.18736 -0.41 0.682
## Year2001 -0.27752 0.20041 -1.38 0.166
## Year2002 -0.29335 0.16763 -1.75 0.080 .
## Year2003 -0.10951 0.18732 -0.58 0.559
## Year2004 -0.32725 0.16531 -1.98 0.048 *
## Year2005 -0.00841 0.15664 -0.05 0.957
## Year2006 -0.15568 0.15285 -1.02 0.309
```

```

## Year2007          -0.12416      0.14271    -0.87      0.385
## Year2008          -0.19732      0.14241    -1.39      0.166
## Year2009          -0.14806      0.14213    -1.04      0.298
## Year2010          -0.18020      0.14674    -1.23      0.220
## Year2011          -0.21499      0.14780    -1.45      0.146
## Year2012          -0.13256      0.14839    -0.89      0.372
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.631
## Multiple R-squared:  0.0317, Adjusted R-squared:  0.0116
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 769 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.306  0.879  0.949  0.912  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.19e-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 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
##   78   73   70   56   62   73   81   66   65   61  106  111  123  114  131
## 2011 2012
##  147  129
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   72   62   55   47   50   62   71   60   58   52   91  102  105  100  103
## 2011 2012

```

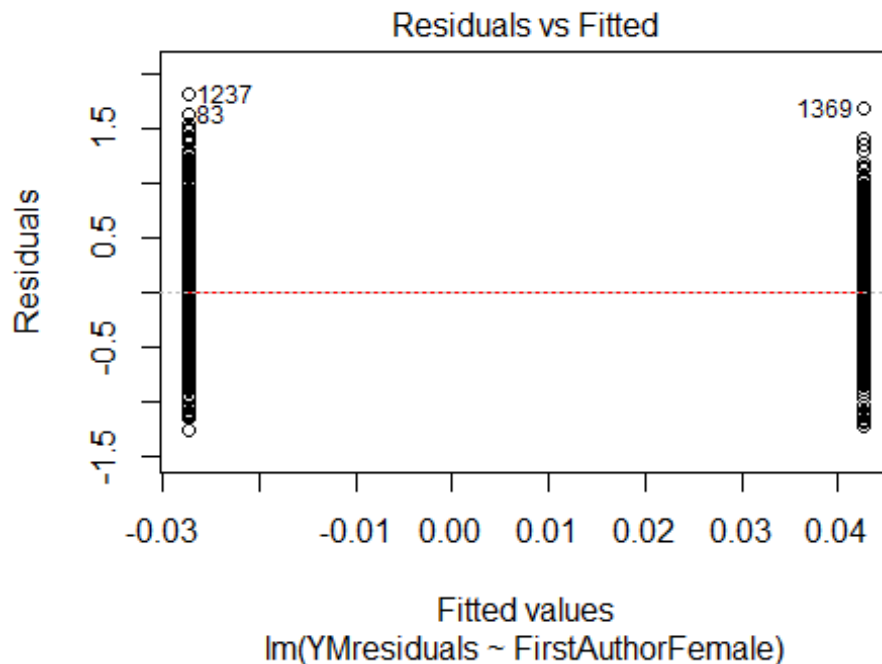
```
## 124 107
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 68 59 51 44 46 59 69 56 56 50 85 96 98 93 99
## 2011 2012
## 111 101
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.5
```



```
##
## 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: 2.04913896933902"
## [1] "Male first author team size 2018 geometric mean: 1.78853554202593"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

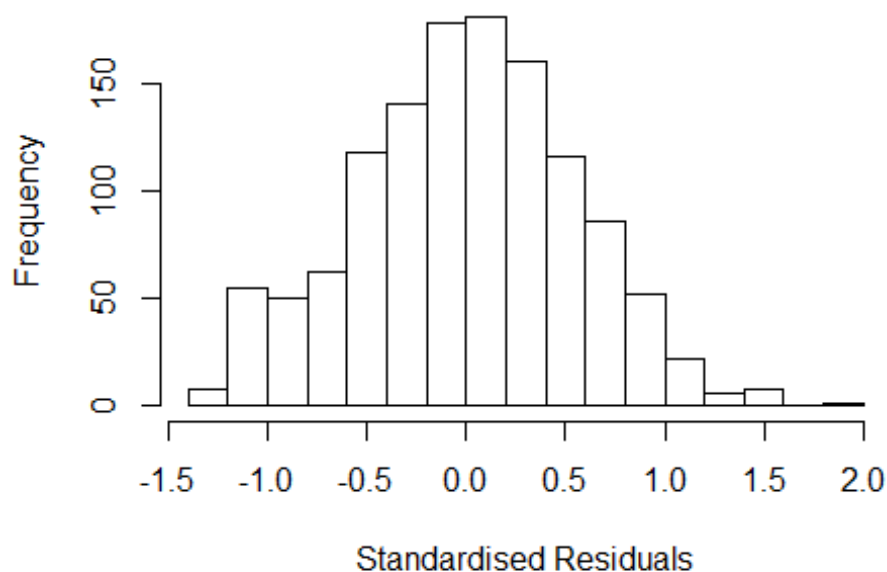
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 920, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.94361251231976"
## [1] "Male last author team size 2018 geometric mean: 1.93486912935791"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 840, 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  1.408 1      1.186
## UniqueAuthors    1.326 4      1.036
## Year              1.486 16     1.012
```

Residuals from first and last author and team size



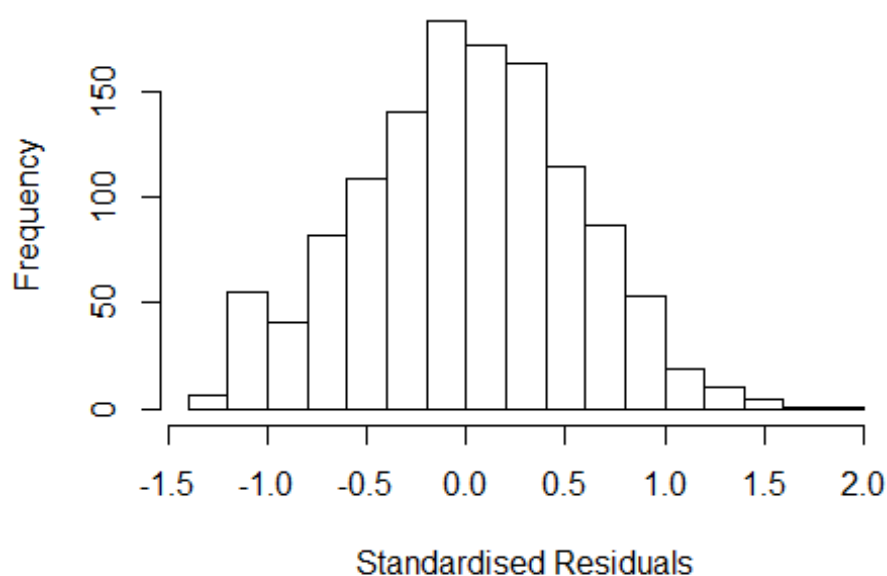
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2680 -0.3722 0.0121 0.3762 1.9107
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.82936 0.07444 11.14 <2e-16 ***
## FirstAuthorFemale1 0.04412 0.03812 1.16 0.2474
## LastAuthorFemale1 0.03623 0.03827 0.95 0.3440
## UniqueAuthors2 0.09791 0.03821 2.56 0.0105 *
## UniqueAuthors3 0.10935 0.04882 2.24 0.0253 *
## UniqueAuthors4 0.23388 0.08719 2.68 0.0074 **
## UniqueAuthors5 0.05154 0.13007 0.40 0.6920
## Year1997 0.11836 0.09854 1.20 0.2300
## Year1998 0.09654 0.11431 0.84 0.3985
## Year1999 0.00869 0.10210 0.09 0.9322
```

```

## Year2000          0.08158      0.12110      0.67      0.5007
## Year2001          0.11106      0.10174      1.09      0.2752
## Year2002          0.26949      0.09389      2.87      0.0042 **
## Year2003          0.28338      0.09650      2.94      0.0034 **
## Year2004          0.33233      0.10597      3.14      0.0018 **
## Year2005          0.22359      0.09902      2.26      0.0241 *
## Year2006          0.19886      0.09615      2.07      0.0388 *
## Year2007          0.16214      0.09529      1.70      0.0891 .
## Year2008          0.12699      0.09353      1.36      0.1748
## Year2009          0.18251      0.09903      1.84      0.0656 .
## Year2010          0.25810      0.09372      2.75      0.0060 **
## Year2011          0.18284      0.09338      1.96      0.0505 .
## Year2012          0.26042      0.09673      2.69      0.0072 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.564
## Multiple R-squared:  0.0408, Adjusted R-squared:  0.0235
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1142 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.227  0.875   0.953   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      8.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "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  1.417 1      1.190
## Year              1.142 16      1.004

```


Residuals from first and last author



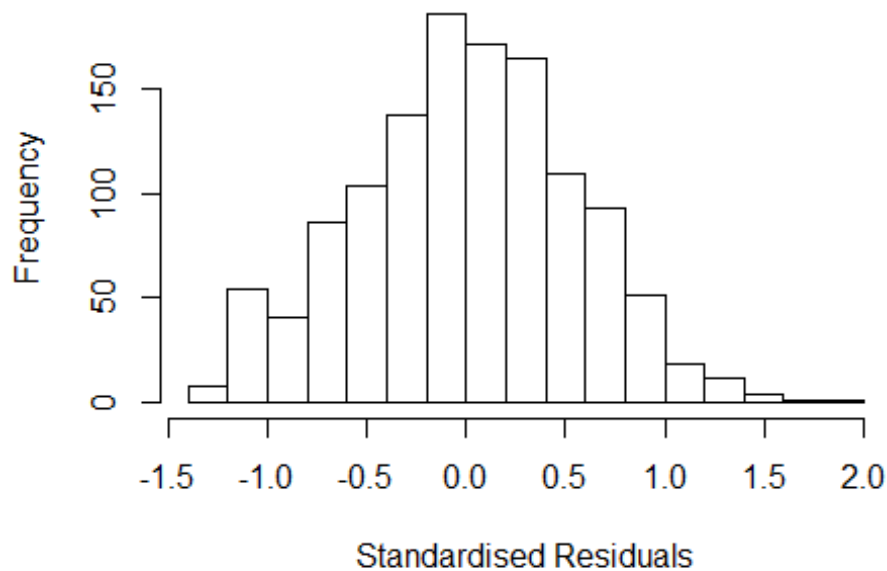
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24268 -0.36853 0.00802 0.36872 1.84681
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8590 0.0738 11.64 < 2e-16 ***
## FirstAuthorFemale1 0.0461 0.0385 1.20 0.23158
## LastAuthorFemale1 0.0357 0.0386 0.92 0.35547
## Year1997 0.1276 0.0983 1.30 0.19465
## Year1998 0.1146 0.1112 1.03 0.30293
## Year1999 0.0356 0.0986 0.36 0.71826
## Year2000 0.1093 0.1223 0.89 0.37168
## Year2001 0.1310 0.1000 1.31 0.19052
## Year2002 0.2885 0.0923 3.13 0.00181 **
## Year2003 0.2956 0.0951 3.11 0.00192 **
## Year2004 0.3604 0.1054 3.42 0.00064 ***
## Year2005 0.2570 0.0998 2.58 0.01012 *
```

```

## Year2006          0.2430      0.0932      2.61  0.00925 **
## Year2007          0.2012      0.0957      2.10  0.03578 *
## Year2008          0.1612      0.0933      1.73  0.08416 .
## Year2009          0.2151      0.0993      2.17  0.03058 *
## Year2010          0.2927      0.0933      3.14  0.00173 **
## Year2011          0.2245      0.0920      2.44  0.01483 *
## Year2012          0.3020      0.0949      3.18  0.00150 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.0301, Adjusted R-squared:  0.0158
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 1149 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.260  0.866  0.953  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      8.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



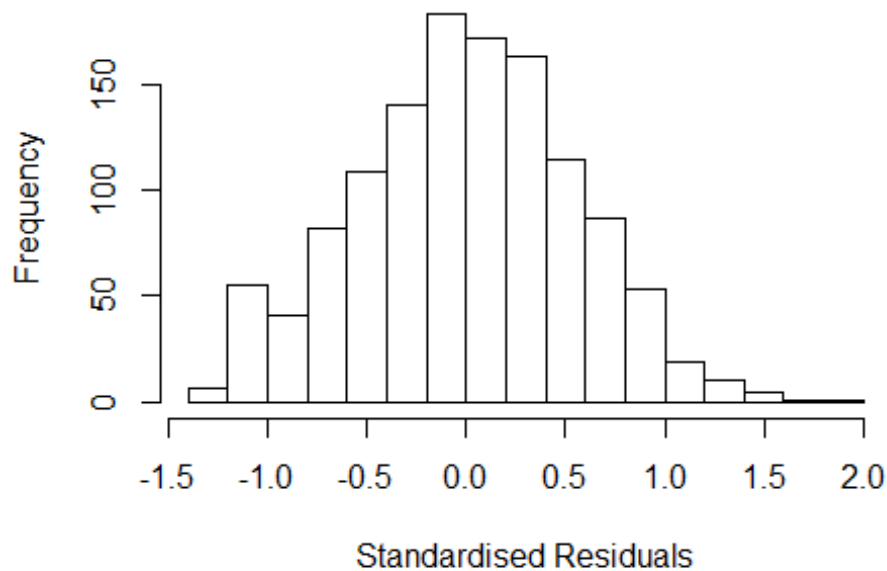
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23118 -0.36462 0.00388 0.36450 1.83886
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8669 0.0724 11.97 < 2e-16 ***
## FirstAuthorFemale1 0.0640 0.0335 1.91 0.05659 .
## Year1997 0.1266 0.0982 1.29 0.19723
## Year1998 0.1114 0.1109 1.01 0.31507
## Year1999 0.0373 0.0984 0.38 0.70497
## Year2000 0.1093 0.1225 0.89 0.37238
## Year2001 0.1274 0.0995 1.28 0.20068
## Year2002 0.2888 0.0922 3.13 0.00178 **
## Year2003 0.2954 0.0951 3.10 0.00195 **
## Year2004 0.3568 0.1052 3.39 0.00072 ***
## Year2005 0.2514 0.0988 2.54 0.01106 *
## Year2006 0.2448 0.0931 2.63 0.00868 **
```

```

## Year2007          0.1996      0.0955      2.09  0.03679 *
## Year2008          0.1613      0.0932      1.73  0.08378 .
## Year2009          0.2126      0.0989      2.15  0.03167 *
## Year2010          0.2958      0.0937      3.16  0.00163 **
## Year2011          0.2232      0.0918      2.43  0.01518 *
## Year2012          0.3003      0.0948      3.17  0.00157 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.0293, Adjusted R-squared:  0.0158
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 1149 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.264  0.866  0.953  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      8.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



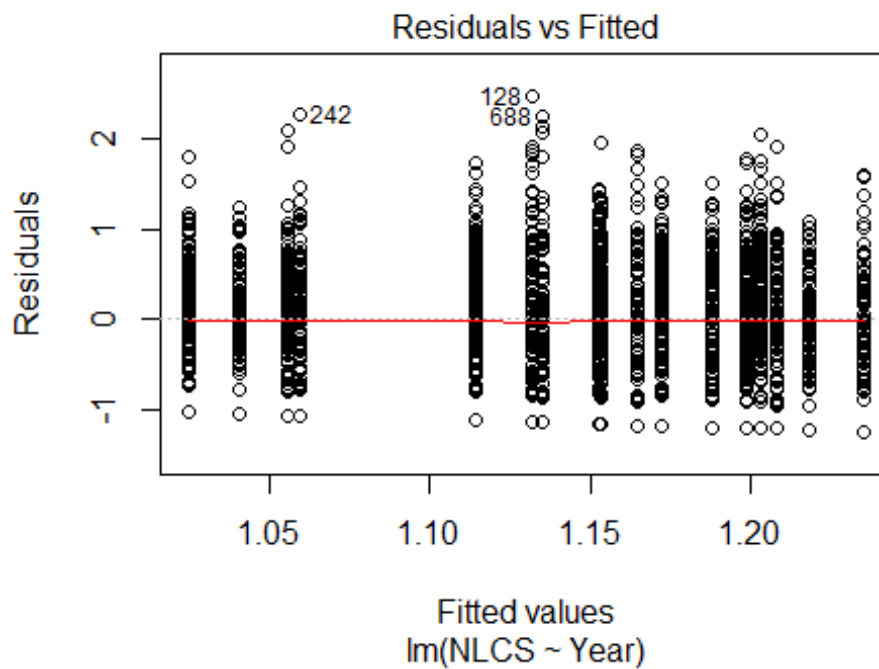
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23573 -0.37750 0.00605 0.36597 1.83352
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8660 0.0738 11.73 < 2e-16 ***
## LastAuthorFemale1 0.0588 0.0336 1.75 0.08039 .
## Year1997 0.1299 0.0984 1.32 0.18721
## Year1998 0.1109 0.1115 0.99 0.32036
## Year1999 0.0362 0.0986 0.37 0.71337
## Year2000 0.1075 0.1221 0.88 0.37903
## Year2001 0.1300 0.1002 1.30 0.19497
## Year2002 0.2890 0.0926 3.12 0.00185 **
## Year2003 0.2962 0.0949 3.12 0.00185 **
## Year2004 0.3697 0.1052 3.51 0.00046 ***
## Year2005 0.2623 0.0996 2.63 0.00854 **
## Year2006 0.2434 0.0935 2.60 0.00936 **
```

```

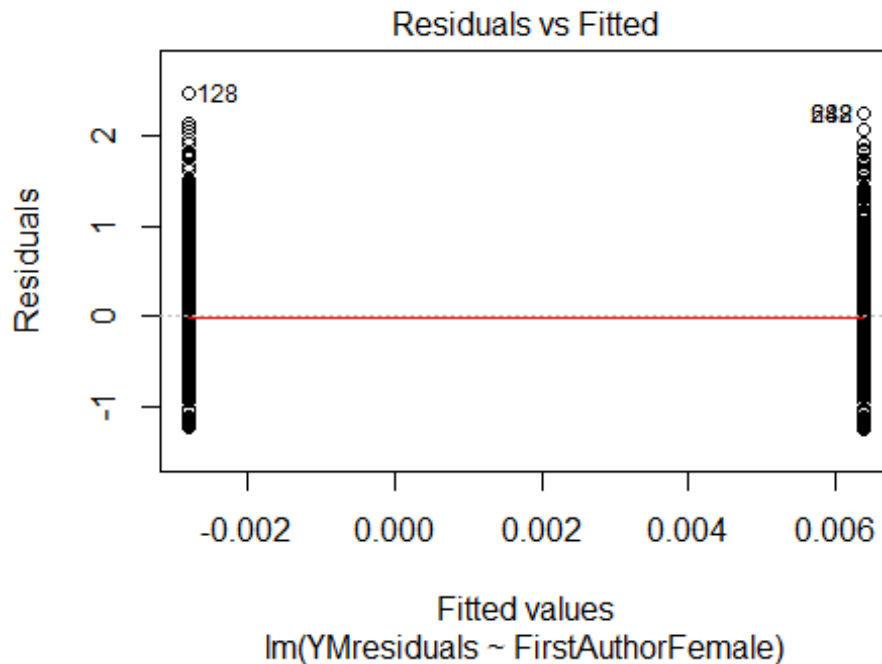
## Year2007          0.2039      0.0958      2.13  0.03356 *
## Year2008          0.1675      0.0935      1.79  0.07349 .
## Year2009          0.2197      0.0994      2.21  0.02726 *
## Year2010          0.2930      0.0929      3.15  0.00165 **
## Year2011          0.2303      0.0920      2.50  0.01241 *
## Year2012          0.3065      0.0949      3.23  0.00127 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.0291, Adjusted R-squared:  0.0156
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 100 weights are ~= 1. The remaining 1141 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.266  0.866  0.953   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      8.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1241"
## [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
## 153 158 163 142 125 145 151 119 151 136 218 218 263 289 274
## 2011 2012
## 261 255
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 113 106 106 97 100 90 112 84 114 93 164 167 196 218 204
## 2011 2012

```

```
## 191 205
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 102 97 96 90 93 84 102 73 102 77 146 153 177 195 185
## 2011 2012
## 174 171
## [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
```

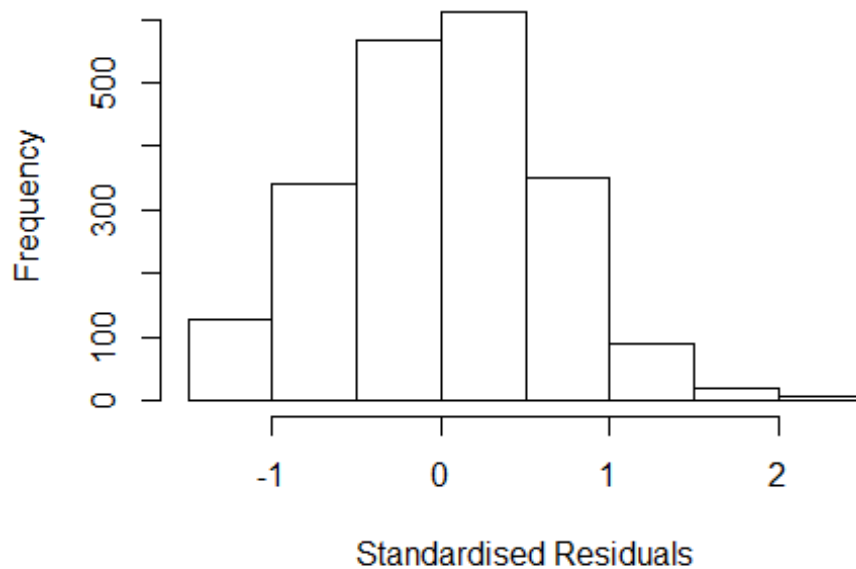


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.00016, df = 1, p-value = 1
```



```
## [1] "Female first author team size 2018 geometric mean: 1.83924867806007"
## [1] "Male first author team size 2018 geometric mean: 2.0538872388072"
##
## 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: 2.05268918825082"
## [1] "Male last author team size 2018 geometric mean: 1.94691627736056"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4400, 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 1.342 1      1.158
## LastAuthorFemale  1.341 1      1.158
## UniqueAuthors    1.162 4      1.019
## Year             1.211 16      1.006
```


Residuals from first and last author and team size



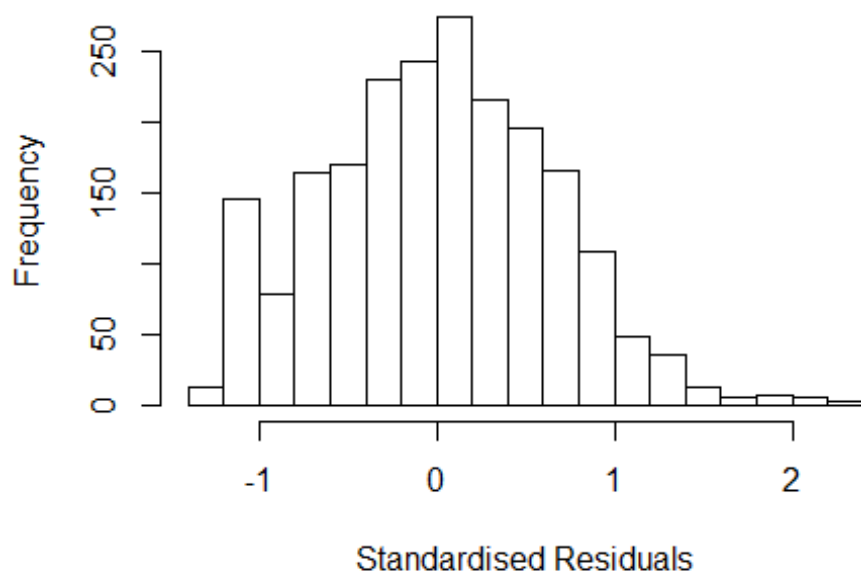
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3331 -0.4336 0.0142 0.4454 2.3824
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.92979 0.08689 10.70 < 2e-16 ***
## FirstAuthorFemale1 -0.03467 0.03491 -0.99 0.32085
## LastAuthorFemale1 0.04288 0.03616 1.19 0.23591
## UniqueAuthors2 0.16220 0.03348 4.84 1.4e-06 ***
## UniqueAuthors3 0.15621 0.04100 3.81 0.00014 ***
## UniqueAuthors4 0.17950 0.07314 2.45 0.01421 *
## UniqueAuthors5 0.34036 0.07729 4.40 1.1e-05 ***
## Year1997 -0.00135 0.11040 -0.01 0.99021
## Year1998 0.11185 0.11213 1.00 0.31861
## Year1999 0.07585 0.10762 0.70 0.48101
```

```

## Year2000      0.01314      0.10688      0.12  0.90217
## Year2001      0.08169      0.11282      0.72  0.46907
## Year2002      0.17774      0.10335      1.72  0.08560 .
## Year2003      0.21929      0.10760      2.04  0.04168 *
## Year2004      0.13622      0.11478      1.19  0.23542
## Year2005      0.23888      0.11246      2.12  0.03377 *
## Year2006      0.03264      0.09696      0.34  0.73647
## Year2007      0.10676      0.09917      1.08  0.28183
## Year2008     -0.04946      0.09756     -0.51  0.61220
## Year2009      0.05329      0.09828      0.54  0.58773
## Year2010      0.12871      0.09942      1.29  0.19561
## Year2011      0.13108      0.09885      1.33  0.18497
## Year2012      0.13005      0.10130      1.28  0.19937
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.648
## Multiple R-squared:  0.0321, Adjusted R-squared:  0.0219
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 174 weights are ~= 1. The remaining 1943 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.148  0.876  0.950  0.911  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 1.297 1      1.139
## LastAuthorFemale  1.320 1      1.149
## Year              1.074 16      1.002

```

Residuals from first and last author



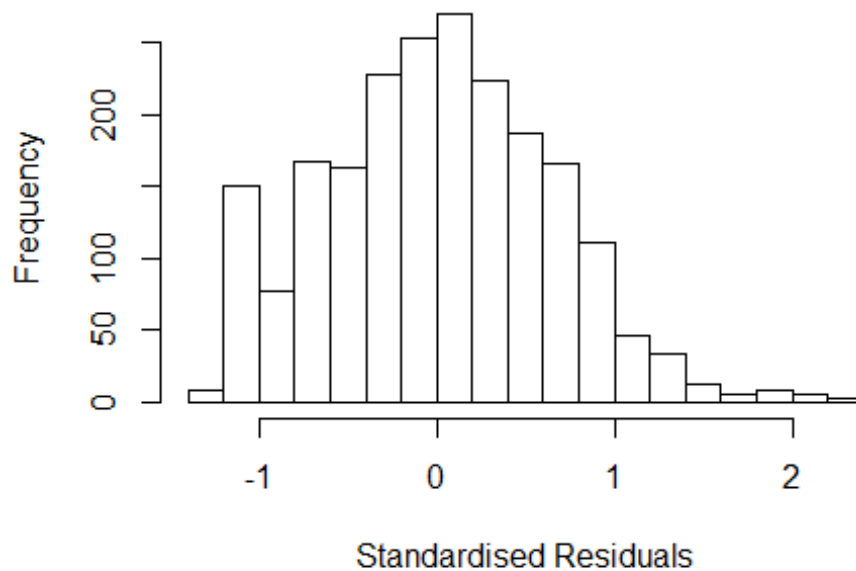
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.300 -0.450 0.014 0.447 2.298
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.005099 0.086051 11.68 <2e-16 ***
## FirstAuthorFemale1 -0.026248 0.034878 -0.75 0.452
## LastAuthorFemale1 0.041032 0.036268 1.13 0.258
## Year1997 0.000934 0.110621 0.01 0.993
## Year1998 0.113167 0.111395 1.02 0.310
## Year1999 0.070321 0.107804 0.65 0.514
## Year2000 0.010636 0.106409 0.10 0.920
## Year2001 0.084881 0.113328 0.75 0.454
## Year2002 0.193051 0.104051 1.86 0.064 .
## Year2003 0.229719 0.107572 2.14 0.033 *
## Year2004 0.144695 0.114573 1.26 0.207
## Year2005 0.280224 0.111783 2.51 0.012 *
```

```

## Year2006          0.067605    0.097033    0.70    0.486
## Year2007          0.127848    0.099466    1.29    0.199
## Year2008         -0.012415    0.098525   -0.13    0.900
## Year2009          0.085827    0.098922    0.87    0.386
## Year2010          0.159425    0.100009    1.59    0.111
## Year2011          0.152426    0.098968    1.54    0.124
## Year2012          0.168323    0.101737    1.65    0.098 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.658
## Multiple R-squared:  0.0139, Adjusted R-squared:  0.00544
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 172 weights are ~= 1. The remaining 1945 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.874  0.949  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      4.72e-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 1.024 1      1.012
## Year              1.024 16      1.001

```

Residuals from first author



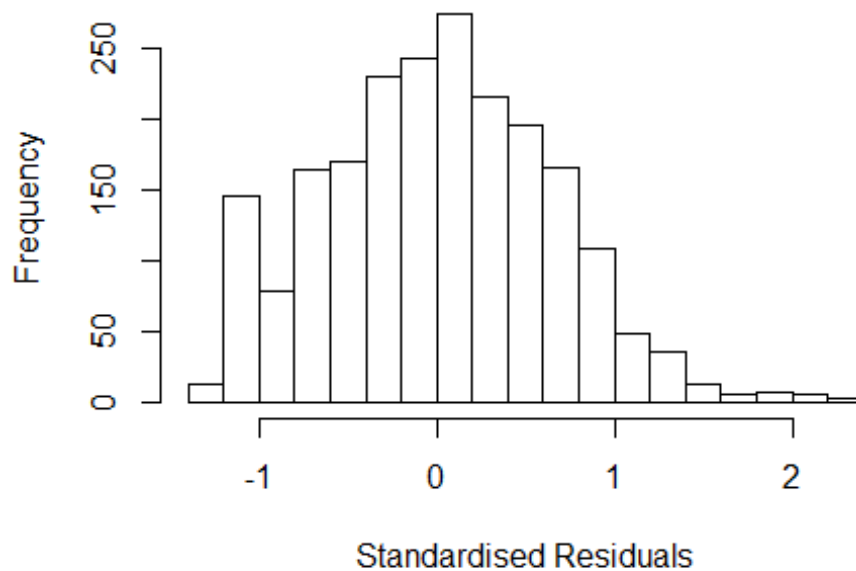
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28058 -0.43979 0.00794 0.45220 2.31060
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01019 0.08617 11.72 <2e-16 ***
## FirstAuthorFemale1 -0.00686 0.03100 -0.22 0.825
## Year1997 0.00507 0.11106 0.05 0.964
## Year1998 0.11089 0.11169 0.99 0.321
## Year1999 0.06758 0.10796 0.63 0.531
## Year2000 0.01508 0.10668 0.14 0.888
## Year2001 0.08587 0.11363 0.76 0.450
## Year2002 0.19347 0.10435 1.85 0.064 .
## Year2003 0.23052 0.10790 2.14 0.033 *
## Year2004 0.14221 0.11447 1.24 0.214
## Year2005 0.27725 0.11178 2.48 0.013 *
## Year2006 0.06889 0.09735 0.71 0.479
```

```

## Year2007          0.12785    0.09981    1.28    0.200
## Year2008         -0.01077    0.09882   -0.11    0.913
## Year2009          0.08559    0.09918    0.86    0.388
## Year2010          0.16060    0.10032    1.60    0.110
## Year2011          0.15606    0.09923    1.57    0.116
## Year2012          0.17061    0.10212    1.67    0.095 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.658
## Multiple R-squared:  0.0132, Adjusted R-squared:  0.00522
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 174 weights are ~= 1. The remaining 1943 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.192  0.874  0.948  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.72e-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 1.042 1          1.021
## Year            1.042 16          1.001

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3065 -0.4427 0.0129 0.4451 2.2881
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.002662 0.086132 11.64 <2e-16 ***
## LastAuthorFemale1 0.028078 0.032263 0.87 0.384
## Year1997 0.000158 0.110971 0.00 0.999
## Year1998 0.112370 0.111482 1.01 0.314
## Year1999 0.067892 0.107719 0.63 0.529
## Year2000 0.009772 0.106564 0.09 0.927
## Year2001 0.083079 0.113525 0.73 0.464
## Year2002 0.192400 0.104135 1.85 0.065 .
## Year2003 0.228504 0.107715 2.12 0.034 *
## Year2004 0.142100 0.114337 1.24 0.214
## Year2005 0.275713 0.111758 2.47 0.014 *
## Year2006 0.065419 0.097088 0.67 0.501
```

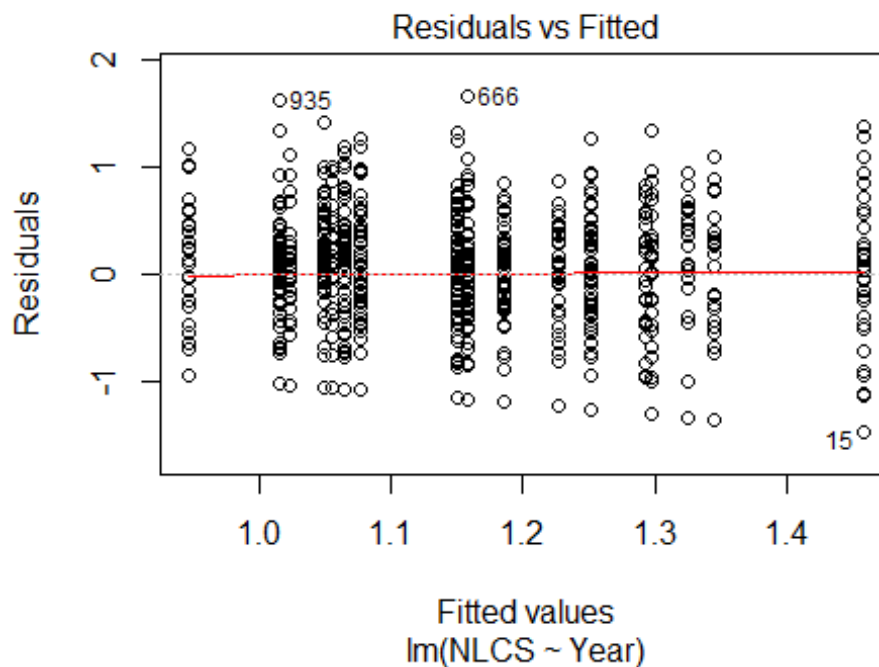
```

## Year2007      0.126010    0.099472    1.27    0.205
## Year2008     -0.015633    0.098523   -0.16    0.874
## Year2009      0.083390    0.098997    0.84    0.400
## Year2010      0.156763    0.099926    1.57    0.117
## Year2011      0.151241    0.098992    1.53    0.127
## Year2012      0.165594    0.101831    1.63    0.104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.658
## Multiple R-squared:  0.0136, Adjusted R-squared:  0.00562
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 1949 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.201  0.875  0.949  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      4.72e-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 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
##   36   31   35   33   38   35   39   44   42   63   75   62   69   66   92
## 2011 2012
##   78   73
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   33   26   26   25   30   26   29   33   31   56   60   54   62   57   80
## 2011 2012

```



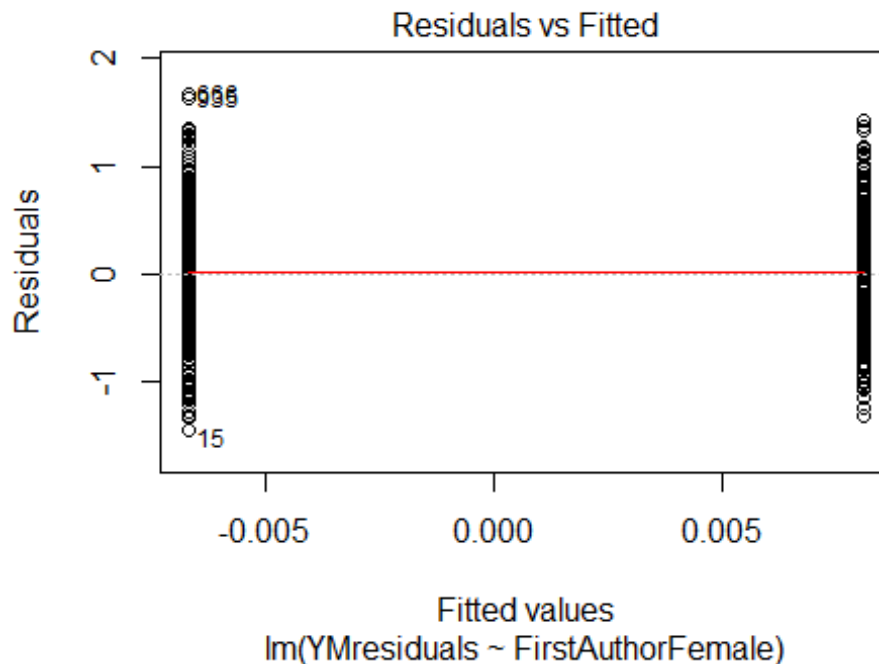
```
## 63 63
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 30 24 25 25 29 23 29 27 30 52 56 50 57 56 75
## 2011 2012
## 58 60
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 18, df = 16, p-value = 0.3
```



```
##
## 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: 2.35716182590798"
## [1] "Male first author team size 2018 geometric mean: 1.91911307189339"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 510, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.39843774948927"
## [1] "Male last author team size 2018 geometric mean: 1.91778803711488"

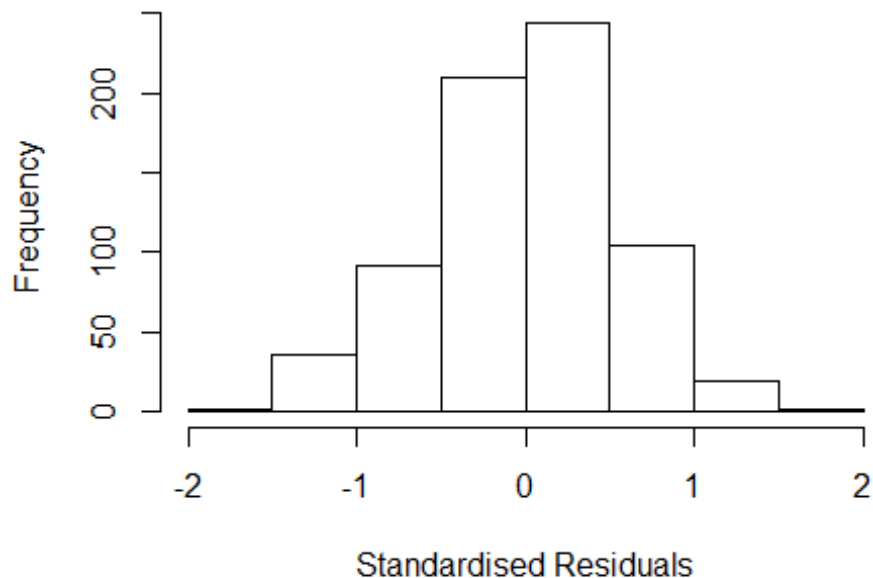
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 510, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.546	1	1.243
LastAuthorFemale	1.541	1	1.241
UniqueAuthors	2.078	4	1.096
Year	2.444	16	1.028

Residuals from first and last author and team size



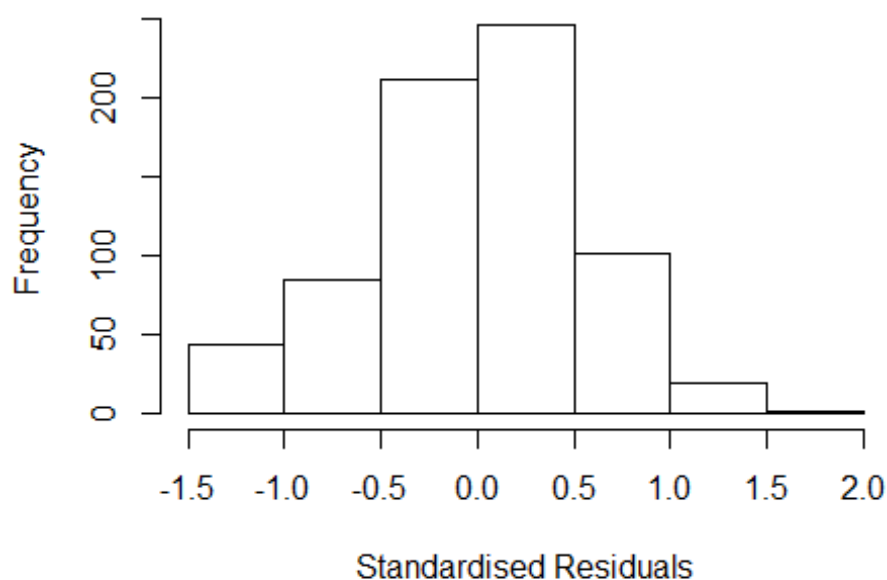
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5016 -0.3693 0.0163 0.3448 1.7789
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4131 0.1446 9.77 < 2e-16 ***
## FirstAuthorFemale1 -0.0173 0.0512 -0.34 0.73637
## LastAuthorFemale1 0.0268 0.0515 0.52 0.60215
## UniqueAuthors2 0.0617 0.0500 1.23 0.21810
## UniqueAuthors3 0.1751 0.0716 2.44 0.01476 *
## UniqueAuthors4 0.3942 0.1118 3.53 0.00045 ***
## UniqueAuthors5 0.2155 0.0881 2.45 0.01473 *
## Year1997 -0.1249 0.1886 -0.66 0.50792
## Year1998 -0.1700 0.1912 -0.89 0.37411
## Year1999 -0.0344 0.2016 -0.17 0.86453
```

```

## Year2000          -0.3936      0.1882    -2.09  0.03687 *
## Year2001          -0.1687      0.2350    -0.72  0.47316
## Year2002          -0.4032      0.1762    -2.29  0.02245 *
## Year2003          -0.5930      0.1903    -3.12  0.00191 **
## Year2004          -0.2193      0.1702    -1.29  0.19799
## Year2005          -0.4448      0.1695    -2.62  0.00889 **
## Year2006          -0.2282      0.1596    -1.43  0.15320
## Year2007          -0.3065      0.1595    -1.92  0.05506 .
## Year2008          -0.3740      0.1616    -2.31  0.02096 *
## Year2009          -0.4033      0.1684    -2.40  0.01686 *
## Year2010          -0.3631      0.1596    -2.28  0.02321 *
## Year2011          -0.4844      0.1602    -3.02  0.00259 **
## Year2012          -0.4165      0.1689    -2.47  0.01392 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.548
## Multiple R-squared:  0.0718, Adjusted R-squared:  0.0419
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 634 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.271  0.858  0.954  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          1.42e-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 1.526 1          1.235
## LastAuthorFemale 1.560 1          1.249
## Year              1.237 16          1.007

```

Residuals from first and last author



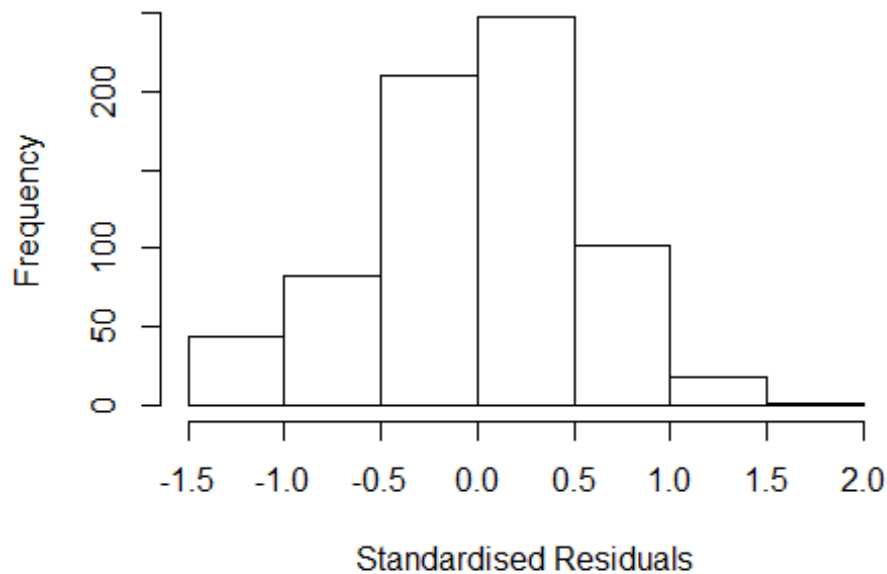
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4594 -0.3873 0.0325 0.3391 1.7226
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.43667 0.13969 10.28 <2e-16 ***
## FirstAuthorFemale1 -0.00249 0.05212 -0.05 0.9619
## LastAuthorFemale1 0.02271 0.05310 0.43 0.6691
## Year1997 -0.10528 0.18285 -0.58 0.5649
## Year1998 -0.16478 0.19052 -0.86 0.3874
## Year1999 -0.03864 0.19675 -0.20 0.8444
## Year2000 -0.37395 0.18368 -2.04 0.0421 *
## Year2001 -0.17658 0.22985 -0.77 0.4426
## Year2002 -0.41581 0.17172 -2.42 0.0157 *
## Year2003 -0.58002 0.18735 -3.10 0.0020 **
## Year2004 -0.21647 0.16750 -1.29 0.1967
## Year2005 -0.41256 0.16667 -2.48 0.0136 *
```

```

## Year2006      -0.22861    0.15766   -1.45    0.1475
## Year2007      -0.26543    0.15616   -1.70    0.0896 .
## Year2008      -0.34129    0.16020   -2.13    0.0335 *
## Year2009      -0.39316    0.16479   -2.39    0.0173 *
## Year2010      -0.30173    0.15527   -1.94    0.0524 .
## Year2011      -0.46155    0.15656   -2.95    0.0033 **
## Year2012      -0.36625    0.16459   -2.23    0.0264 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.0522, Adjusted R-squared:  0.0274
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 637 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.308  0.855   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      1.42e-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 1.137 1         1.066
## Year              1.137 16         1.004

```

Residuals from first author



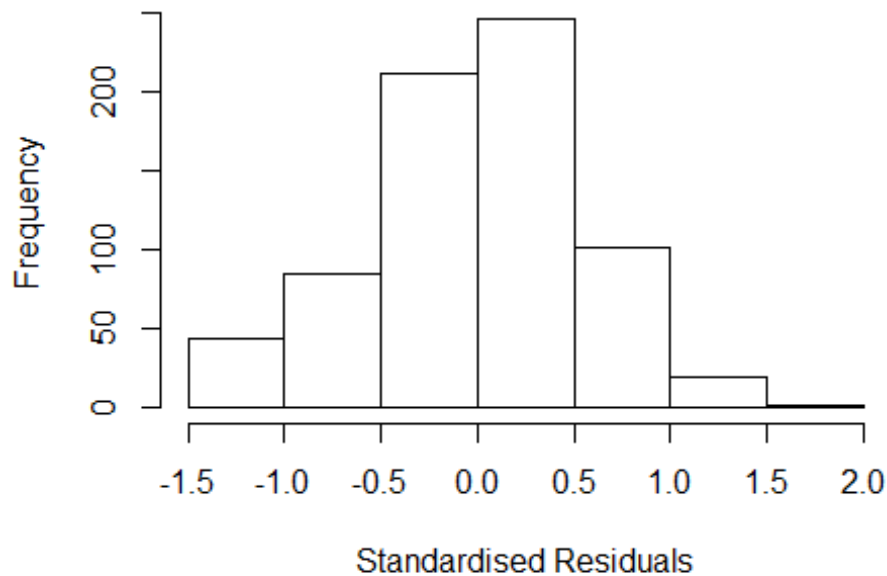
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4372 -0.3901  0.0353  0.3393  1.7191
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4372     0.1398   10.28  <2e-16 ***
## FirstAuthorFemale1  0.0107     0.0451    0.24  0.8130
## Year1997         -0.1041     0.1828   -0.57  0.5694
## Year1998         -0.1649     0.1907   -0.86  0.3875
## Year1999         -0.0386     0.1968   -0.20  0.8446
## Year2000         -0.3736     0.1835   -2.04  0.0422 *
## Year2001         -0.1715     0.2279   -0.75  0.4520
## Year2002         -0.4132     0.1710   -2.42  0.0160 *
## Year2003         -0.5794     0.1868   -3.10  0.0020 **
## Year2004         -0.2145     0.1672   -1.28  0.2000
## Year2005         -0.4100     0.1658   -2.47  0.0136 *
## Year2006         -0.2251     0.1562   -1.44  0.1499
```

```

## Year2007          -0.2617      0.1550    -1.69    0.0917 .
## Year2008          -0.3383      0.1593    -2.12    0.0340 *
## Year2009          -0.3909      0.1642    -2.38    0.0176 *
## Year2010          -0.2981      0.1541    -1.93    0.0535 .
## Year2011          -0.4570      0.1547    -2.95    0.0032 **
## Year2012          -0.3625      0.1635    -2.22    0.0269 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.0518, Adjusted R-squared:  0.0283
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 637 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.312  0.854   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      1.42e-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 1.16 1      1.077
## Year      1.16 16      1.005

```


Residuals from last author



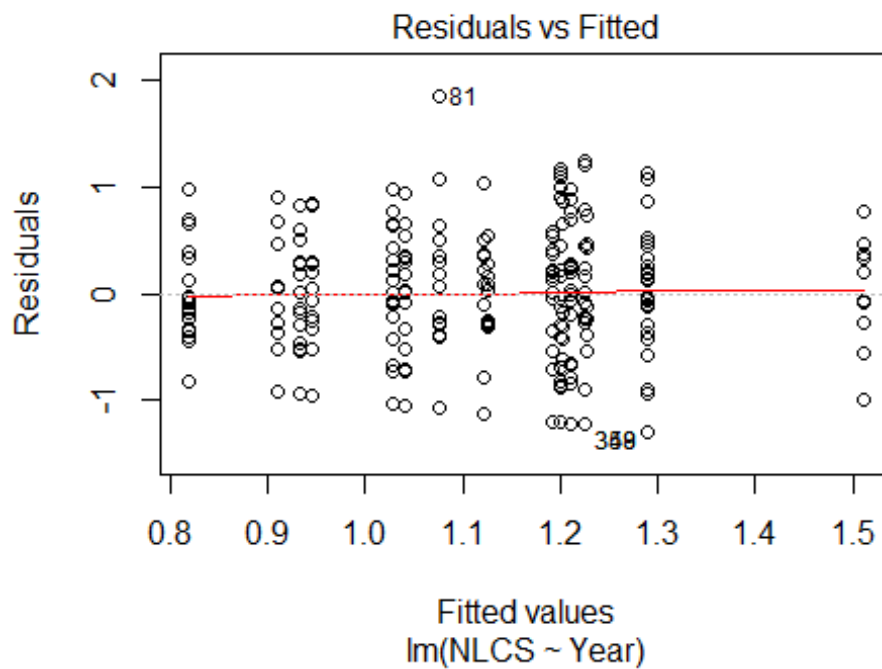
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4577 -0.3869 0.0337 0.3387 1.7233
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4364 0.1394 10.30 <2e-16 ***
## LastAuthorFemale1 0.0213 0.0459 0.46 0.6435
## Year1997 -0.1056 0.1829 -0.58 0.5640
## Year1998 -0.1647 0.1906 -0.86 0.3877
## Year1999 -0.0391 0.1960 -0.20 0.8420
## Year2000 -0.3740 0.1838 -2.04 0.0422 *
## Year2001 -0.1767 0.2300 -0.77 0.4425
## Year2002 -0.4160 0.1717 -2.42 0.0156 *
## Year2003 -0.5804 0.1874 -3.10 0.0020 **
## Year2004 -0.2167 0.1674 -1.29 0.1960
## Year2005 -0.4125 0.1667 -2.47 0.0136 *
## Year2006 -0.2289 0.1578 -1.45 0.1474
```

```

## Year2007          -0.2657      0.1561    -1.70    0.0891 .
## Year2008          -0.3416      0.1603    -2.13    0.0335 *
## Year2009          -0.3935      0.1647    -2.39    0.0172 *
## Year2010          -0.3021      0.1551    -1.95    0.0519 .
## Year2011          -0.4620      0.1565    -2.95    0.0033 **
## Year2012          -0.3667      0.1645    -2.23    0.0262 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.0522, Adjusted R-squared:  0.0288
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 638 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.306  0.854  0.949  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      1.42e-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: 706"
## [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
##   14   17   17   23   15    9   12   16   18   12   20   22   23   32   36
## 2011 2012
##   34   22
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   10   11   14   17   13    6    9   12   16    8   16   19   18   18   28
## 2011 2012

```

```
## 27 15
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 10 10 12 16 13 6 9 11 15 7 14 18 17 18 27
## 2011 2012
## 25 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 = 18, df = 16, p-value = 0.3
```



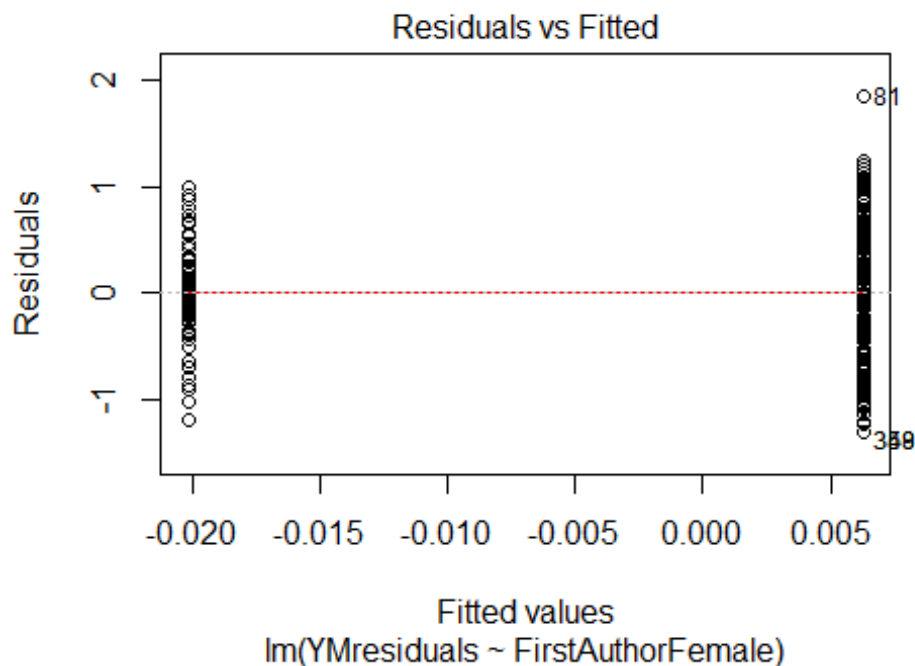
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.6, df = 1, p-value = 0.02

## [1] "Female first author team size 2018 geometric mean: 1.63609610059322"
## [1] "Male first author team size 2018 geometric mean: 1.31950791077289"

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

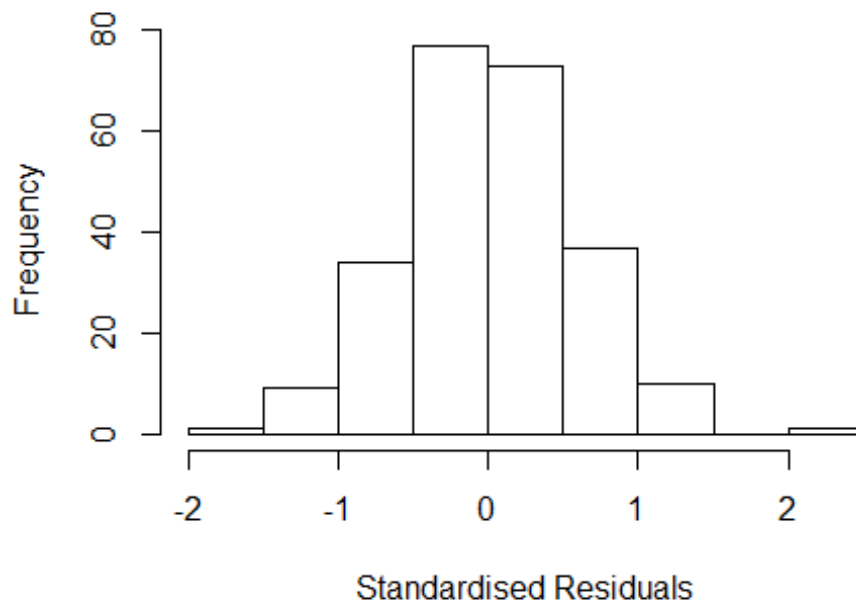
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 51, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.42349781425291"
## [1] "Male last author team size 2018 geometric mean: 1.49561152357164"

## Warning in wilcox.test.default(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 = 0.9
## 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 1.904 1      1.380
## LastAuthorFemale  2.061 1      1.436
## UniqueAuthors    2.090 4      1.097
## Year              2.918 16     1.034
```

Residuals from first and last author and team size



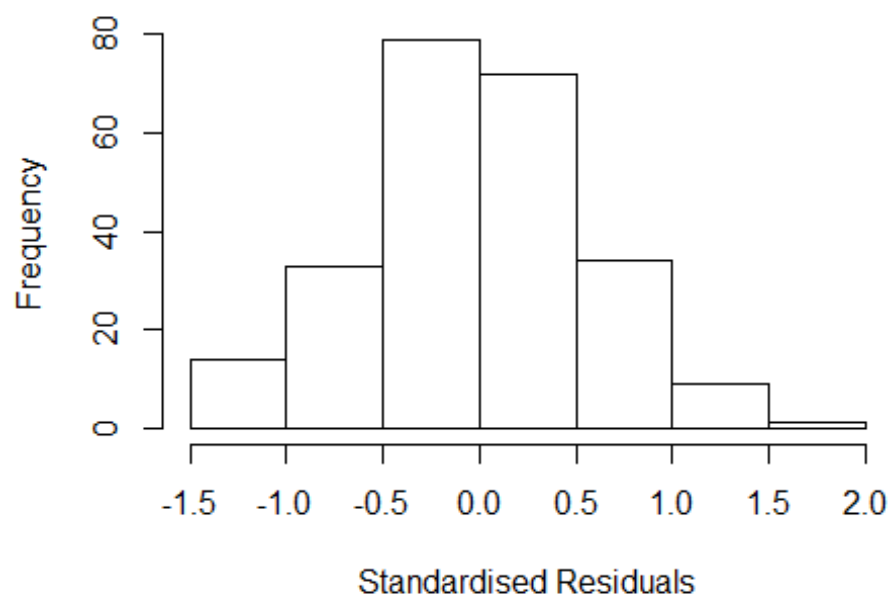
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.60477 -0.34207 0.00492 0.38695 2.00488
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6940 0.2348 2.96 0.0035 **
## FirstAuthorFemale1 -0.0415 0.1037 -0.40 0.6894
## LastAuthorFemale1 0.1371 0.1073 1.28 0.2028
## UniqueAuthors2 0.4116 0.0919 4.48 1.2e-05 ***
## UniqueAuthors3 0.3318 0.1083 3.06 0.0025 **
## UniqueAuthors4 0.3514 0.1838 1.91 0.0572 .
## UniqueAuthors5 0.2995 0.1812 1.65 0.0998 .
## Year1997 0.1045 0.2888 0.36 0.7179
## Year1998 0.3421 0.3604 0.95 0.3436
## Year1999 0.2321 0.2931 0.79 0.4293
```

```

## Year2000          -0.0423      0.2657    -0.16    0.8737
## Year2001           0.3587      0.2881     1.25    0.2144
## Year2002           0.1503      0.2767     0.54    0.5876
## Year2003           0.6385      0.2729     2.34    0.0202 *
## Year2004           0.1738      0.2813     0.62    0.5375
## Year2005           0.3581      0.2955     1.21    0.2268
## Year2006           0.2346      0.2524     0.93    0.3537
## Year2007           0.1647      0.2639     0.62    0.5332
## Year2008           0.2622      0.3157     0.83    0.4072
## Year2009           0.2651      0.2738     0.97    0.3340
## Year2010           0.3620      0.2582     1.40    0.1623
## Year2011          -0.0341      0.2535    -0.13    0.8932
## Year2012          -0.0538      0.2556    -0.21    0.8336
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.573
## Multiple R-squared:  0.187, Adjusted R-squared:  0.105
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 231 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.196  0.887  0.959   0.914   0.989   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.931 1      1.390
## LastAuthorFemale  1.772 1      1.331
## Year              1.456 16      1.012

```

Residuals from first and last author



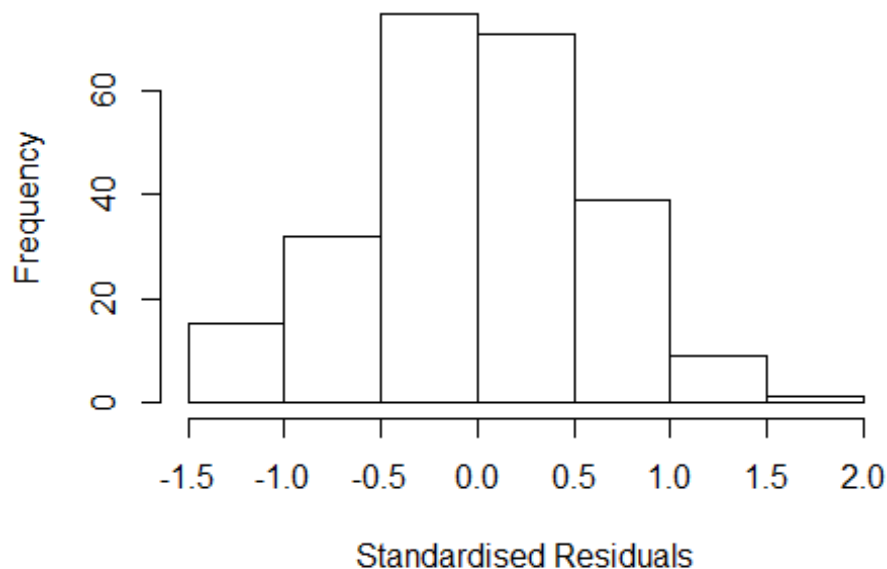
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4384 -0.3641 -0.0115 0.3531 1.8962
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.89684 0.18291 4.90 1.8e-06 ***
## FirstAuthorFemale1 -0.12281 0.10500 -1.17 0.2434
## LastAuthorFemale1 0.16827 0.10432 1.61 0.1082
## Year1997 0.13321 0.28267 0.47 0.6379
## Year1998 0.28107 0.34156 0.82 0.4114
## Year1999 0.13798 0.27000 0.51 0.6098
## Year2000 0.00351 0.22910 0.02 0.9878
## Year2001 0.26146 0.31757 0.82 0.4112
## Year2002 0.18579 0.20801 0.89 0.3727
## Year2003 0.68661 0.22399 3.07 0.0024 **
## Year2004 0.14792 0.24575 0.60 0.5479
## Year2005 0.37168 0.26232 1.42 0.1579
```

```

## Year2006      0.32103      0.20713      1.55      0.1226
## Year2007      0.09925      0.23315      0.43      0.6707
## Year2008      0.23684      0.32669      0.72      0.4692
## Year2009      0.31063      0.24259      1.28      0.2017
## Year2010      0.37333      0.21590      1.73      0.0852 .
## Year2011     -0.08713      0.21174     -0.41      0.6811
## Year2012     -0.02525      0.22180     -0.11      0.9095
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.579
## Multiple R-squared:  0.0977, Adjusted R-squared:  0.0249
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 218 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.262  0.839   0.957   0.903   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.13e-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 1.148 1      1.071
## Year      1.148 16      1.004

```


Residuals from first author



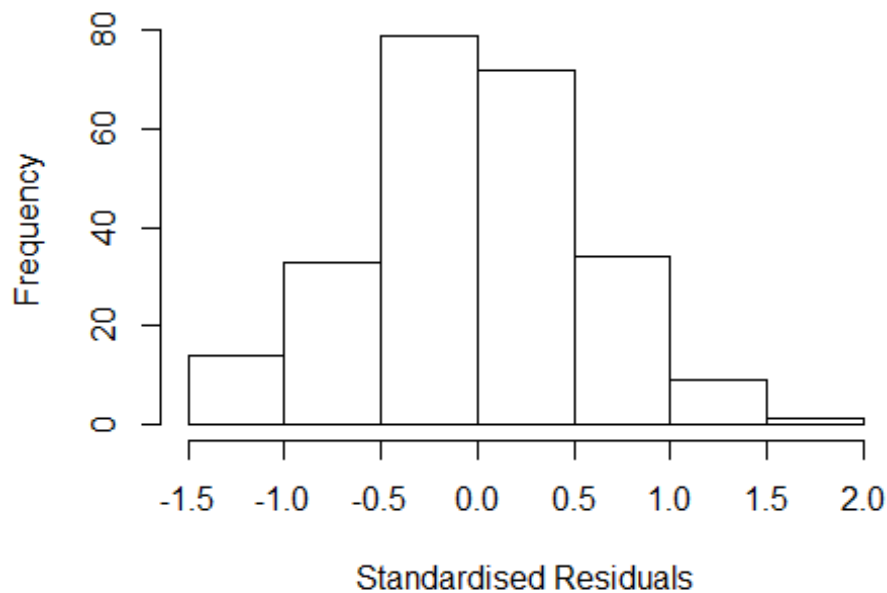
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2970 -0.3604 -0.0147 0.3478 1.9040
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90526 0.18226 4.97 1.3e-06 ***
## FirstAuthorFemale1 -0.01713 0.08294 -0.21 0.8365
## Year1997 0.16883 0.28943 0.58 0.5603
## Year1998 0.28653 0.34466 0.83 0.4067
## Year1999 0.12179 0.26814 0.45 0.6501
## Year2000 0.03819 0.22701 0.17 0.8666
## Year2001 0.26875 0.31446 0.85 0.3937
## Year2002 0.22427 0.20644 1.09 0.2785
## Year2003 0.70093 0.22359 3.13 0.0019 **
## Year2004 0.15346 0.23963 0.64 0.5226
## Year2005 0.37358 0.25906 1.44 0.1507
## Year2006 0.33117 0.21139 1.57 0.1186
```

```

## Year2007          0.11354    0.23124    0.49    0.6239
## Year2008          0.23806    0.31856    0.75    0.4557
## Year2009          0.32544    0.24214    1.34    0.1803
## Year2010          0.39179    0.21595    1.81    0.0710 .
## Year2011         -0.06188    0.21419   -0.29    0.7729
## Year2012         -0.00328    0.22455   -0.01    0.9883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.0869, Adjusted R-squared:  0.0176
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 224 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.287  0.859  0.960  0.910  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.13e-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 1.13 1      1.063
## Year              1.13 16      1.004

```

Residuals from last author



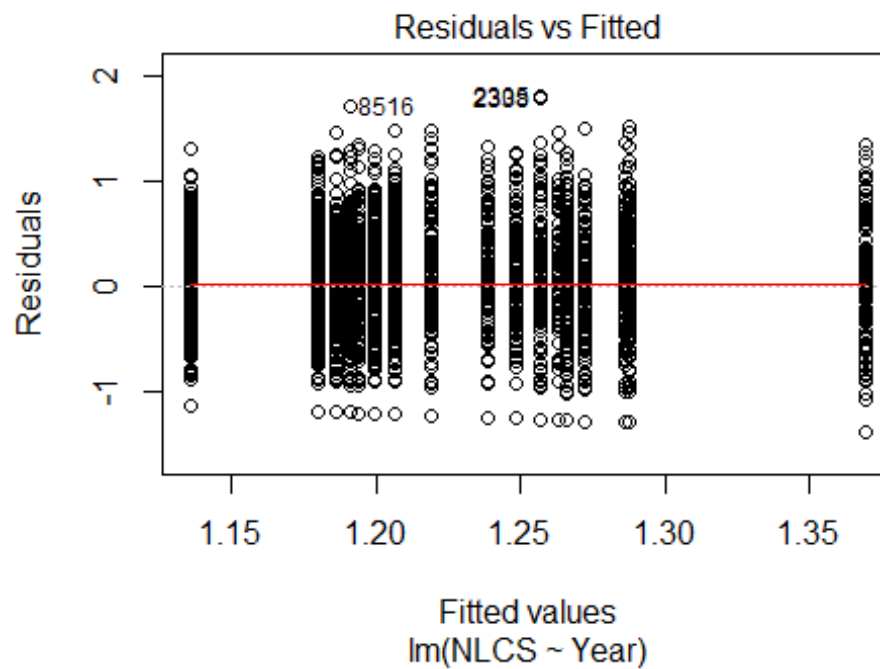
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3606 -0.3452 -0.0149 0.3608 1.9191
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8903 0.1825 4.88 2e-06 ***
## LastAuthorFemale1 0.0973 0.0846 1.15 0.2515
## Year1997 0.1493 0.2852 0.52 0.6012
## Year1998 0.2768 0.3351 0.83 0.4096
## Year1999 0.1215 0.2650 0.46 0.6469
## Year2000 0.0120 0.2285 0.05 0.9580
## Year2001 0.2573 0.3187 0.81 0.4204
## Year2002 0.1757 0.2095 0.84 0.4026
## Year2003 0.6958 0.2251 3.09 0.0022 **
## Year2004 0.1341 0.2391 0.56 0.5756
## Year2005 0.3702 0.2643 1.40 0.1627
## Year2006 0.3239 0.2080 1.56 0.1208
```

```

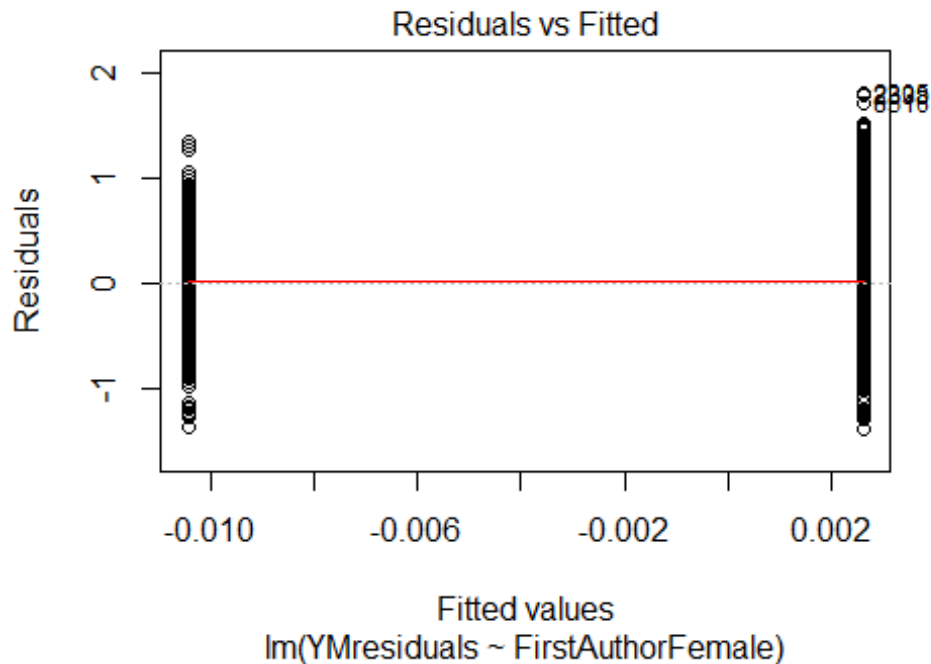
## Year2007          0.0865      0.2347      0.37      0.7126
## Year2008          0.2172      0.3226      0.67      0.5014
## Year2009          0.3095      0.2421      1.28      0.2025
## Year2010          0.3730      0.2171      1.72      0.0872 .
## Year2011         -0.0901      0.2123     -0.42      0.6718
## Year2012         -0.0302      0.2227     -0.14      0.8921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.581
## Multiple R-squared:  0.0927, Adjusted R-squared:  0.0238
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 220 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.253  0.850  0.958  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      4.13e-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: 242"
## [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
## 470 430 400 408 381 378 345 332 401 446 478 465 561 587 624
## 2011 2012
## 623 658
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 160 143 139 139 120 98 130 130 167 175 164 162 252 248 306
## 2011 2012

```

```
## 304 310
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 141 118 124 120 103 86 99 101 140 136 118 133 193 195 237
## 2011 2012
## 231 252
## [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
```

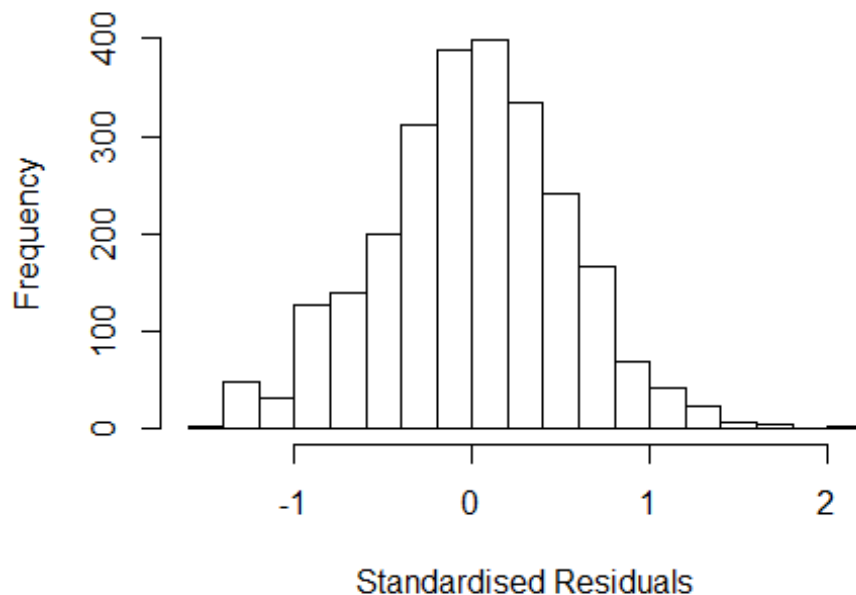


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.1, df = 1, p-value = 0.003
```



```
## [1] "Female first author team size 2018 geometric mean: 3.59569810154277"
## [1] "Male first author team size 2018 geometric mean: 3.03954963147402"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7600, p-value = 0.009
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.4278695311649"
## [1] "Male last author team size 2018 geometric mean: 3.16981038145948"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4400, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.061 1          1.030
## LastAuthorFemale  1.056 1          1.027
## UniqueAuthors     1.164 4          1.019
## Year              1.227 16          1.006
```

Residuals from first and last author and team size



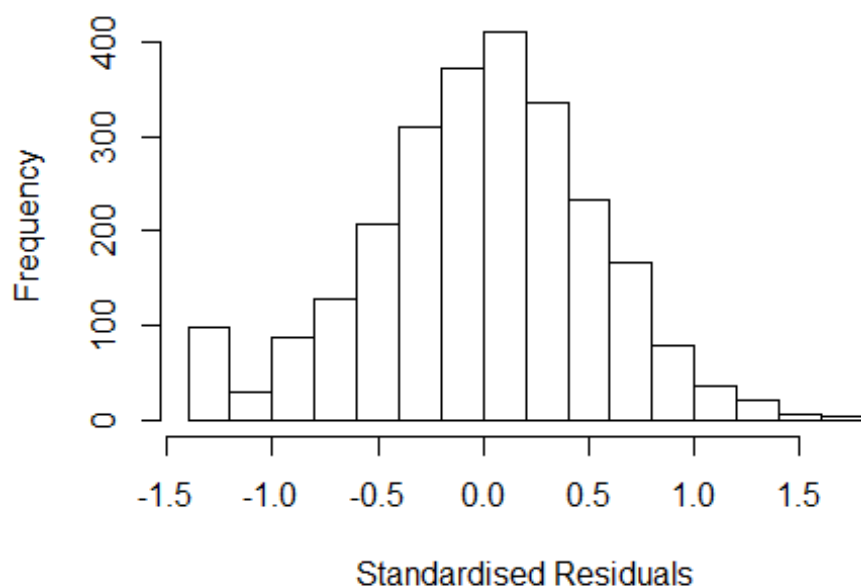
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.503 -0.337 0.018 0.336 2.058
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0720 0.0640 16.74 < 2e-16 ***
## FirstAuthorFemale1 -0.0548 0.0255 -2.15 0.0317 *
## LastAuthorFemale1 -0.0250 0.0381 -0.66 0.5117
## UniqueAuthors2 0.3231 0.0492 6.57 6.1e-11 ***
## UniqueAuthors3 0.3572 0.0487 7.34 3.0e-13 ***
## UniqueAuthors4 0.4309 0.0526 8.20 3.9e-16 ***
## UniqueAuthors5 0.4301 0.0566 7.60 4.2e-14 ***
## Year1997 -0.0477 0.0691 -0.69 0.4895
## Year1998 -0.1333 0.0702 -1.90 0.0578 .
## Year1999 -0.0893 0.0687 -1.30 0.1940
```

```

## Year2000          -0.0946      0.0881    -1.07    0.2833
## Year2001          -0.1672      0.0793    -2.11    0.0351 *
## Year2002          -0.1179      0.0745    -1.58    0.1137
## Year2003          -0.0689      0.0734    -0.94    0.3476
## Year2004          -0.1847      0.0709    -2.60    0.0093 **
## Year2005          -0.1222      0.0654    -1.87    0.0617 .
## Year2006          -0.1676      0.0695    -2.41    0.0159 *
## Year2007          -0.1915      0.0696    -2.75    0.0060 **
## Year2008          -0.1926      0.0596    -3.23    0.0013 **
## Year2009          -0.1928      0.0635    -3.04    0.0024 **
## Year2010          -0.2695      0.0599    -4.50    7.1e-06 ***
## Year2011          -0.2385      0.0579    -4.12    3.9e-05 ***
## Year2012          -0.2672      0.0584    -4.57    5.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.506
## Multiple R-squared:  0.0627, Adjusted R-squared:  0.0544
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 218 weights are ~= 1. The remaining 2309 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0607 0.8610 0.9510 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.96e-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 1.044 1 1.022
## LastAuthorFemale 1.049 1 1.024
## Year 1.095 16 1.003

```


Residuals from first and last author



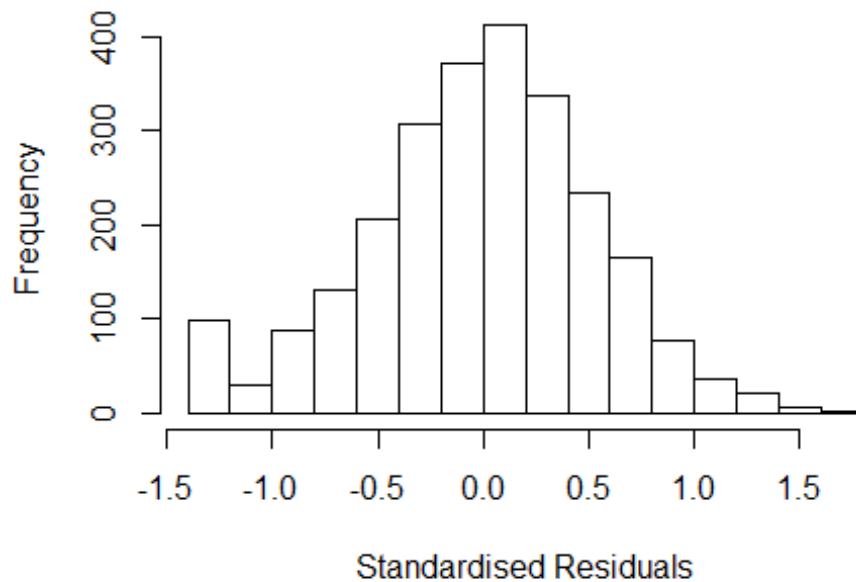
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.364 -0.336 0.012 0.340 1.790
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36352 0.04848 28.12 < 2e-16 ***
## FirstAuthorFemale1 -0.01507 0.02586 -0.58 0.56000
## LastAuthorFemale1 -0.00823 0.03936 -0.21 0.83434
## Year1997 -0.05680 0.07058 -0.80 0.42101
## Year1998 -0.11652 0.07147 -1.63 0.10315
## Year1999 -0.08457 0.07020 -1.20 0.22844
## Year2000 -0.09597 0.08921 -1.08 0.28214
## Year2001 -0.14302 0.07961 -1.80 0.07254 .
## Year2002 -0.08479 0.07597 -1.12 0.26446
## Year2003 -0.06724 0.07308 -0.92 0.35759
## Year2004 -0.16008 0.07238 -2.21 0.02708 *
## Year2005 -0.08625 0.06610 -1.30 0.19207
```

```

## Year2006      -0.13214    0.07095   -1.86  0.06266 .
## Year2007      -0.15612    0.07045   -2.22  0.02678 *
## Year2008      -0.15474    0.06084   -2.54  0.01103 *
## Year2009      -0.15359    0.06452   -2.38  0.01737 *
## Year2010      -0.22963    0.05963   -3.85  0.00012 ***
## Year2011      -0.18263    0.05807   -3.14  0.00168 **
## Year2012      -0.22473    0.05859   -3.84  0.00013 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.51
## Multiple R-squared:  0.0139, Adjusted R-squared:  0.00678
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 211 weights are ~= 1. The remaining 2316 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.193  0.860  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      3.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.046 1      1.023
## Year              1.046 16      1.001

```

Residuals from first author



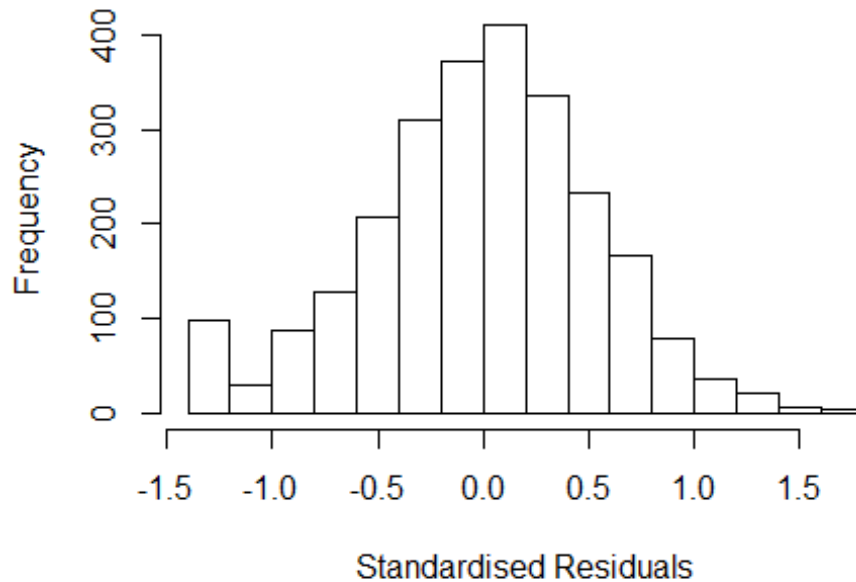
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.363 -0.335 0.012 0.339 1.791
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3634 0.0485 28.09 < 2e-16 ***
## FirstAuthorFemale1 -0.0155 0.0260 -0.60 0.55012
## Year1997 -0.0569 0.0706 -0.81 0.42008
## Year1998 -0.1169 0.0714 -1.64 0.10187
## Year1999 -0.0845 0.0702 -1.20 0.22866
## Year2000 -0.0967 0.0891 -1.09 0.27765
## Year2001 -0.1440 0.0792 -1.82 0.06931 .
## Year2002 -0.0848 0.0760 -1.12 0.26467
## Year2003 -0.0684 0.0727 -0.94 0.34713
## Year2004 -0.1610 0.0719 -2.24 0.02511 *
## Year2005 -0.0866 0.0660 -1.31 0.18964
## Year2006 -0.1323 0.0709 -1.87 0.06229 .
```

```

## Year2007          -0.1563      0.0704   -2.22  0.02642 *
## Year2008          -0.1551      0.0607   -2.55  0.01070 *
## Year2009          -0.1541      0.0644   -2.39  0.01670 *
## Year2010          -0.2304      0.0594   -3.88  0.00011 ***
## Year2011          -0.1834      0.0578   -3.17  0.00153 **
## Year2012          -0.2256      0.0584   -3.86  0.00012 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.51
## Multiple R-squared:  0.0138, Adjusted R-squared:  0.00716
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 213 weights are ~= 1. The remaining 2314 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.192  0.859  0.951  0.897  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.049 1          1.024
## Year            1.049 16          1.001

```

Residuals from last author



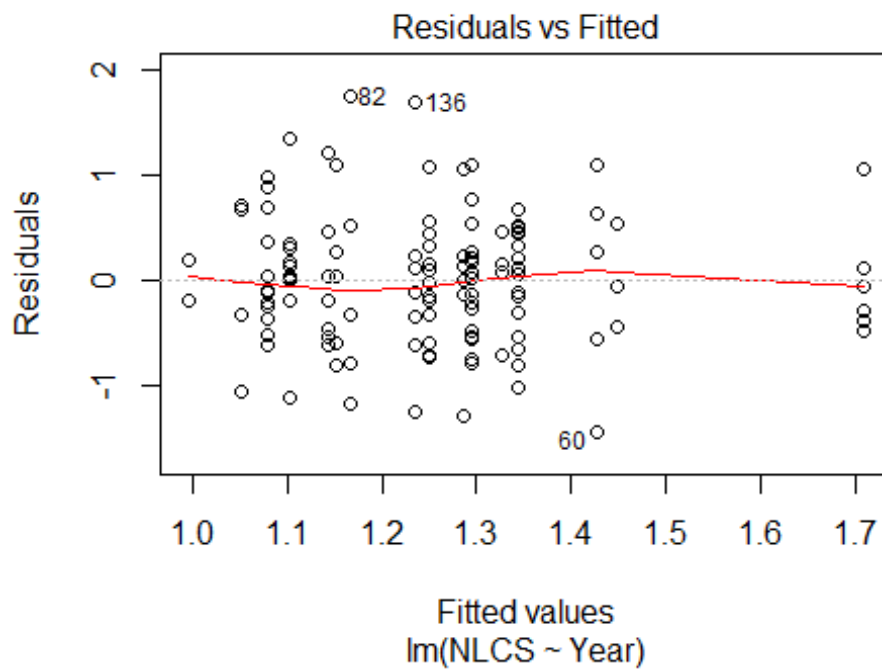
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3619 -0.3348  0.0135  0.3406  1.7921
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.36186    0.04855   28.05  < 2e-16 ***
## LastAuthorFemale1 -0.00985    0.03953   -0.25  0.80329
## Year1997        -0.05781    0.07045   -0.82  0.41196
## Year1998        -0.11779    0.07129   -1.65  0.09858 .
## Year1999        -0.08501    0.07008   -1.21  0.22523
## Year2000        -0.09597    0.08916   -1.08  0.28187
## Year2001        -0.14556    0.07935   -1.83  0.06671 .
## Year2002        -0.08670    0.07567   -1.15  0.25201
## Year2003        -0.06908    0.07303   -0.95  0.34429
## Year2004        -0.16168    0.07232   -2.24  0.02546 *
## Year2005        -0.08825    0.06599   -1.34  0.18124
## Year2006        -0.13320    0.07081   -1.88  0.06007 .
```

```

## Year2007          -0.15854      0.06990      -2.27   0.02341 *
## Year2008          -0.15704      0.06055      -2.59   0.00956 **
## Year2009          -0.15519      0.06433      -2.41   0.01591 *
## Year2010          -0.23138      0.05944      -3.89   0.00010 ***
## Year2011          -0.18366      0.05801      -3.17   0.00156 **
## Year2012          -0.22588      0.05847      -3.86   0.00011 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.51
## Multiple R-squared:  0.0137, Adjusted R-squared:  0.00706
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 216 weights are ~= 1. The remaining 2311 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.192  0.860  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      3.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2527"
## [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
##   24   19   20   19   24   13   22   13   22   26   35   33   20    5    6
## 2011 2012
##    6    1
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    6    6    5    5    8    4    8    5   11   13   17   19   15    4    2
## 2011 2012

```

```
##      4      0
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      4      3      4      5      6      3      7      5      9     12     14     18     11      3      2
## 2011 2012
##      2      0
## [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
```



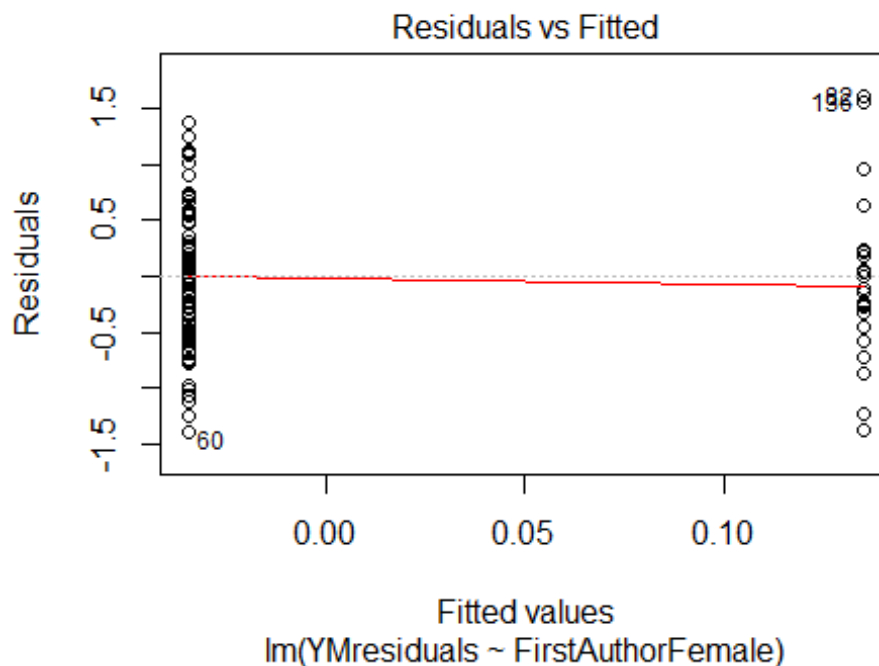
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.2, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 3.13016916014657"
## [1] "Male first author team size 2018 geometric mean: 3.5567021666885"

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

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.47213595499958"
## [1] "Male last author team size 2018 geometric mean: 3.19364008301972"

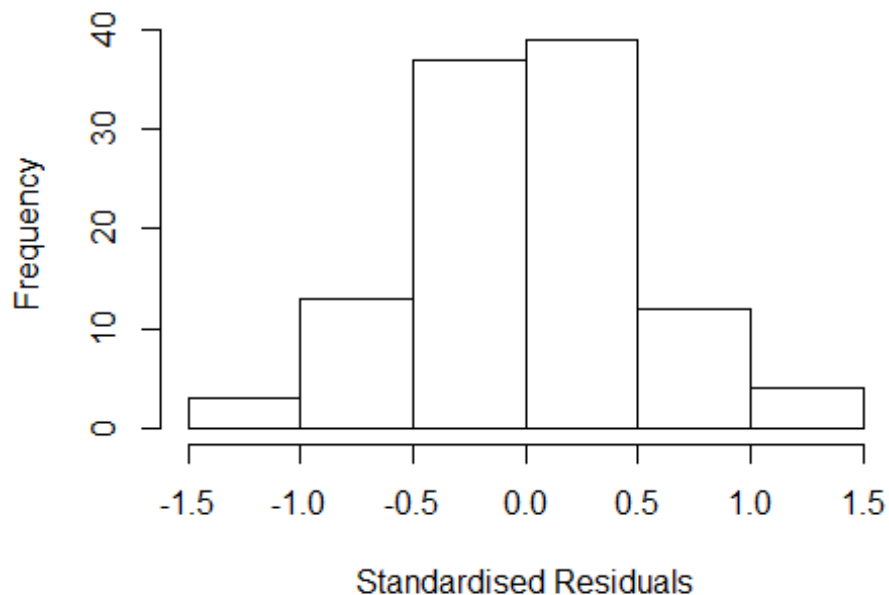
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	7.051	1	2.655
LastAuthorFemale	13.415	1	3.663
UniqueAuthors	276.984	4	2.020
Year	5334.959	15	1.331

Residuals from first and last author and team size



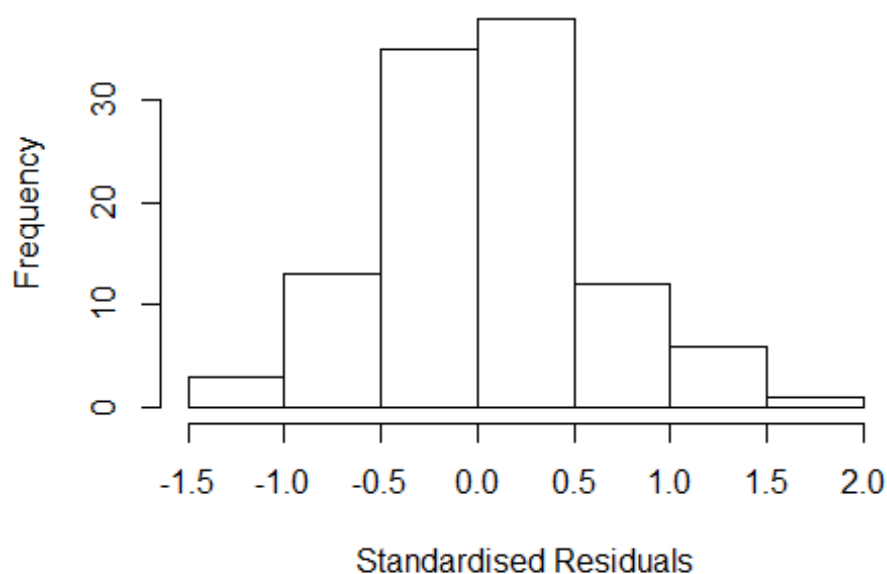
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1839 -0.2798  0.0154  0.2955  1.3975
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.058      0.196    5.39 6.1e-07 ***
## FirstAuthorFemale1  0.204      0.180    1.13  0.2617
## LastAuthorFemale1 -0.257      0.286   -0.90  0.3714
## UniqueAuthors2     0.491      0.206    2.38  0.0193 *
## UniqueAuthors3     0.662      0.216    3.07  0.0029 **
## UniqueAuthors4     0.665      0.323    2.06  0.0427 *
## UniqueAuthors5     0.127      0.291    0.44  0.6630
## Year1997          -0.228      0.185   -1.23  0.2202
## Year1998           0.126      0.494    0.25  0.7999
## Year1999          -0.227      0.451   -0.50  0.6155
```

```

## Year2000          -0.399      0.276    -1.44    0.1522
## Year2001          -0.696      0.646    -1.08    0.2842
## Year2002          -0.212      0.340    -0.62    0.5342
## Year2003          -0.377      0.377    -1.00    0.3194
## Year2004          -0.626      0.194    -3.23    0.0017 **
## Year2005          -0.338      0.202    -1.67    0.0982 .
## Year2006          -0.423      0.192    -2.20    0.0303 *
## Year2007          -0.253      0.181    -1.40    0.1657
## Year2008          -0.729      0.163    -4.47    2.4e-05 ***
## Year2009           0.100      0.313     0.32    0.7499
## Year2010          -0.586      0.342    -1.71    0.0905 .
## Year2011          -0.275      0.264    -1.04    0.3004
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.537
## Multiple R-squared:  0.261, Adjusted R-squared:  0.08
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 96 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.478  0.859   0.966   0.912   0.990   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.26e-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  7.125  1      2.669
## LastAuthorFemale   8.151  1      2.855
## Year               36.855 15      1.128

```

Residuals from first and last author



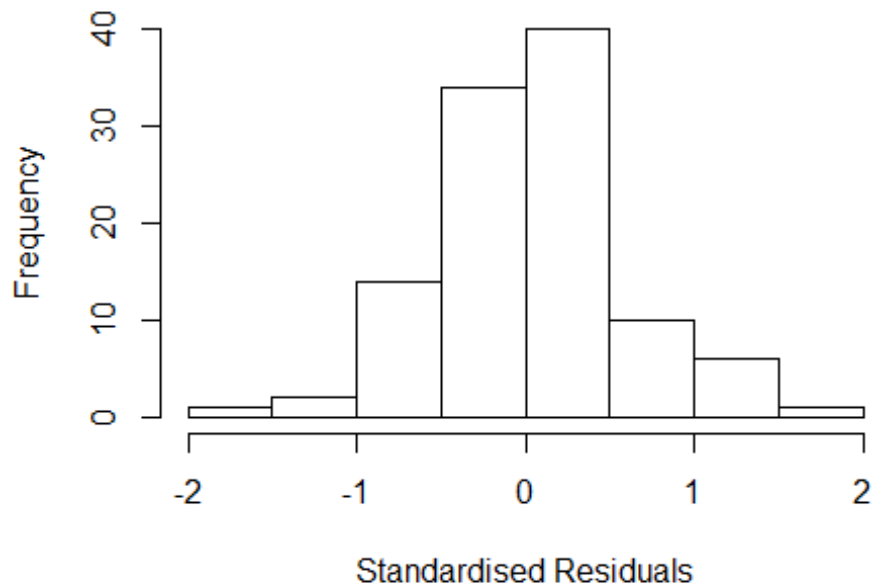
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2776 -0.3607 0.0322 0.2731 1.8071
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4992 0.0892 16.81 < 2e-16 ***
## FirstAuthorFemale1 0.2546 0.2449 1.04 0.3014
## LastAuthorFemale1 -0.2615 0.3708 -0.71 0.4825
## Year1997 -0.1781 0.1245 -1.43 0.1563
## Year1998 -0.2216 0.6869 -0.32 0.7477
## Year1999 -0.6488 0.7073 -0.92 0.3614
## Year2000 -0.5029 0.2679 -1.88 0.0638 .
## Year2001 -0.8603 0.6075 -1.42 0.1602
## Year2002 -0.2943 0.3532 -0.83 0.4070
## Year2003 -0.3951 0.3677 -1.07 0.2855
## Year2004 -0.5938 0.2148 -2.76 0.0069 **
## Year2005 -0.3181 0.1949 -1.63 0.1062
```

```

## Year2006          -0.3067      0.1469   -2.09   0.0396 *
## Year2007          -0.2383      0.1418   -1.68   0.0963 .
## Year2008          -0.5800      0.1299   -4.46  2.3e-05 ***
## Year2009          -0.0403      0.2680   -0.15   0.8809
## Year2010          -0.4973      0.4720   -1.05   0.2949
## Year2011          -0.0507      0.0937   -0.54   0.5898
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.543
## Multiple R-squared:  0.136, Adjusted R-squared:  -0.0271
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 99 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.245  0.883  0.960  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      9.26e-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 8.099 1          2.846
## Year              8.099 15          1.072

```

Residuals from first author

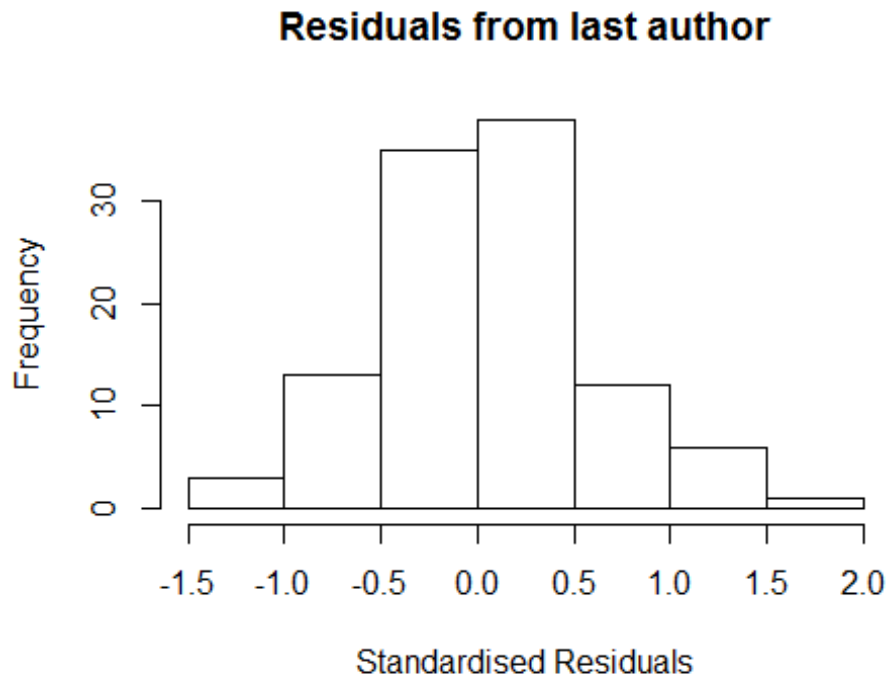


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5019 -0.3546 0.0444 0.2742 1.7727
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4942 0.0924 16.17 < 2e-16 ***
## FirstAuthorFemale1 0.2741 0.2662 1.03 0.3059
## Year1997 -0.1731 0.1269 -1.36 0.1757
## Year1998 -0.2214 0.6649 -0.33 0.7399
## Year1999 -0.6290 0.7065 -0.89 0.3756
## Year2000 -0.5535 0.2741 -2.02 0.0464 *
## Year2001 -0.8642 0.6036 -1.43 0.1556
## Year2002 -0.2665 0.3907 -0.68 0.4969
## Year2003 -0.3892 0.3753 -1.04 0.3025
## Year2004 -0.6336 0.1978 -3.20 0.0019 **
## Year2005 -0.3180 0.1933 -1.65 0.1033
## Year2006 -0.3240 0.1396 -2.32 0.0225 *
```

```

## Year2007          -0.2642      0.1414    -1.87    0.0650 .
## Year2008          -0.5765      0.1300    -4.44    2.6e-05 ***
## Year2009          -0.0352      0.2681    -0.13    0.8957
## Year2010          -0.7734      0.2512    -3.08    0.0027 **
## Year2011          -0.0457      0.0968    -0.47    0.6375
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.547
## Multiple R-squared:  0.132, Adjusted R-squared:  -0.0201
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 99 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.273  0.884  0.964  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      9.26e-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 5.18 1          2.276
## Year            5.18 15          1.056

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2834 -0.3768 0.0316 0.3027 2.1445
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5623 0.0995 15.70 < 2e-16 ***
## LastAuthorFemale1 -0.2649 0.2809 -0.94 0.34817
## Year1997 -0.2412 0.1321 -1.83 0.07113 .
## Year1998 -0.2789 0.8208 -0.34 0.73484
## Year1999 -0.7948 0.5749 -1.38 0.17019
## Year2000 -0.5033 0.2742 -1.84 0.06973 .
## Year2001 -0.7705 0.5665 -1.36 0.17720
## Year2002 -0.3804 0.2715 -1.40 0.16463
## Year2003 -0.4554 0.3685 -1.24 0.21972
## Year2004 -0.5187 0.2183 -2.38 0.01959 *
## Year2005 -0.3090 0.1897 -1.63 0.10670
## Year2006 -0.3058 0.1701 -1.80 0.07553 .
```

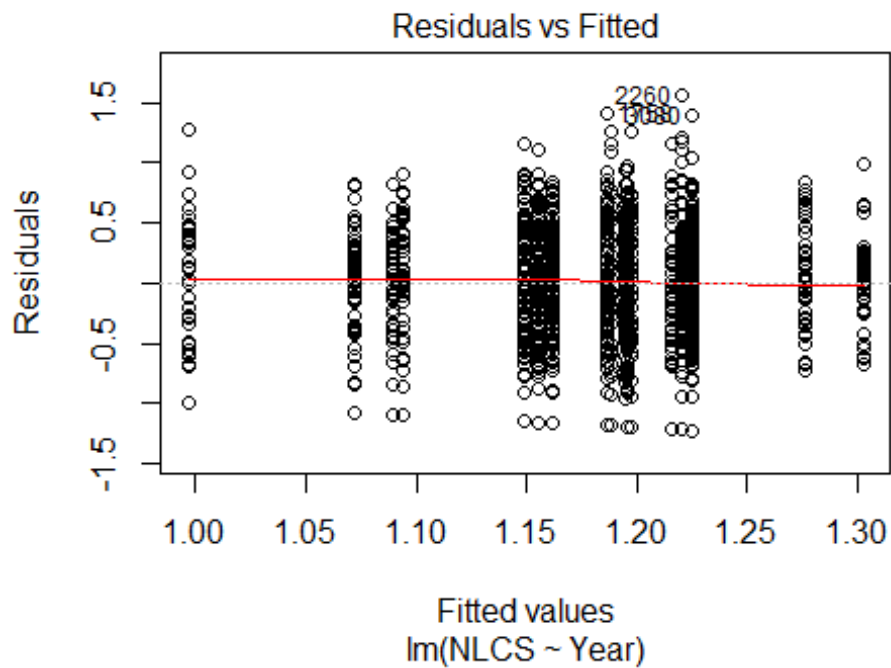
```

## Year2007          -0.2622      0.1572    -1.67   0.09877 .
## Year2008          -0.6191      0.1544    -4.01   0.00012 ***
## Year2009          -0.1031      0.2691    -0.38   0.70247
## Year2010          -0.3024      0.3300    -0.92   0.36183
## Year2011          -0.1138      0.1036    -1.10   0.27478
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.114, Adjusted R-squared:  -0.0415
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 100 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.103  0.907  0.958  0.905  0.990  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.26e-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: 108"
## [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
##  118  126  104  102   97   93  104   88  120  174  206  200  219  240  303
## 2011 2012
##  345  332
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   39   52   36   48   42   37   48   45   63   89  111  104  115  146  182
## 2011 2012
##  205  208

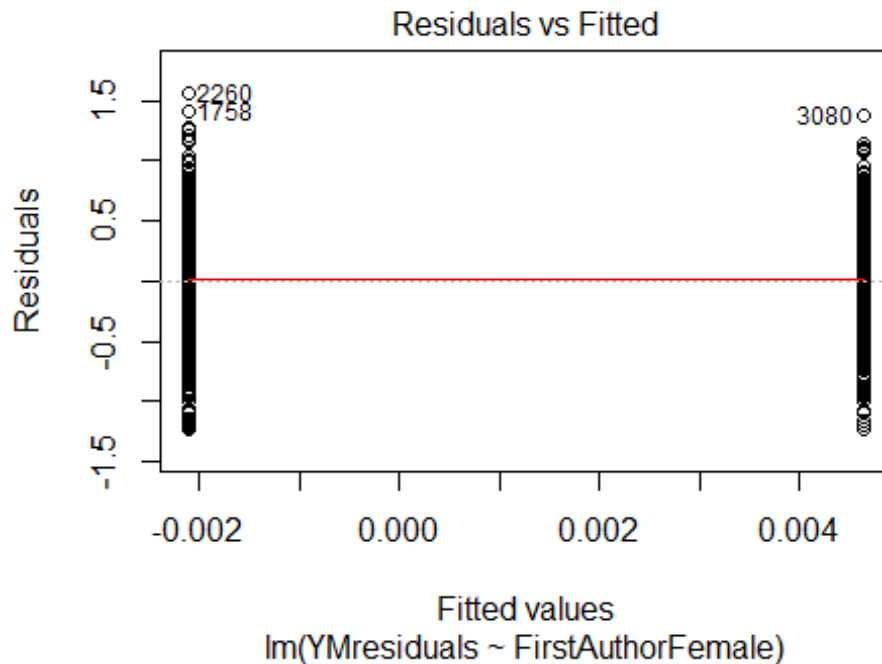
```



```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   35   43   31   39   39   33   36   37   54   74   99   83   99  125  156
## 2011 2012
##  156  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 = 33, df = 16, p-value = 0.008
```

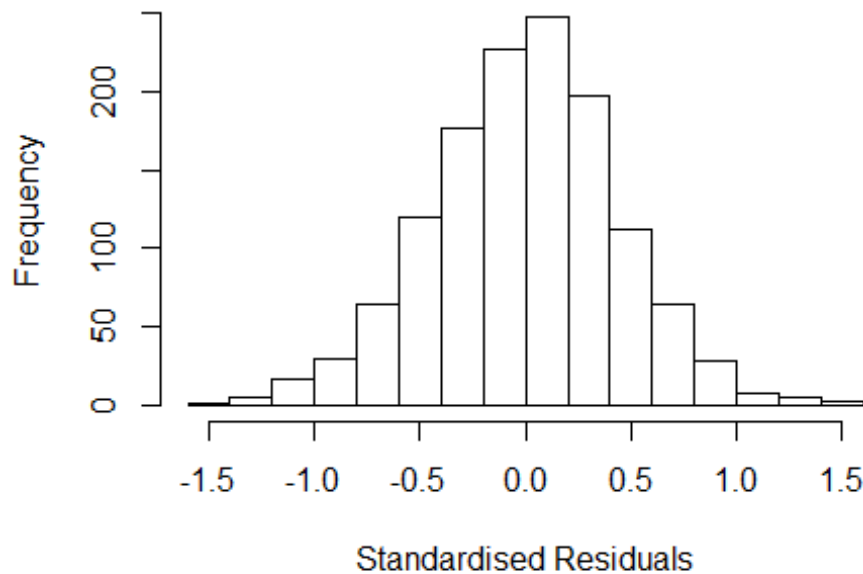


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.6, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 4.20743871836068"
## [1] "Male first author team size 2018 geometric mean: 3.93375263893209"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.82131764125252"
## [1] "Male last author team size 2018 geometric mean: 3.84017953012043"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "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  1.078 1          1.038
## UniqueAuthors     1.447 4          1.047
## Year               1.623 16         1.015
```

Residuals from first and last author and team size



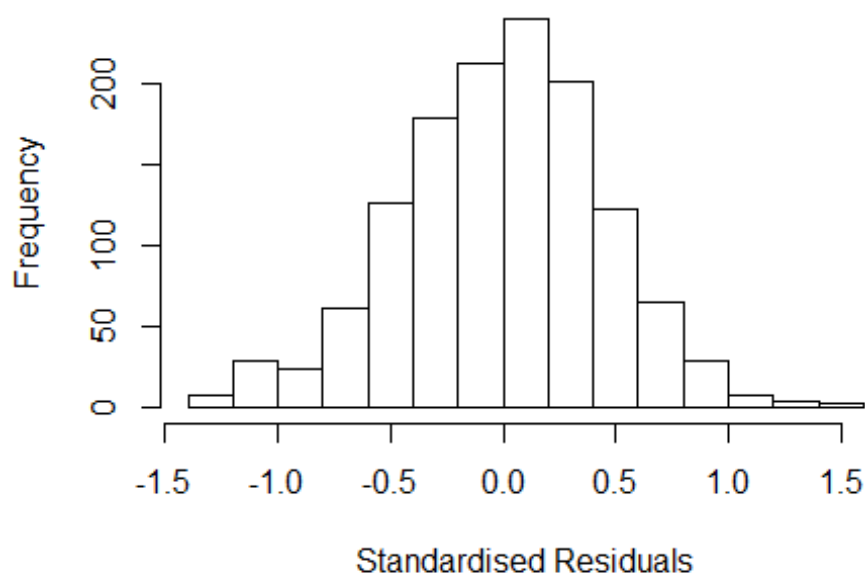
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4347 -0.2851 0.0107 0.2761 1.4955
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9754 0.1022 9.54 < 2e-16 ***
## FirstAuthorFemale1 -0.0274 0.0273 -1.00 0.3164
## LastAuthorFemale1 0.0611 0.0299 2.04 0.0415 *
## UniqueAuthors2 0.3509 0.0830 4.23 2.5e-05 ***
## UniqueAuthors3 0.3396 0.0810 4.19 2.9e-05 ***
## UniqueAuthors4 0.4496 0.0825 5.45 6.1e-08 ***
## UniqueAuthors5 0.4289 0.0811 5.29 1.5e-07 ***
## Year1997 -0.0514 0.1119 -0.46 0.6459
## Year1998 -0.1638 0.1287 -1.27 0.2032
## Year1999 -0.1649 0.1040 -1.58 0.1133
```

```

## Year2000          -0.2445      0.0908    -2.69    0.0072 **
## Year2001          -0.2508      0.1123    -2.23    0.0257 *
## Year2002          -0.1249      0.1080    -1.16    0.2479
## Year2003          -0.2132      0.1096    -1.94    0.0521 .
## Year2004          -0.2606      0.0855    -3.05    0.0023 **
## Year2005          -0.1391      0.0827    -1.68    0.0927 .
## Year2006          -0.1787      0.0805    -2.22    0.0267 *
## Year2007          -0.1345      0.0891    -1.51    0.1316
## Year2008          -0.1878      0.0815    -2.30    0.0214 *
## Year2009          -0.1198      0.0792    -1.51    0.1307
## Year2010          -0.1960      0.0740    -2.65    0.0082 **
## Year2011          -0.1429      0.0742    -1.93    0.0543 .
## Year2012          -0.1695      0.0749    -2.26    0.0237 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.426
## Multiple R-squared:  0.062, Adjusted R-squared:  0.0459
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1208 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.192  0.866  0.953   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          7.65e-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 1.099 1          1.048
## LastAuthorFemale  1.078 1          1.038
## Year              1.122 16          1.004

```

Residuals from first and last author



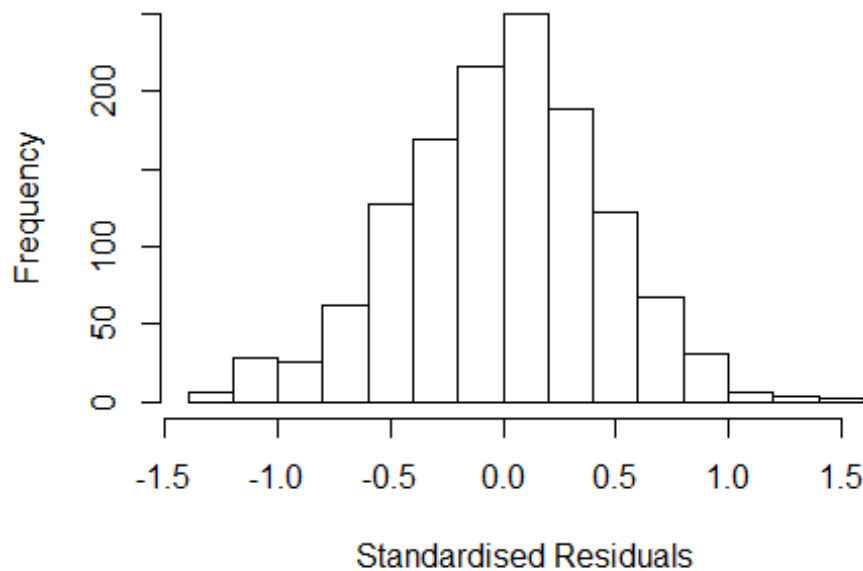
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2613 -0.2869  0.0119  0.2820  1.5400
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2980     0.0627   20.70  <2e-16 ***
## FirstAuthorFemale1 -0.0118     0.0277   -0.43    0.671
## LastAuthorFemale1  0.0616     0.0306    2.01    0.045 *
## Year1997          -0.1062     0.1053   -1.01    0.313
## Year1998          -0.1851     0.1345   -1.38    0.169
## Year1999          -0.1374     0.1031   -1.33    0.183
## Year2000          -0.2061     0.0918   -2.24    0.025 *
## Year2001          -0.2639     0.1307   -2.02    0.044 *
## Year2002          -0.0951     0.1057   -0.90    0.368
## Year2003          -0.1631     0.1114   -1.46    0.143
## Year2004          -0.2138     0.0855   -2.50    0.013 *
## Year2005          -0.1088     0.0804   -1.35    0.176
```

```

## Year2006          -0.1204      0.0773   -1.56    0.120
## Year2007          -0.0774      0.0858   -0.90    0.367
## Year2008          -0.1325      0.0784   -1.69    0.091 .
## Year2009          -0.0580      0.0740   -0.78    0.433
## Year2010          -0.1341      0.0708   -1.89    0.058 .
## Year2011          -0.0983      0.0715   -1.37    0.170
## Year2012          -0.1100      0.0711   -1.55    0.122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.0146, Adjusted R-squared:  0.000841
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1187 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.180  0.865  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      7.65e-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 1.072 1      1.036
## Year              1.072 16      1.002

```

Residuals from first author



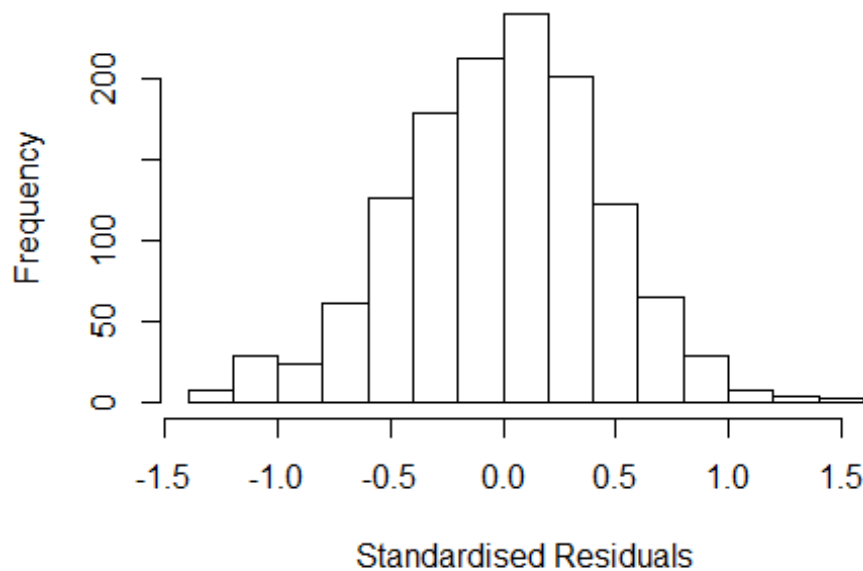
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2518 -0.2894 0.0103 0.2782 1.5282
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31160 0.06301 20.82 <2e-16 ***
## FirstAuthorFemale1 -0.00175 0.02734 -0.06 0.949
## Year1997 -0.11426 0.10438 -1.09 0.274
## Year1998 -0.19446 0.13472 -1.44 0.149
## Year1999 -0.14622 0.10403 -1.41 0.160
## Year2000 -0.21032 0.09207 -2.28 0.023 *
## Year2001 -0.26536 0.13247 -2.00 0.045 *
## Year2002 -0.10173 0.10616 -0.96 0.338
## Year2003 -0.16605 0.11029 -1.51 0.132
## Year2004 -0.21591 0.08642 -2.50 0.013 *
## Year2005 -0.11316 0.08120 -1.39 0.164
## Year2006 -0.12748 0.07782 -1.64 0.102
```

```

## Year2007      -0.08284    0.08635   -0.96    0.338
## Year2008      -0.13918    0.07887   -1.76    0.078 .
## Year2009      -0.05975    0.07449   -0.80    0.423
## Year2010      -0.13554    0.07140   -1.90    0.058 .
## Year2011      -0.09872    0.07203   -1.37    0.171
## Year2012      -0.11393    0.07157   -1.59    0.112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.0116, Adjusted R-squared:  -0.00148
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1189 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.188  0.863  0.949  0.899  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.65e-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 1.052 1          1.026
## Year            1.052 16          1.002

```


Residuals from last author



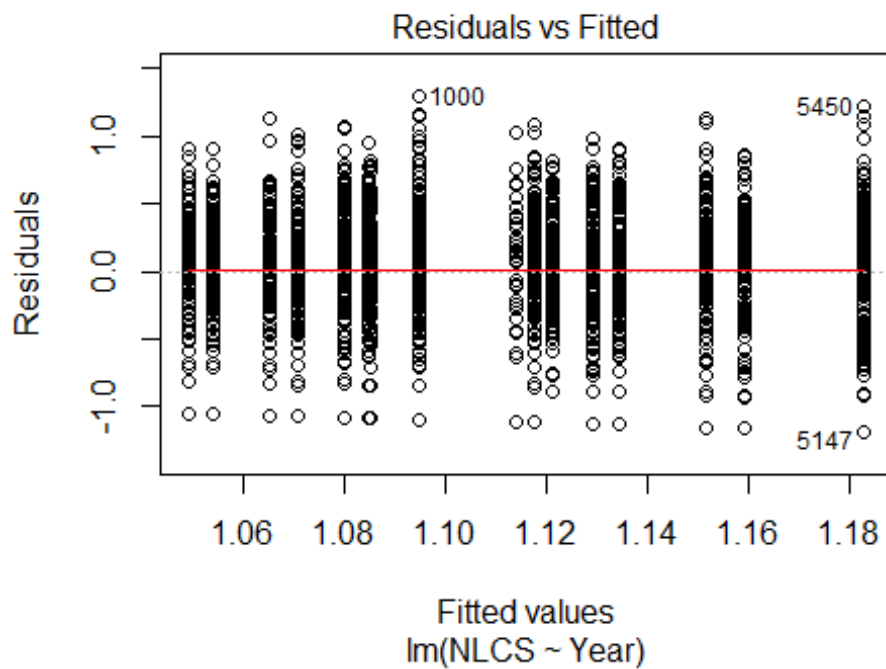
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2547 -0.2892 0.0122 0.2811 1.5440
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2968 0.0627 20.68 <2e-16 ***
## LastAuthorFemale1 0.0591 0.0303 1.95 0.051 .
## Year1997 -0.1088 0.1052 -1.03 0.301
## Year1998 -0.1883 0.1347 -1.40 0.162
## Year1999 -0.1390 0.1028 -1.35 0.177
## Year2000 -0.2069 0.0918 -2.25 0.024 *
## Year2001 -0.2646 0.1308 -2.02 0.043 *
## Year2002 -0.0981 0.1054 -0.93 0.352
## Year2003 -0.1664 0.1110 -1.50 0.134
## Year2004 -0.2163 0.0847 -2.55 0.011 *
## Year2005 -0.1111 0.0798 -1.39 0.164
## Year2006 -0.1225 0.0770 -1.59 0.112
```

```

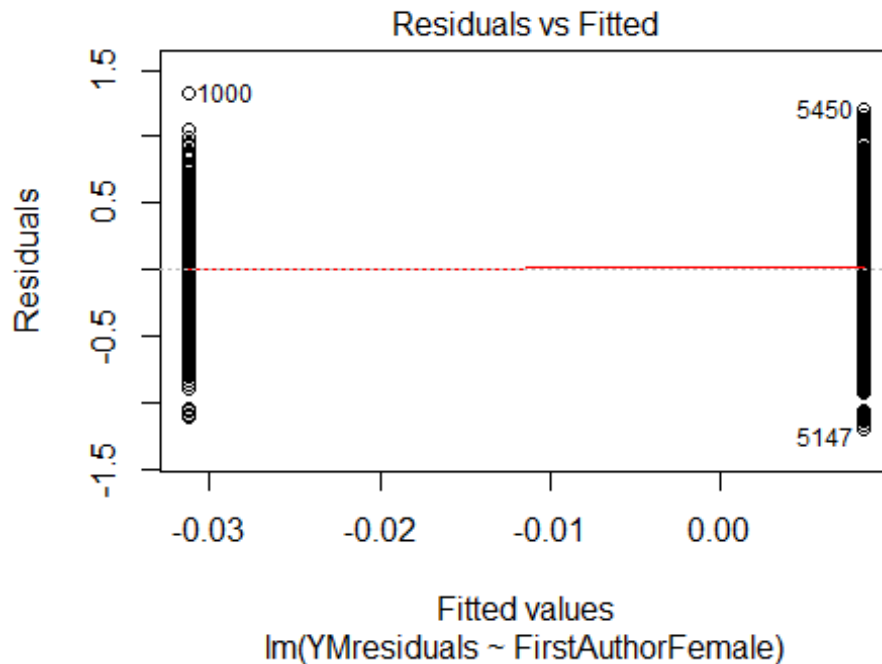
## Year2007          -0.0801      0.0855   -0.94    0.349
## Year2008          -0.1338      0.0782   -1.71    0.087 .
## Year2009          -0.0609      0.0734   -0.83    0.407
## Year2010          -0.1355      0.0706   -1.92    0.055 .
## Year2011          -0.1013      0.0707   -1.43    0.153
## Year2012          -0.1113      0.0710   -1.57    0.117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0145, Adjusted R-squared:  0.00146
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1185 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.178  0.865  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      7.65e-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: 1307"
## [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
##  359  256  267  228  239  239  243  219  242  249  259  235  280  326  331
## 2011 2012
##  347  289
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  235  157  174  154  103   45  181  159  161  154  160  136  176  209  229
## 2011 2012

```

```
## 235 202
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 180 127 142 119 88 31 160 129 141 131 125 120 146 181 191
## 2011 2012
## 200 162
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 39, df = 16, p-value = 0.001
```

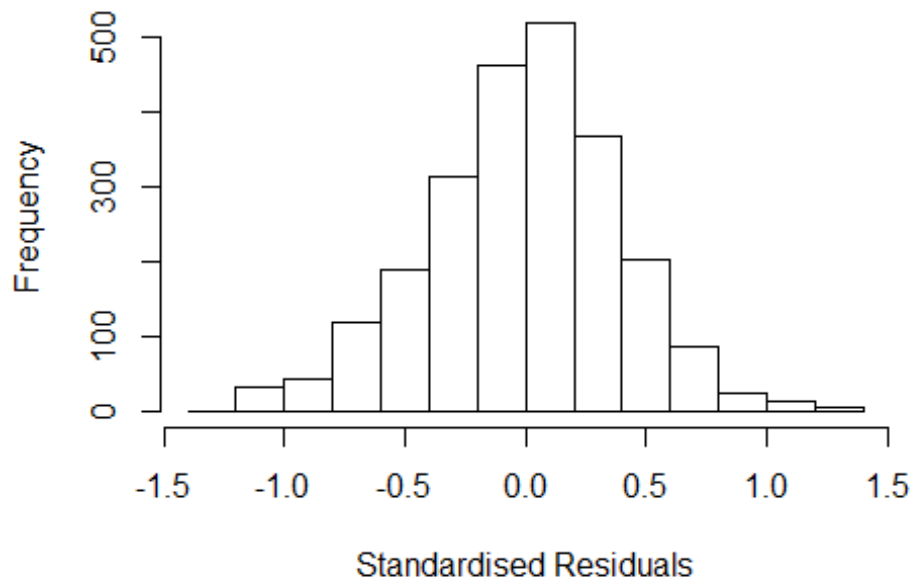


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.9, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.79595871845508"
## [1] "Male first author team size 2018 geometric mean: 3.56343764221344"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.33548566357632"
## [1] "Male last author team size 2018 geometric mean: 3.48601687718908"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2600, 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 1.041 1 1.020
## LastAuthorFemale 1.023 1 1.012
## UniqueAuthors 1.157 4 1.018
## Year 1.207 16 1.006
```

Residuals from first and last author and team size



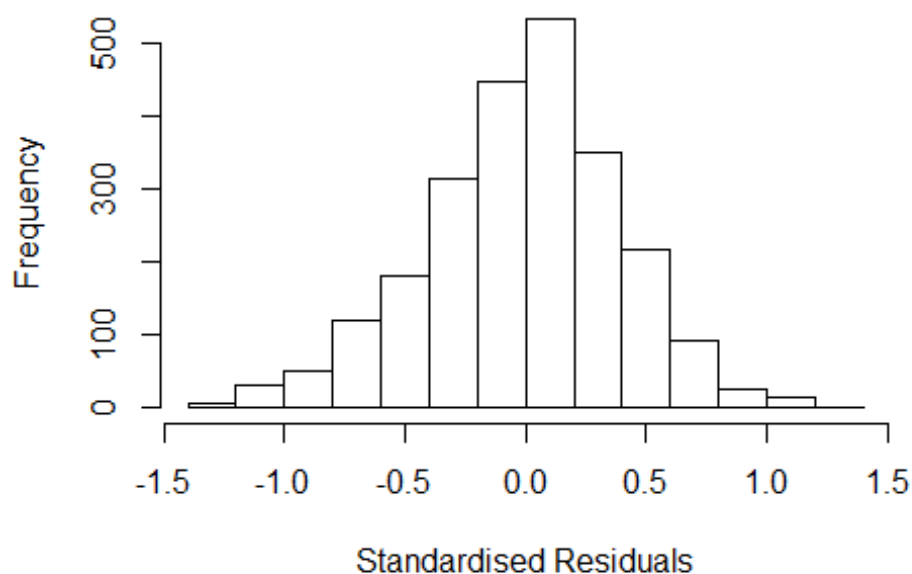
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2601 -0.2483  0.0146  0.2541  1.2804
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.97122    0.05774   16.82 < 2e-16 ***
## FirstAuthorFemale1 -0.05404    0.02019   -2.68  0.00749 **
## LastAuthorFemale1 -0.01179    0.02763   -0.43  0.66970
## UniqueAuthors2     0.13769    0.05312    2.59  0.00960 **
## UniqueAuthors3     0.15894    0.05284    3.01  0.00266 **
## UniqueAuthors4     0.20814    0.05457    3.81  0.00014 ***
## UniqueAuthors5     0.20657    0.05438    3.80  0.00015 ***
## Year1997          -0.01600    0.04234   -0.38  0.70556
## Year1998           0.00844    0.04635    0.18  0.85555
## Year1999           0.01742    0.05167    0.34  0.73602
```

```

## Year2000      -0.02823    0.04824   -0.59  0.55856
## Year2001      0.03771    0.08638    0.44  0.66245
## Year2002     -0.07000    0.04042   -1.73  0.08345 .
## Year2003     -0.00468    0.04572   -0.10  0.91852
## Year2004     -0.06472    0.04380   -1.48  0.13962
## Year2005     -0.05964    0.04505   -1.32  0.18571
## Year2006     -0.03288    0.04748   -0.69  0.48874
## Year2007      0.06335    0.05054    1.25  0.21013
## Year2008      0.06306    0.04420    1.43  0.15375
## Year2009      0.05316    0.04475    1.19  0.23496
## Year2010      0.01908    0.04151    0.46  0.64585
## Year2011     -0.03276    0.04272   -0.77  0.44333
## Year2012      0.08071    0.04524    1.78  0.07454 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0288, Adjusted R-squared:  0.0197
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 181 weights are ~= 1. The remaining 2192 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.228  0.863  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      4.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1      1.018
## LastAuthorFemale  1.021 1      1.011
## Year              1.050 16      1.002

```

Residuals from first and last author



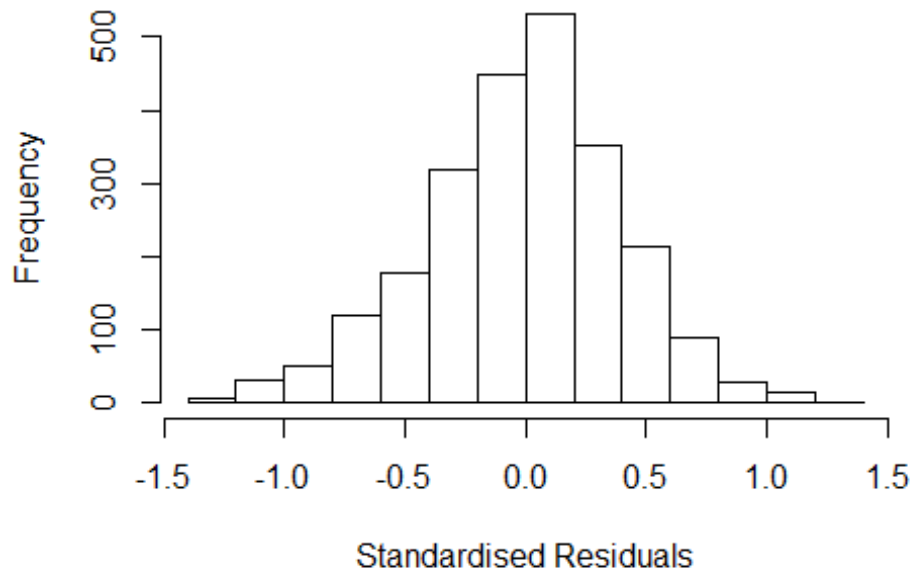
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2279 -0.2492  0.0106  0.2524  1.2918
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12457    0.03027   37.15  <2e-16 ***
## FirstAuthorFemale1 -0.05187    0.02023   -2.56   0.010 *
## LastAuthorFemale1 -0.01152    0.02795   -0.41   0.680
## Year1997        -0.01808    0.04250   -0.43   0.671
## Year1998         0.00862    0.04594    0.19   0.851
## Year1999         0.02355    0.05168    0.46   0.649
## Year2000        -0.02284    0.04924   -0.46   0.643
## Year2001         0.03916    0.08736    0.45   0.654
## Year2002        -0.05416    0.04010   -1.35   0.177
## Year2003         0.00997    0.04561    0.22   0.827
## Year2004        -0.05234    0.04368   -1.20   0.231
## Year2005        -0.04510    0.04534   -0.99   0.320
```

```

## Year2006      -0.02500    0.04776   -0.52    0.601
## Year2007      0.08167    0.05018    1.63    0.104
## Year2008      0.08043    0.04396    1.83    0.067 .
## Year2009      0.06379    0.04490    1.42    0.156
## Year2010      0.03616    0.04156    0.87    0.384
## Year2011     -0.01269    0.04262   -0.30    0.766
## Year2012      0.10338    0.04513    2.29    0.022 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0173, Adjusted R-squared:  0.00975
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 198 weights are ~= 1. The remaining 2175 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.223  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      4.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000      0
##      psi            subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1      1.016
## Year              1.032 16      1.001

```


Residuals from first author



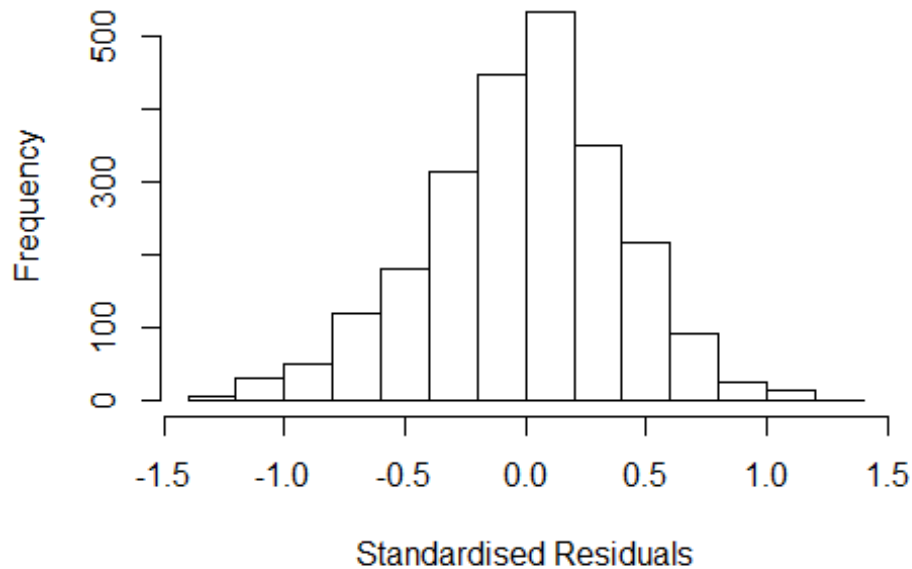
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2265 -0.2477 0.0112 0.2539 1.2935
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12369 0.03018 37.23 <2e-16 ***
## FirstAuthorFemale1 -0.05243 0.02018 -2.60 0.0094 **
## Year1997 -0.01811 0.04247 -0.43 0.6699
## Year1998 0.00778 0.04591 0.17 0.8655
## Year1999 0.02324 0.05175 0.45 0.6534
## Year2000 -0.02256 0.04919 -0.46 0.6465
## Year2001 0.03972 0.08748 0.45 0.6499
## Year2002 -0.05395 0.04011 -1.35 0.1787
## Year2003 0.01014 0.04562 0.22 0.8242
## Year2004 -0.05201 0.04367 -1.19 0.2338
## Year2005 -0.04513 0.04533 -1.00 0.3195
## Year2006 -0.02507 0.04781 -0.52 0.6000
```

```

## Year2007          0.08149    0.05020    1.62    0.1047
## Year2008          0.08038    0.04395    1.83    0.0676 .
## Year2009          0.06337    0.04482    1.41    0.1576
## Year2010          0.03575    0.04155    0.86    0.3897
## Year2011         -0.01300    0.04261   -0.31    0.7604
## Year2012          0.10277    0.04517    2.28    0.0230 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0172, Adjusted R-squared:  0.0101
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 199 weights are ~= 1. The remaining 2174 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.222  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      4.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.016 1          1.008
## Year              1.016 16          1.001

```

Residuals from last author



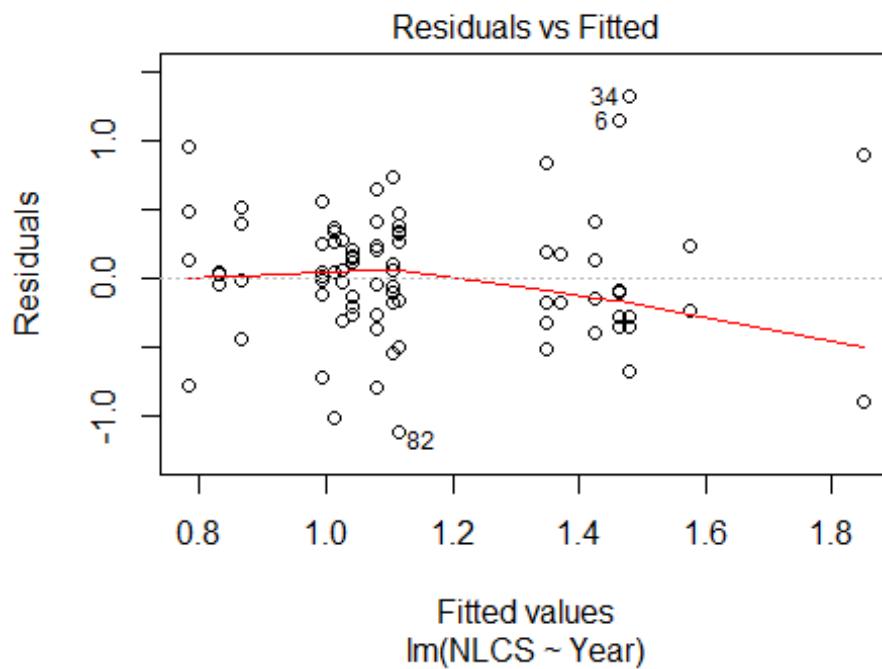
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2127 -0.2477 0.0138 0.2494 1.2474
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11589 0.03002 37.18 <2e-16 ***
## LastAuthorFemale1 -0.01630 0.02799 -0.58 0.560
## Year1997 -0.02267 0.04265 -0.53 0.595
## Year1998 0.00964 0.04624 0.21 0.835
## Year1999 0.02470 0.05174 0.48 0.633
## Year2000 -0.02297 0.04923 -0.47 0.641
## Year2001 0.03501 0.08677 0.40 0.687
## Year2002 -0.05504 0.04003 -1.38 0.169
## Year2003 0.01322 0.04552 0.29 0.771
## Year2004 -0.05526 0.04369 -1.26 0.206
## Year2005 -0.04866 0.04566 -1.07 0.287
## Year2006 -0.02754 0.04796 -0.57 0.566
```

```

## Year2007      0.07497      0.05036      1.49      0.137
## Year2008      0.07792      0.04440      1.75      0.079 .
## Year2009      0.06067      0.04484      1.35      0.176
## Year2010      0.03489      0.04164      0.84      0.402
## Year2011     -0.01582      0.04262     -0.37      0.711
## Year2012      0.09685      0.04505      2.15      0.032 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0144, Adjusted R-squared:  0.00728
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 197 weights are ~= 1. The remaining 2176 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.259  0.865  0.950  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      4.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2373"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1504"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    7    3    7    8   11   10    4    5    8   14   12   12    6    6
## 2011 2012
##    6    7
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2    6    2    7    5    4    4    2    5    5    8    8    7    5    5
## 2011 2012

```

```
##      4      7
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      2      6      2      6      4      3      4      2      3      5      7      8      7      5      4
## 2011 2012
##      4      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 = 31, df = 16, p-value = 0.01
```



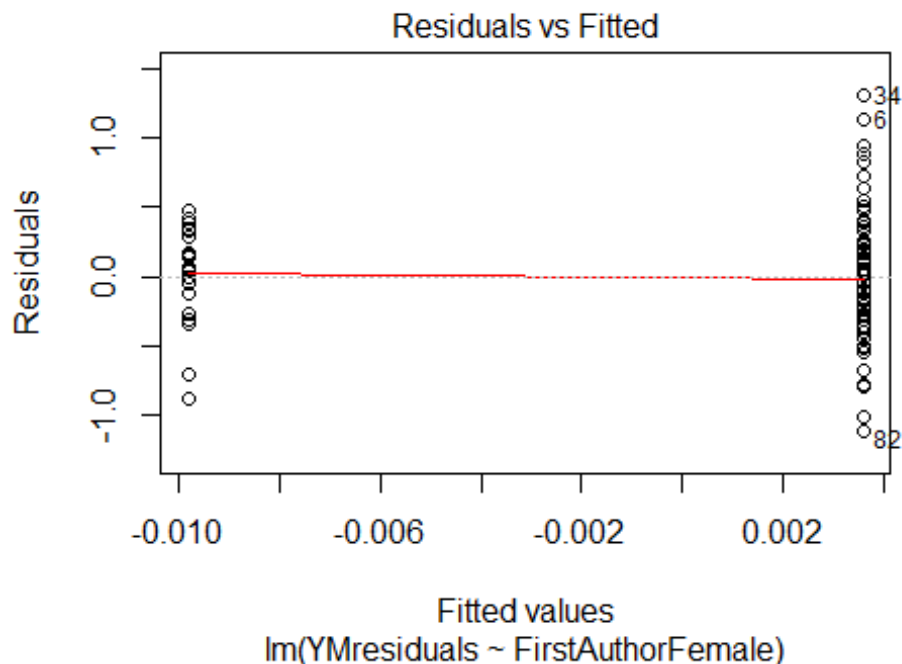
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.8, df = 1, p-value = 0.05
##
## [1] "Female first author team size 2018 geometric mean: 3.47602664488645"
## [1] "Male first author team size 2018 geometric mean: 4"
##
## Wilcoxon 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: 2.82842712474619"
## [1] "Male last author team size 2018 geometric mean: 4.58257569495584"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

## Warning in lmrob.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 outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'Year2003'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.
```



```
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 10.50 1          3.241
```

```

## LastAuthorFemale 12.85 1 3.585
## Year 20.27 16 1.099

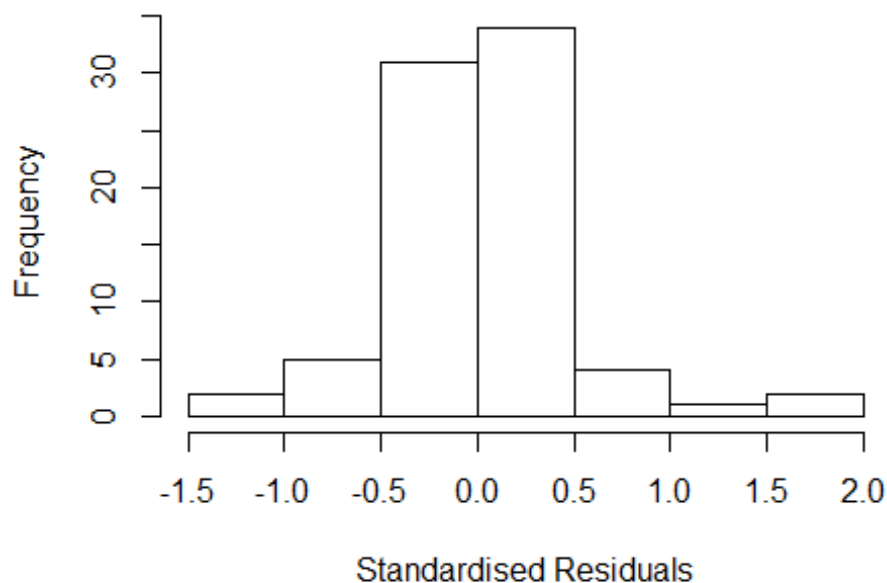
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1334 -0.1962 0.0117 0.2281 1.9343
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37100 0.12884 10.64 1.9e-15 ***
## FirstAuthorFemale1 0.12969 0.08790 1.48 0.1453
## LastAuthorFemale1 -0.06995 0.09751 -0.72 0.4760
## Year1997 -0.08168 0.17750 -0.46 0.6471
## Year1998 0.23997 0.20489 1.17 0.2461
## Year1999 -0.29543 0.17159 -1.72 0.0903 .
## Year2000 -0.24078 0.42476 -0.57 0.5729
## Year2001 -0.44137 0.16610 -2.66 0.0101 *
## Year2002 0.00892 0.19019 0.05 0.9627
## Year2003 -0.47674 0.15773 -3.02 0.0037 **
## Year2004 -0.42012 0.48232 -0.87 0.3872
## Year2005 -0.11007 0.28391 -0.39 0.6996
## Year2006 -0.23761 0.25635 -0.93 0.3577
## Year2007 -0.27073 0.21045 -1.29 0.2032
## Year2008 -0.29823 0.18591 -1.60 0.1139
## Year2009 -0.56347 0.13563 -4.15 0.0001 ***
## Year2010 -0.40161 0.28967 -1.39 0.1707
## Year2011 -0.39162 0.16371 -2.39 0.0199 *
## Year2012 -0.36330 0.15514 -2.34 0.0225 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared: 0.196, Adjusted R-squared: -0.0449
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(22,28) are outliers with |weight| = 0 ( < 0.0013);
## 7 weights are ~1. The remaining 70 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.200 0.929 0.967 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          1.27e-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"

## Warning in outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'Year2003'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.
```

Residuals from first and last author



```
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 6.154 1          2.481
## Year              6.154 16          1.058

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS          Year          OneField Fields          residuals
```



```

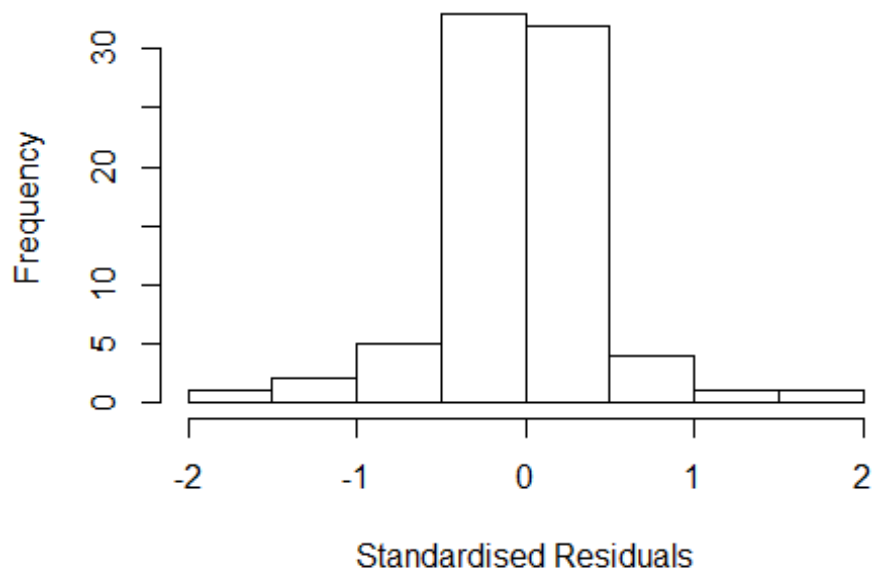
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8924 -0.1939 -0.0124  0.2115  1.8857
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.37100    0.12914   10.62 1.7e-15 ***
## FirstAuthorFemale1 0.10236    0.07424    1.38  0.1730
## Year1997         -0.11718    0.16404   -0.71  0.4777
## Year1998          0.20500    0.22358    0.92  0.3628
## Year1999         -0.30022    0.16921   -1.77  0.0810 .
## Year2000         -0.22601    0.42443   -0.53  0.5963
## Year2001         -0.46268    0.15565   -2.97  0.0042 **
## Year2002          0.00308    0.19084    0.02  0.9872
## Year2003          1.37300    0.12914   10.63 1.6e-15 ***
## Year2004         -0.45369    0.46051   -0.99  0.3284
## Year2005         -0.12741    0.28845   -0.44  0.6603
## Year2006         -0.22830    0.25795   -0.89  0.3796
## Year2007         -0.28743    0.22582   -1.27  0.2079
## Year2008         -0.30754    0.18531   -1.66  0.1021
## Year2009         -0.58059    0.13357   -4.35 5.3e-05 ***
## Year2010         -0.40789    0.27431   -1.49  0.1422
## Year2011         -0.39440    0.16054   -2.46  0.0169 *
## Year2012         -0.37163    0.15178   -2.45  0.0172 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.356, Adjusted R-squared:  0.176
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 2 observations c(22,29) are outliers with |weight| = 0 ( < 0.0013);
## 8 weights are ~ 1. The remaining 69 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.148  0.914  0.965  0.899  0.986  0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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"

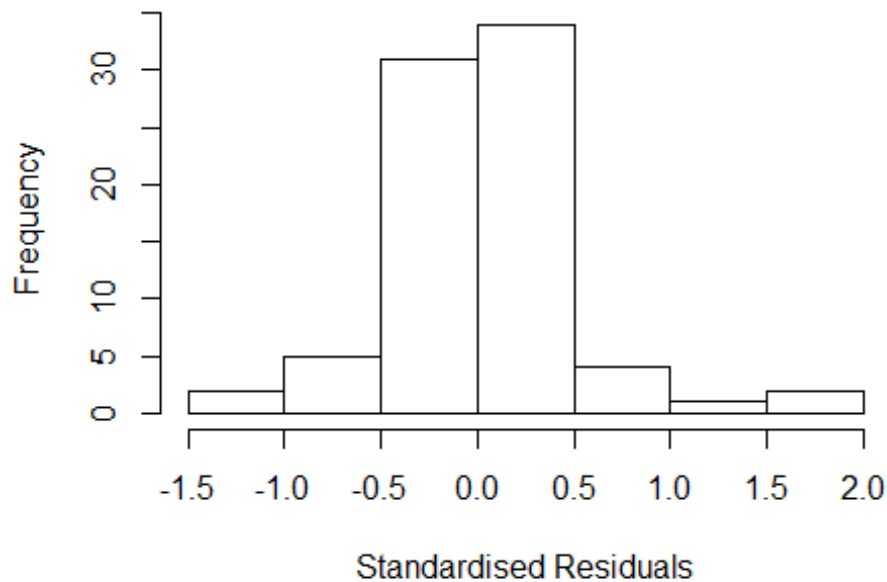
## Warning in outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'Year2003'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.
```

Residuals from first author



```
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 10.58 1 3.252
## Year 10.58 16 1.076
```

Residuals from last author



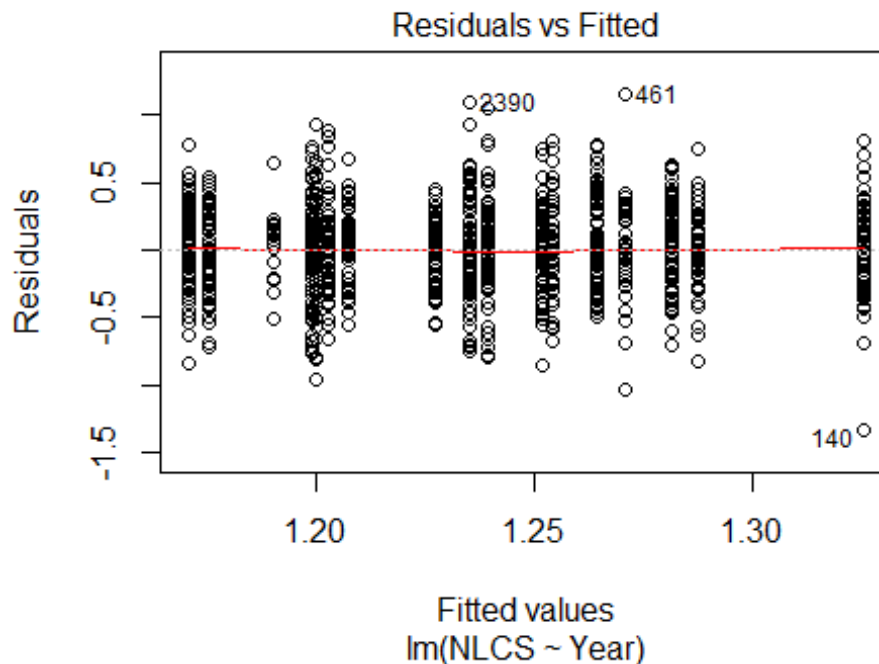
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2067 -0.1832 0.0216 0.1989 1.8445
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3710 0.1290 10.63 1.6e-15 ***
## LastAuthorFemale1 -0.0199 0.0851 -0.23 0.81562
## Year1997 -0.0797 0.1758 -0.45 0.65199
## Year1998 0.2150 0.2209 0.97 0.33425
## Year1999 -0.2619 0.1602 -1.63 0.10723
## Year2000 -0.2327 0.4257 -0.55 0.58663
## Year2001 -0.4015 0.1849 -2.17 0.03374 *
## Year2002 0.0598 0.2157 0.28 0.78259
## Year2003 -0.3971 0.1545 -2.57 0.01265 *
## Year2004 -0.4459 0.4812 -0.93 0.35773
## Year2005 -0.0908 0.2768 -0.33 0.74390
## Year2006 -0.1643 0.2691 -0.61 0.54368
```

```

## Year2007          -0.2632      0.2143    -1.23   0.22421
## Year2008          -0.2845      0.1849    -1.54   0.12906
## Year2009          -0.5316      0.1348    -3.94   0.00021 ***
## Year2010          -0.3706      0.2578    -1.44   0.15559
## Year2011          -0.3381      0.1700    -1.99   0.05113 .
## Year2012          -0.3207      0.1513    -2.12   0.03809 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.183, Adjusted R-squared:  -0.0444
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 2 observations c(22,28) are outliers with |weight| = 0 ( < 0.0013);
## 5 weights are ~= 1. The remaining 72 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.187  0.906  0.971  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          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 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
##  132   84  100   84   97   82  110  135  109  139  122  123  144  137  136
## 2011 2012
##  130  122
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   65   45   51   36   48   15   66   82   73   79   73   72   88   84   93

```

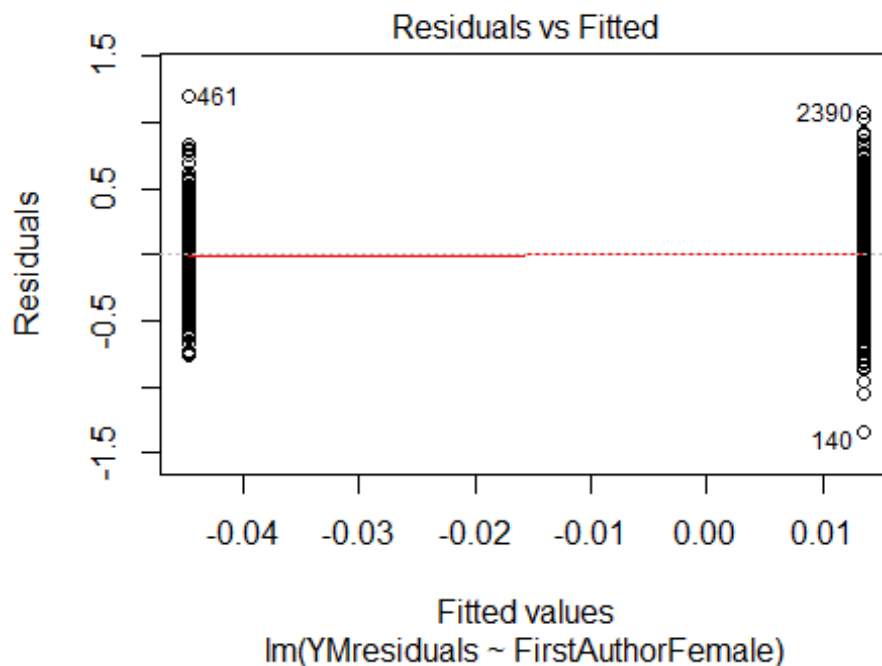
```
## 2011 2012
## 88 82
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 30 45 21 39 11 57 62 62 66 55 56 77 67 79
## 2011 2012
## 71 59
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.03
```



```
##
## 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: 4.38975904959877"
## [1] "Male first author team size 2018 geometric mean: 3.49895527551486"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

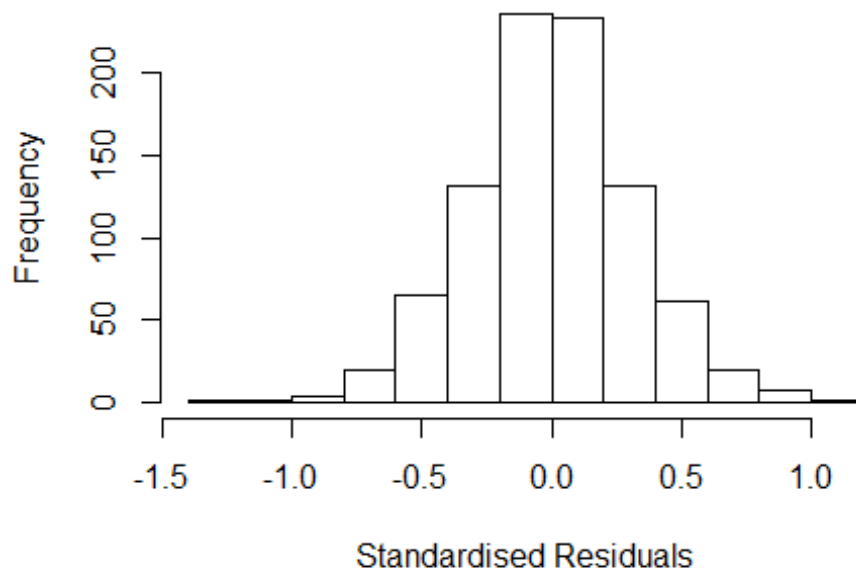
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 210, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.58964584275858"
## [1] "Male last author team size 2018 geometric mean: 3.74557731057464"

## Warning in wilcox.test.default(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 1.171 1      1.082
## LastAuthorFemale  1.155 1      1.075
## UniqueAuthors    1.433 4      1.046
## Year              1.625 16     1.015
```

Residuals from first and last author and team size



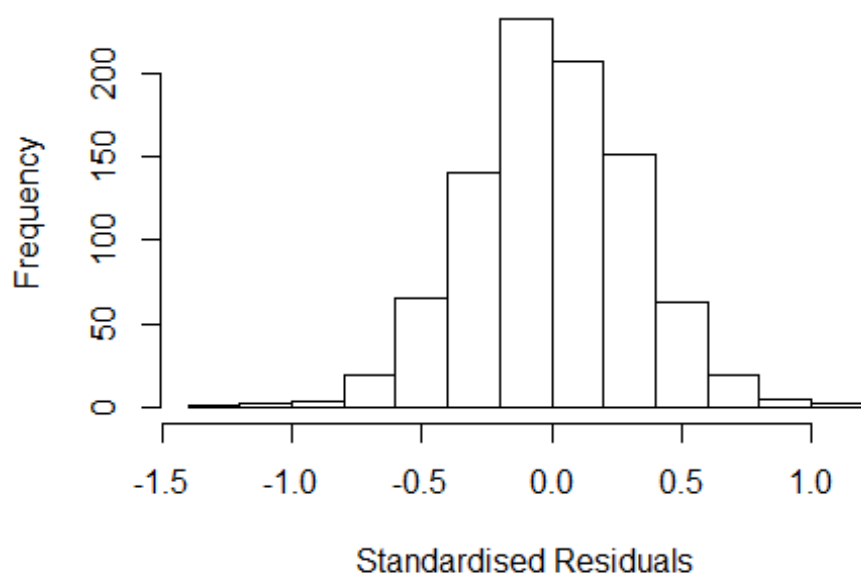
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.325166 -0.189669 -0.000739 0.195363 1.017605
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17201 0.07157 16.38 < 2e-16 ***
## FirstAuthorFemale1 -0.07456 0.02792 -2.67 0.00770 **
## LastAuthorFemale1 0.00433 0.03450 0.13 0.90003
## UniqueAuthors2 0.18849 0.06043 3.12 0.00187 **
## UniqueAuthors3 0.15316 0.06119 2.50 0.01249 *
## UniqueAuthors4 0.23814 0.06318 3.77 0.00017 ***
## UniqueAuthors5 0.23030 0.06217 3.70 0.00023 ***
## Year1997 -0.09731 0.08192 -1.19 0.23521
## Year1998 -0.07344 0.06232 -1.18 0.23896
## Year1999 -0.07855 0.06746 -1.16 0.24458
```

```

## Year2000      -0.13241    0.06182    -2.14    0.03248 *
## Year2001      -0.18240    0.08257    -2.21    0.02743 *
## Year2002      -0.14217    0.05202    -2.73    0.00640 **
## Year2003      -0.12722    0.05850    -2.17    0.02992 *
## Year2004      -0.15615    0.05969    -2.62    0.00904 **
## Year2005      -0.13793    0.05782    -2.39    0.01727 *
## Year2006      -0.14547    0.06045    -2.41    0.01631 *
## Year2007      -0.08181    0.06301    -1.30    0.19453
## Year2008      -0.09209    0.05344    -1.72    0.08518 .
## Year2009      -0.03030    0.05759    -0.53    0.59889
## Year2010      -0.08350    0.05743    -1.45    0.14631
## Year2011      -0.11740    0.05729    -2.05    0.04073 *
## Year2012      -0.09391    0.07143    -1.31    0.18894
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.296
## Multiple R-squared:  0.0533, Adjusted R-squared:  0.0298
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 834 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0072 0.8620 0.9500 0.8970 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.10e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.133 1 1.064
## LastAuthorFemale 1.124 1 1.060
## Year 1.152 16 1.004

```


Residuals from first and last author



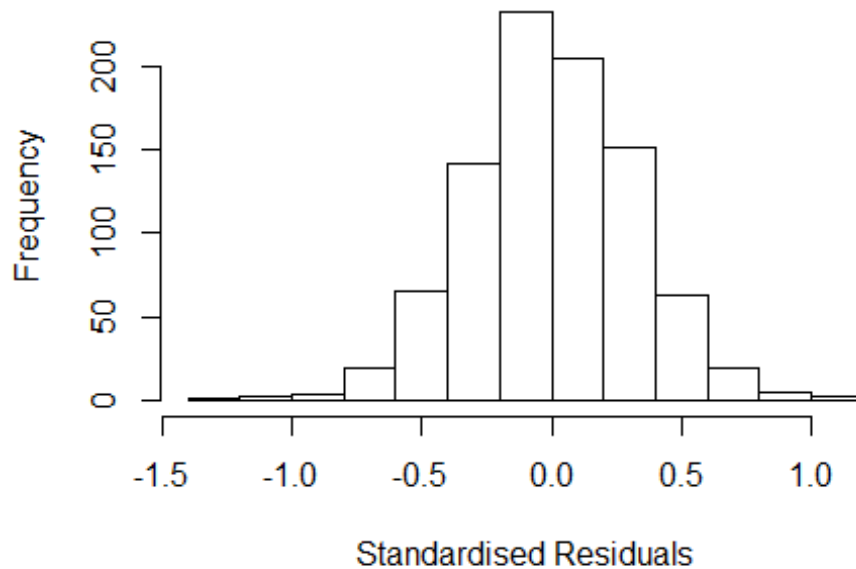
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35459 -0.20658 -0.00509  0.21301  1.04295
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.35459    0.04334   31.25  <2e-16 ***
## FirstAuthorFemale1 -0.07641    0.02797   -2.73   0.0064 **
## LastAuthorFemale1 -0.00483    0.03408   -0.14   0.8874
## Year1997        -0.08581    0.08132   -1.06   0.2917
## Year1998        -0.06341    0.06435   -0.99   0.3247
## Year1999        -0.06952    0.06599   -1.05   0.2924
## Year2000        -0.13942    0.06411   -2.17   0.0299 *
## Year2001        -0.18717    0.09405   -1.99   0.0469 *
## Year2002        -0.12985    0.05241   -2.48   0.0134 *
## Year2003        -0.11615    0.06018   -1.93   0.0539 .
## Year2004        -0.15210    0.06034   -2.52   0.0119 *
## Year2005        -0.13350    0.05869   -2.27   0.0232 *
```

```

## Year2006      -0.14400    0.06306   -2.28   0.0226 *
## Year2007      -0.06288    0.06159   -1.02   0.3075
## Year2008      -0.08705    0.05437   -1.60   0.1097
## Year2009      -0.01566    0.05768   -0.27   0.7861
## Year2010      -0.06486    0.05667   -1.14   0.2527
## Year2011      -0.10092    0.05742   -1.76   0.0792 .
## Year2012      -0.07154    0.07074   -1.01   0.3122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.307
## Multiple R-squared:  0.0308, Adjusted R-squared:  0.0113
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 832 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0128 0.8680 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      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.075 1      1.037
## Year      1.075 16      1.002

```

Residuals from first author



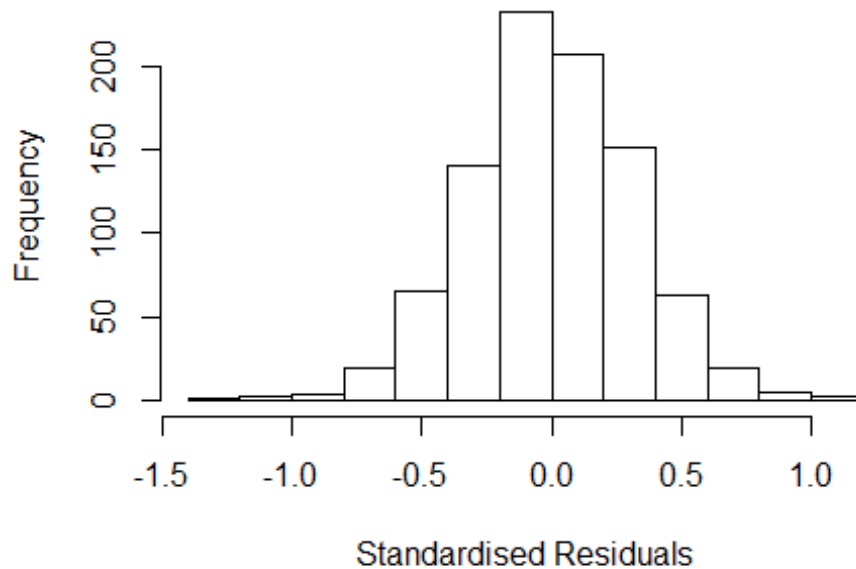
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35433 -0.20586 -0.00457 0.21266 1.04339
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3543 0.0432 31.33 <2e-16 ***
## FirstAuthorFemale1 -0.0770 0.0274 -2.81 0.005 **
## Year1997 -0.0857 0.0813 -1.05 0.292
## Year1998 -0.0643 0.0641 -1.00 0.316
## Year1999 -0.0698 0.0660 -1.06 0.291
## Year2000 -0.1394 0.0641 -2.18 0.030 *
## Year2001 -0.1868 0.0940 -1.99 0.047 *
## Year2002 -0.1300 0.0524 -2.48 0.013 *
## Year2003 -0.1162 0.0602 -1.93 0.054 .
## Year2004 -0.1520 0.0602 -2.52 0.012 *
## Year2005 -0.1340 0.0587 -2.28 0.023 *
## Year2006 -0.1439 0.0630 -2.28 0.023 *
```

```

## Year2007          -0.0631      0.0616   -1.02    0.306
## Year2008          -0.0873      0.0544   -1.61    0.109
## Year2009          -0.0161      0.0576   -0.28    0.780
## Year2010          -0.0651      0.0567   -1.15    0.251
## Year2011          -0.1011      0.0574   -1.76    0.078 .
## Year2012          -0.0717      0.0707   -1.01    0.311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.307
## Multiple R-squared:  0.0308, Adjusted R-squared:  0.0123
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 832 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0131 0.8680 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      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.06 1      1.030
## Year      1.06 16      1.002

```

Residuals from last author



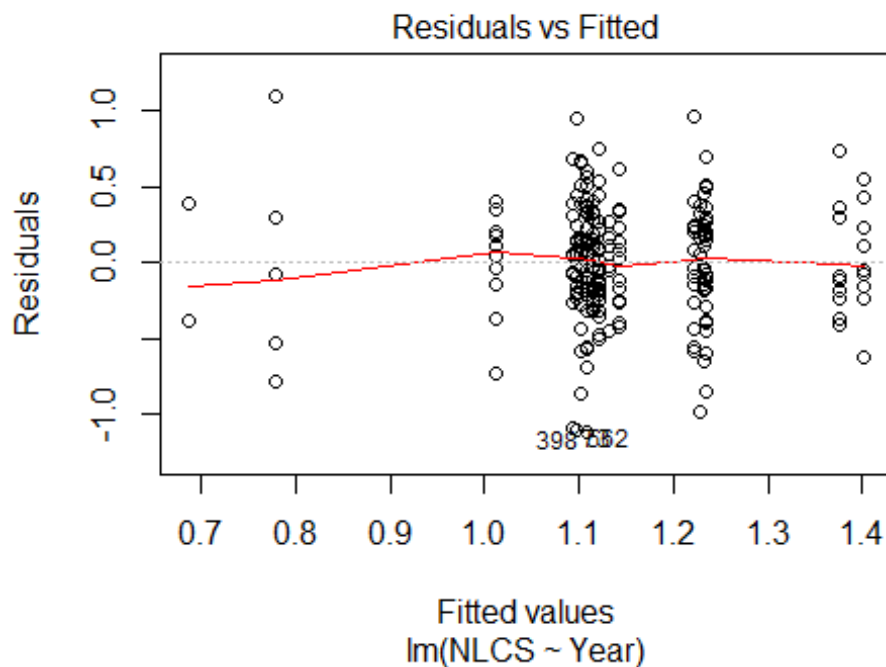
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34360 -0.20191 -0.00326 0.21568 1.06854
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3436 0.0443 30.36 <2e-16 ***
## LastAuthorFemale1 -0.0189 0.0341 -0.56 0.579
## Year1997 -0.0865 0.0829 -1.04 0.297
## Year1998 -0.0560 0.0650 -0.86 0.389
## Year1999 -0.0652 0.0657 -0.99 0.321
## Year2000 -0.1514 0.0652 -2.32 0.021 *
## Year2001 -0.1902 0.0944 -2.01 0.044 *
## Year2002 -0.1289 0.0533 -2.42 0.016 *
## Year2003 -0.1115 0.0611 -1.82 0.068 .
## Year2004 -0.1601 0.0611 -2.62 0.009 **
## Year2005 -0.1342 0.0605 -2.22 0.027 *
## Year2006 -0.1513 0.0636 -2.38 0.018 *
```

```

## Year2007          -0.0752      0.0629   -1.20    0.232
## Year2008          -0.0937      0.0561   -1.67    0.095 .
## Year2009          -0.0162      0.0584   -0.28    0.781
## Year2010          -0.0714      0.0572   -1.25    0.212
## Year2011          -0.1046      0.0581   -1.80    0.072 .
## Year2012          -0.0861      0.0714   -1.21    0.228
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.309
## Multiple R-squared:  0.0212, Adjusted R-squared:  0.00258
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 833 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0195 0.8700 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      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 911"
## [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   19   23   19   21   28   23   13   31   39   39   27   52   41   27
## 2011 2012
##   27   35
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2    6    9    9    6   10   10    5   13   19   19   13   26   22   19
## 2011 2012

```

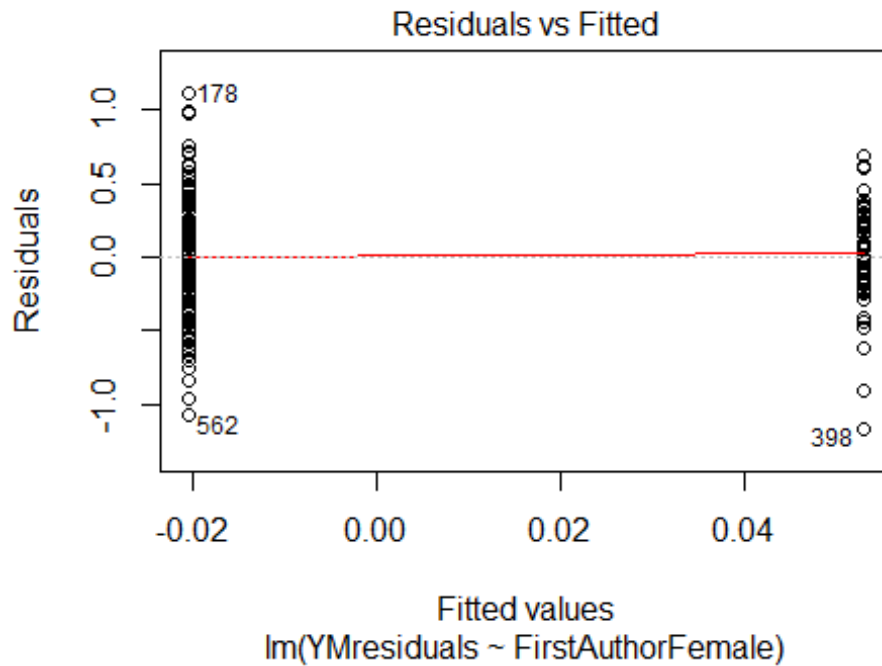
```
## 14 12
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2 1 7 7 6 10 9 4 12 19 15 10 19 18 12
## 2011 2012
## 12 10
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 21, 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: 3.3658654363386"
## [1] "Male first author team size 2018 geometric mean: 2.2456485239871"

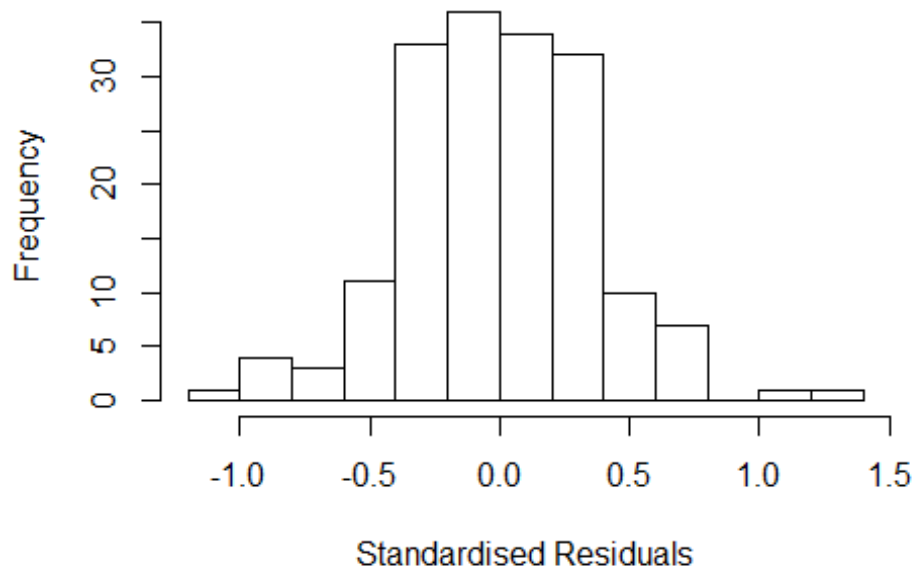
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24, p-value = 0.3
## 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: 2.65811042998648"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
## FirstAuthorFemale	5.147	1	2.269
## LastAuthorFemale	2.384	1	1.544
## UniqueAuthors	51.172	4	1.635
## Year	194.208	16	1.179

Residuals from first and last author and team size



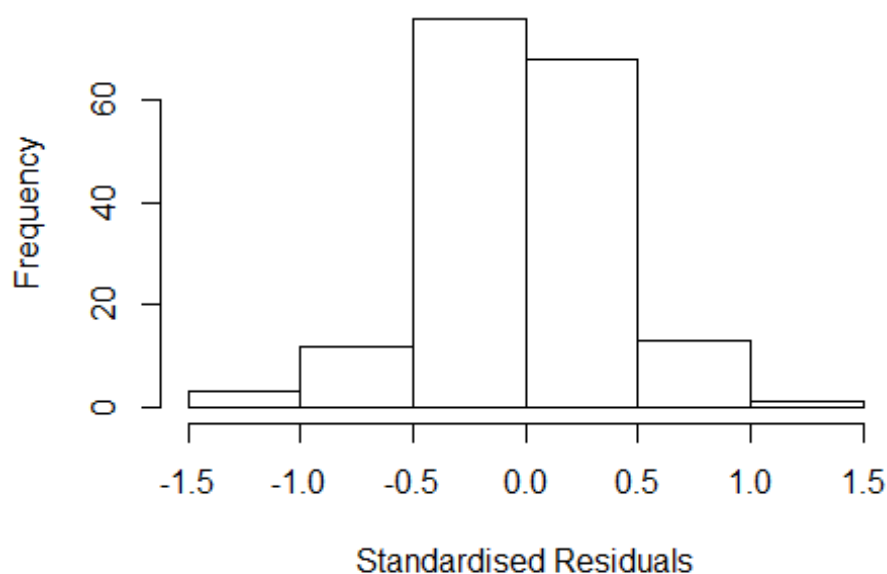
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.14770 -0.25300 -0.00385  0.26040  1.24853
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8711    0.3800    2.29   0.023 *
## FirstAuthorFemale1 0.1070    0.0681    1.57   0.118
## LastAuthorFemale1 -0.1012    0.0901   -1.12   0.263
## UniqueAuthors2    -0.1539    0.2316   -0.66   0.507
## UniqueAuthors3    -0.1161    0.2199   -0.53   0.598
## UniqueAuthors4    -0.0506    0.2372   -0.21   0.831
## UniqueAuthors5     0.0710    0.3311    0.21   0.830
## Year1997          0.2848    0.3206    0.89   0.376
## Year1998          0.5068    0.3304    1.53   0.127
## Year1999          0.2828    0.3451    0.82   0.414
```

```

## Year2000          0.4729      0.3992      1.18      0.238
## Year2001          0.6225      0.3338      1.87      0.064 .
## Year2002          0.2376      0.3300      0.72      0.473
## Year2003         -0.3197      0.7188     -0.44      0.657
## Year2004          0.4541      0.3381      1.34      0.181
## Year2005          0.3370      0.3132      1.08      0.284
## Year2006          0.3849      0.3298      1.17      0.245
## Year2007          0.4708      0.3232      1.46      0.147
## Year2008          0.2857      0.3247      0.88      0.380
## Year2009          0.2287      0.3182      0.72      0.473
## Year2010          0.2683      0.3277      0.82      0.414
## Year2011          0.3362      0.3335      1.01      0.315
## Year2012          0.2658      0.3876      0.69      0.494
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.365
## Multiple R-squared:  0.164, Adjusted R-squared:  0.041
## Convergence in 31 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 156 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.219  0.898  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      5.78e-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 4.574 1      2.139
## LastAuthorFemale  2.238 1      1.496
## Year              8.513 16      1.069

```

Residuals from first and last author



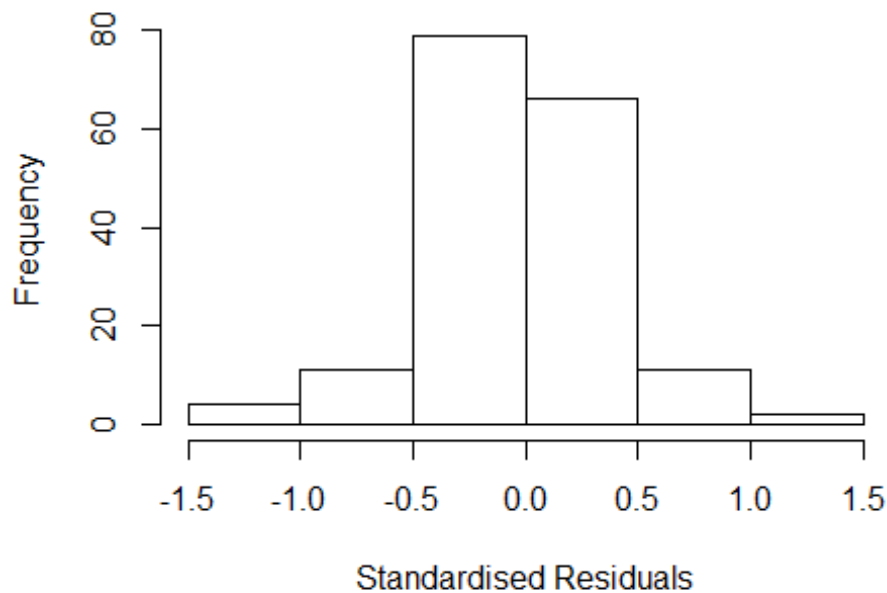
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1571 -0.2234 -0.0131 0.2305 1.4006
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7405 0.2805 2.64 0.0091 **
## FirstAuthorFemale1 0.0927 0.0660 1.40 0.1621
## LastAuthorFemale1 -0.1100 0.0801 -1.37 0.1714
## Year1997 0.2758 0.2911 0.95 0.3449
## Year1998 0.5281 0.3070 1.72 0.0874 .
## Year1999 0.3034 0.3202 0.95 0.3448
## Year2000 0.6301 0.3020 2.09 0.0386 *
## Year2001 0.6649 0.2994 2.22 0.0278 *
## Year2002 0.2624 0.3015 0.87 0.3854
## Year2003 -0.2701 0.7379 -0.37 0.7149
## Year2004 0.4571 0.3100 1.47 0.1424
## Year2005 0.3700 0.2848 1.30 0.1957
```

```

## Year2006          0.3982      0.3045      1.31      0.1929
## Year2007          0.4830      0.2950      1.64      0.1036
## Year2008          0.3239      0.2971      1.09      0.2774
## Year2009          0.2706      0.2885      0.94      0.3499
## Year2010          0.3164      0.3072      1.03      0.3047
## Year2011          0.4122      0.2973      1.39      0.1676
## Year2012          0.4081      0.3162      1.29      0.1988
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.366
## Multiple R-squared:  0.152, Adjusted R-squared:  0.0524
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 153 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.110  0.892  0.954  0.899  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      5.78e-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 4.28 1      2.069
## Year              4.28 16      1.046

```

Residuals from first author



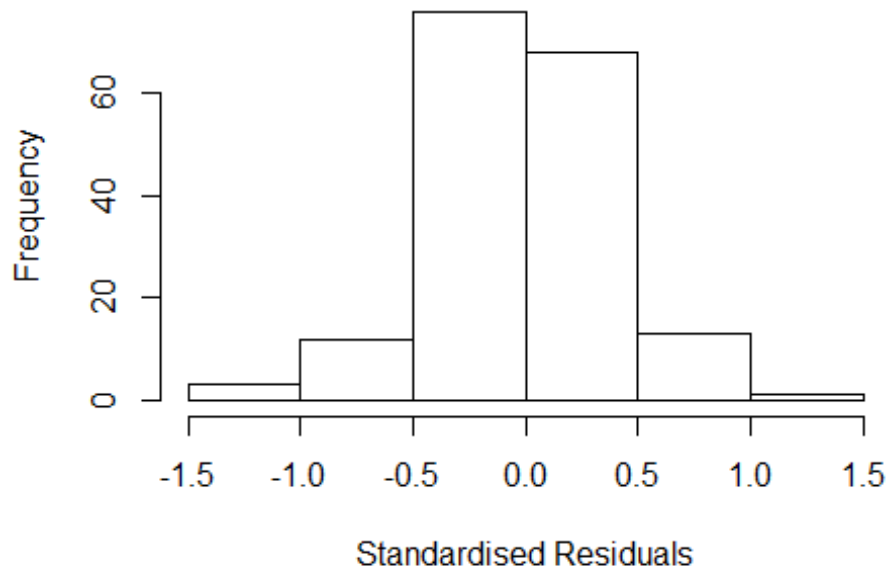
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1248 -0.2231 -0.0279 0.2399 1.3967
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6855 0.3474 1.97 0.050 .
## FirstAuthorFemale1 0.0819 0.0639 1.28 0.202
## Year1997 0.3416 0.3534 0.97 0.335
## Year1998 0.5864 0.3682 1.59 0.113
## Year1999 0.3734 0.3702 1.01 0.315
## Year2000 0.6579 0.3673 1.79 0.075 .
## Year2001 0.7072 0.3646 1.94 0.054 .
## Year2002 0.3083 0.3663 0.84 0.401
## Year2003 -0.2112 0.7756 -0.27 0.786
## Year2004 0.4973 0.3714 1.34 0.182
## Year2005 0.4141 0.3521 1.18 0.241
## Year2006 0.4462 0.3695 1.21 0.229
```

```

## Year2007          0.5275      0.3596      1.47      0.144
## Year2008          0.3573      0.3627      0.99      0.326
## Year2009          0.3227      0.3532      0.91      0.362
## Year2010          0.3606      0.3714      0.97      0.333
## Year2011          0.4525      0.3607      1.25      0.212
## Year2012          0.4377      0.3789      1.16      0.250
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.142, Adjusted R-squared:  0.0474
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 158 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.115  0.893  0.959  0.902  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.78e-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 2.024 1      1.423
## Year              2.024 16      1.022

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.196421 -0.239965 -0.000797 0.229628 1.377262
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7327 0.2874 2.55 0.012 *
## LastAuthorFemale1 -0.0945 0.0760 -1.24 0.216
## Year1997 0.3763 0.2874 1.31 0.192
## Year1998 0.5644 0.3080 1.83 0.069 .
## Year1999 0.3569 0.3305 1.08 0.282
## Year2000 0.6297 0.3134 2.01 0.046 *
## Year2001 0.6804 0.3043 2.24 0.027 *
## Year2002 0.3130 0.3075 1.02 0.310
## Year2003 -0.2390 0.7512 -0.32 0.751
## Year2004 0.4846 0.3135 1.55 0.124
## Year2005 0.3917 0.2911 1.35 0.180
## Year2006 0.4271 0.3092 1.38 0.169
```

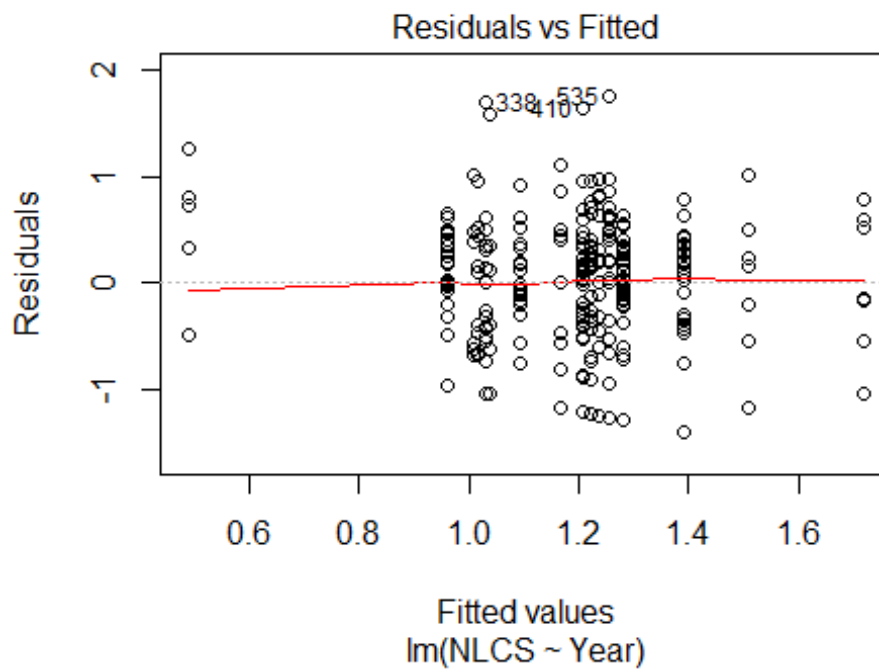
```

## Year2007          0.5090      0.3008      1.69      0.093 .
## Year2008          0.3604      0.3037      1.19      0.237
## Year2009          0.3050      0.2936      1.04      0.301
## Year2010          0.3394      0.3168      1.07      0.286
## Year2011          0.4456      0.3003      1.48      0.140
## Year2012          0.4637      0.3133      1.48      0.141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.138, Adjusted R-squared:  0.0431
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 160 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.141  0.894  0.959  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      5.78e-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: 173"
## [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
##   43   42   38   36   29   57   37   62   48   64   62   79   83   49   48
## 2011 2012
##   55   65
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    9   12   12    6    6    7    7   16   16   24   14   27   25   25   20
## 2011 2012

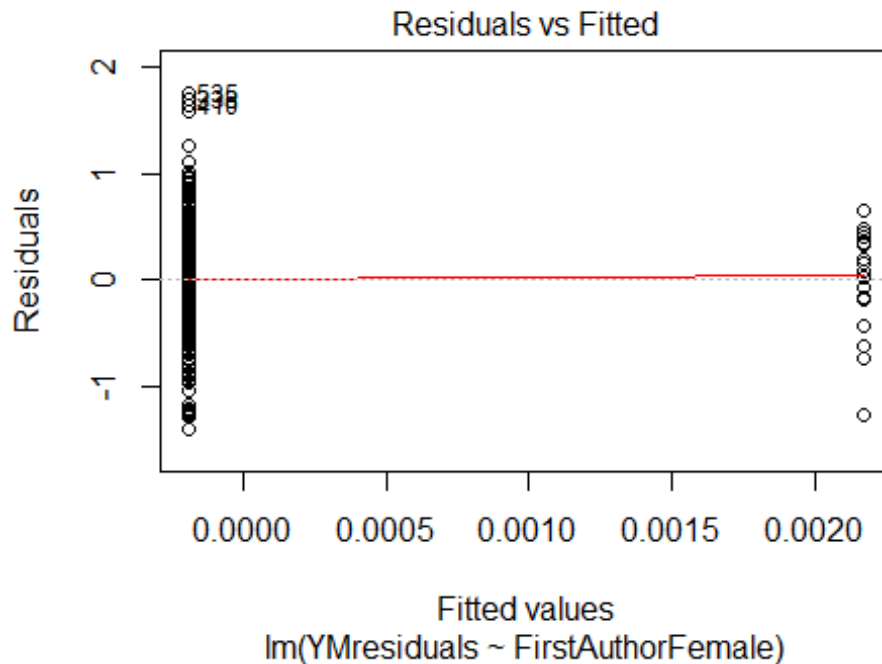
```



```
## 23 33
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 8 9 6 5 3 5 12 14 21 11 18 19 16 17
## 2011 2012
## 21 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 = 28, df = 16, p-value = 0.03
```

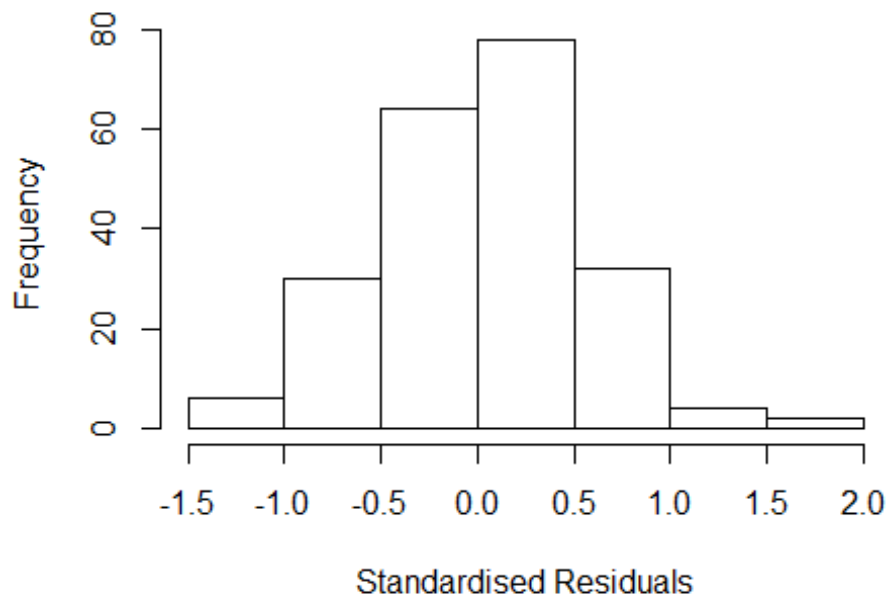


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.3, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 2.37956557896877"
## [1] "Male first author team size 2018 geometric mean: 2.64580400631004"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.22064303492292"
## [1] "Male last author team size 2018 geometric mean: 2.65160038737395"
##
## 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 2.086 1          1.444
## LastAuthorFemale 2.145 1          1.464
## UniqueAuthors    7.158 4          1.279
## Year             10.723 16         1.077
```

Residuals from first and last author and team size



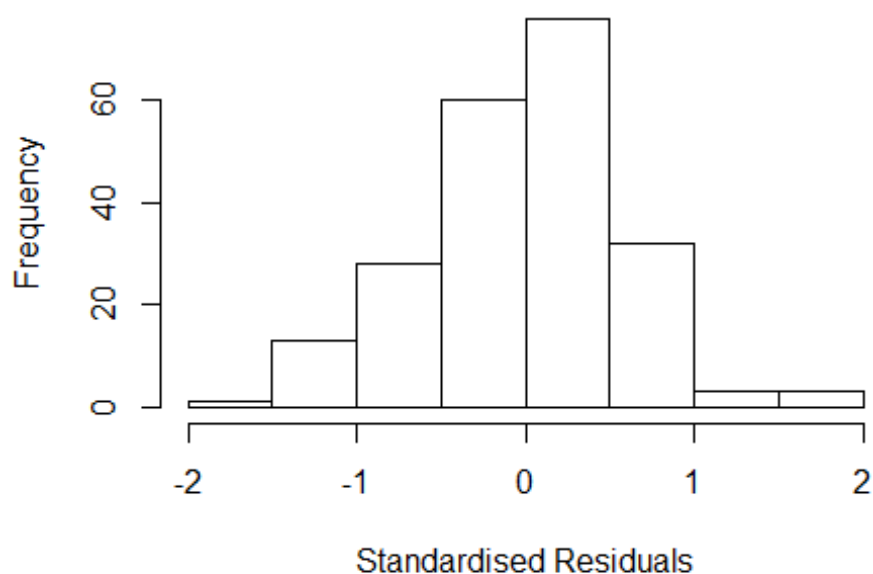
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4670 -0.4054  0.0465  0.3726  1.6337
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5244    0.3985    1.32  0.18977
## FirstAuthorFemale1 -0.1092    0.1481   -0.74  0.46173
## LastAuthorFemale1  0.1203    0.1610    0.75  0.45569
## UniqueAuthors2    0.5421    0.1632    3.32  0.00107 **
## UniqueAuthors3    0.5574    0.1655    3.37  0.00091 ***
## UniqueAuthors4    0.7528    0.1691    4.45  1.4e-05 ***
## UniqueAuthors5    0.7862    0.2713    2.90  0.00420 **
## Year1997          0.5267    0.4607    1.14  0.25436
## Year1998         -0.2676    0.4558   -0.59  0.55785
## Year1999         -0.0217    0.4357   -0.05  0.96032
```

```

## Year2000          0.2595      0.4210      0.62  0.53847
## Year2001          1.4841      0.6323      2.35  0.01994 *
## Year2002          0.5743      0.6233      0.92  0.35803
## Year2003         -0.0109      0.4069     -0.03  0.97873
## Year2004          0.1378      0.4049      0.34  0.73400
## Year2005          0.2801      0.4170      0.67  0.50250
## Year2006          0.2660      0.4246      0.63  0.53173
## Year2007          0.2344      0.4018      0.58  0.56040
## Year2008         -0.0254      0.3871     -0.07  0.94774
## Year2009          0.3324      0.3841      0.87  0.38794
## Year2010          0.0221      0.3866      0.06  0.95439
## Year2011          0.1829      0.3928      0.47  0.64191
## Year2012          0.1792      0.3872      0.46  0.64408
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.513
## Multiple R-squared:  0.283, Adjusted R-squared:  0.202
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 202 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.289  0.876  0.942  0.896  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.63e-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 1.481 1      1.217
## LastAuthorFemale  1.920 1      1.386
## Year              1.769 16      1.018

```

Residuals from first and last author



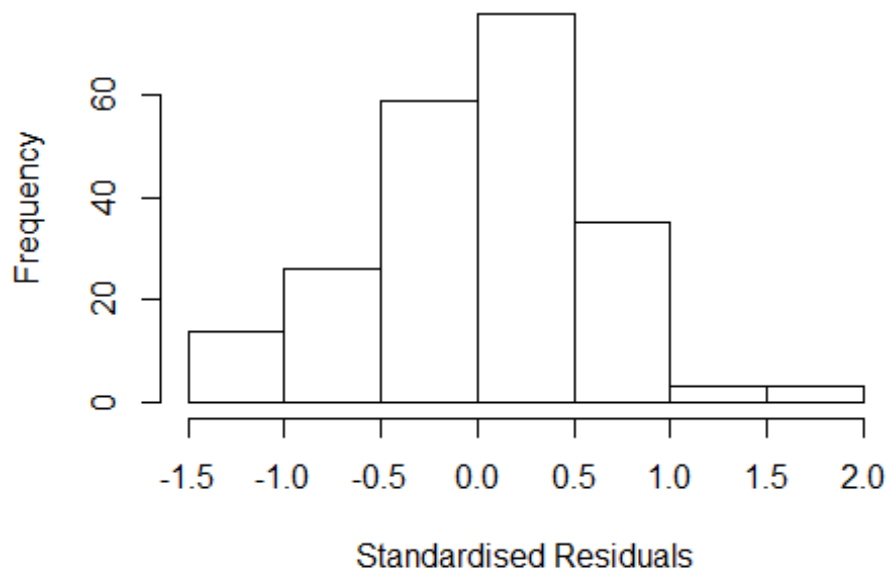
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5207 -0.3867 0.0614 0.3770 1.8795
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8931 0.2548 3.51 0.00057 ***
## FirstAuthorFemale1 -0.0796 0.1330 -0.60 0.54996
## LastAuthorFemale1 0.1705 0.1747 0.98 0.33023
## Year1997 0.4688 0.3925 1.19 0.23381
## Year1998 -0.4588 0.3363 -1.36 0.17407
## Year1999 -0.0611 0.4497 -0.14 0.89206
## Year2000 0.2361 0.4370 0.54 0.58967
## Year2001 1.2225 0.3535 3.46 0.00067 ***
## Year2002 0.8011 0.4761 1.68 0.09401 .
## Year2003 -0.1039 0.2900 -0.36 0.72042
## Year2004 0.2102 0.3041 0.69 0.49030
## Year2005 0.2335 0.3464 0.67 0.50116
```

```

## Year2006          0.3175      0.3636      0.87  0.38367
## Year2007          0.3281      0.3080      1.07  0.28812
## Year2008          0.0629      0.2940      0.21  0.83067
## Year2009          0.4869      0.2818      1.73  0.08564 .
## Year2010          0.2280      0.2729      0.84  0.40444
## Year2011          0.3484      0.2771      1.26  0.21011
## Year2012          0.3826      0.2692      1.42  0.15684
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.169, Adjusted R-squared:  0.0929
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 200 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.230  0.872  0.949  0.897  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      4.63e-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 1.212 1      1.101
## Year              1.212 16      1.006

```

Residuals from first author



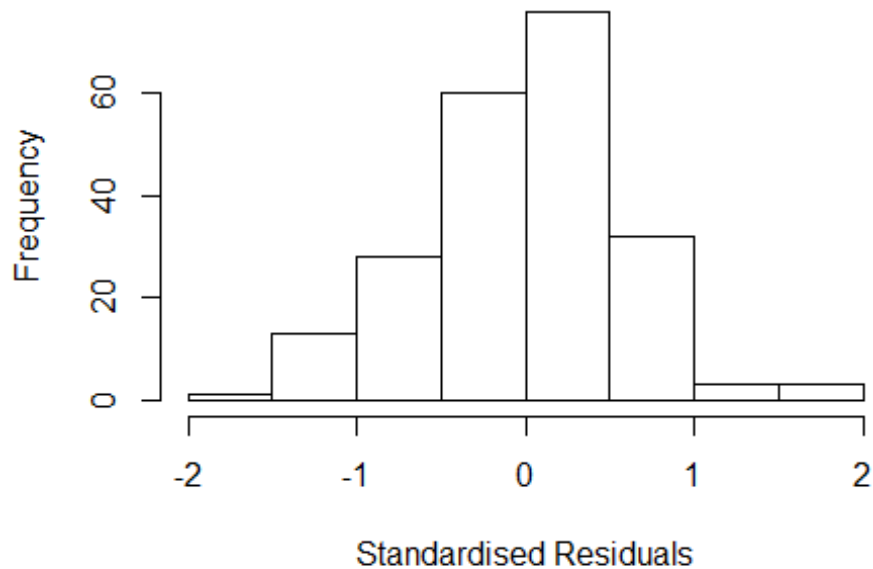
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3980 -0.3969 0.0676 0.3610 1.8688
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8918 0.2557 3.49 0.00060 ***
## FirstAuthorFemale1 -0.0321 0.1192 -0.27 0.78770
## Year1997 0.4939 0.3963 1.25 0.21417
## Year1998 -0.4455 0.3478 -1.28 0.20171
## Year1999 -0.0335 0.4347 -0.08 0.93864
## Year2000 0.2504 0.4687 0.53 0.59383
## Year2001 1.2241 0.3544 3.45 0.00068 ***
## Year2002 0.7847 0.4562 1.72 0.08699 .
## Year2003 -0.0980 0.2953 -0.33 0.74042
## Year2004 0.2554 0.2960 0.86 0.38929
## Year2005 0.2454 0.3502 0.70 0.48425
## Year2006 0.3205 0.3652 0.88 0.38121
```

```

## Year2007          0.3211      0.3079      1.04  0.29829
## Year2008          0.0629      0.2944      0.21  0.83092
## Year2009          0.5062      0.2813      1.80  0.07348 .
## Year2010          0.2343      0.2740      0.86  0.39339
## Year2011          0.3646      0.2789      1.31  0.19254
## Year2012          0.3988      0.2703      1.48  0.14164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.164, Adjusted R-squared:  0.0918
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 205 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.224  0.862  0.953  0.897  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      4.63e-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 1.553 1      1.246
## Year              1.553 16      1.014

```


Residuals from last author



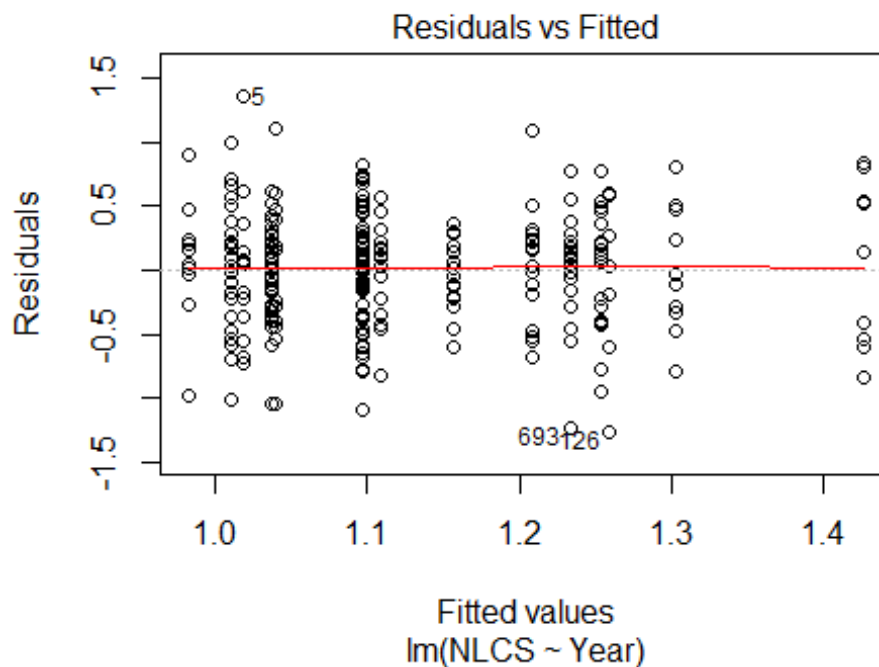
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4958 -0.3875 0.0672 0.3634 1.8786
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8922 0.2554 3.49 0.00059 ***
## LastAuthorFemale1 0.1465 0.1645 0.89 0.37414
## Year1997 0.4596 0.3916 1.17 0.24187
## Year1998 -0.4569 0.3383 -1.35 0.17832
## Year1999 -0.0778 0.4560 -0.17 0.86468
## Year2000 0.2212 0.4303 0.51 0.60781
## Year2001 1.2236 0.3542 3.45 0.00067 ***
## Year2002 0.8011 0.4745 1.69 0.09291 .
## Year2003 -0.1085 0.2922 -0.37 0.71083
## Year2004 0.2090 0.3046 0.69 0.49337
## Year2005 0.2353 0.3500 0.67 0.50229
## Year2006 0.3195 0.3647 0.88 0.38207
```

```

## Year2007          0.3137      0.3054      1.03  0.30556
## Year2008          0.0604      0.2939      0.21  0.83732
## Year2009          0.4809      0.2830      1.70  0.09082 .
## Year2010          0.2198      0.2725      0.81  0.42087
## Year2011          0.3473      0.2778      1.25  0.21271
## Year2012          0.3740      0.2699      1.39  0.16737
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.168, Adjusted R-squared:  0.0969
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 206 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.223  0.873  0.955  0.898  0.981  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.63e-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: 216"
## [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
##   25   24   27   15   19   34   27   28   34   31   41   45   57   41   58
## 2011 2012
##   48   45
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   14   12   10   11    8   12    9   12   17   18   17   18   26   16   33
## 2011 2012

```

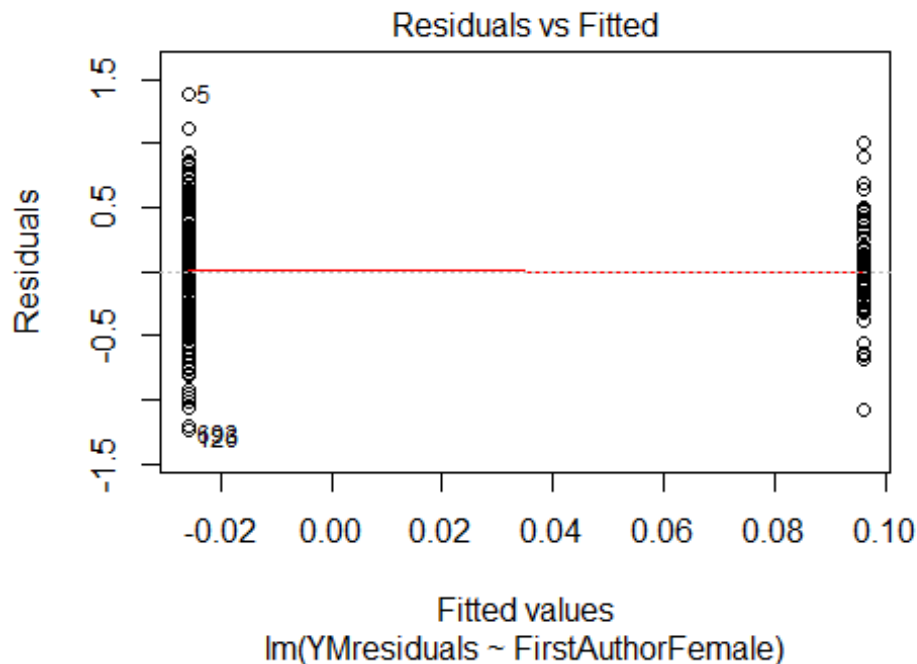
```
## 27 18
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 8 9 8 6 8 7 7 14 16 15 15 21 12 26
## 2011 2012
## 22 14
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```



```
##
## 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: 3.14939492487456"
## [1] "Male first author team size 2018 geometric mean: 3.217496968773"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 160, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.99255573947769"
## [1] "Male last author team size 2018 geometric mean: 3.23157545694402"

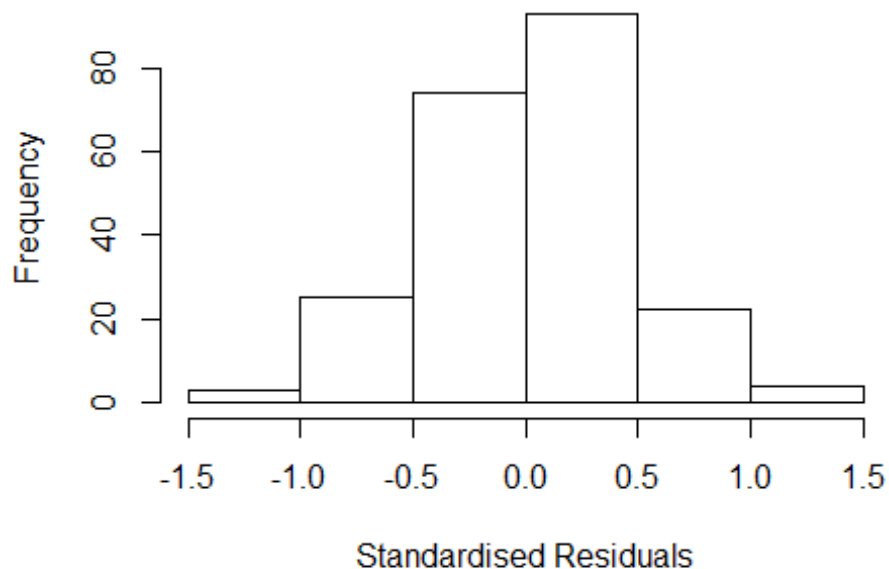
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 94, p-value = 0.7
## 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	1.341	1	1.158
LastAuthorFemale	1.823	1	1.350
UniqueAuthors	4.080	4	1.192
Year	6.690	16	1.061

Residuals from first and last author and team size



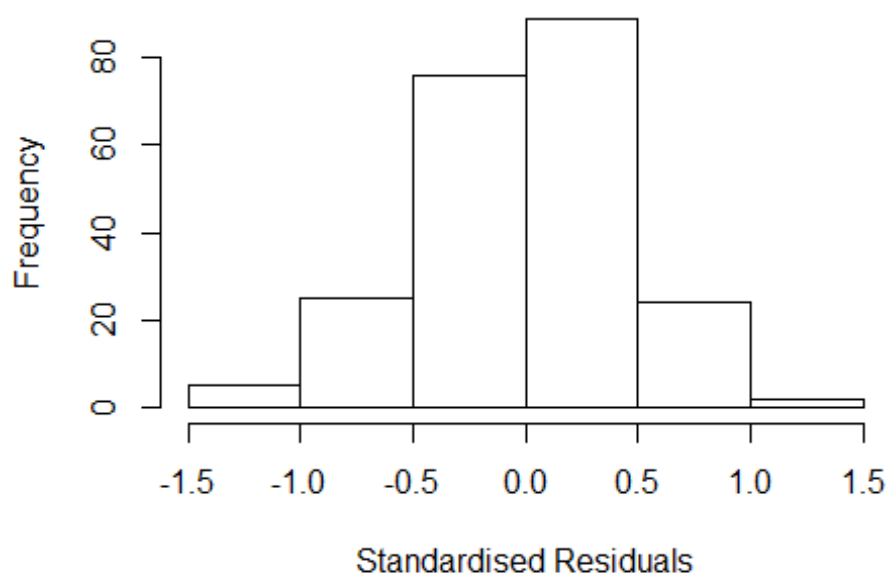
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2036 -0.2879 0.0272 0.2644 1.3747
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4223 0.1996 2.12 0.03563 *
## FirstAuthorFemale1 0.0949 0.0676 1.40 0.16194
## LastAuthorFemale1 0.0406 0.0982 0.41 0.67945
## UniqueAuthors2 0.5033 0.1452 3.47 0.00065 ***
## UniqueAuthors3 0.5780 0.1479 3.91 0.00013 ***
## UniqueAuthors4 0.5862 0.1457 4.02 8.2e-05 ***
## UniqueAuthors5 0.6366 0.1601 3.98 9.8e-05 ***
## Year1997 0.2109 0.1816 1.16 0.24693
## Year1998 0.2857 0.2155 1.33 0.18641
## Year1999 0.5181 0.3637 1.42 0.15589
```

```

## Year2000          0.3012      0.2716      1.11  0.26882
## Year2001          0.0725      0.2411      0.30  0.76388
## Year2002          0.0876      0.2039      0.43  0.66794
## Year2003          0.2032      0.2349      0.87  0.38794
## Year2004          0.1518      0.2117      0.72  0.47413
## Year2005          0.3641      0.1895      1.92  0.05613 .
## Year2006          0.1368      0.1778      0.77  0.44242
## Year2007          0.0941      0.1817      0.52  0.60507
## Year2008         -0.0130      0.2048     -0.06  0.94932
## Year2009          0.1769      0.1895      0.93  0.35177
## Year2010          0.1366      0.1790      0.76  0.44628
## Year2011          0.1280      0.1782      0.72  0.47316
## Year2012          0.1600      0.2032      0.79  0.43202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.19,   Adjusted R-squared:  0.0999
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 197 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.284  0.872  0.955  0.909  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.52e-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 1.225 1      1.107
## LastAuthorFemale  1.777 1      1.333
## Year              1.965 16      1.021

```

Residuals from first and last author



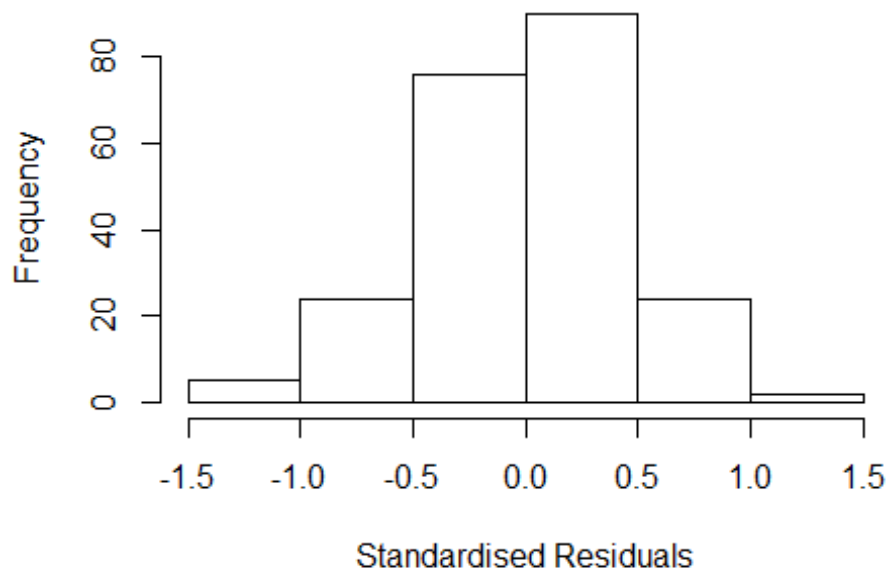
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1778 -0.3072 0.0308 0.2649 1.4715
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9035 0.1403 6.44 8.5e-10 ***
## FirstAuthorFemale1 0.1379 0.0675 2.04 0.042 *
## LastAuthorFemale1 0.0264 0.1061 0.25 0.804
## Year1997 0.2025 0.1580 1.28 0.201
## Year1998 0.2900 0.2090 1.39 0.167
## Year1999 0.5928 0.3428 1.73 0.085 .
## Year2000 0.2293 0.3557 0.64 0.520
## Year2001 -0.0900 0.2434 -0.37 0.712
## Year2002 0.0800 0.1932 0.41 0.679
## Year2003 0.1759 0.2407 0.73 0.466
## Year2004 0.2197 0.1959 1.12 0.263
## Year2005 0.3527 0.1948 1.81 0.072 .
```

```

## Year2006          0.1935      0.1622      1.19      0.234
## Year2007          0.1438      0.1777      0.81      0.419
## Year2008          0.0587      0.1957      0.30      0.765
## Year2009          0.2699      0.1649      1.64      0.103
## Year2010          0.1486      0.1659      0.90      0.371
## Year2011          0.2090      0.1608      1.30      0.195
## Year2012          0.2742      0.1769      1.55      0.123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.448
## Multiple R-squared:  0.0861, Adjusted R-squared:  0.00465
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 198 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.258  0.874  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      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 1.16 1      1.077
## Year              1.16 16      1.005

```


Residuals from first author



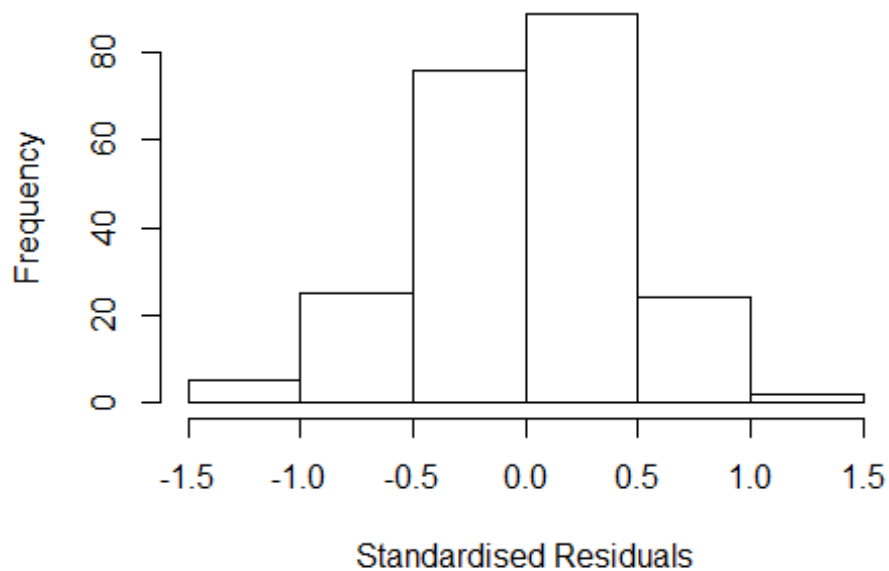
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1924 -0.3082 0.0253 0.2636 1.4718
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9032 0.1392 6.49 6.5e-10 ***
## FirstAuthorFemale1 0.1404 0.0656 2.14 0.034 *
## Year1997 0.2022 0.1573 1.29 0.200
## Year1998 0.2900 0.2086 1.39 0.166
## Year1999 0.5937 0.3385 1.75 0.081 .
## Year2000 0.2300 0.3566 0.64 0.520
## Year2001 -0.0895 0.2447 -0.37 0.715
## Year2002 0.0797 0.1926 0.41 0.679
## Year2003 0.1770 0.2399 0.74 0.461
## Year2004 0.2270 0.1908 1.19 0.236
## Year2005 0.3601 0.1907 1.89 0.060 .
## Year2006 0.1930 0.1615 1.19 0.234
```

```

## Year2007          0.1471      0.1763      0.83      0.405
## Year2008          0.0601      0.1952      0.31      0.759
## Year2009          0.2750      0.1615      1.70      0.090 .
## Year2010          0.1505      0.1649      0.91      0.363
## Year2011          0.2148      0.1566      1.37      0.172
## Year2012          0.2892      0.1612      1.79      0.074 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.444
## Multiple R-squared:  0.0866, Adjusted R-squared:  0.0101
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 197 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.250  0.873  0.950  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      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 1.665 1      1.291
## Year              1.665 16      1.016

```

Residuals from last author



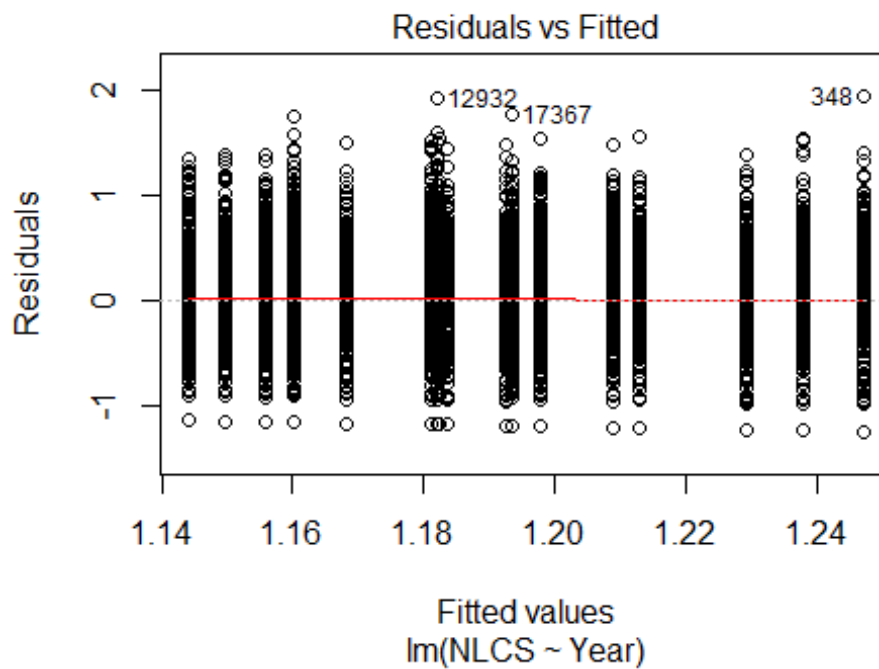
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1992 -0.2934 0.0216 0.2573 1.4361
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9389 0.1428 6.58 4e-10 ***
## LastAuthorFemale1 0.0524 0.1029 0.51 0.611
## Year1997 0.2009 0.1556 1.29 0.198
## Year1998 0.2725 0.2072 1.32 0.190
## Year1999 0.5580 0.3464 1.61 0.109
## Year2000 0.2602 0.3739 0.70 0.487
## Year2001 -0.0917 0.2457 -0.37 0.709
## Year2002 0.0665 0.1926 0.35 0.730
## Year2003 0.1413 0.2422 0.58 0.560
## Year2004 0.2092 0.2028 1.03 0.304
## Year2005 0.3377 0.1973 1.71 0.088 .
## Year2006 0.2065 0.1646 1.25 0.211
```

```

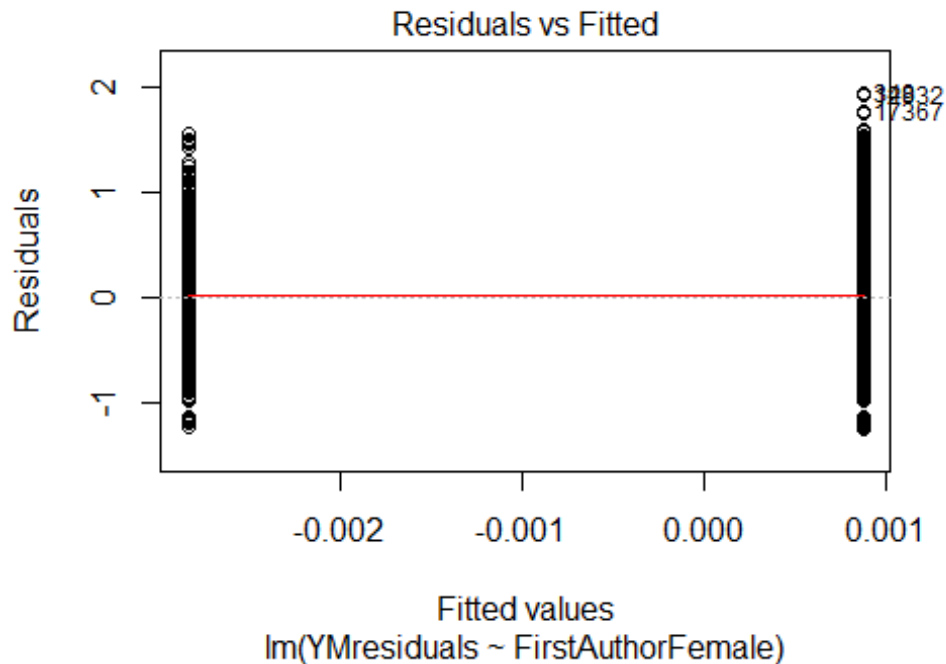
## Year2007          0.1402      0.1794      0.78      0.435
## Year2008          0.0585      0.2015      0.29      0.772
## Year2009          0.2395      0.1698      1.41      0.160
## Year2010          0.1469      0.1684      0.87      0.384
## Year2011          0.2121      0.1681      1.26      0.208
## Year2012          0.2598      0.1808      1.44      0.152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.444
## Multiple R-squared:  0.0718, Adjusted R-squared:  -0.00594
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 197 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.274  0.863  0.950  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      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 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
##  884  753  728  653  683  649  650  700  747  808  811  839  941 1048 1083
## 2011 2012
## 1158 1118
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  451  350  365  338  292  181  369  399  420  424  434  455  544  603  631
## 2011 2012

```

```
## 681 697
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 361 296 294 286 245 152 316 326 361 340 356 391 452 515 495
## 2011 2012
## 546 572
## [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
```

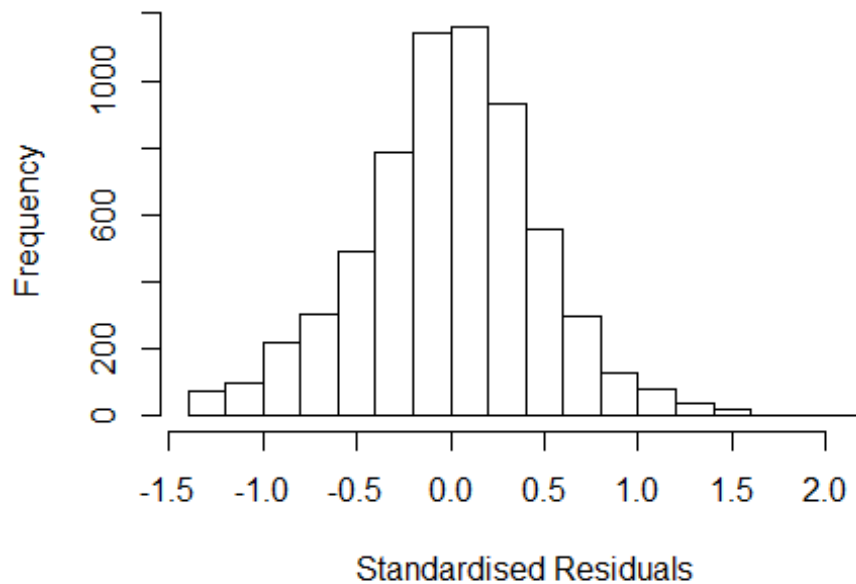


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 23, df = 1, p-value = 1e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 3.94999986981236"
## [1] "Male first author team size 2018 geometric mean: 3.48597833968186"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 30000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.96271217781538"
## [1] "Male last author team size 2018 geometric mean: 3.57676529086559"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 18000, 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 1.034 1      1.017
## LastAuthorFemale  1.021 1      1.010
## UniqueAuthors     1.084 4      1.010
## Year              1.114 16      1.003
```

Residuals from first and last author and team size



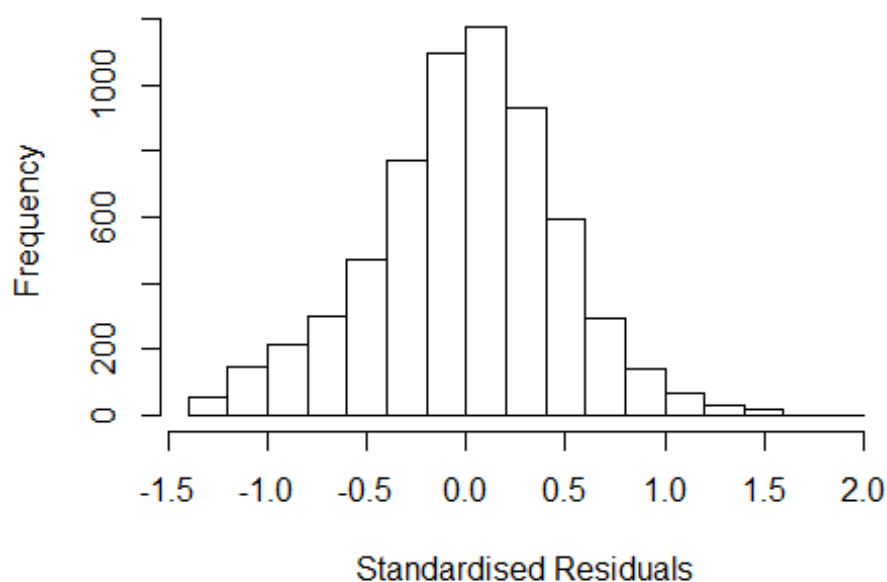
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37575 -0.28791  0.00799  0.29212  2.15887
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03902    0.03817   27.22 < 2e-16 ***
## FirstAuthorFemale1 -0.01144    0.01357   -0.84  0.39916
## LastAuthorFemale1  0.01651    0.01873    0.88  0.37797
## UniqueAuthors2     0.23084    0.03243    7.12 1.2e-12 ***
## UniqueAuthors3     0.22985    0.03239    7.10 1.4e-12 ***
## UniqueAuthors4     0.28397    0.03334    8.52 < 2e-16 ***
## UniqueAuthors5     0.32910    0.03311    9.94 < 2e-16 ***
## Year1997         -0.03510    0.03648   -0.96  0.33604
## Year1998         -0.05682    0.03829   -1.48  0.13789
## Year1999         -0.00888    0.03749   -0.24  0.81267
```

```

## Year2000      -0.07727      0.03970      -1.95      0.05165 .
## Year2001      -0.05140      0.04691      -1.10      0.27327
## Year2002      -0.08883      0.03478      -2.55      0.01068 *
## Year2003      -0.06810      0.03540      -1.92      0.05446 .
## Year2004      -0.10092      0.03423      -2.95      0.00320 **
## Year2005      -0.11650      0.03650      -3.19      0.00142 **
## Year2006      -0.11978      0.03782      -3.17      0.00155 **
## Year2007      -0.10751      0.03629      -2.96      0.00306 **
## Year2008      -0.11524      0.03456      -3.33      0.00086 ***
## Year2009      -0.08490      0.03462      -2.45      0.01423 *
## Year2010      -0.09124      0.03339      -2.73      0.00631 **
## Year2011      -0.13718      0.03314      -4.14      3.5e-05 ***
## Year2012      -0.09399      0.03202      -2.94      0.00334 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0309, Adjusted R-squared:  0.0275
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 4596 is an outlier with |weight| = 0 ( < 1.6e-05);
## 570 weights are ~= 1. The remaining 5733 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0113 0.8570 0.9490 0.8900 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.59e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1 1.014
## LastAuthorFemale 1.016 1 1.008
## Year 1.042 16 1.001

```


Residuals from first and last author



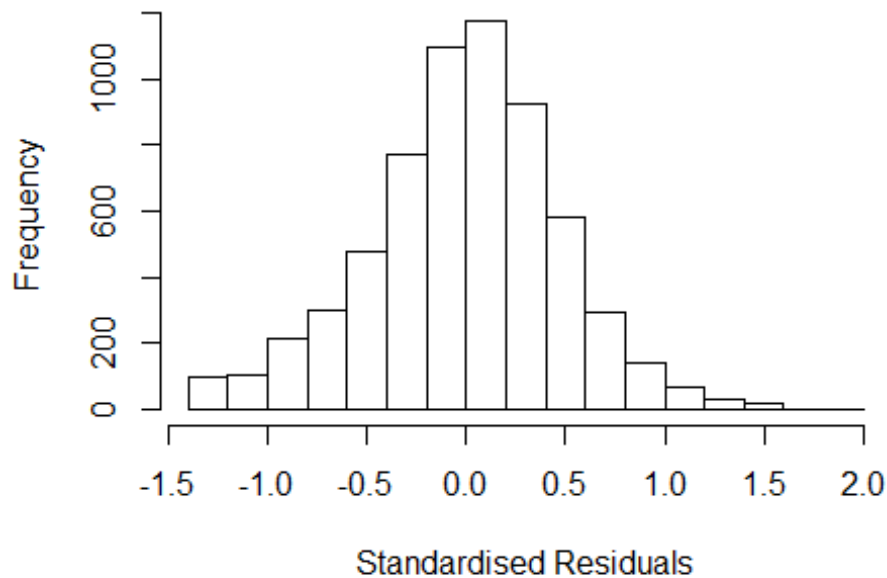
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2714 -0.2919  0.0164  0.3012  1.9364
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.25362    0.02571   48.76  <2e-16 ***
## FirstAuthorFemale1 0.00355    0.01367    0.26  0.7949
## LastAuthorFemale1 0.01979    0.01900    1.04  0.2978
## Year1997        -0.02700    0.03678   -0.73  0.4630
## Year1998        -0.04523    0.03820   -1.18  0.2365
## Year1999        -0.00198    0.03763   -0.05  0.9581
## Year2000        -0.05709    0.03986   -1.43  0.1522
## Year2001        -0.04194    0.04787   -0.88  0.3810
## Year2002        -0.05488    0.03441   -1.59  0.1108
## Year2003        -0.04321    0.03559   -1.21  0.2248
## Year2004        -0.07556    0.03457   -2.19  0.0289 *
## Year2005        -0.08815    0.03660   -2.41  0.0161 *
```

```

## Year2006      -0.09468    0.03781   -2.50    0.0123 *
## Year2007      -0.07954    0.03643   -2.18    0.0290 *
## Year2008      -0.07603    0.03471   -2.19    0.0285 *
## Year2009      -0.05580    0.03476   -1.61    0.1085
## Year2010      -0.05460    0.03353   -1.63    0.1035
## Year2011      -0.09973    0.03312   -3.01    0.0026 **
## Year2012      -0.05633    0.03197   -1.76    0.0781 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.00364,    Adjusted R-squared:  0.000791
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 565 weights are ~= 1. The remaining 5739 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0138 0.8580 0.9490 0.8900 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.59e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.028 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
## -1.2549 -0.2930 0.0168 0.3008 1.9351
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25494 0.02565 48.93 <2e-16 ***
## FirstAuthorFemale1 0.00459 0.01367 0.34 0.7368
## Year1997 -0.02662 0.03676 -0.72 0.4691
## Year1998 -0.04415 0.03815 -1.16 0.2472
## Year1999 -0.00175 0.03757 -0.05 0.9629
## Year2000 -0.05599 0.03979 -1.41 0.1595
## Year2001 -0.04217 0.04783 -0.88 0.3780
## Year2002 -0.05507 0.03439 -1.60 0.1094
## Year2003 -0.04271 0.03557 -1.20 0.2299
## Year2004 -0.07545 0.03454 -2.18 0.0290 *
## Year2005 -0.08799 0.03658 -2.41 0.0162 *
## Year2006 -0.09457 0.03776 -2.50 0.0123 *
```

```

## Year2007          -0.07873    0.03637   -2.16    0.0305 *
## Year2008          -0.07503    0.03466   -2.16    0.0304 *
## Year2009          -0.05464    0.03471   -1.57    0.1156
## Year2010          -0.05317    0.03346   -1.59    0.1121
## Year2011          -0.09878    0.03308   -2.99    0.0028 **
## Year2012          -0.05486    0.03191   -1.72    0.0856 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.00346,    Adjusted R-squared:  0.000766
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 560 weights are ~= 1. The remaining 5744 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0142 0.8580 0.9490 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.59e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.015 1          1.007
## Year          1.015 16          1.000
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.2725 -0.2920  0.0168  0.3016  1.9357

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2543    0.0256  49.06  <2e-16 ***
## LastAuthorFemale1  0.0201    0.0190   1.06   0.2894
## Year1997         -0.0269    0.0368  -0.73   0.4642
## Year1998         -0.0453    0.0382  -1.19   0.2360
## Year1999         -0.0020    0.0376  -0.05   0.9576
## Year2000         -0.0573    0.0399  -1.44   0.1508
## Year2001         -0.0419    0.0479  -0.87   0.3819
## Year2002         -0.0548    0.0344  -1.59   0.1116
## Year2003         -0.0433    0.0356  -1.22   0.2240
## Year2004         -0.0754    0.0346  -2.18   0.0292 *
## Year2005         -0.0880    0.0366  -2.40   0.0162 *
## Year2006         -0.0946    0.0378  -2.50   0.0123 *
## Year2007         -0.0792    0.0364  -2.18   0.0293 *
## Year2008         -0.0758    0.0347  -2.19   0.0288 *
## Year2009         -0.0556    0.0347  -1.60   0.1093
## Year2010         -0.0545    0.0335  -1.63   0.1042
## Year2011         -0.0996    0.0331  -3.01   0.0026 **
## Year2012         -0.0560    0.0319  -1.76   0.0791 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.00363,    Adjusted R-squared:  0.000938
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 568 weights are ~= 1. The remaining 5736 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0139 0.8580 0.9490 0.8900 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.59e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 6304"
## [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
## 10 9 11 11 14 14 17 13 16 22 43 41 18 7 11
## 2011 2012
## 12 14
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2 1 2 4 5 7 9 8 12 11 25 23 13 5 7
## 2011 2012
## 8 8
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1 0 2 3 5 4 7 7 5 10 21 17 11 5 6
## 2011 2012
## 7 7
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4.07673907388117"
## [1] "Male first author team size 2018 geometric mean: 2.84731468735758"

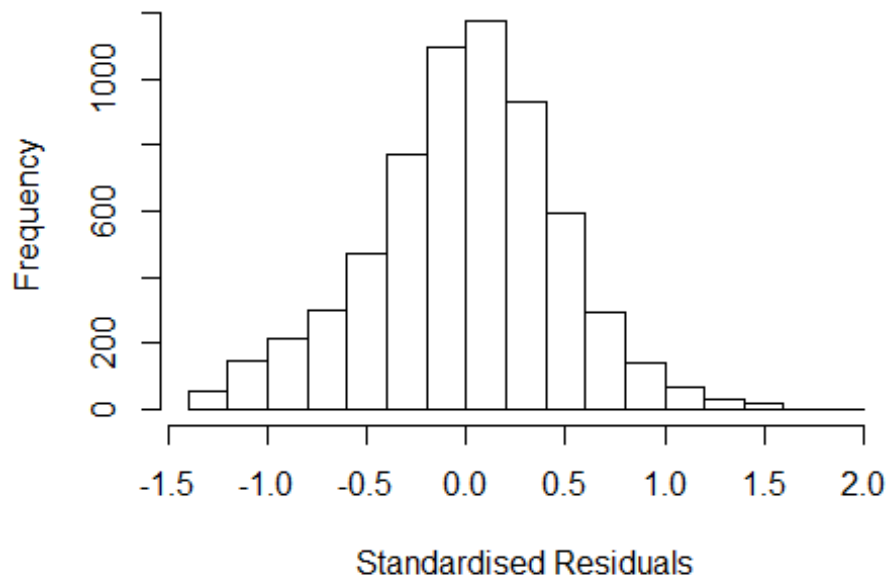
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## 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: 3.94870097166964"
## [1] "Male last author team size 2018 geometric mean: 3.2521387269927"

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

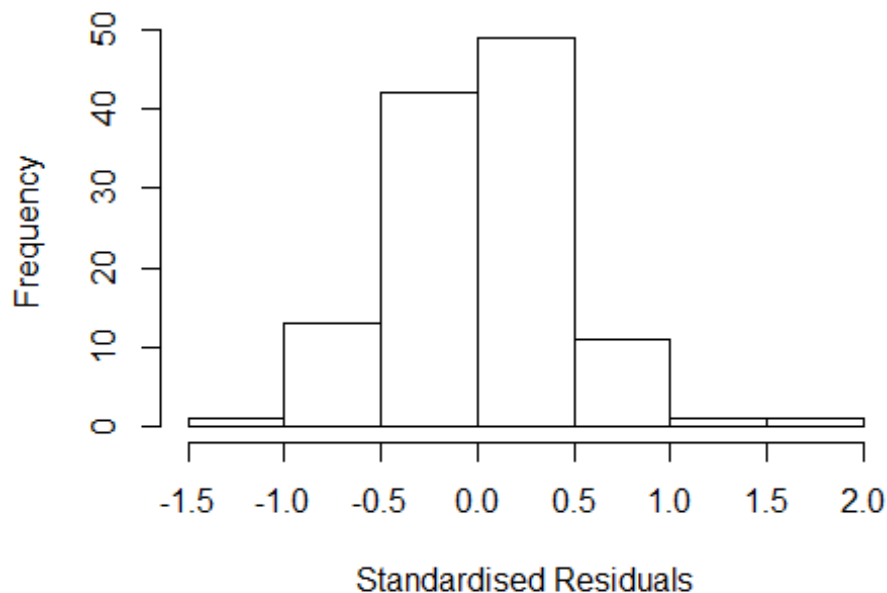
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 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  6.091  1      2.468
## LastAuthorFemale   3.915  1      1.979
## UniqueAuthors     62.981  4      1.678
## Year               327.252 15      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
## -1.3038 -0.2925 0.0601 0.2279 1.5038
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.6672 0.2025 8.23 9.1e-13 ***
## FirstAuthorFemale1 0.1076 0.0957 1.12 0.26348
## LastAuthorFemale1 0.0458 0.1269 0.36 0.71900
## UniqueAuthors2 0.2298 0.2025 1.13 0.25929
## UniqueAuthors3 0.1615 0.1839 0.88 0.38220
## UniqueAuthors4 0.4378 0.2262 1.94 0.05588 .
## UniqueAuthors5 0.4973 0.1938 2.57 0.01183 *
## Year1998 -0.5472 0.1781 -3.07 0.00277 **
## Year1999 -0.6411 0.1793 -3.58 0.00055 ***
## Year2000 -0.8624 0.2111 -4.09 9.1e-05 ***
```

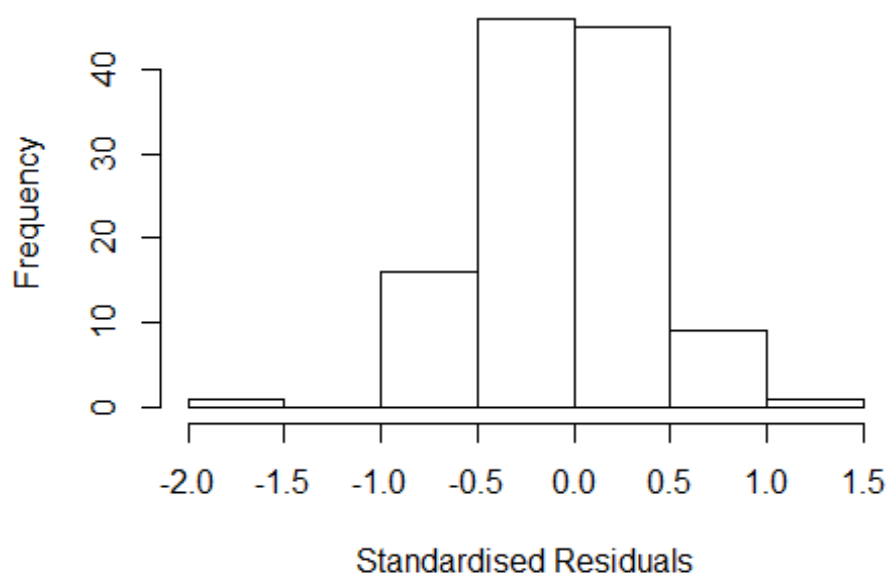


```

## Year2001          -0.6126      0.1521    -4.03  0.00011 ***
## Year2002          -0.7454      0.1490    -5.00  2.6e-06 ***
## Year2003          -0.3624      0.1658    -2.19  0.03131 *
## Year2004          -0.5249      0.3649    -1.44  0.15353
## Year2005          -0.7189      0.1435    -5.01  2.5e-06 ***
## Year2006          -0.9075      0.1517    -5.98  3.8e-08 ***
## Year2007          -0.8628      0.1442    -5.98  3.8e-08 ***
## Year2008          -0.9156      0.2509    -3.65  0.00043 ***
## Year2009          -0.8827      0.1977    -4.47  2.2e-05 ***
## Year2010          -0.4928      0.2425    -2.03  0.04490 *
## Year2011          -1.0313      0.1615    -6.39  6.0e-09 ***
## Year2012          -1.0104      0.1217    -8.30  6.5e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.435
## Multiple R-squared:  0.257, Adjusted R-squared:  0.095
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 112 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.207  0.885   0.961   0.917   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           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 2.978 1 1.726
## LastAuthorFemale 2.621 1 1.619
## Year 5.540 15 1.059

```

Residuals from first and last author



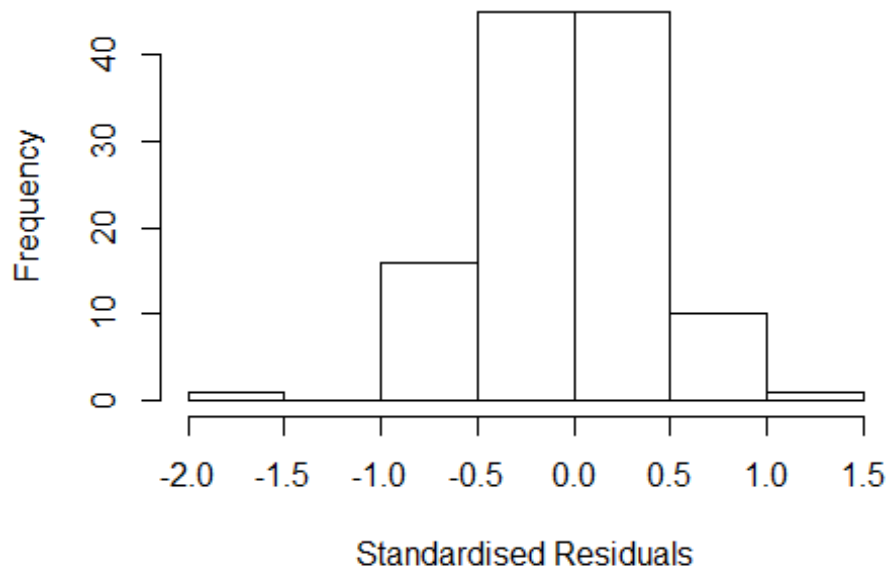
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5225 -0.2436 -0.0484 0.3065 1.4031
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.90e+00 3.33e-08 5.69e+07 < 2e-16 ***
## FirstAuthorFemale1 8.75e-02 8.96e-02 9.80e-01 0.33128
## LastAuthorFemale1 8.78e-02 1.24e-01 7.10e-01 0.48114
## Year1998 -4.28e-01 2.06e-01 -2.08e+00 0.04021 *
## Year1999 -6.61e-01 1.55e-01 -4.26e+00 4.6e-05 ***
## Year2000 -8.74e-01 2.94e-01 -2.97e+00 0.00371 **
## Year2001 -6.21e-01 1.53e-01 -4.07e+00 9.3e-05 ***
## Year2002 -8.12e-01 1.18e-01 -6.90e+00 4.9e-10 ***
## Year2003 -3.76e-01 1.66e-01 -2.26e+00 0.02591 *
## Year2004 -3.75e-01 3.22e-01 -1.16e+00 0.24781
## Year2005 -7.13e-01 1.33e-01 -5.38e+00 4.8e-07 ***
## Year2006 -7.87e-01 1.14e-01 -6.93e+00 4.2e-10 ***
```

```

## Year2007          -8.02e-01    9.50e-02 -8.45e+00    2.5e-13 ***
## Year2008          -8.57e-01    2.35e-01 -3.65e+00    0.00042 ***
## Year2009          -9.16e-01    1.93e-01 -4.75e+00    6.7e-06 ***
## Year2010          -4.48e-01    1.70e-01 -2.63e+00    0.00985 **
## Year2011          -1.01e+00    1.71e-01 -5.90e+00    4.9e-08 ***
## Year2012          -1.03e+00    1.41e-01 -7.30e+00    7.0e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.165, Adjusted R-squared:  0.0235
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 109 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.301  0.907  0.962  0.923  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      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 2.759 1          1.661
## Year              2.759 15          1.034

```

Residuals from first author



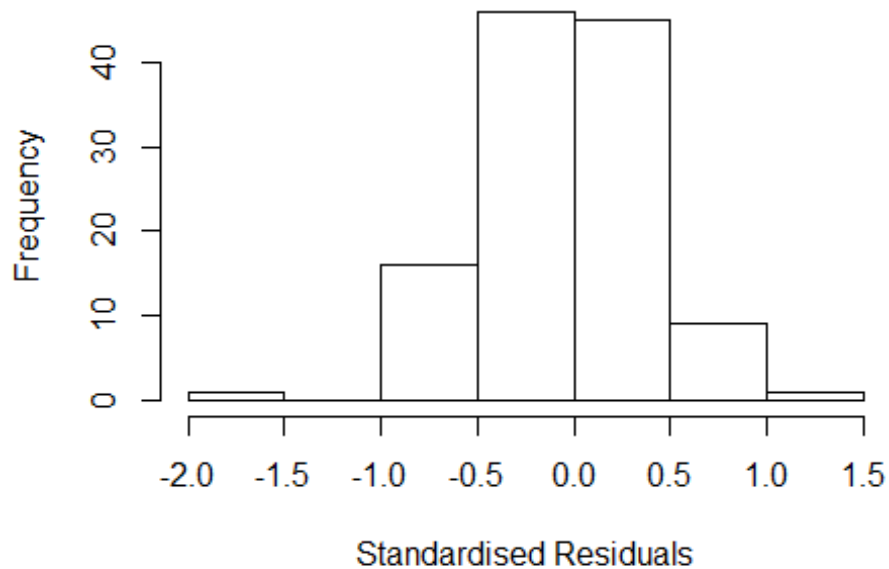
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5632 -0.2459 -0.0212 0.3038 1.4701
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.90e+00 2.76e-08 6.87e+07 < 2e-16 ***
## FirstAuthorFemale1 9.56e-02 8.67e-02 1.10e+00 0.27293
## Year1998 -4.36e-01 2.04e-01 -2.13e+00 0.03558 *
## Year1999 -6.63e-01 1.57e-01 -4.21e+00 5.5e-05 ***
## Year2000 -8.57e-01 2.82e-01 -3.04e+00 0.00299 **
## Year2001 -5.99e-01 1.48e-01 -4.04e+00 0.00010 ***
## Year2002 -8.02e-01 1.15e-01 -6.97e+00 3.4e-10 ***
## Year2003 -3.77e-01 1.66e-01 -2.27e+00 0.02522 *
## Year2004 -3.34e-01 3.31e-01 -1.01e+00 0.31561
## Year2005 -6.79e-01 1.27e-01 -5.36e+00 5.2e-07 ***
## Year2006 -7.86e-01 1.12e-01 -7.01e+00 2.8e-10 ***
## Year2007 -7.99e-01 9.43e-02 -8.47e+00 2.1e-13 ***
```

```

## Year2008          -8.36e-01    2.38e-01 -3.51e+00    0.00067 ***
## Year2009          -8.82e-01    1.88e-01 -4.68e+00    8.8e-06 ***
## Year2010          -4.34e-01    1.70e-01 -2.55e+00    0.01214 *
## Year2011          -9.99e-01    1.76e-01 -5.68e+00    1.3e-07 ***
## Year2012          -9.72e-01    1.17e-01 -8.33e+00    4.1e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.482
## Multiple R-squared:  0.164, Adjusted R-squared:  0.0312
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 110 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.272  0.904   0.962   0.923   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          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 2.564 1        1.601
## Year            2.564 15        1.032

```

Residuals from last author



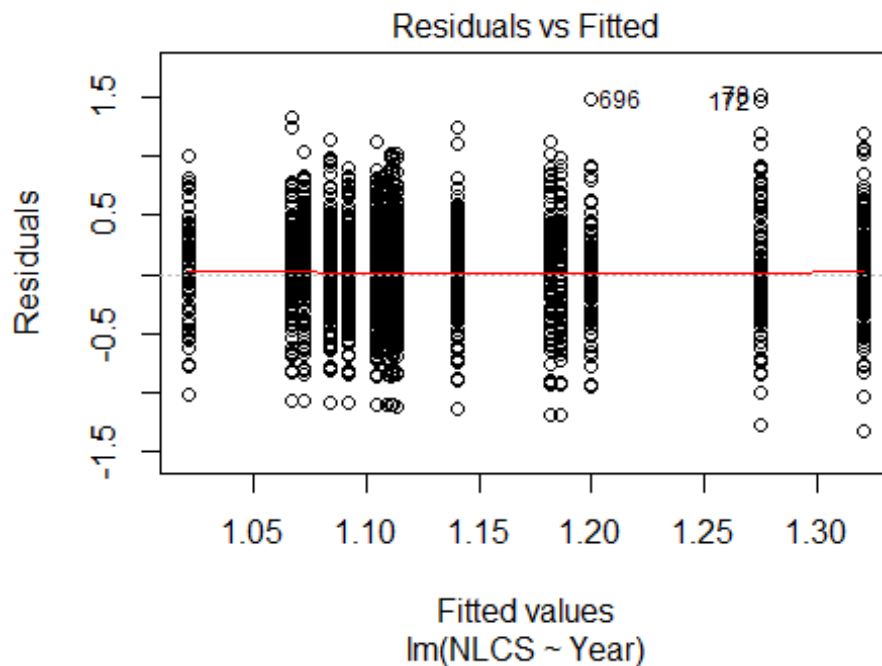
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5568 -0.2402 -0.0553 0.2938 1.3504
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.90e+00 1.78e-08 1.06e+08 < 2e-16 ***
## LastAuthorFemale1 1.00e-01 1.26e-01 8.00e-01 0.42761
## Year1998 -3.40e-01 1.85e-01 -1.83e+00 0.06957 .
## Year1999 -6.34e-01 1.29e-01 -4.90e+00 3.6e-06 ***
## Year2000 -8.33e-01 2.86e-01 -2.91e+00 0.00441 **
## Year2001 -6.03e-01 1.68e-01 -3.59e+00 0.00051 ***
## Year2002 -7.74e-01 1.12e-01 -6.91e+00 4.4e-10 ***
## Year2003 -3.61e-01 1.69e-01 -2.13e+00 0.03522 *
## Year2004 -3.40e-01 3.11e-01 -1.09e+00 0.27627
## Year2005 -6.98e-01 1.32e-01 -5.28e+00 7.4e-07 ***
## Year2006 -7.63e-01 1.10e-01 -6.93e+00 4.1e-10 ***
## Year2007 -7.77e-01 9.26e-02 -8.39e+00 3.1e-13 ***
```

```

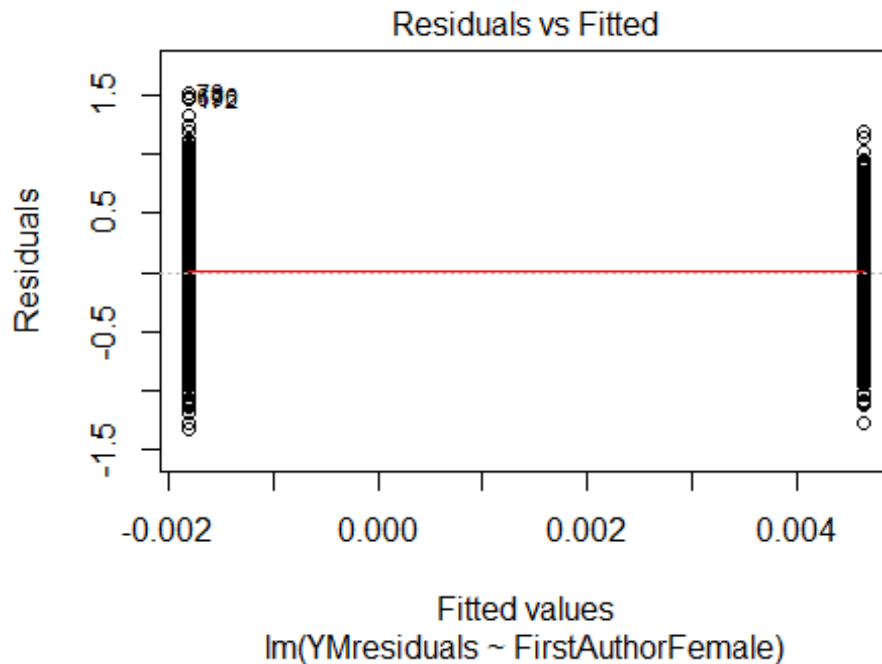
## Year2008          -8.17e-01    2.29e-01 -3.57e+00  0.00055 ***
## Year2009          -8.63e-01    1.80e-01 -4.78e+00  5.9e-06 ***
## Year2010          -4.18e-01    1.60e-01 -2.61e+00  0.01043 *
## Year2011          -9.87e-01    1.74e-01 -5.67e+00  1.4e-07 ***
## Year2012          -9.92e-01    1.55e-01 -6.38e+00  5.3e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.478
## Multiple R-squared:  0.161, Adjusted R-squared:  0.0281
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 5 weights are ~= 1. The remaining 113 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.266  0.895  0.965  0.923  0.994  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+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 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
##  211  239  202  194  169  213  212  216  221  257  275  296  274  303  293
## 2011 2012
##  323  338
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  108  132  108  114   87   75  116  133  131  151  147  162  163  187  198
## 2011 2012
##  206  204

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   91  111   89   98   74   63  101  115  105  129  128  145  130  157  157
## 2011 2012
##  173  171
## [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
```

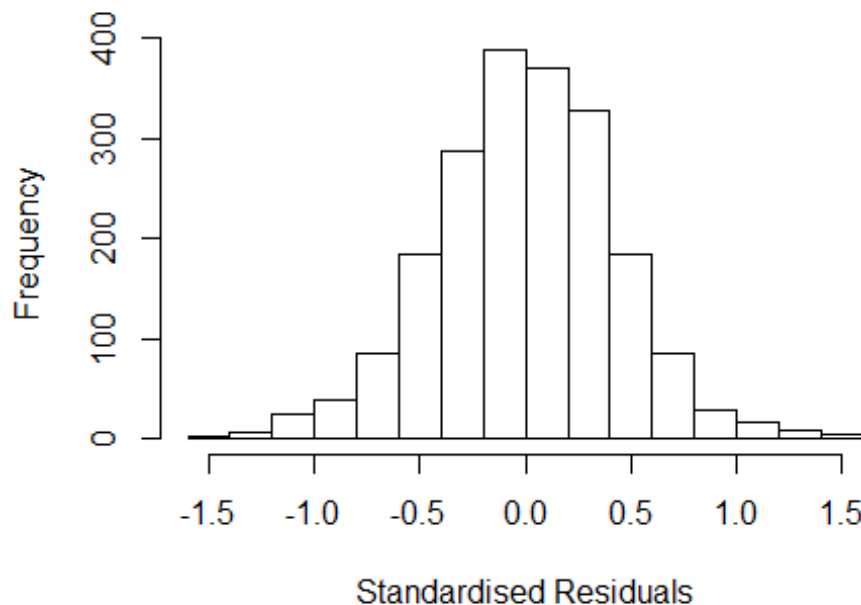


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 1e-04
```

```
## [1] "Female first author team size 2018 geometric mean: 3.27532857866553"
## [1] "Male first author team size 2018 geometric mean: 3.42799332092805"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.00430903290024"
## [1] "Male last author team size 2018 geometric mean: 3.53266904012188"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, 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 1.050 1      1.025
## LastAuthorFemale  1.047 1      1.023
## UniqueAuthors    1.241 4      1.027
## Year              1.333 16      1.009
```

Residuals from first and last author and team size



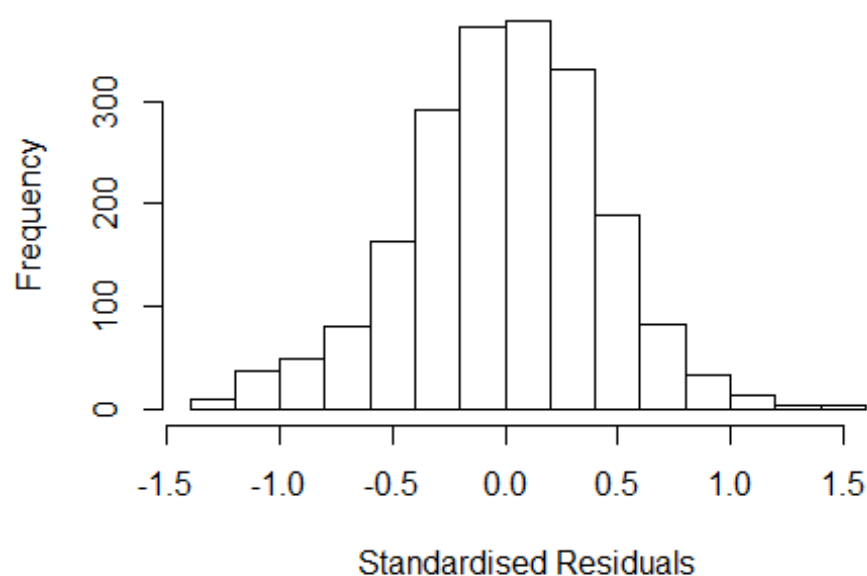
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41583 -0.27163  0.00125  0.27442  1.50310
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9608    0.0776   12.37 < 2e-16 ***
## FirstAuthorFemale1 -0.0172    0.0196   -0.87  0.3817
## LastAuthorFemale1 -0.0351    0.0248   -1.42  0.1570
## UniqueAuthors2     0.3261    0.0576    5.66 1.7e-08 ***
## UniqueAuthors3     0.3608    0.0578    6.24 5.2e-10 ***
## UniqueAuthors4     0.3615    0.0589    6.14 9.8e-10 ***
## UniqueAuthors5     0.4019    0.0599    6.71 2.5e-11 ***
## Year1997           0.0943    0.0779    1.21  0.2263
## Year1998          -0.0756    0.0798   -0.95  0.3438
## Year1999          -0.0880    0.0790   -1.11  0.2655
```

```

## Year2000          -0.0932      0.0878    -1.06    0.2884
## Year2001          -0.2198      0.0899    -2.45    0.0146 *
## Year2002          -0.1839      0.0755    -2.43    0.0150 *
## Year2003          -0.1758      0.0778    -2.26    0.0239 *
## Year2004          -0.2322      0.0736    -3.15    0.0016 **
## Year2005          -0.1850      0.0704    -2.63    0.0087 **
## Year2006          -0.1509      0.0724    -2.08    0.0372 *
## Year2007          -0.1667      0.0689    -2.42    0.0157 *
## Year2008          -0.2199      0.0730    -3.01    0.0026 **
## Year2009          -0.1596      0.0690    -2.31    0.0207 *
## Year2010          -0.2148      0.0692    -3.11    0.0019 **
## Year2011          -0.2067      0.0693    -2.98    0.0029 **
## Year2012          -0.2146      0.0687    -3.12    0.0018 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.0766, Adjusted R-squared:  0.0665
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 177 weights are ~= 1. The remaining 1860 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.134  0.872  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      4.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.050 1      1.025
## LastAuthorFemale  1.045 1      1.022
## Year              1.096 16      1.003

```

Residuals from first and last author



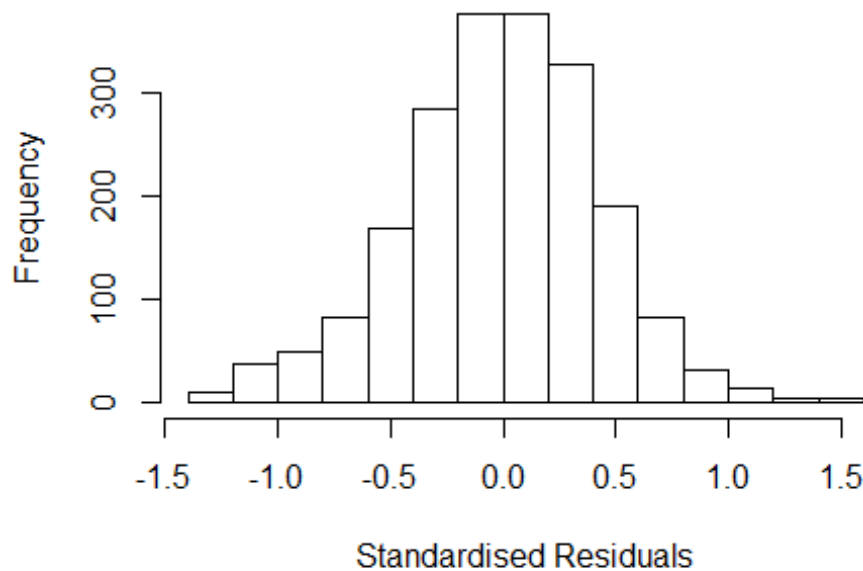
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34589 -0.27411  0.00793  0.27642  1.54242
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.247575   0.062308   20.02  <2e-16 ***
## FirstAuthorFemale1  0.000824   0.020006    0.04   0.967
## LastAuthorFemale1 -0.031858   0.025433   -1.25   0.211
## Year1997         0.098316   0.077261    1.27   0.203
## Year1998        -0.066318   0.077119   -0.86   0.390
## Year1999        -0.078995   0.078562   -1.01   0.315
## Year2000        -0.032827   0.089395   -0.37   0.713
## Year2001        -0.186373   0.089285   -2.09   0.037 *
## Year2002        -0.136325   0.076805   -1.77   0.076 .
## Year2003        -0.124390   0.077964   -1.60   0.111
## Year2004        -0.184374   0.074748   -2.47   0.014 *
## Year2005        -0.129908   0.070924   -1.83   0.067 .
```

```

## Year2006      -0.115748    0.071937   -1.61    0.108
## Year2007      -0.116298    0.069089   -1.68    0.092 .
## Year2008      -0.164741    0.074095   -2.22    0.026 *
## Year2009      -0.104468    0.068539   -1.52    0.128
## Year2010      -0.163143    0.068985   -2.36    0.018 *
## Year2011      -0.144810    0.068781   -2.11    0.035 *
## Year2012      -0.149172    0.068595   -2.17    0.030 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0251, Adjusted R-squared:  0.0164
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 158 weights are ~= 1. The remaining 1879 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.126  0.873   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      4.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.051 1      1.025
## Year              1.051 16      1.002

```

Residuals from first author



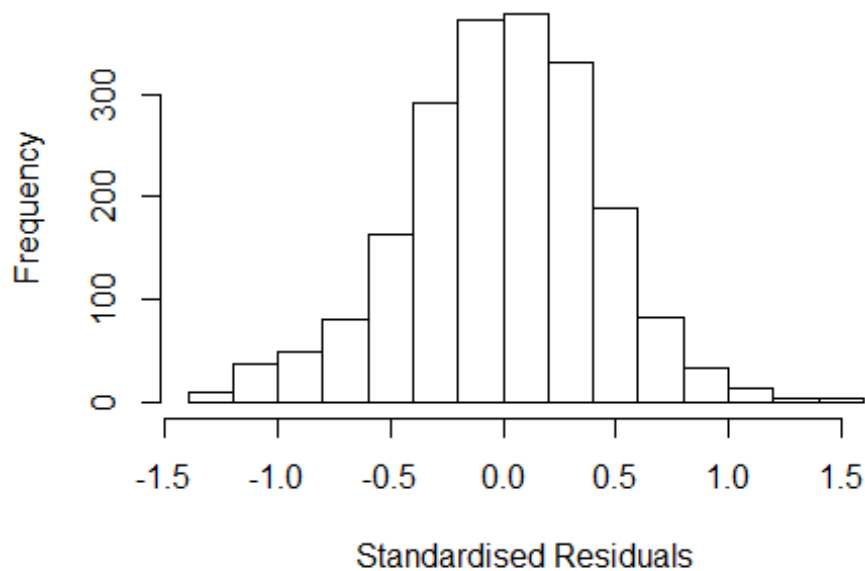
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34332 -0.27543 0.00756 0.27951 1.54938
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.240623 0.061633 20.13 <2e-16 ***
## FirstAuthorFemale1 -0.000211 0.020057 -0.01 0.992
## Year1997 0.102693 0.077061 1.33 0.183
## Year1998 -0.061713 0.076688 -0.80 0.421
## Year1999 -0.073859 0.078264 -0.94 0.345
## Year2000 -0.029132 0.089263 -0.33 0.744
## Year2001 -0.184906 0.088971 -2.08 0.038 *
## Year2002 -0.132876 0.076650 -1.73 0.083 .
## Year2003 -0.122071 0.077861 -1.57 0.117
## Year2004 -0.181969 0.074547 -2.44 0.015 *
## Year2005 -0.125585 0.070633 -1.78 0.076 .
## Year2006 -0.113970 0.071826 -1.59 0.113
```

```

## Year2007          -0.114348    0.068967   -1.66    0.097 .
## Year2008          -0.164178    0.074030   -2.22    0.027 *
## Year2009          -0.101194    0.068414   -1.48    0.139
## Year2010          -0.160685    0.068805   -2.34    0.020 *
## Year2011          -0.141543    0.068472   -2.07    0.039 *
## Year2012          -0.147330    0.068460   -2.15    0.032 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0244, Adjusted R-squared:  0.0162
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 172 weights are ~= 1. The remaining 1865 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.122  0.872  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      4.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.044 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
## -1.34601 -0.27437  0.00799  0.27623  1.54225
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2478    0.0623   20.04  <2e-16 ***
## LastAuthorFemale1 -0.0318    0.0255   -1.25    0.212
## Year1997          0.0983    0.0771    1.27    0.203
## Year1998         -0.0663    0.0771   -0.86    0.390
## Year1999         -0.0790    0.0785   -1.01    0.314
## Year2000         -0.0329    0.0894   -0.37    0.712
## Year2001         -0.1865    0.0893   -2.09    0.037 *
## Year2002         -0.1363    0.0767   -1.78    0.076 .
## Year2003         -0.1244    0.0777   -1.60    0.110
## Year2004         -0.1844    0.0747   -2.47    0.014 *
## Year2005         -0.1299    0.0709   -1.83    0.067 .
## Year2006         -0.1157    0.0717   -1.61    0.107
```

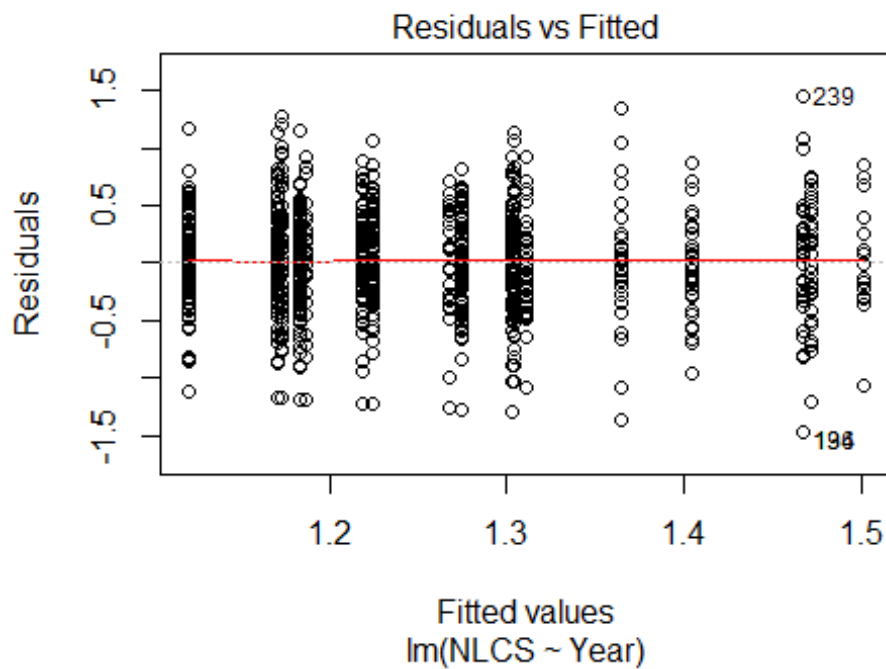


```

## Year2007          -0.1163      0.0690   -1.68    0.092 .
## Year2008          -0.1647      0.0741   -2.22    0.026 *
## Year2009          -0.1044      0.0684   -1.53    0.127
## Year2010          -0.1630      0.0687   -2.37    0.018 *
## Year2011          -0.1447      0.0686   -2.11    0.035 *
## Year2012          -0.1491      0.0685   -2.18    0.030 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0251, Adjusted R-squared:  0.0169
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1880 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.127  0.874  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      4.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2037"
## [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
##  104  98  109  113  105  101  102  119  146  154  162  206  193  197  222
## 2011 2012
##  201  244
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   29  34  39  32  28  17  30  41  64  64  54  89  100  94  107
## 2011 2012

```

```
## 109 125
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 27 29 29 23 23 14 21 32 55 52 41 71 73 65 80
## 2011 2012
## 80 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 = 35, df = 16, p-value = 0.004
```



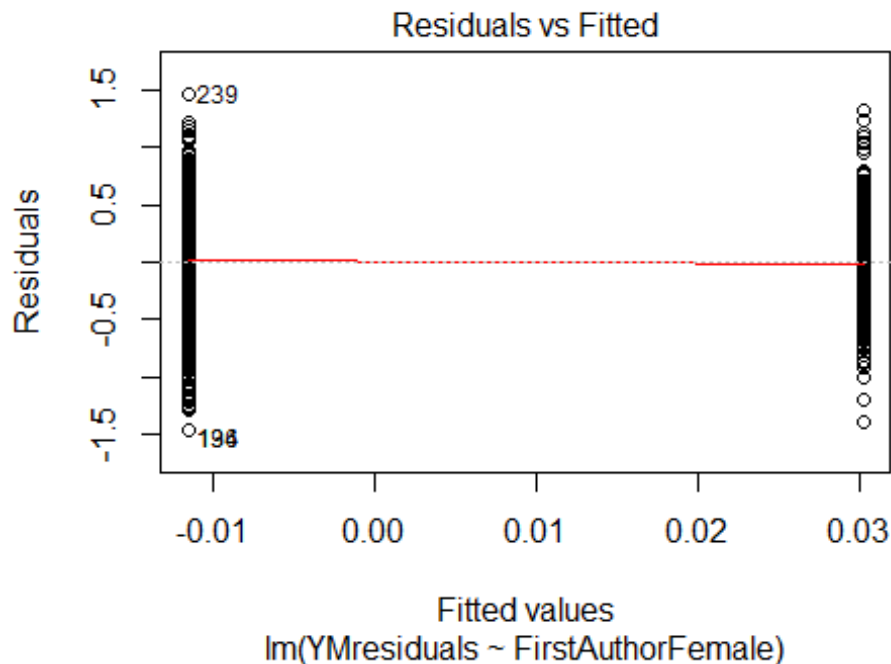
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.2, df = 1, p-value = 0.3

## [1] "Female first author team size 2018 geometric mean: 3.18473785886878"
## [1] "Male first author team size 2018 geometric mean: 3.80622047035429"

## Warning in wilcox.test.default(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 = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.12389697014727"
## [1] "Male last author team size 2018 geometric mean: 3.81100370704497"

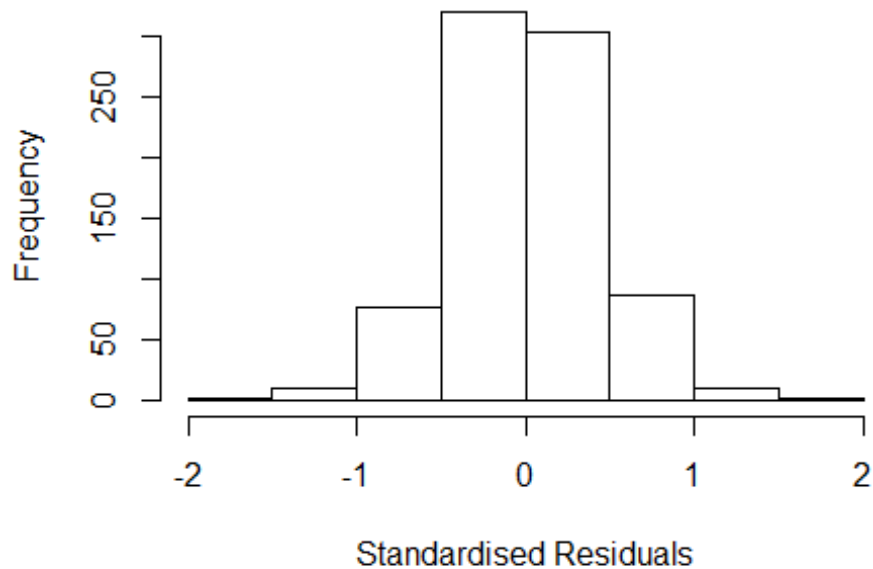
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## 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	1.161	1	1.078
LastAuthorFemale	1.228	1	1.108
UniqueAuthors	1.553	4	1.057
Year	1.820	16	1.019

Residuals from first and last author and team size



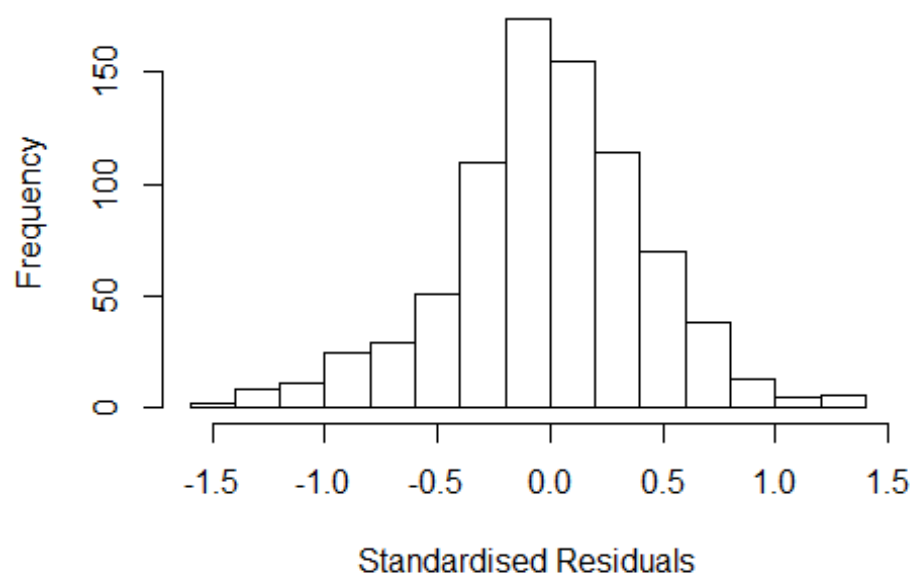
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.59980 -0.25579 -0.00369 0.25218 1.75502
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9668 0.1258 7.69 4.5e-14 ***
## FirstAuthorFemale1 -0.0323 0.0338 -0.96 0.3393
## LastAuthorFemale1 0.0204 0.0421 0.48 0.6279
## UniqueAuthors2 0.3241 0.1006 3.22 0.0013 **
## UniqueAuthors3 0.3622 0.1019 3.55 0.0004 ***
## UniqueAuthors4 0.4188 0.1041 4.02 6.3e-05 ***
## UniqueAuthors5 0.4832 0.1029 4.70 3.1e-06 ***
## Year1997 0.2707 0.1222 2.22 0.0270 *
## Year1998 0.1002 0.1180 0.85 0.3958
## Year1999 0.2043 0.1366 1.49 0.1354
```

```

## Year2000          -0.0781      0.1298    -0.60    0.5475
## Year2001           0.0797      0.1460     0.55    0.5851
## Year2002          -0.1736      0.1199    -1.45    0.1481
## Year2003          -0.0295      0.1136    -0.26    0.7950
## Year2004          -0.1725      0.1106    -1.56    0.1192
## Year2005          -0.0733      0.1069    -0.69    0.4930
## Year2006           0.0191      0.1182     0.16    0.8716
## Year2007          -0.0761      0.1000    -0.76    0.4468
## Year2008          -0.0392      0.1010    -0.39    0.6979
## Year2009          -0.1229      0.1046    -1.17    0.2407
## Year2010          -0.2473      0.0996    -2.48    0.0132 *
## Year2011          -0.2128      0.0961    -2.21    0.0271 *
## Year2012          -0.2110      0.0979    -2.15    0.0315 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.131, Adjusted R-squared:  0.107
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 733 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0043 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          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 1.122 1          1.059
## LastAuthorFemale 1.176 1          1.084
## Year              1.307 16          1.008

```

Residuals from first and last author



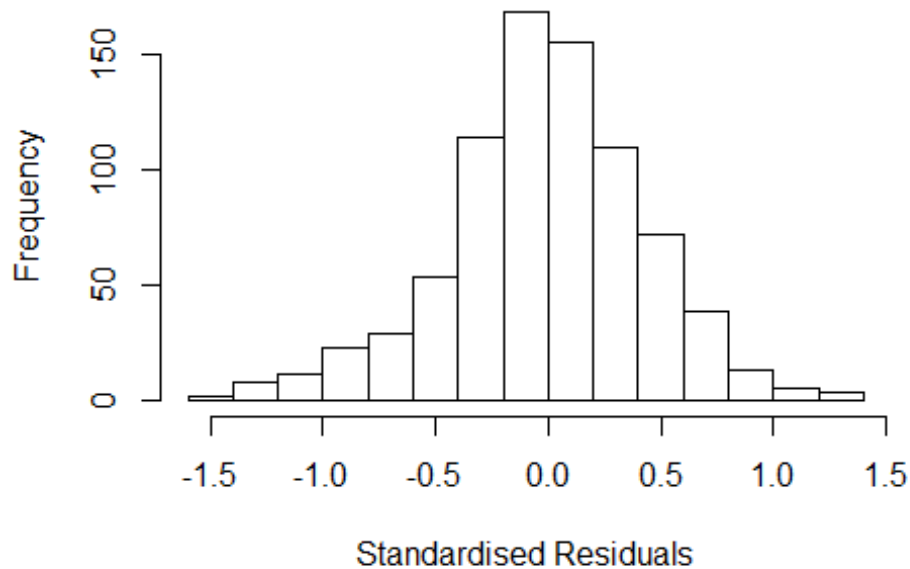
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.55593 -0.25295 -0.00609 0.25136 1.35146
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31266 0.09221 14.24 <2e-16 ***
## FirstAuthorFemale1 0.00315 0.03427 0.09 0.927
## LastAuthorFemale1 0.04272 0.04109 1.04 0.299
## Year1997 0.20055 0.12648 1.59 0.113
## Year1998 0.06265 0.11719 0.53 0.593
## Year1999 0.16704 0.14201 1.18 0.240
## Year2000 -0.06898 0.13791 -0.50 0.617
## Year2001 0.08383 0.14613 0.57 0.566
## Year2002 -0.17540 0.12854 -1.36 0.173
## Year2003 -0.03499 0.11339 -0.31 0.758
## Year2004 -0.18559 0.11781 -1.58 0.116
## Year2005 -0.08295 0.11583 -0.72 0.474
```

```

## Year2006          0.02785      0.12186      0.23      0.819
## Year2007          -0.08198      0.10486     -0.78      0.435
## Year2008          -0.02028      0.10507     -0.19      0.847
## Year2009          -0.09588      0.10898     -0.88      0.379
## Year2010          -0.23425      0.10344     -2.26      0.024 *
## Year2011          -0.19204      0.10126     -1.90      0.058 .
## Year2012          -0.18131      0.10174     -1.78      0.075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.0689, Adjusted R-squared:  0.0477
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 735 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0813 0.8460 0.9490 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      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 1.119 1      1.058
## Year              1.119 16      1.004

```

Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.51308 -0.25860 -0.00678  0.25328  1.39039
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.31507    0.09087   14.47  <2e-16 ***
## FirstAuthorFemale1 0.00454    0.03429    0.13   0.895
## Year1997        0.19801    0.12569    1.58   0.116
## Year1998        0.06122    0.11581    0.53   0.597
## Year1999        0.17127    0.14079    1.22   0.224
## Year2000       -0.06037    0.13627   -0.44   0.658
## Year2001        0.09315    0.14261    0.65   0.514
## Year2002       -0.17737    0.12683   -1.40   0.162
## Year2003       -0.03036    0.11179   -0.27   0.786
## Year2004       -0.18011    0.11658   -1.54   0.123
## Year2005       -0.07521    0.11400   -0.66   0.510
## Year2006        0.02943    0.12060    0.24   0.807
```

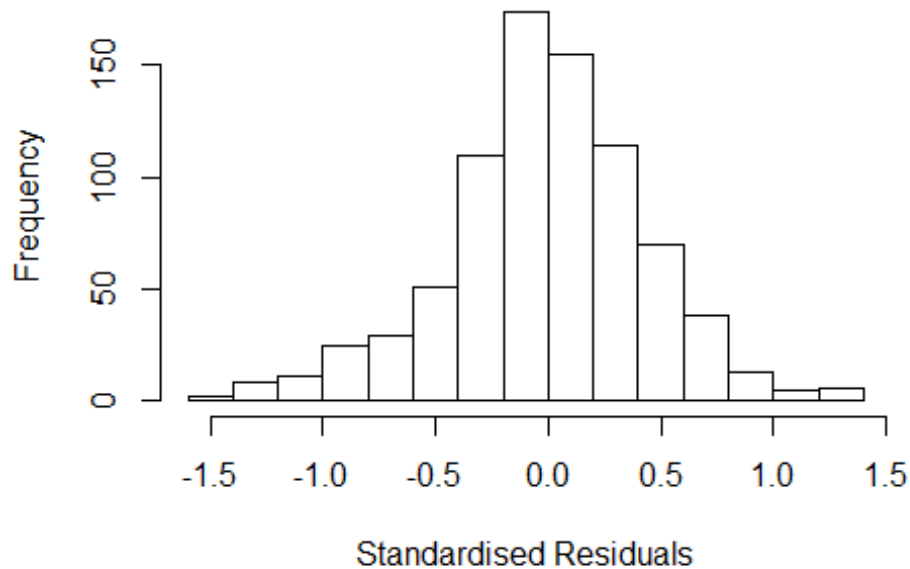


```

## Year2007      -0.07786    0.10375   -0.75    0.453
## Year2008      -0.01581    0.10347   -0.15    0.879
## Year2009      -0.09239    0.10791   -0.86    0.392
## Year2010      -0.22931    0.10204   -2.25    0.025 *
## Year2011      -0.18751    0.09998   -1.88    0.061 .
## Year2012      -0.17914    0.10038   -1.78    0.075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.0677, Adjusted R-squared:  0.0476
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 734 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.106  0.849  0.951  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      1.24e-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 1.17 1          1.082
## Year            1.17 16          1.005

```

Residuals from last author



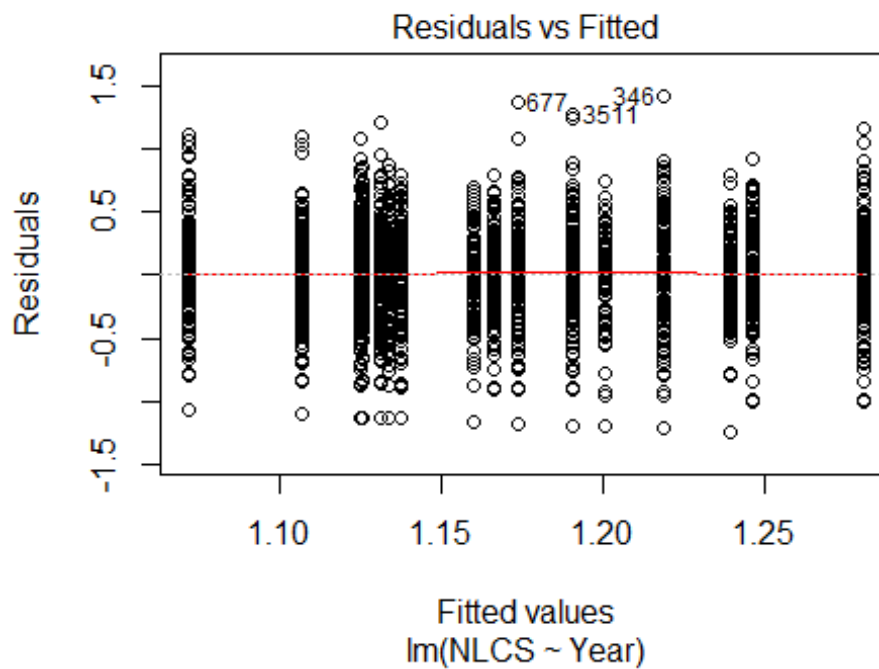
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.55609 -0.25156 -0.00466 0.25358 1.35406
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3130 0.0924 14.21 <2e-16 ***
## LastAuthorFemale1 0.0429 0.0410 1.05 0.295
## Year1997 0.2001 0.1265 1.58 0.114
## Year1998 0.0630 0.1170 0.54 0.590
## Year1999 0.1673 0.1416 1.18 0.238
## Year2000 -0.0685 0.1370 -0.50 0.617
## Year2001 0.0838 0.1462 0.57 0.567
## Year2002 -0.1747 0.1279 -1.37 0.172
## Year2003 -0.0340 0.1122 -0.30 0.762
## Year2004 -0.1845 0.1166 -1.58 0.114
## Year2005 -0.0825 0.1158 -0.71 0.476
## Year2006 0.0280 0.1216 0.23 0.818
```

```

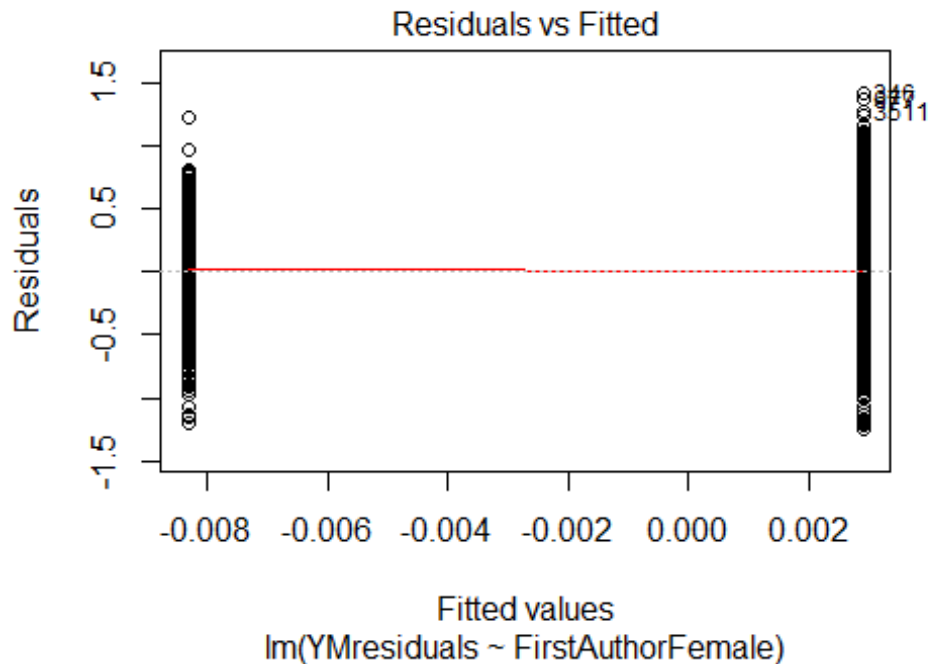
## Year2007          -0.0813      0.1038    -0.78      0.434
## Year2008          -0.0197      0.1046    -0.19      0.851
## Year2009          -0.0953      0.1085    -0.88      0.380
## Year2010          -0.2339      0.1031    -2.27      0.024 *
## Year2011          -0.1917      0.1011    -1.90      0.058 .
## Year2012          -0.1809      0.1016    -1.78      0.075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.0687, Adjusted R-squared:  0.0487
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 734 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0834 0.8470 0.9490 0.8880 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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: 808"
## [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
## 286 243 225 202 187 208 184 217 194 201 221 190 205 231 216
## 2011 2012
## 254 268
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 175 147 148 138 110 91 114 147 129 130 147 136 140 153 149
## 2011 2012

```

```
## 189 196
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 142 126 125 121 97 78 98 120 111 117 122 128 113 124 129
## 2011 2012
## 158 172
## [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
```

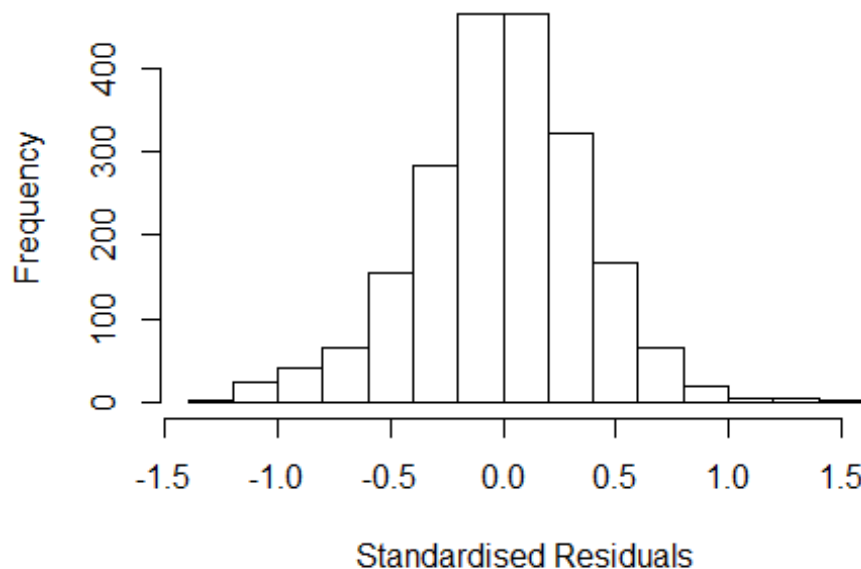


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 3.83373366275646"
## [1] "Male first author team size 2018 geometric mean: 3.2415096159721"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.12983057622487"
## [1] "Male last author team size 2018 geometric mean: 3.21092951904601"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, 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 1.046 1      1.023
## LastAuthorFemale  1.038 1      1.019
## UniqueAuthors    1.284 4      1.032
## Year              1.345 16     1.009
```

Residuals from first and last author and team size



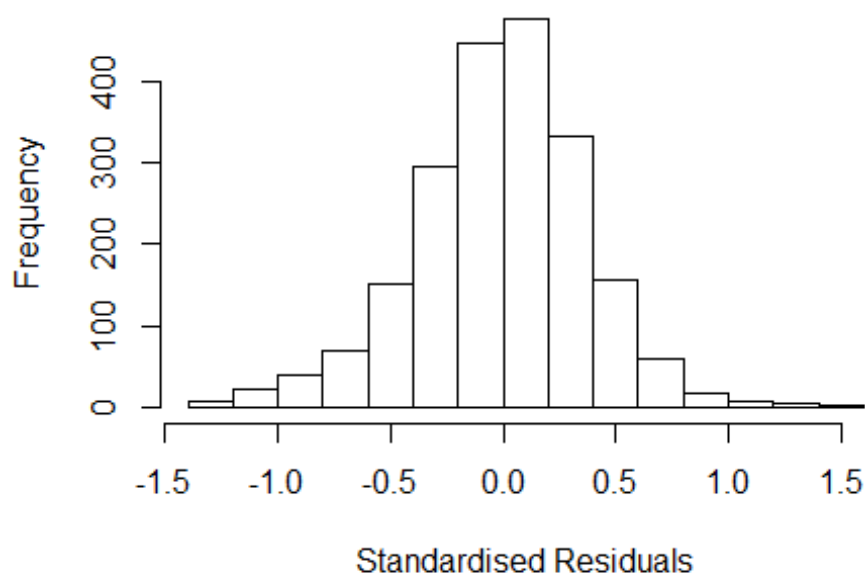
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.249474 -0.229151 0.000526 0.229689 1.426519
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21500 0.07689 15.80 < 2e-16 ***
## FirstAuthorFemale1 0.00678 0.01892 0.36 0.72021
## LastAuthorFemale1 0.00468 0.02781 0.17 0.86639
## UniqueAuthors2 0.04092 0.06969 0.59 0.55720
## UniqueAuthors3 0.07524 0.06898 1.09 0.27551
## UniqueAuthors4 0.09605 0.06934 1.39 0.16612
## UniqueAuthors5 0.13796 0.06966 1.98 0.04778 *
## Year1997 -0.08076 0.05528 -1.46 0.14420
## Year1998 -0.06429 0.04860 -1.32 0.18604
## Year1999 -0.01585 0.04557 -0.35 0.72807
```

```

## Year2000      -0.10349    0.05037    -2.05    0.04005 *
## Year2001      -0.07787    0.05087    -1.53    0.12595
## Year2002      -0.04592    0.04673    -0.98    0.32587
## Year2003      -0.15309    0.04358    -3.51    0.00045 ***
## Year2004      -0.16512    0.04760    -3.47    0.00053 ***
## Year2005      -0.18934    0.04446    -4.26    2.2e-05 ***
## Year2006      -0.19923    0.04688    -4.25    2.2e-05 ***
## Year2007      -0.20552    0.04254    -4.83    1.5e-06 ***
## Year2008      -0.15184    0.04574    -3.32    0.00092 ***
## Year2009      -0.11753    0.04759    -2.47    0.01361 *
## Year2010      -0.18074    0.04768    -3.79    0.00015 ***
## Year2011      -0.23021    0.04624    -4.98    6.9e-07 ***
## Year2012      -0.15561    0.04548    -3.42    0.00063 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.0418, Adjusted R-squared:  0.0316
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 160 weights are ~= 1. The remaining 1921 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0417 0.8630 0.9510 0.8930 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.81e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1 1.022
## LastAuthorFemale 1.027 1 1.014
## Year 1.068 16 1.002

```

Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.25963 -0.22646  0.00464  0.22850  1.42459
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.29050    0.03327   38.79 < 2e-16 ***
## FirstAuthorFemale1 0.00971    0.01886    0.51  0.60676
## LastAuthorFemale1 0.00615    0.02753    0.22  0.82334
## Year1997        -0.07909    0.05514   -1.43  0.15164
## Year1998        -0.05920    0.04928   -1.20  0.22977
## Year1999        -0.02233    0.04487   -0.50  0.61878
## Year2000        -0.09851    0.05049   -1.95  0.05118 .
## Year2001        -0.08122    0.05110   -1.59  0.11212
## Year2002        -0.03702    0.04653   -0.80  0.42633
## Year2003        -0.15141    0.04369   -3.47  0.00054 ***
## Year2004        -0.15034    0.04748   -3.17  0.00157 **
## Year2005        -0.17999    0.04456   -4.04  5.6e-05 ***
```

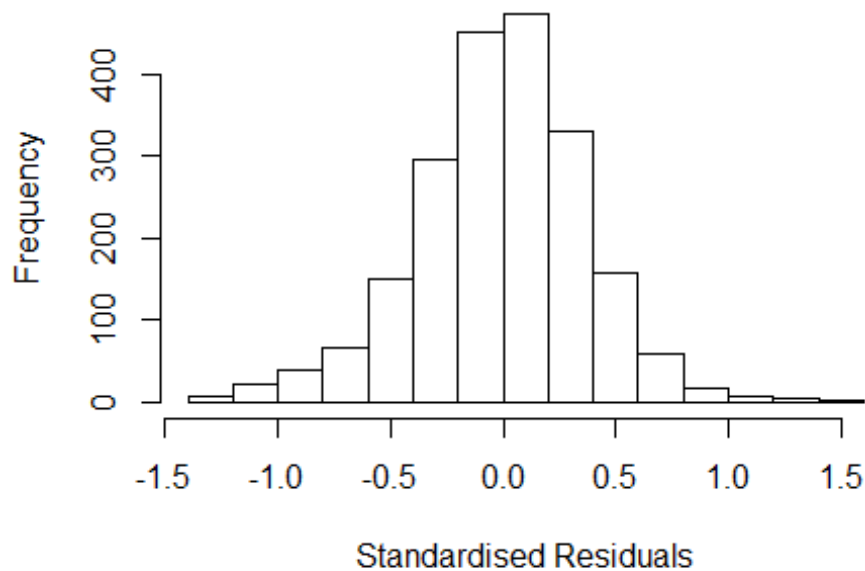


```

## Year2006      -0.18359    0.04670   -3.93  8.7e-05 ***
## Year2007      -0.19656    0.04269   -4.60  4.4e-06 ***
## Year2008      -0.13987    0.04609   -3.03  0.00244 **
## Year2009      -0.10668    0.04812   -2.22  0.02674 *
## Year2010      -0.17373    0.04780   -3.63  0.00029 ***
## Year2011      -0.22230    0.04625   -4.81  1.6e-06 ***
## Year2012      -0.14505    0.04476   -3.24  0.00121 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.342
## Multiple R-squared:  0.0319, Adjusted R-squared:  0.0235
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 161 weights are ~= 1. The remaining 1920 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0433 0.8660 0.9530 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.81e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1      1.021
## Year      1.042 16      1.001

```

Residuals from first author



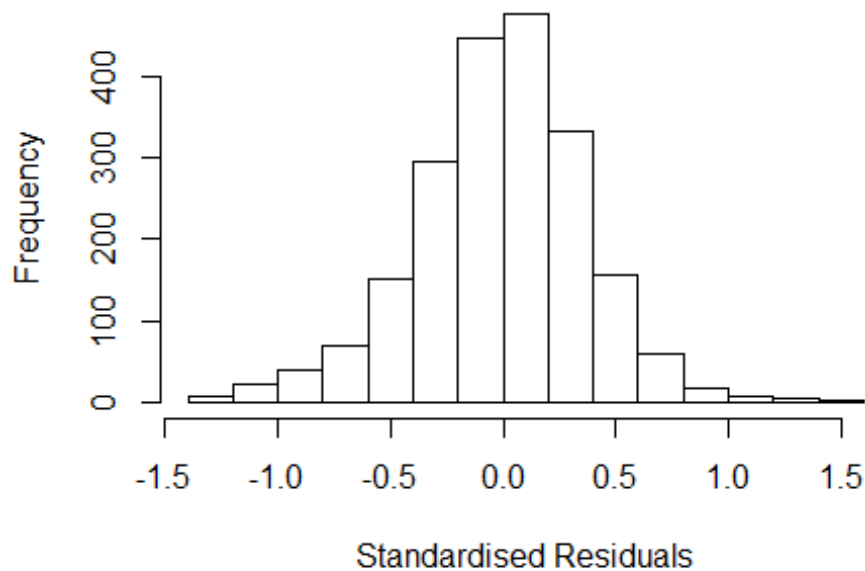
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.25366 -0.22733 0.00442 0.22940 1.42407
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29087 0.03322 38.86 < 2e-16 ***
## FirstAuthorFemale1 0.00995 0.01886 0.53 0.59799
## Year1997 -0.07895 0.05509 -1.43 0.15199
## Year1998 -0.05923 0.04926 -1.20 0.22936
## Year1999 -0.02204 0.04482 -0.49 0.62288
## Year2000 -0.09853 0.05048 -1.95 0.05108 .
## Year2001 -0.08099 0.05106 -1.59 0.11285
## Year2002 -0.03721 0.04651 -0.80 0.42372
## Year2003 -0.15128 0.04366 -3.46 0.00054 ***
## Year2004 -0.15032 0.04747 -3.17 0.00157 **
## Year2005 -0.17989 0.04457 -4.04 5.6e-05 ***
## Year2006 -0.18328 0.04670 -3.92 9.0e-05 ***
```

```

## Year2007          -0.19633    0.04271   -4.60  4.5e-06 ***
## Year2008          -0.13989    0.04611   -3.03  0.00244 **
## Year2009          -0.10611    0.04808   -2.21  0.02742 *
## Year2010          -0.17324    0.04771   -3.63  0.00029 ***
## Year2011          -0.22184    0.04615   -4.81  1.6e-06 ***
## Year2012          -0.14455    0.04464   -3.24  0.00122 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.342
## Multiple R-squared:  0.0319, Adjusted R-squared:  0.0239
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1924 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0438 0.8660 0.9530 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.81e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.025 1          1.013
## Year            1.025 16          1.001

```

Residuals from last author



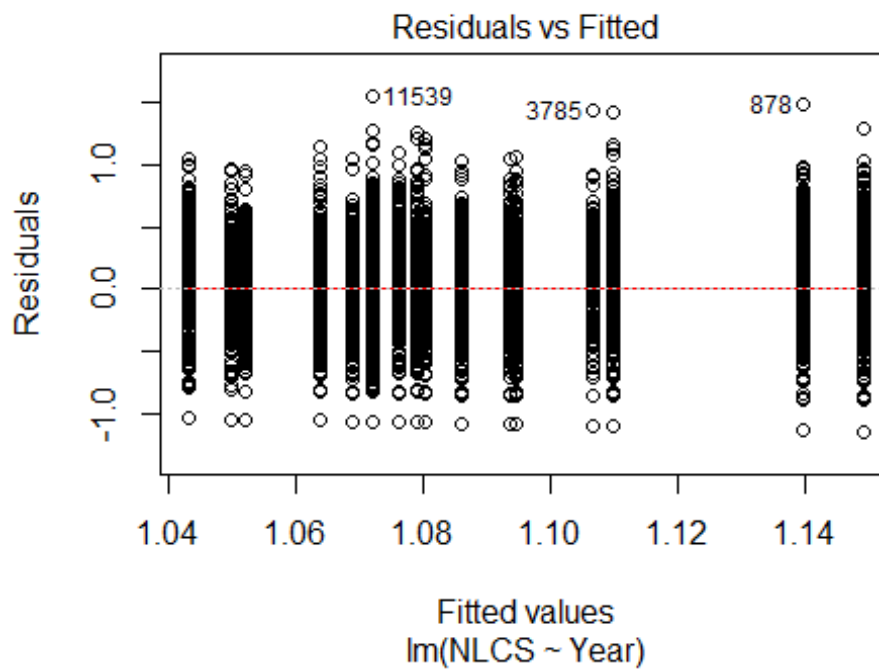
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26228 -0.22851 0.00466 0.22899 1.42244
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29220 0.03305 39.10 < 2e-16 ***
## LastAuthorFemale1 0.00685 0.02758 0.25 0.80377
## Year1997 -0.07863 0.05511 -1.43 0.15377
## Year1998 -0.05885 0.04921 -1.20 0.23182
## Year1999 -0.02322 0.04483 -0.52 0.60446
## Year2000 -0.09851 0.05050 -1.95 0.05121 .
## Year2001 -0.08130 0.05105 -1.59 0.11142
## Year2002 -0.03677 0.04643 -0.79 0.42852
## Year2003 -0.15095 0.04360 -3.46 0.00055 ***
## Year2004 -0.14919 0.04740 -3.15 0.00167 **
## Year2005 -0.17886 0.04446 -4.02 5.9e-05 ***
## Year2006 -0.18255 0.04667 -3.91 9.5e-05 ***
```

```

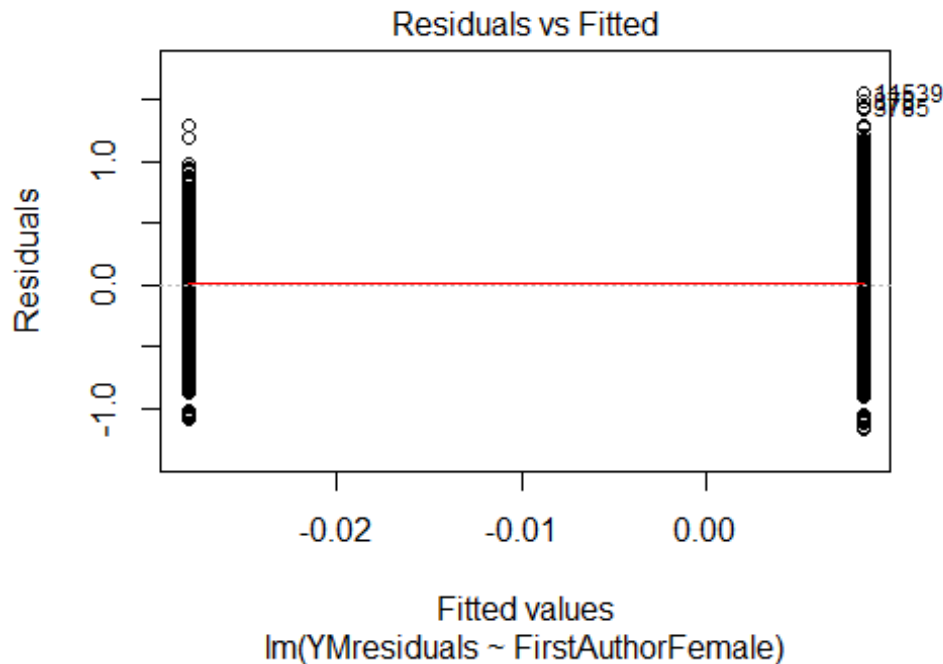
## Year2007          -0.19517      0.04257    -4.59  4.8e-06 ***
## Year2008          -0.13885      0.04602    -3.02  0.00258 **
## Year2009          -0.10568      0.04805    -2.20  0.02794 *
## Year2010          -0.17287      0.04774    -3.62  0.00030 ***
## Year2011          -0.22066      0.04602    -4.80  1.7e-06 ***
## Year2012          -0.14416      0.04461    -3.23  0.00125 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.342
## Multiple R-squared:  0.0317, Adjusted R-squared:  0.0237
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 160 weights are ~= 1. The remaining 1921 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0458 0.8660 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      4.81e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2081"
## [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
## 762 599 640 586 587 560 530 544 559 565 625 584 628 652 662
## 2011 2012
## 675 661
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 487 365 395 376 297 238 341 374 371 360 397 374 417 443 444
## 2011 2012

```

```
## 465 444
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 377 294 319 302 244 197 274 293 303 300 335 313 349 383 370
## 2011 2012
## 397 374
## [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
```

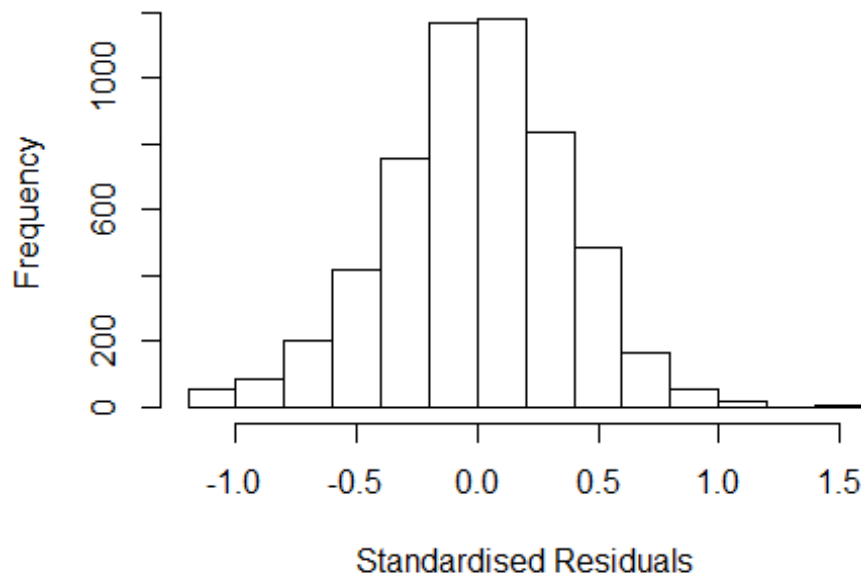


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.3, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 3.79543290316665"
## [1] "Male first author team size 2018 geometric mean: 3.45511967734032"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6900, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.16278955321671"
## [1] "Male last author team size 2018 geometric mean: 3.43812374861329"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5200, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.026 1      1.013
## LastAuthorFemale  1.020 1      1.010
## UniqueAuthors     1.112 4      1.013
## Year              1.139 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
## -1.18765 -0.23693  0.00518  0.23371  1.49615
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06493    0.04382   24.30  < 2e-16 ***
## FirstAuthorFemale1 -0.04344    0.01188   -3.66  0.00026 ***
## LastAuthorFemale1 -0.01020    0.01599   -0.64  0.52358
## UniqueAuthors2     0.05313    0.03891    1.37  0.17218
## UniqueAuthors3     0.08293    0.03909    2.12  0.03390 *
## UniqueAuthors4     0.11151    0.03967    2.81  0.00496 **
## UniqueAuthors5     0.10662    0.03944    2.70  0.00688 **
## Year1997          0.01121    0.03219    0.35  0.72772
## Year1998         -0.00237    0.03278   -0.07  0.94242
## Year1999         -0.03641    0.03127   -1.16  0.24439
```

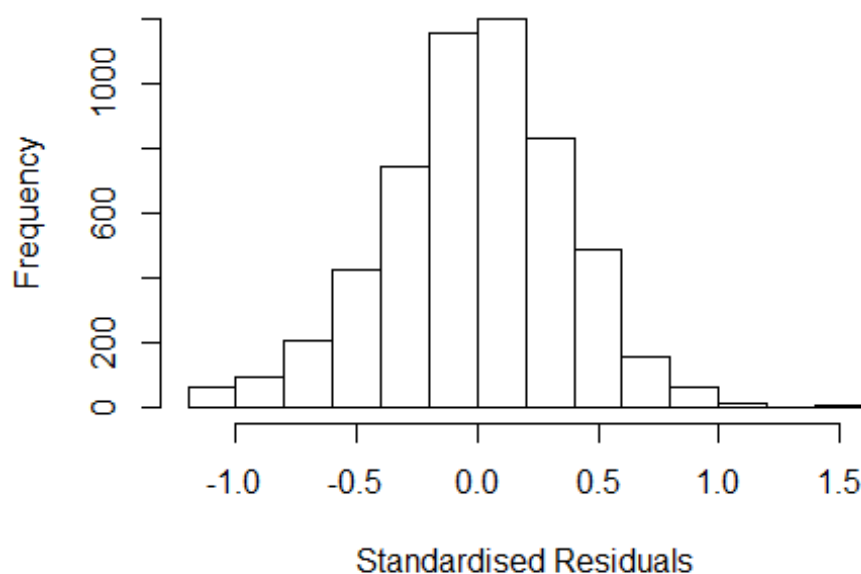


```

## Year2000      -0.01479      0.03251      -0.45      0.64922
## Year2001      -0.00908      0.03358      -0.27      0.78676
## Year2002      -0.04578      0.03204      -1.43      0.15315
## Year2003      -0.06305      0.03049      -2.07      0.03873 *
## Year2004      -0.07502      0.03117      -2.41      0.01613 *
## Year2005      -0.08711      0.03024      -2.88      0.00398 **
## Year2006      -0.04783      0.02938      -1.63      0.10365
## Year2007      -0.05645      0.02979      -1.90      0.05813 .
## Year2008      -0.03038      0.02978      -1.02      0.30772
## Year2009      -0.03469      0.02979      -1.16      0.24426
## Year2010      -0.05583      0.03017      -1.85      0.06426 .
## Year2011      -0.09277      0.03021      -3.07      0.00215 **
## Year2012      -0.05311      0.03077      -1.73      0.08439 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.355
## Multiple R-squared:  0.0138, Adjusted R-squared:  0.00976
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 459 weights are ~= 1. The remaining 4965 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0363 0.8640 0.9510 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          1.84e-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 1.024 1      1.012
## LastAuthorFemale  1.013 1      1.006
## Year              1.035 16      1.001

```

Residuals from first and last author



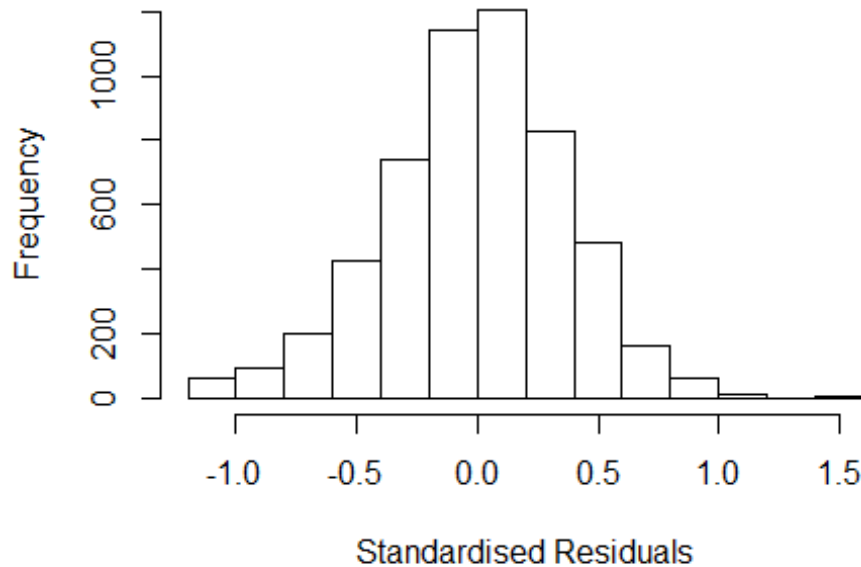
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.15240 -0.23707  0.00474  0.23683  1.48360
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.138238   0.022868  49.77 < 2e-16 ***
## FirstAuthorFemale1 -0.041099   0.011910  -3.45  0.00056 ***
## LastAuthorFemale1 -0.007648   0.016091  -0.48  0.63460
## Year1997         0.014164   0.032296   0.44  0.66099
## Year1998        -0.000921   0.032814  -0.03  0.97760
## Year1999        -0.034902   0.031336  -1.11  0.26541
## Year2000        -0.010450   0.032715  -0.32  0.74942
## Year2001        -0.001867   0.033845  -0.06  0.95602
## Year2002        -0.038188   0.031808  -1.20  0.22996
## Year2003        -0.059773   0.030459  -1.96  0.04977 *
## Year2004        -0.065958   0.031275  -2.11  0.03499 *
## Year2005        -0.075582   0.030253  -2.50  0.01251 *
```

```

## Year2006          -0.039660    0.029403    -1.35    0.17745
## Year2007          -0.045931    0.029823    -1.54    0.12359
## Year2008          -0.018380    0.029851    -0.62    0.53810
## Year2009          -0.024343    0.029678    -0.82    0.41212
## Year2010          -0.044193    0.030140    -1.47    0.14263
## Year2011          -0.083271    0.030341    -2.74    0.00608 **
## Year2012          -0.040103    0.030750    -1.30    0.19223
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.00828,    Adjusted R-squared:  0.00497
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 435 weights are ~= 1. The remaining 4989 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0438 0.8630 0.9530 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      1.84e-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 1.023 1      1.012
## Year              1.023 16      1.001

```

Residuals from first author



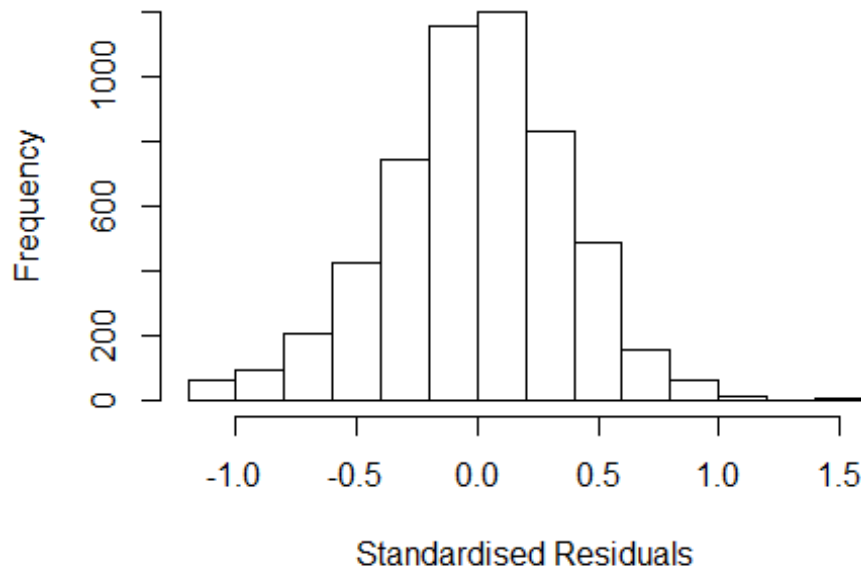
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15159 -0.23650 0.00537 0.23659 1.48441
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.137524 0.022769 49.96 < 2e-16 ***
## FirstAuthorFemale1 -0.041446 0.011911 -3.48 0.00051 ***
## Year1997 0.014063 0.032294 0.44 0.66325
## Year1998 -0.000666 0.032797 -0.02 0.98380
## Year1999 -0.034911 0.031340 -1.11 0.26535
## Year2000 -0.010502 0.032724 -0.32 0.74827
## Year2001 -0.001748 0.033849 -0.05 0.95882
## Year2002 -0.038102 0.031798 -1.20 0.23088
## Year2003 -0.059700 0.030457 -1.96 0.05004 .
## Year2004 -0.066070 0.031278 -2.11 0.03470 *
## Year2005 -0.075455 0.030247 -2.49 0.01264 *
## Year2006 -0.039802 0.029411 -1.35 0.17601
```

```

## Year2007          -0.046127    0.029814    -1.55    0.12188
## Year2008          -0.018397    0.029846    -0.62    0.53765
## Year2009          -0.024461    0.029684    -0.82    0.40994
## Year2010          -0.044329    0.030155    -1.47    0.14160
## Year2011          -0.083339    0.030348    -2.75    0.00605 **
## Year2012          -0.040197    0.030763    -1.31    0.19138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.00823,    Adjusted R-squared:  0.00511
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 437 weights are ~= 1. The remaining 4987 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0433 0.8630 0.9520 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      1.84e-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 1.013 1          1.006
## Year            1.013 16          1.000

```

Residuals from last author



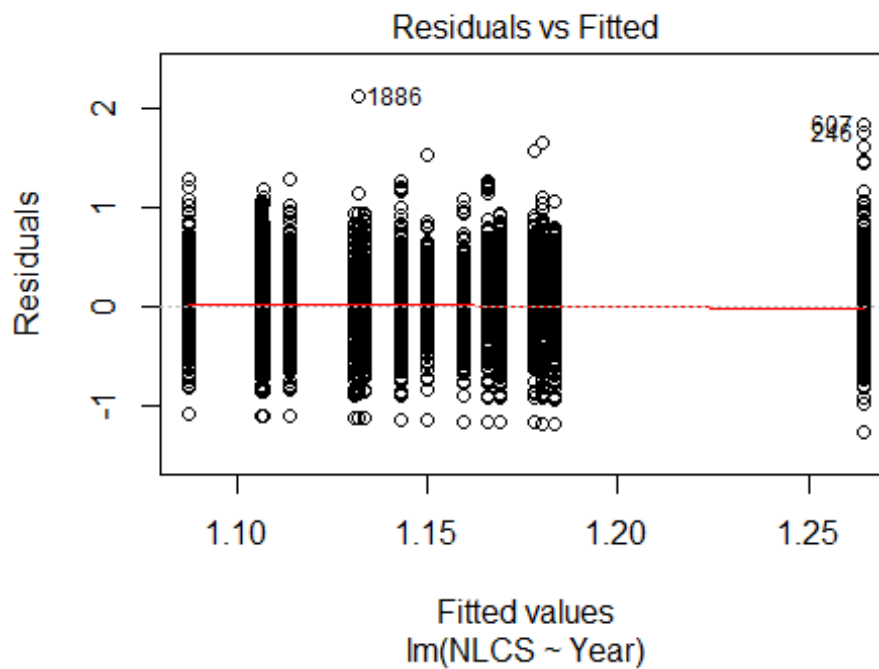
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.14426 -0.23549  0.00617  0.23822  1.49174
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13259    0.02287   49.52  <2e-16 ***
## LastAuthorFemale1 -0.01114    0.01608   -0.69   0.4886
## Year1997         0.01167    0.03236    0.36   0.7183
## Year1998        -0.00424    0.03290   -0.13   0.8975
## Year1999        -0.03645    0.03145   -1.16   0.2465
## Year2000        -0.01249    0.03273   -0.38   0.7027
## Year2001        -0.00459    0.03396   -0.14   0.8925
## Year2002        -0.04103    0.03196   -1.28   0.1993
## Year2003        -0.06241    0.03044   -2.05   0.0404 *
## Year2004        -0.07176    0.03129   -2.29   0.0219 *
## Year2005        -0.07861    0.03037   -2.59   0.0097 **
## Year2006        -0.04404    0.02936   -1.50   0.1336
```

```

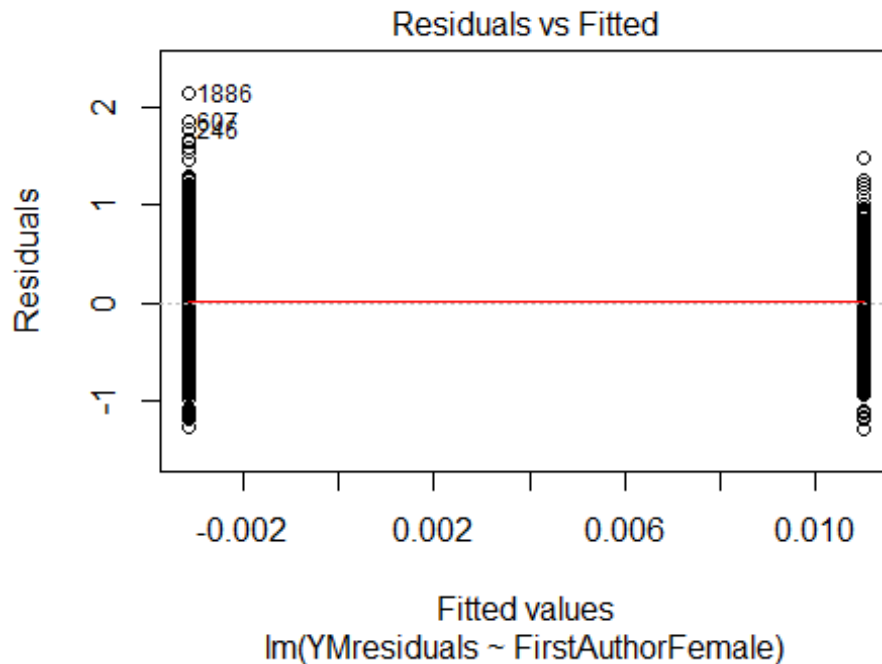
## Year2007          -0.05167      0.02990    -1.73    0.0840 .
## Year2008          -0.02249      0.02996    -0.75    0.4529
## Year2009          -0.03029      0.02959    -1.02    0.3059
## Year2010          -0.04974      0.03016    -1.65    0.0992 .
## Year2011          -0.09046      0.03030    -2.99    0.0028 **
## Year2012          -0.04462      0.03075    -1.45    0.1468
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.357
## Multiple R-squared:  0.00606,    Adjusted R-squared:  0.00293
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 448 weights are ~= 1. The remaining 4976 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.041  0.865  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      1.84e-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: 5424"
## [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
## 633 586 556 549 609 514 508 512 555 644 718 678 742 726 692
## 2011 2012
## 769 698
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 309 283 281 272 310 211 267 299 337 376 406 395 441 454 430
## 2011 2012

```

```
## 479 443
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 257 217 234 222 268 172 219 247 281 320 345 323 352 372 342
## 2011 2012
## 381 362
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 85, df = 16, p-value = 2e-11
```

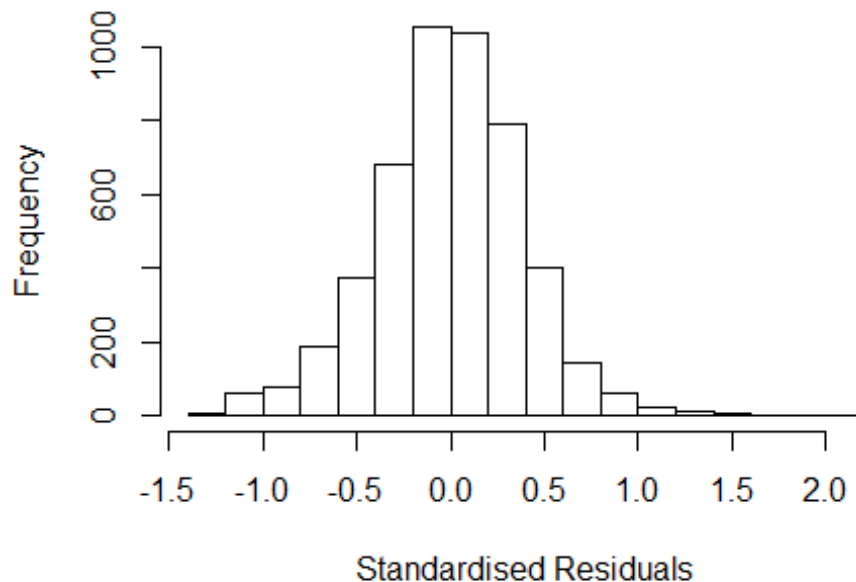


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 22, df = 1, p-value = 2e-06
```

```
## [1] "Female first author team size 2018 geometric mean: 3.72653341017803"
## [1] "Male first author team size 2018 geometric mean: 2.80449497434341"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, p-value = 5e-05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.62321628401151"
## [1] "Male last author team size 2018 geometric mean: 2.87942063592905"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8400, 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 1.036 1      1.018
## LastAuthorFemale  1.021 1      1.011
## UniqueAuthors    1.067 4      1.008
## Year              1.102 16     1.003
```

Residuals from first and last author and team size



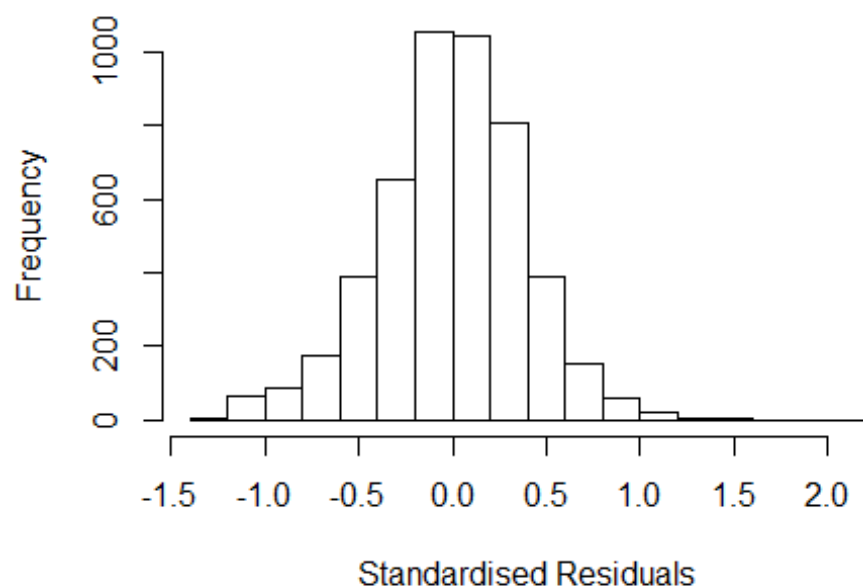
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27859 -0.23475  0.00384  0.24235  2.05859
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12830    0.03949   28.57 < 2e-16 ***
## FirstAuthorFemale1  0.00888    0.01256    0.71  0.47953
## LastAuthorFemale1 -0.01027    0.01762   -0.58  0.56001
## UniqueAuthors2     0.11259    0.02963    3.80  0.00015 ***
## UniqueAuthors3     0.14141    0.02972    4.76  2.0e-06 ***
## UniqueAuthors4     0.17089    0.03047    5.61  2.1e-08 ***
## UniqueAuthors5     0.19252    0.03113    6.18  6.8e-10 ***
## Year1997          -0.08333    0.04134   -2.02  0.04392 *
## Year1998          -0.09678    0.03954   -2.45  0.01443 *
## Year1999          -0.07182    0.03857   -1.86  0.06268 .
```

```

## Year2000      -0.07454      0.03657      -2.04      0.04157 *
## Year2001      -0.10134      0.03888      -2.61      0.00918 **
## Year2002      -0.05994      0.03916      -1.53      0.12591
## Year2003      -0.13721      0.03590      -3.82      0.00013 ***
## Year2004      -0.09735      0.03565      -2.73      0.00634 **
## Year2005      -0.15447      0.03662      -4.22      2.5e-05 ***
## Year2006      -0.12251      0.03591      -3.41      0.00065 ***
## Year2007      -0.14502      0.03509      -4.13      3.6e-05 ***
## Year2008      -0.13610      0.03495      -3.89      0.00010 ***
## Year2009      -0.08842      0.03483      -2.54      0.01115 *
## Year2010      -0.13658      0.03544      -3.85      0.00012 ***
## Year2011      -0.15671      0.03481      -4.50      6.9e-06 ***
## Year2012      -0.15414      0.03487      -4.42      1.0e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.359
## Multiple R-squared:  0.0259, Adjusted R-squared:  0.0215
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 3 observations c(222,292,621) are outliers with |weight| = 0 ( < 2e-05);
## 397 weights are ~= 1. The remaining 4514 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0049 0.8670 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           2.04e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1 1.012
## LastAuthorFemale 1.021 1 1.010
## Year 1.044 16 1.001

```

Residuals from first and last author



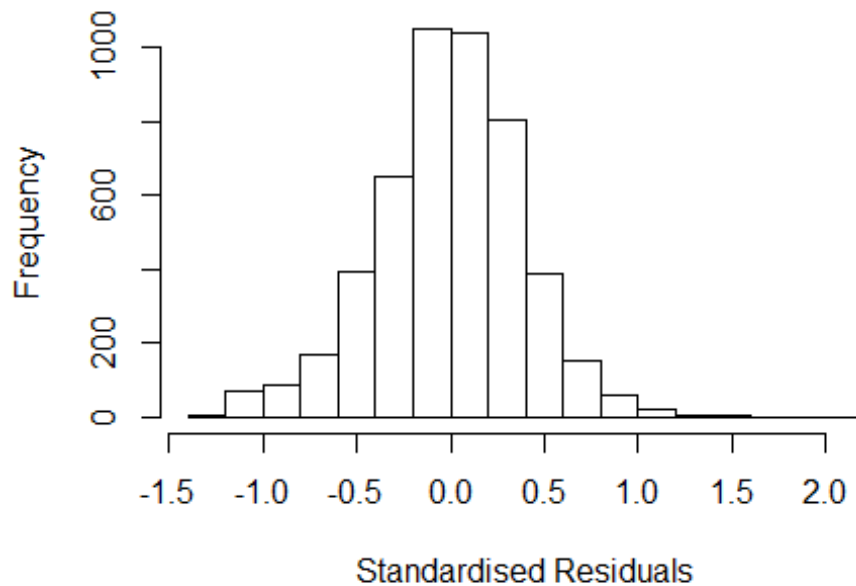
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27530 -0.23387 0.00626 0.24117 2.10689
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25583 0.02980 42.14 < 2e-16 ***
## FirstAuthorFemale1 0.01948 0.01257 1.55 0.12121
## LastAuthorFemale1 -0.00577 0.01764 -0.33 0.74344
## Year1997 -0.09141 0.04219 -2.17 0.03031 *
## Year1998 -0.10172 0.04046 -2.51 0.01197 *
## Year1999 -0.07661 0.03919 -1.95 0.05064 .
## Year2000 -0.07427 0.03749 -1.98 0.04762 *
## Year2001 -0.10890 0.03985 -2.73 0.00630 **
## Year2002 -0.05562 0.03931 -1.42 0.15709
## Year2003 -0.13730 0.03664 -3.75 0.00018 ***
## Year2004 -0.09498 0.03639 -2.61 0.00908 **
## Year2005 -0.14940 0.03748 -3.99 6.8e-05 ***
```

```

## Year2006      -0.11903    0.03676   -3.24  0.00121 **
## Year2007      -0.13929    0.03573   -3.90  9.8e-05 ***
## Year2008      -0.12867    0.03578   -3.60  0.00033 ***
## Year2009      -0.08018    0.03556   -2.25  0.02419 *
## Year2010      -0.12902    0.03630   -3.55  0.00038 ***
## Year2011      -0.14571    0.03553   -4.10  4.2e-05 ***
## Year2012      -0.14203    0.03554   -4.00  6.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.0101, Adjusted R-squared:  0.00648
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(222,621) are outliers with |weight| = 0 ( < 2e-05);
## 414 weights are ~ = 1. The remaining 4498 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0002 0.8640 0.9510 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1          1.011
## Year              1.023 16          1.001

```

Residuals from first author



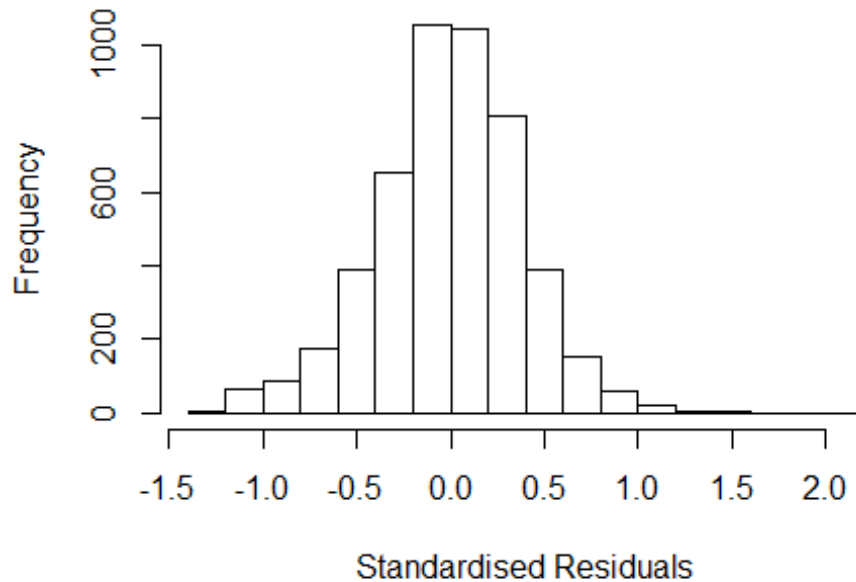
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27467 -0.23326 0.00606 0.24082 2.10724
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2554 0.0298 42.18 < 2e-16 ***
## FirstAuthorFemale1 0.0192 0.0126 1.53 0.12594
## Year1997 -0.0916 0.0422 -2.17 0.03000 *
## Year1998 -0.1017 0.0405 -2.51 0.01203 *
## Year1999 -0.0767 0.0392 -1.96 0.05046 .
## Year2000 -0.0743 0.0375 -1.98 0.04762 *
## Year2001 -0.1089 0.0399 -2.73 0.00630 **
## Year2002 -0.0557 0.0393 -1.42 0.15661
## Year2003 -0.1374 0.0367 -3.75 0.00018 ***
## Year2004 -0.0950 0.0364 -2.61 0.00905 **
## Year2005 -0.1493 0.0375 -3.98 6.9e-05 ***
## Year2006 -0.1191 0.0368 -3.24 0.00120 **
```

```

## Year2007          -0.1396      0.0357   -3.91  9.5e-05 ***
## Year2008          -0.1289      0.0358   -3.60  0.00032 ***
## Year2009          -0.0805      0.0356   -2.26  0.02361 *
## Year2010          -0.1295      0.0363   -3.57  0.00036 ***
## Year2011          -0.1462      0.0355   -4.12  3.9e-05 ***
## Year2012          -0.1422      0.0356   -4.00  6.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.0101, Adjusted R-squared:  0.00665
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(222,621) are outliers with |weight| = 0 ( < 2e-05);
## 407 weights are ~= 1. The remaining 4505 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8650 0.9520 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.02 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
## -1.25833 -0.23400 0.00531 0.24170 2.10404
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25833 0.02971 42.36 < 2e-16 ***
## LastAuthorFemale1 -0.00428 0.01769 -0.24 0.80887
## Year1997 -0.09025 0.04223 -2.14 0.03266 *
## Year1998 -0.10137 0.04045 -2.51 0.01225 *
## Year1999 -0.07628 0.03919 -1.95 0.05166 .
## Year2000 -0.07338 0.03746 -1.96 0.05017 .
## Year2001 -0.10688 0.03975 -2.69 0.00720 **
## Year2002 -0.05414 0.03928 -1.38 0.16817
## Year2003 -0.13547 0.03661 -3.70 0.00022 ***
## Year2004 -0.09236 0.03631 -2.54 0.01101 *
## Year2005 -0.14731 0.03741 -3.94 8.3e-05 ***
## Year2006 -0.11639 0.03667 -3.17 0.00151 **
```

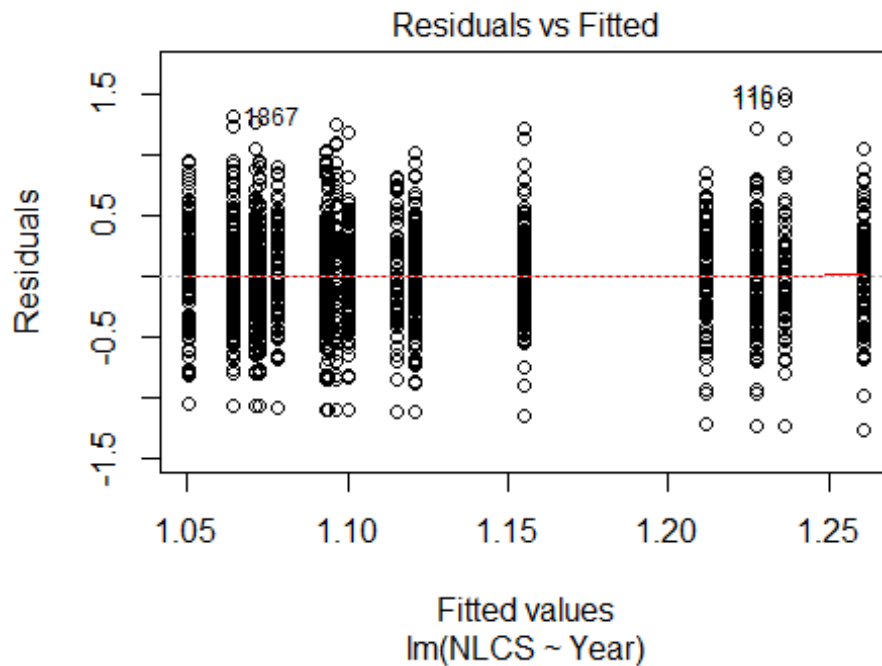


```

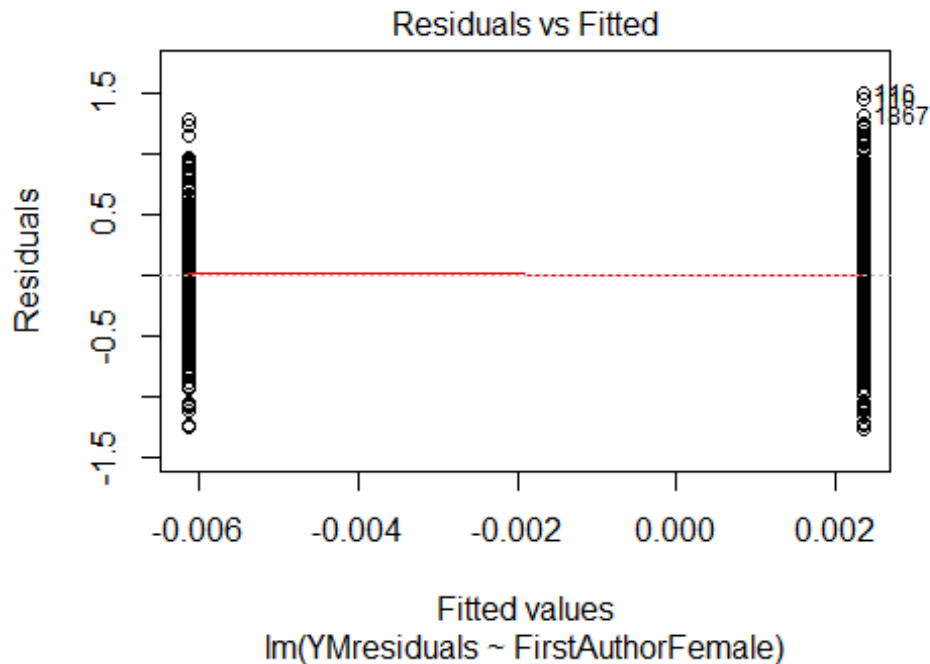
## Year2007          -0.13661      0.03563    -3.83   0.00013 ***
## Year2008          -0.12718      0.03573    -3.56   0.00038 ***
## Year2009          -0.07813      0.03553    -2.20   0.02791 *
## Year2010          -0.12751      0.03625    -3.52   0.00044 ***
## Year2011          -0.14300      0.03541    -4.04   5.5e-05 ***
## Year2012          -0.13979      0.03544    -3.94   8.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.00964,    Adjusted R-squared:  0.0062
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(222,621) are outliers with |weight| = 0 ( < 2e-05);
## 413 weights are ~ = 1. The remaining 4499 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0003 0.8650 0.9510 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4914"
## [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
## 168 194 168 195 172 181 184 185 173 227 194 214 189 234 227
## 2011 2012
## 233 230
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 98 75 100 89 71 97 113 108 129 101 123 104 132 135

```

```
## 2011 2012
## 135 141
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 55 82 57 74 75 58 80 93 94 110 86 107 88 100 108
## 2011 2012
## 114 114
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 41, df = 16, p-value = 6e-04
```

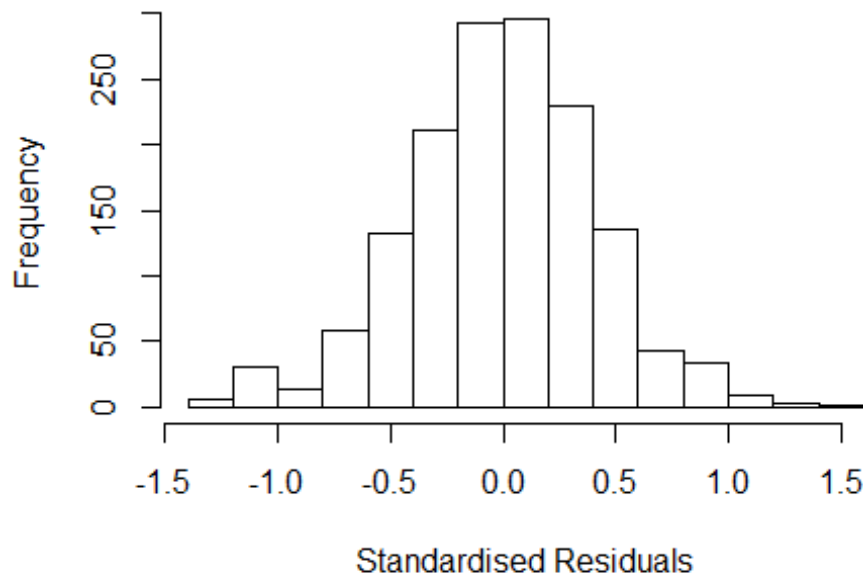


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.19, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 3.80276026971024"
## [1] "Male first author team size 2018 geometric mean: 3.63119864464823"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.56281580252378"
## [1] "Male last author team size 2018 geometric mean: 3.7820261837918"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, 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 1.064 1      1.031
## LastAuthorFemale  1.088 1      1.043
## UniqueAuthors    1.347 4      1.038
## Year              1.348 16     1.009
```

Residuals from first and last author and team size



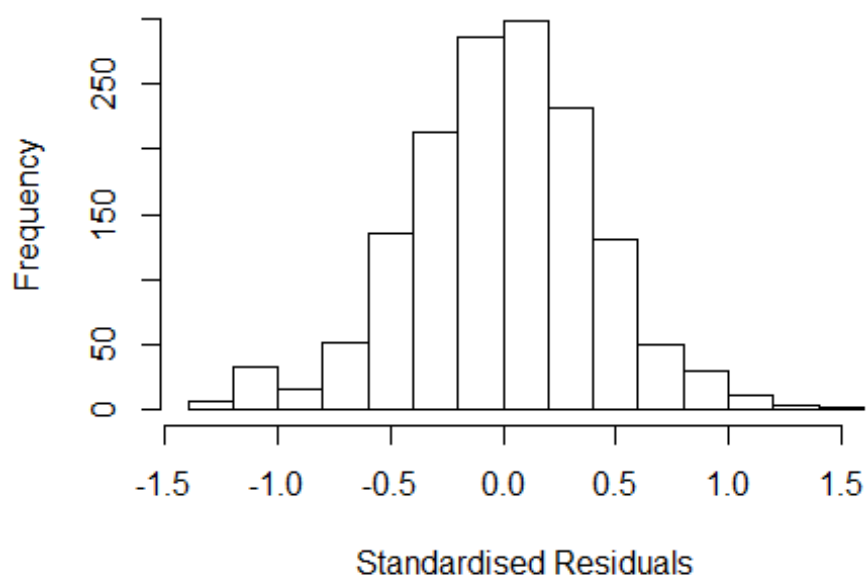
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.299268 -0.266096 0.000387 0.262703 1.529886
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17137 0.07186 16.30 <2e-16 ***
## FirstAuthorFemale1 -0.02401 0.02417 -0.99 0.321
## LastAuthorFemale1 0.01148 0.03129 0.37 0.714
## UniqueAuthors2 -0.00721 0.05557 -0.13 0.897
## UniqueAuthors3 0.03574 0.05696 0.63 0.530
## UniqueAuthors4 0.04360 0.05737 0.76 0.447
## UniqueAuthors5 0.06638 0.05874 1.13 0.259
## Year1997 0.08791 0.07810 1.13 0.261
## Year1998 -0.05219 0.08944 -0.58 0.560
## Year1999 0.11616 0.07955 1.46 0.144
```

```

## Year2000      0.06077      0.08167      0.74      0.457
## Year2001     -0.11763      0.08631     -1.36      0.173
## Year2002     -0.12502      0.07536     -1.66      0.097 .
## Year2003     -0.10908      0.07669     -1.42      0.155
## Year2004     -0.14727      0.06876     -2.14      0.032 *
## Year2005     -0.07786      0.06984     -1.11      0.265
## Year2006     -0.12128      0.07235     -1.68      0.094 .
## Year2007     -0.12049      0.06879     -1.75      0.080 .
## Year2008     -0.13170      0.07559     -1.74      0.082 .
## Year2009     -0.04158      0.06810     -0.61      0.542
## Year2010     -0.09196      0.06784     -1.36      0.175
## Year2011     -0.12940      0.06718     -1.93      0.054 .
## Year2012     -0.13935      0.06834     -2.04      0.042 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.0386, Adjusted R-squared:  0.0242
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 137 weights are ~= 1. The remaining 1358 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0962 0.8670 0.9510 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.69e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.028 1      1.014
## LastAuthorFemale 1.053 1      1.026
## Year              1.077 16      1.002

```

Residuals from first and last author



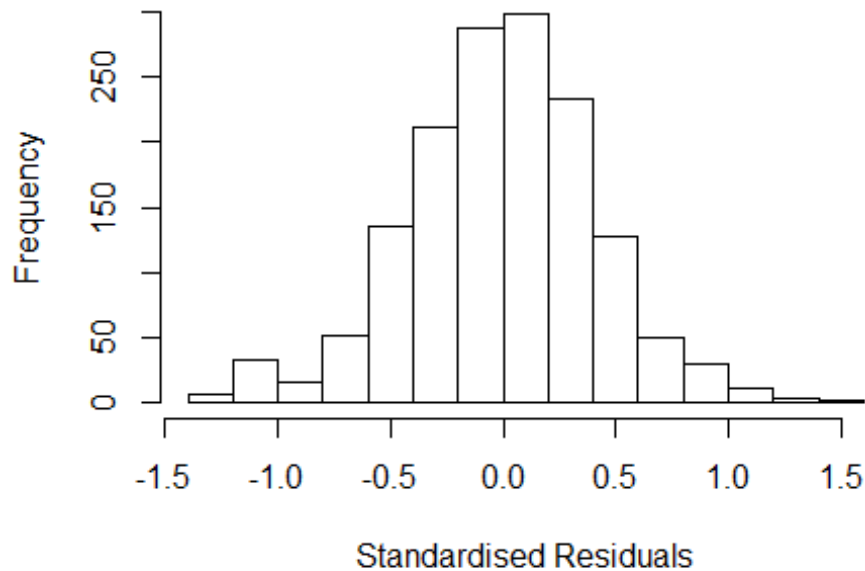
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30068 -0.26825  0.00335  0.26396  1.54490
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1921    0.0577   20.67  <2e-16 ***
## FirstAuthorFemale1 -0.0191    0.0238   -0.80    0.422
## LastAuthorFemale1  0.0103    0.0311    0.33    0.741
## Year1997          0.0855    0.0778    1.10    0.272
## Year1998         -0.0511    0.0908   -0.56    0.574
## Year1999          0.1086    0.0796    1.36    0.173
## Year2000          0.0685    0.0818    0.84    0.403
## Year2001         -0.1194    0.0874   -1.37    0.172
## Year2002         -0.1187    0.0748   -1.59    0.113
## Year2003         -0.1071    0.0765   -1.40    0.162
## Year2004         -0.1477    0.0695   -2.13    0.034 *
## Year2005         -0.0722    0.0695   -1.04    0.299
```

```

## Year2006          -0.1180      0.0724   -1.63    0.103
## Year2007          -0.1148      0.0688   -1.67    0.095 .
## Year2008          -0.1241      0.0758   -1.64    0.102
## Year2009          -0.0386      0.0681   -0.57    0.571
## Year2010          -0.0829      0.0677   -1.23    0.221
## Year2011          -0.1180      0.0666   -1.77    0.077 .
## Year2012          -0.1314      0.0682   -1.93    0.054 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.0346, Adjusted R-squared:  0.0228
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 140 weights are ~= 1. The remaining 1355 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0871 0.8670 0.9490 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.025 1      1.013
## 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
## -1.30206 -0.26833 0.00474 0.26315 1.54396
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1930 0.0576 20.71 <2e-16 ***
## FirstAuthorFemale1 -0.0194 0.0238 -0.81 0.416
## Year1997 0.0852 0.0778 1.09 0.274
## Year1998 -0.0509 0.0910 -0.56 0.576
## Year1999 0.1090 0.0796 1.37 0.171
## Year2000 0.0695 0.0817 0.85 0.395
## Year2001 -0.1186 0.0874 -1.36 0.175
## Year2002 -0.1188 0.0747 -1.59 0.112
## Year2003 -0.1066 0.0764 -1.39 0.163
## Year2004 -0.1468 0.0694 -2.11 0.035 *
## Year2005 -0.0722 0.0695 -1.04 0.299
## Year2006 -0.1177 0.0723 -1.63 0.104
```

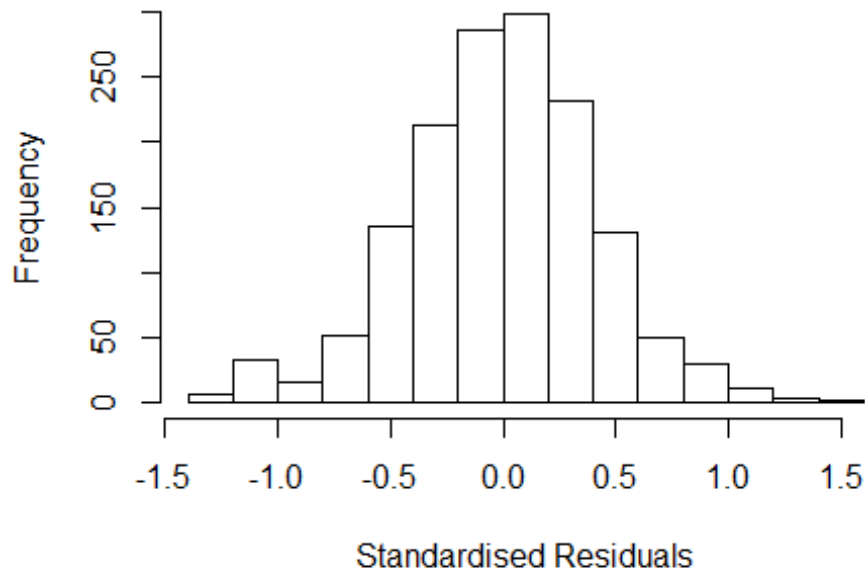


```

## Year2007          -0.1138      0.0687   -1.66    0.098 .
## Year2008          -0.1224      0.0754   -1.62    0.105
## Year2009          -0.0384      0.0681   -0.56    0.572
## Year2010          -0.0817      0.0675   -1.21    0.227
## Year2011          -0.1176      0.0666   -1.77    0.078 .
## Year2012          -0.1305      0.0681   -1.91    0.056 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.0345, Adjusted R-squared:  0.0234
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 142 weights are ~= 1. The remaining 1353 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0869 0.8660 0.9490 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 1.051 1          1.025
## Year            1.051 16          1.002

```

Residuals from last author



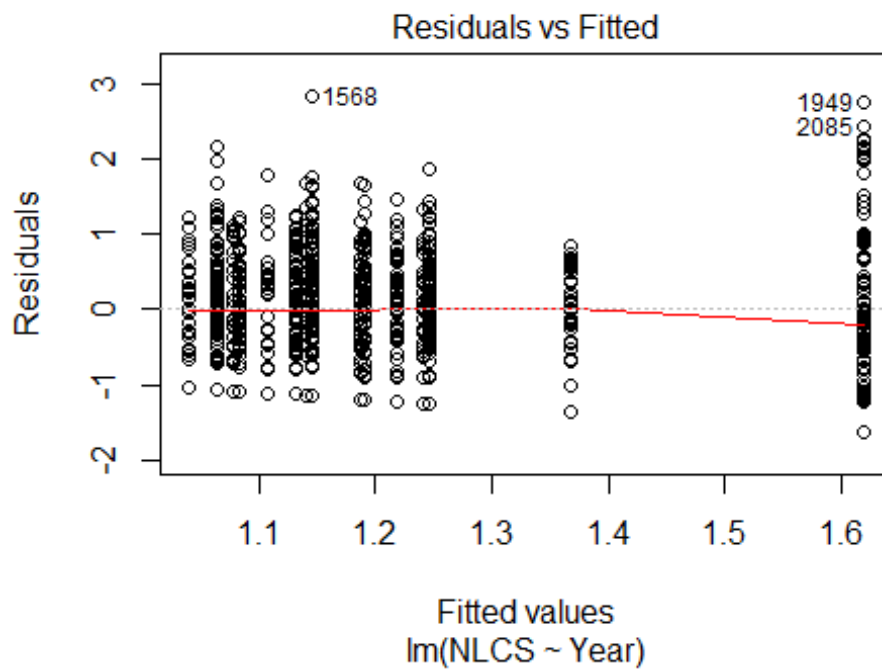
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29653 -0.27063  0.00711  0.26389  1.54761
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1894     0.0575  20.67  <2e-16 ***
## LastAuthorFemale1  0.0112     0.0310   0.36   0.719
## Year1997          0.0844     0.0775   1.09   0.277
## Year1998         -0.0534     0.0907  -0.59   0.556
## Year1999          0.1071     0.0795   1.35   0.178
## Year2000          0.0660     0.0816   0.81   0.419
## Year2001         -0.1221     0.0870  -1.40   0.161
## Year2002         -0.1214     0.0744  -1.63   0.103
## Year2003         -0.1103     0.0761  -1.45   0.147
## Year2004         -0.1522     0.0689  -2.21   0.027 *
## Year2005         -0.0751     0.0692  -1.09   0.278
## Year2006         -0.1210     0.0719  -1.68   0.093 .
```

```

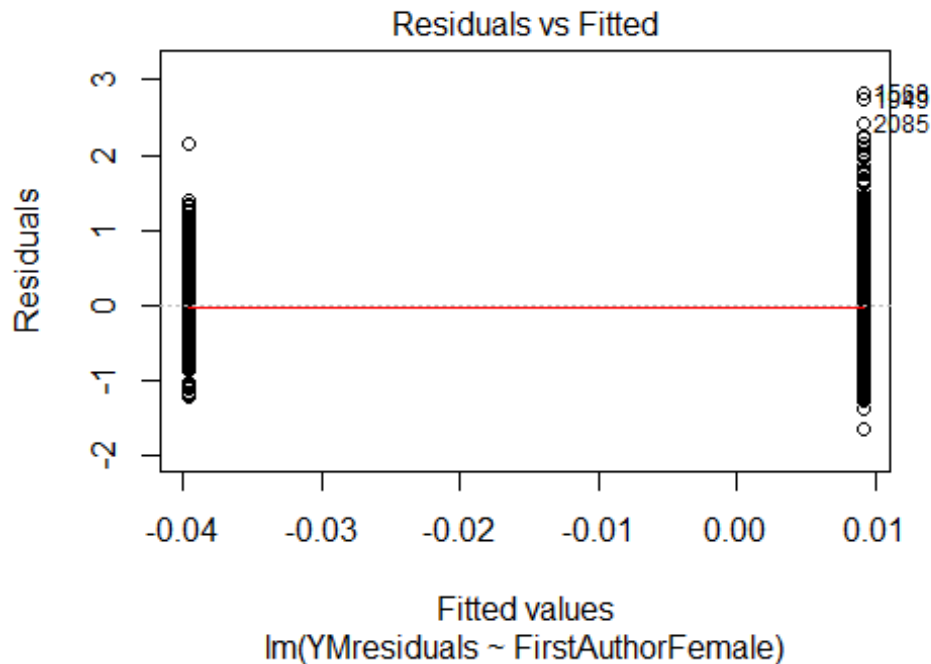
## Year2007          -0.1183      0.0683    -1.73      0.083 .
## Year2008          -0.1262      0.0754    -1.67      0.094 .
## Year2009          -0.0416      0.0676    -0.62      0.538
## Year2010          -0.0865      0.0672    -1.29      0.198
## Year2011          -0.1200      0.0663    -1.81      0.070 .
## Year2012          -0.1352      0.0677    -2.00      0.046 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.0342, Adjusted R-squared:  0.0231
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 144 weights are ~= 1. The remaining 1351 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0854 0.8660 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      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 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
##   74   58   65   52   75   80   62   54   44   58  110  107  109  144  176
## 2011 2012
##  153  179
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   31   31   30   44   52   35   42   29   30   66   79   73   99  129
## 2011 2012

```

```
## 109 107
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 26 25 28 26 38 45 31 37 25 25 55 64 62 83 108
## 2011 2012
## 94 92
## [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 = 3e-10
```

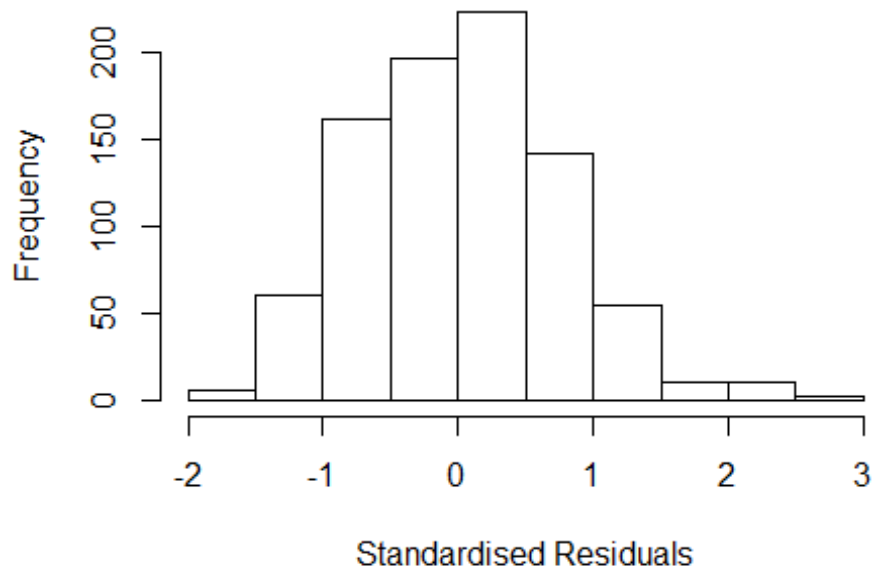


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.3, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.70023082362171"
## [1] "Male first author team size 2018 geometric mean: 2.56870027948565"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.84005938026132"
## [1] "Male last author team size 2018 geometric mean: 2.55380990115084"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 880, 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 1.154 1      1.074
## LastAuthorFemale  1.124 1      1.060
## UniqueAuthors    1.378 4      1.041
## Year              1.463 16     1.012
```

Residuals from first and last author and team size



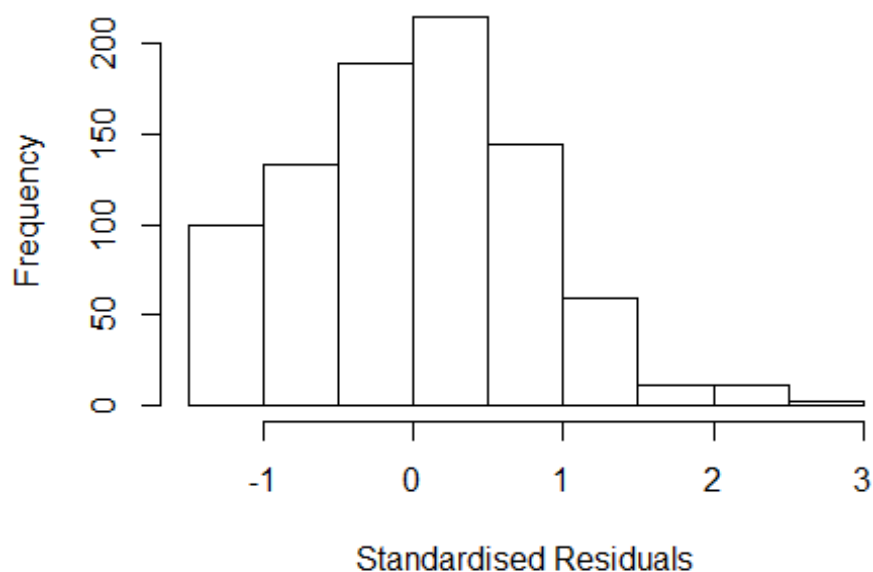
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1568 77956064159 3.971 2010      1700      1      2.991
## 2085 84857647437 4.039 2012      1700      1      2.697
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6805 -0.5149  0.0312  0.5046  2.9913
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06615    0.19424     5.49  5.4e-08 ***
## FirstAuthorFemale1 -0.07978    0.06565    -1.22   0.2246
## LastAuthorFemale1 -0.02116    0.07432    -0.28   0.7759
## UniqueAuthors2     0.09306    0.06566     1.42   0.1568
## UniqueAuthors3     0.43127    0.07773     5.55  3.9e-08 ***
## UniqueAuthors4     0.30762    0.09391     3.28   0.0011 **
## UniqueAuthors5     0.39112    0.15114     2.59   0.0098 **
## Year1997        -0.08227    0.23870    -0.34   0.7304
## Year1998        -0.16757    0.23622    -0.71   0.4783
```

```

## Year1999      0.14431      0.22573      0.64      0.5228
## Year2000     -0.14044      0.23664     -0.59      0.5530
## Year2001     -0.04082      0.21232     -0.19      0.8476
## Year2002     -0.08995      0.24291     -0.37      0.7113
## Year2003      0.00804      0.22605      0.04      0.9716
## Year2004      0.03518      0.23411      0.15      0.8806
## Year2005      0.23416      0.22913      1.02      0.3071
## Year2006     -0.15464      0.20700     -0.75      0.4552
## Year2007      0.02226      0.21192      0.11      0.9164
## Year2008      0.01249      0.20965      0.06      0.9525
## Year2009     -0.08917      0.20267     -0.44      0.6601
## Year2010     -0.17951      0.20636     -0.87      0.3846
## Year2011     -0.20772      0.20783     -1.00      0.3178
## Year2012      0.18310      0.23075      0.79      0.4277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.75
## Multiple R-squared:  0.0786, Adjusted R-squared:  0.0545
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 795 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.076  0.871  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           1.16e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.123 1 1.060
## LastAuthorFemale 1.116 1 1.057
## Year 1.091 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
## 1568 77956064159 3.971 2010      1700      1      2.886
## 2085 84857647437 4.039 2012      1700      1      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
## -1.4679 -0.5587  0.0356  0.5182  2.8859
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2214     0.2010   6.08 1.9e-09 ***
## FirstAuthorFemale1 -0.0486     0.0675  -0.72   0.47
## LastAuthorFemale1 -0.0180     0.0776  -0.23   0.82
## Year1997          -0.1125     0.2490  -0.45   0.65
## Year1998          -0.2277     0.2488  -0.92   0.36
## Year1999           0.0661     0.2367   0.28   0.78
## Year2000          -0.1218     0.2436  -0.50   0.62
## Year2001          -0.0384     0.2199  -0.17   0.86
## Year2002          -0.1514     0.2507  -0.60   0.55
## Year2003           0.0277     0.2285   0.12   0.90
## Year2004          -0.0104     0.2404  -0.04   0.97
```

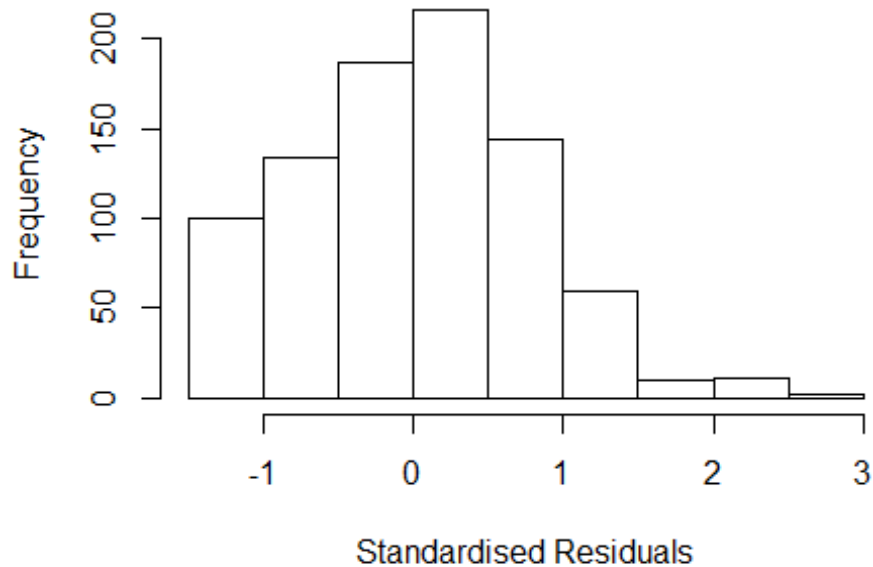


```

## Year2005          0.2441      0.2362      1.03      0.30
## Year2006          -0.1417      0.2181     -0.65      0.52
## Year2007           0.0226      0.2226      0.10      0.92
## Year2008           0.0194      0.2200      0.09      0.93
## Year2009          -0.0758      0.2122     -0.36      0.72
## Year2010          -0.1364      0.2177     -0.63      0.53
## Year2011          -0.1795      0.2178     -0.82      0.41
## Year2012           0.2464      0.2414      1.02      0.31
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.788
## Multiple R-squared:  0.0302, Adjusted R-squared:  0.00955
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 802 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.151  0.866  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      1.16e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1      1.021
## Year              1.042 16      1.001

```

Residuals from first author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1568 77956064159 3.971 2010      1700      1      2.886
## 2085 84857647437 4.039 2012      1700      1      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
## -1.466 -0.557  0.037  0.517  2.888
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2191     0.2004   6.08 1.8e-09 ***
## FirstAuthorFemale1 -0.0531     0.0651  -0.81   0.42
## Year1997          -0.1122     0.2488  -0.45   0.65
## Year1998          -0.2267     0.2484  -0.91   0.36
## Year1999           0.0670     0.2363   0.28   0.78
## Year2000          -0.1220     0.2432  -0.50   0.62
## Year2001          -0.0388     0.2195  -0.18   0.86
## Year2002          -0.1500     0.2503  -0.60   0.55
## Year2003           0.0296     0.2279   0.13   0.90
## Year2004          -0.0096     0.2399  -0.04   0.97
## Year2005           0.2459     0.2355   1.04   0.30
```

```

## Year2006          -0.1415      0.2178   -0.65      0.52
## Year2007           0.0240      0.2222    0.11      0.91
## Year2008           0.0203      0.2196    0.09      0.93
## Year2009          -0.0757      0.2118   -0.36      0.72
## Year2010          -0.1366      0.2173   -0.63      0.53
## Year2011          -0.1797      0.2174   -0.83      0.41
## Year2012           0.2470      0.2409    1.03      0.31
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.787
## Multiple R-squared:  0.0302, Adjusted R-squared:  0.0107
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 803 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.149  0.867  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      1.16e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.037 1      1.018
## Year              1.037 16      1.001
##
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1568 77956064159 3.971 2010      1700      1      2.886
## 2085 84857647437 4.039 2012      1700      1      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
## -1.4601 -0.5591  0.0345  0.5212  2.8905
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2181      0.2033   5.99 3.1e-09 ***
## LastAuthorFemale1 -0.0331      0.0752  -0.44   0.66
## Year1997          -0.1114      0.2510  -0.44   0.66
## Year1998          -0.2307      0.2506  -0.92   0.36
## Year1999           0.0634      0.2394   0.26   0.79
## Year2000          -0.1183      0.2456  -0.48   0.63
## Year2001          -0.0436      0.2222  -0.20   0.84
## Year2002          -0.1536      0.2531  -0.61   0.54
## Year2003           0.0208      0.2305   0.09   0.93
## Year2004          -0.0111      0.2419  -0.05   0.96
## Year2005           0.2365      0.2374   1.00   0.32
## Year2006          -0.1468      0.2201  -0.67   0.50
## Year2007           0.0158      0.2243   0.07   0.94
## Year2008           0.0142      0.2219   0.06   0.95
## Year2009          -0.0822      0.2142  -0.38   0.70
## Year2010          -0.1376      0.2197  -0.63   0.53
## Year2011          -0.1856      0.2202  -0.84   0.40
## Year2012           0.2420      0.2435   0.99   0.32
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.787
## Multiple R-squared:  0.0295, Adjusted R-squared:  0.00996
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 804 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      0.148  0.868  0.950  0.914  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.16e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000          0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)

```

```

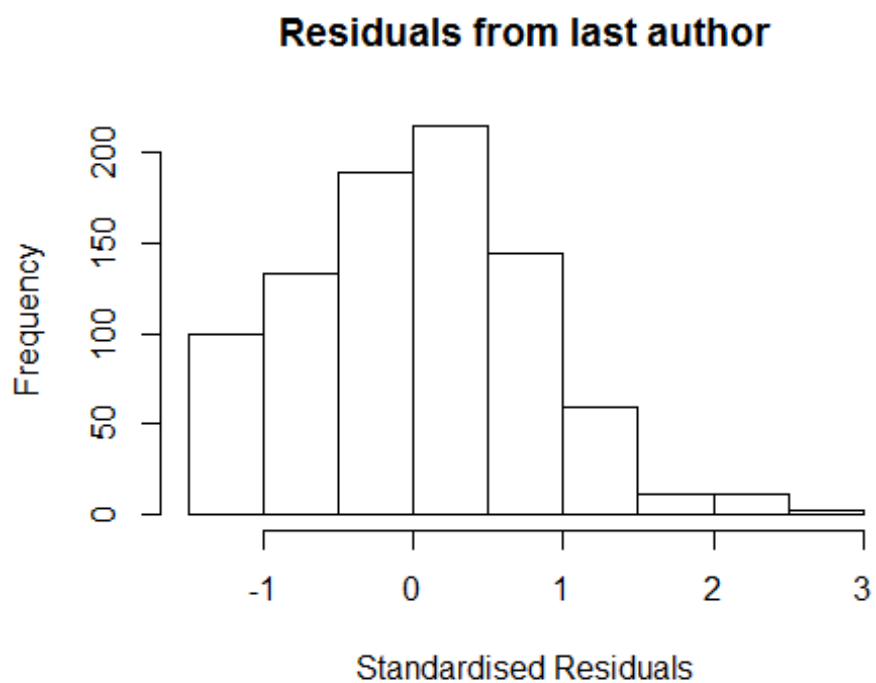
## [1] "Sample size for the above analysis: 864"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1701"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1999 2000 2001 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    1    1    2    4    8    8   23   23   14   18    9   21   16
##
## 1999 2000 2001 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    0    0    2    1    5    7   16   11    9   14    7   16   10
##
## 1999 2000 2001 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    0    0    2    1    5    4   15   10    8   10    5   14    8
## [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: 1.69838132956495"

## Warning in wilcox.test.default(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 = 0.9
## 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: 1.91947121957741"

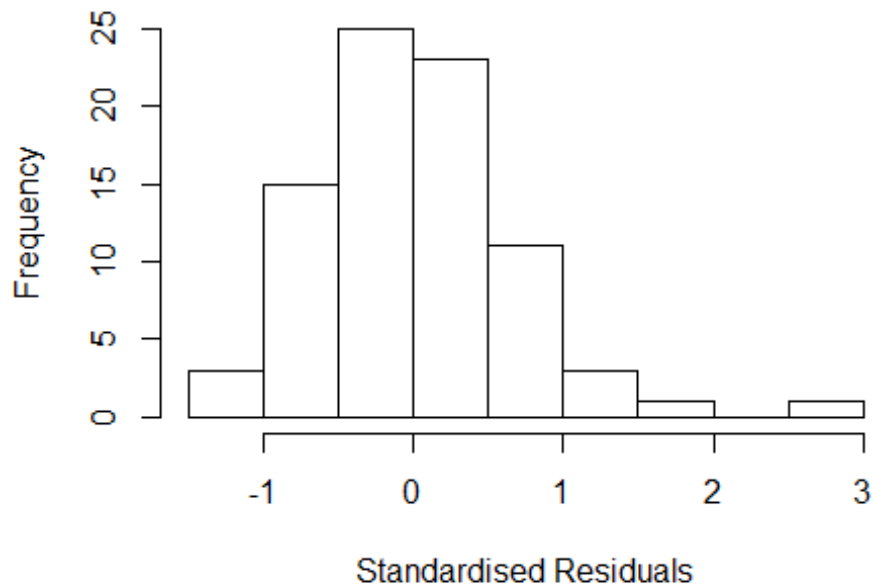
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1, p-value = 0.4
## 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  2.720  1      1.649
## LastAuthorFemale   2.166  1      1.472
## UniqueAuthors    297.834  4      2.038
## Year              455.844 10      1.358
```

Residuals from first and last author and team size



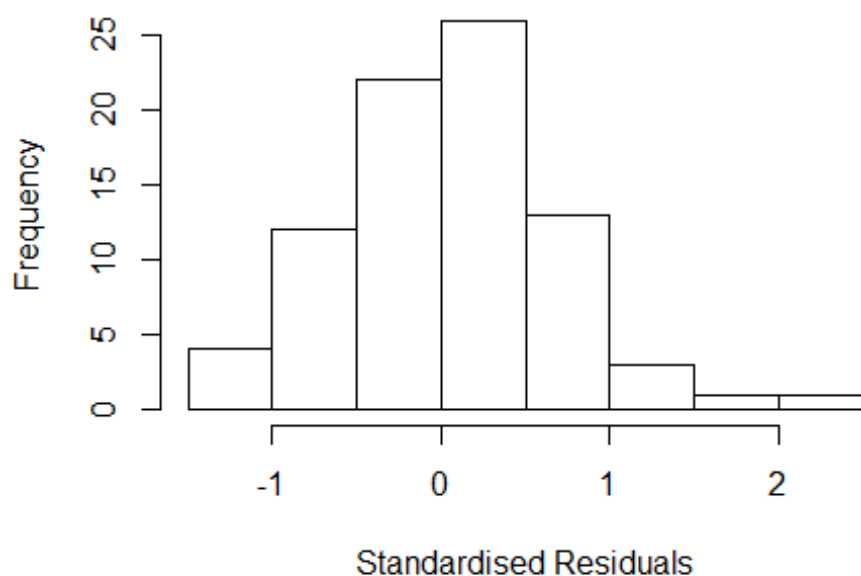
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 171 79960142338 3.392 2011      1701      1      2.524
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1798 -0.3803 -0.0415  0.3225  2.5243
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6625    0.2430     2.73  0.0082 **
## FirstAuthorFemale1  0.0863    0.1603     0.54  0.5923
## LastAuthorFemale1 -0.4483    0.1846    -2.43  0.0179 *
## UniqueAuthors2    -0.1206    0.2114    -0.57  0.5703
## UniqueAuthors3     0.1516    0.2287     0.66  0.5097
## UniqueAuthors4     0.0574    0.3626     0.16  0.8747
## UniqueAuthors5     0.2621    0.2603     1.01  0.3176
## Year2003          0.7671    0.3248     2.36  0.0212 *
## Year2004          0.4865    0.3718     1.31  0.1953
## Year2005          0.5585    0.4457     1.25  0.2147
```

```

## Year2006          0.4307      0.3221      1.34      0.1858
## Year2007          0.4971      0.3759      1.32      0.1907
## Year2008          0.9962      0.3172      3.14      0.0025 **
## Year2009          0.6379      0.2928      2.18      0.0330 *
## Year2010          0.5712      0.3572      1.60      0.1146
## Year2011          0.2052      0.4104      0.50      0.6188
## Year2012          0.1062      0.2905      0.37      0.7158
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.672
## Multiple R-squared:  0.223, Adjusted R-squared:  0.0321
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 74 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.127  0.893   0.962   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      1.22e-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 2.275 1      1.508
## LastAuthorFemale  1.789 1      1.338
## Year              2.575 10      1.048

```


Residuals from first and last author



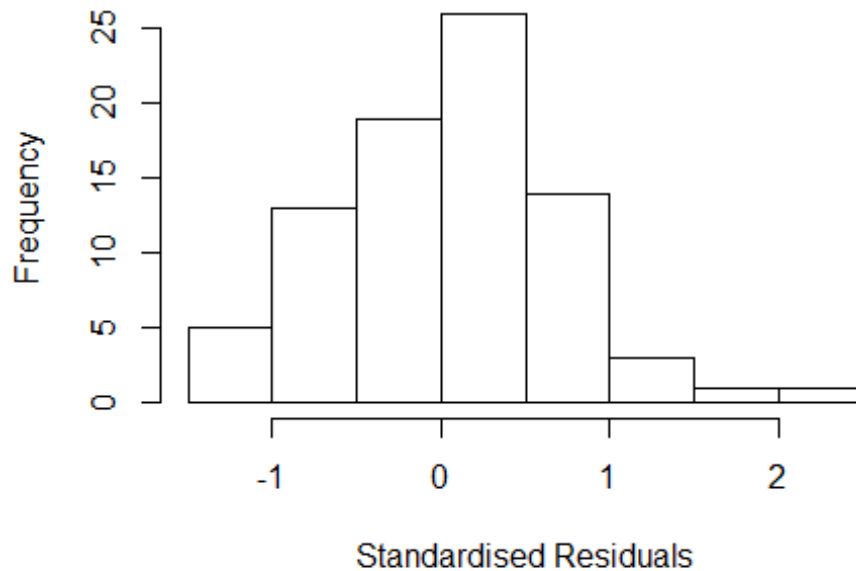
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3330 -0.3611 0.0137 0.4234 2.4346
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6991 0.2143 3.26 0.0017 **
## FirstAuthorFemale1 0.0550 0.1529 0.36 0.7202
## LastAuthorFemale1 -0.4902 0.1612 -3.04 0.0033 **
## Year2003 0.6099 0.2143 2.85 0.0058 **
## Year2004 0.4156 0.3360 1.24 0.2203
## Year2005 0.5213 0.4074 1.28 0.2049
## Year2006 0.3943 0.2726 1.45 0.1525
## Year2007 0.4897 0.3127 1.57 0.1219
## Year2008 0.9027 0.2698 3.35 0.0013 **
## Year2009 0.6339 0.2574 2.46 0.0163 *
## Year2010 0.5293 0.3160 1.67 0.0985 .
## Year2011 0.2583 0.2684 0.96 0.3393
```

```

## Year2012          0.0714      0.2508      0.28      0.7768
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.636
## Multiple R-squared:  0.207, Adjusted R-squared:  0.0694
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 73 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.110  0.889  0.951  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          1.22e-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 1.958 1          1.399
## Year              1.958 10          1.034

```

Residuals from first author



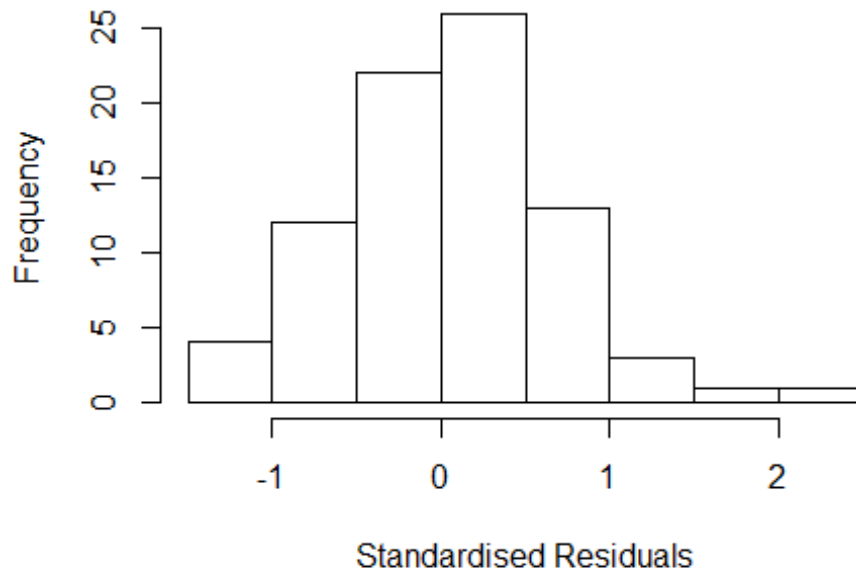
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2648 -0.3754 0.0852 0.4252 2.4728
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.569 0.307 1.85 0.0685 .
## FirstAuthorFemale1 -0.174 0.165 -1.06 0.2930
## Year2003 0.740 0.307 2.41 0.0186 *
## Year2004 0.555 0.400 1.39 0.1698
## Year2005 0.773 0.451 1.72 0.0905 .
## Year2006 0.495 0.367 1.35 0.1816
## Year2007 0.496 0.374 1.33 0.1892
## Year2008 1.028 0.360 2.86 0.0056 **
## Year2009 0.696 0.340 2.05 0.0443 *
## Year2010 0.656 0.384 1.71 0.0924 .
## Year2011 0.351 0.367 0.96 0.3422
## Year2012 0.107 0.336 0.32 0.7519
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.701
## Multiple R-squared:  0.141, Adjusted R-squared:  0.00572
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 76 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.188  0.894  0.963  0.920  0.987  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.22e-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 1.724 1           1.313
## Year           1.724 10           1.028

```

Residuals from last author



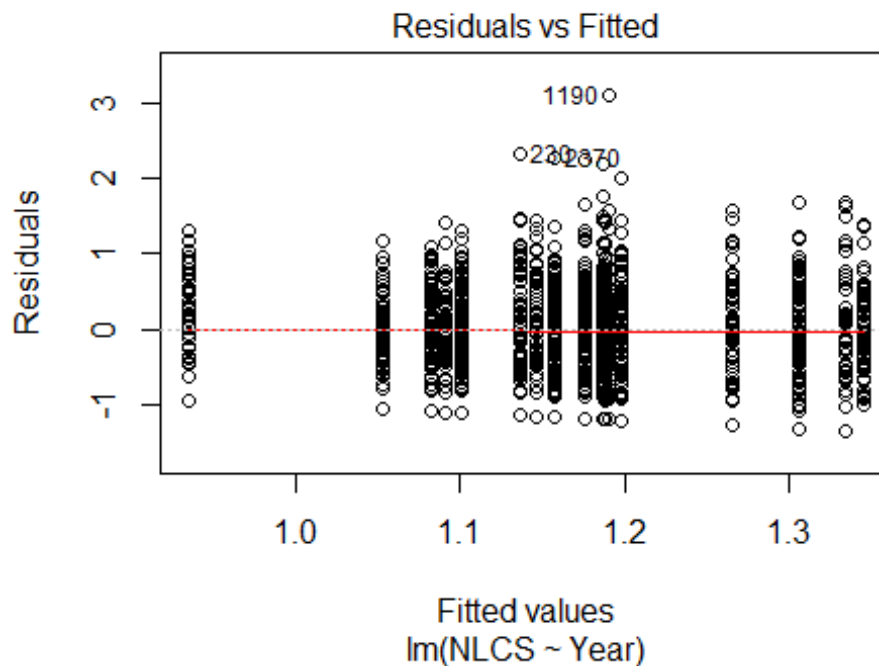
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34794 -0.35701 0.00262 0.41208 2.42699
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7125 0.1986 3.59 0.00062 ***
## LastAuthorFemale1 -0.4620 0.1575 -2.93 0.00452 **
## Year2003 0.5965 0.1986 3.00 0.00370 **
## Year2004 0.4019 0.3255 1.23 0.22106
## Year2005 0.5377 0.3972 1.35 0.18018
## Year2006 0.3855 0.2668 1.45 0.15286
## Year2007 0.4889 0.3079 1.59 0.11685
## Year2008 0.9013 0.2636 3.42 0.00105 **
## Year2009 0.6354 0.2480 2.56 0.01255 *
## Year2010 0.5161 0.3059 1.69 0.09606 .
## Year2011 0.2525 0.2621 0.96 0.33873
## Year2012 0.0584 0.2381 0.25 0.80705
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.633
## Multiple R-squared:  0.208, Adjusted R-squared:  0.083
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 73 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110  0.887  0.946  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           1.22e-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: 82"
## [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
## 110 104 100 100 138 127 137 92 106 123 151 162 199 175 180
## 2011 2012
## 169 187
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 62 64 49 58 73 79 82 59 69 74 93 95 120 112 117
## 2011 2012
## 118 121
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 53 58 44 51 59 67 66 50 60 64 76 75 95 97 98
## 2011 2012
## 91 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 = 32, df = 16, p-value = 0.01
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.2, df = 1, p-value = 0.3

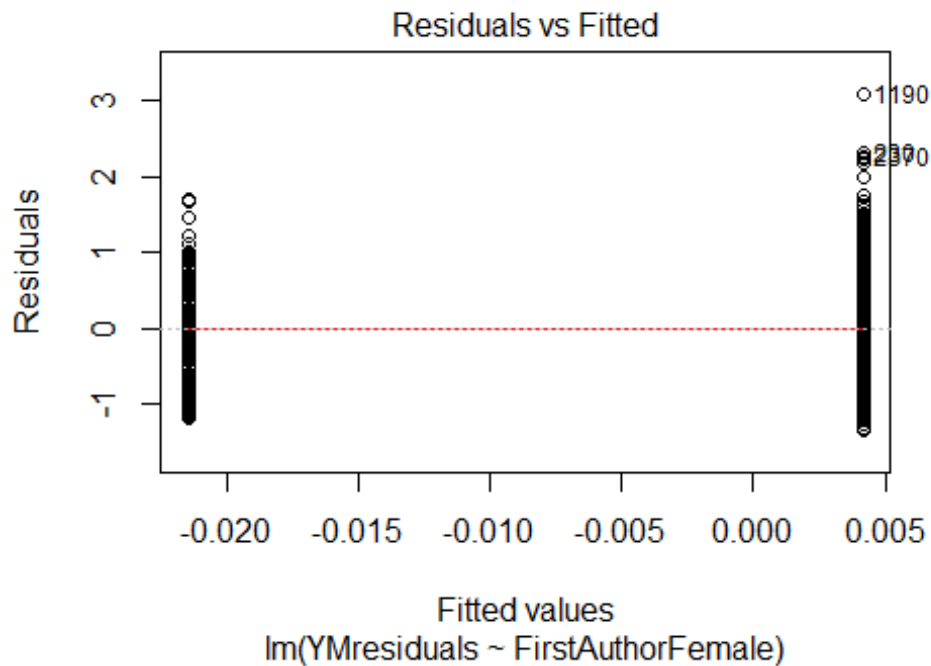
## [1] "Female first author team size 2018 geometric mean: 2"
## [1] "Male first author team size 2018 geometric mean: 2.19921403011256"

## Warning in wilcox.test.default(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
##
```

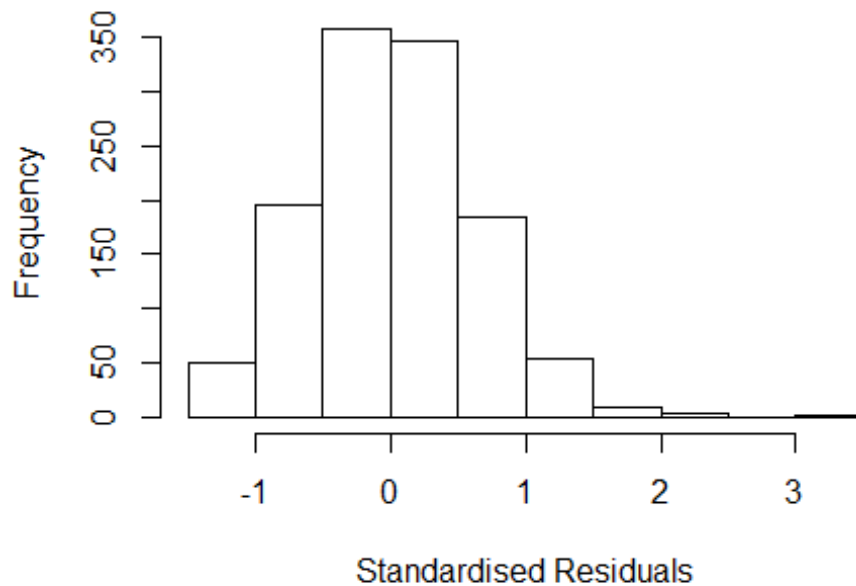
```
## [1] "Female last author team size 2018 geometric mean: 3.16227766016838"
## [1] "Male last author team size 2018 geometric mean: 2.03753853024301"

## Warning in wilcox.test.default(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 = 0.3
## 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 1.197 1      1.094
## LastAuthorFemale  1.207 1      1.098
## UniqueAuthors     1.424 4      1.045
## Year              1.514 16     1.013
```


Residuals from first and last author and team size



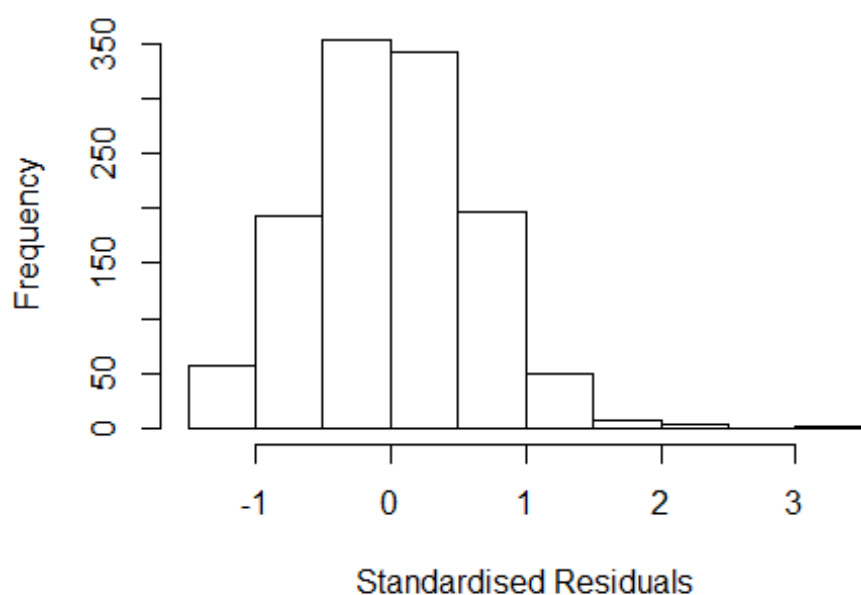
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1190 3042535216 4.284 2004      1702      3      3.189
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.36243 -0.40339 -0.00153  0.42045  3.18947
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1811    0.0944   12.51 < 2e-16 ***
## FirstAuthorFemale1  0.0316    0.0520    0.61  0.54333
## LastAuthorFemale1 -0.1042    0.0528   -1.97  0.04869 *
## UniqueAuthors2     0.1178    0.0443    2.66  0.00796 **
## UniqueAuthors3     0.2010    0.0548    3.67  0.00025 ***
## UniqueAuthors4     0.2315    0.0819    2.82  0.00481 **
## UniqueAuthors5     0.3057    0.0927    3.30  0.00100 ***
## Year1997          -0.1240    0.1483   -0.84  0.40335
## Year1998           0.0482    0.1515    0.32  0.75019
## Year1999          -0.3268    0.1323   -2.47  0.01363 *
```

```

## Year2000          -0.2318      0.1145      -2.02      0.04311 *
## Year2001           0.0793      0.1132       0.70      0.48399
## Year2002          -0.0629      0.1265      -0.50      0.61921
## Year2003          -0.1620      0.1193      -1.36      0.17489
## Year2004          -0.0866      0.1147      -0.75      0.45068
## Year2005          -0.1637      0.1202      -1.36      0.17348
## Year2006          -0.0513      0.1213      -0.42      0.67250
## Year2007          -0.2959      0.1118      -2.65      0.00822 **
## Year2008          -0.0521      0.1125      -0.46      0.64302
## Year2009          -0.1513      0.1130      -1.34      0.18064
## Year2010          -0.1200      0.1081      -1.11      0.26725
## Year2011          -0.1831      0.1088      -1.68      0.09263 .
## Year2012          -0.2003      0.1121      -1.79      0.07432 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.0461, Adjusted R-squared:  0.0283
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 460 is an outlier with |weight| = 0 ( < 8.3e-05);
## 99 weights are ~= 1. The remaining 1103 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0893 0.8770 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          8.31e-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 1.184 1          1.088
## LastAuthorFemale 1.176 1          1.085
## Year          1.077 16          1.002

```

Residuals from first and last author

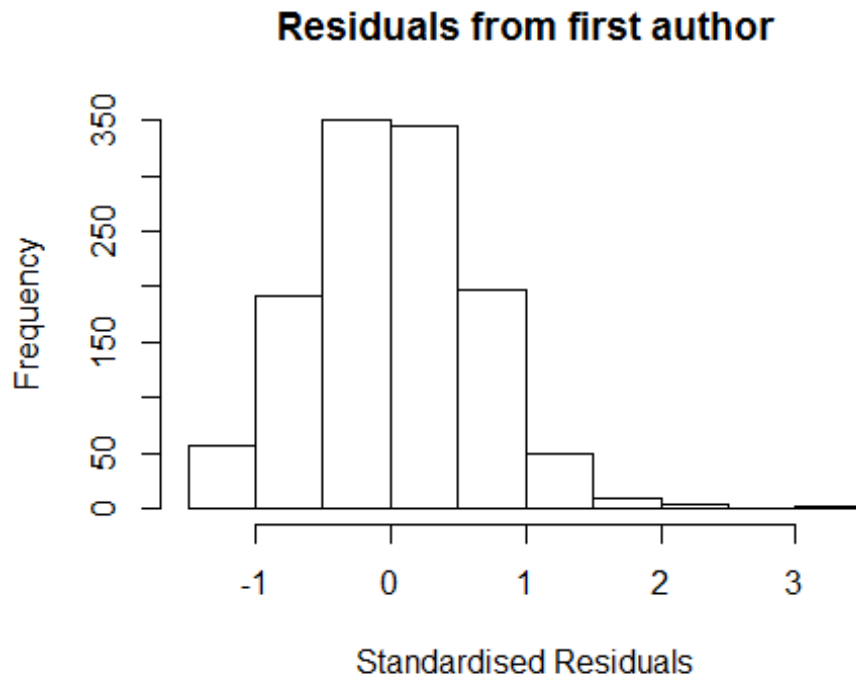


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1190 3042535216 4.284 2004      1702      3      3.095
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q   Median      3Q      Max
## -1.317210 -0.417759  0.000795  0.426795  3.095255
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2744     0.0959   13.29  <2e-16 ***
## FirstAuthorFemale1  0.0297     0.0519    0.57   0.567
## LastAuthorFemale1 -0.1090     0.0531   -2.05   0.040 *
## Year1997          -0.1315     0.1506   -0.87   0.383
## Year1998           0.0428     0.1521    0.28   0.778
## Year1999          -0.3243     0.1359   -2.39   0.017 *
## Year2000          -0.2324     0.1175   -1.98   0.048 *
## Year2001           0.0828     0.1169    0.71   0.479
## Year2002          -0.0339     0.1284   -0.26   0.792
## Year2003          -0.1604     0.1205   -1.33   0.184
## Year2004          -0.0856     0.1175   -0.73   0.466
## Year2005          -0.1197     0.1215   -0.99   0.325
```

```

## Year2006          -0.0253      0.1248   -0.20    0.840
## Year2007          -0.2545      0.1147   -2.22    0.027 *
## Year2008          -0.0405      0.1155   -0.35    0.726
## Year2009          -0.1216      0.1165   -1.04    0.297
## Year2010          -0.0977      0.1130   -0.86    0.387
## Year2011          -0.1514      0.1114   -1.36    0.174
## Year2012          -0.1496      0.1141   -1.31    0.190
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.622
## Multiple R-squared:  0.0268, Adjusted R-squared:  0.012
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 460 is an outlier with |weight| = 0 ( < 8.3e-05);
## 97 weights are ~= 1. The remaining 1105 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.132  0.875   0.951   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          8.31e-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 1.043 1          1.021
## Year              1.043 16          1.001

```

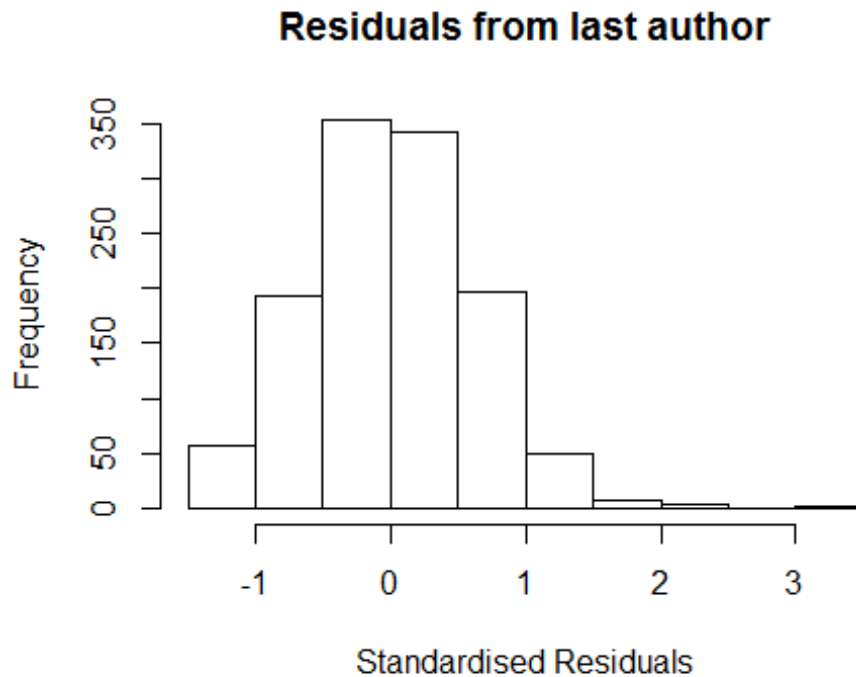


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1190 3042535216 4.284 2004      1702      3      3.095
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30797 -0.41417  0.00289  0.42182  3.10930
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2662     0.0952   13.30  <2e-16 ***
## FirstAuthorFemale1 -0.0172     0.0493   -0.35    0.728
## Year1997         -0.1317     0.1503   -0.88    0.381
## Year1998          0.0417     0.1523    0.27    0.784
## Year1999         -0.3201     0.1364   -2.35    0.019 *
## Year2000         -0.2257     0.1171   -1.93    0.054 .
## Year2001          0.0889     0.1161    0.77    0.444
## Year2002         -0.0309     0.1278   -0.24    0.809
## Year2003         -0.1655     0.1200   -1.38    0.168
## Year2004         -0.0915     0.1164   -0.79    0.432
## Year2005         -0.1205     0.1211   -0.99    0.320
## Year2006         -0.0268     0.1245   -0.22    0.830
```

```

## Year2007          -0.2544      0.1147    -2.22      0.027 *
## Year2008          -0.0410      0.1155    -0.36      0.723
## Year2009          -0.1264      0.1160    -1.09      0.276
## Year2010          -0.1026      0.1128    -0.91      0.363
## Year2011          -0.1515      0.1112    -1.36      0.173
## Year2012          -0.1475      0.1137    -1.30      0.195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.623
## Multiple R-squared:  0.0237, Adjusted R-squared:  0.00972
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 460 is an outlier with |weight| = 0 ( < 8.3e-05);
## 99 weights are ~= 1. The remaining 1103 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.129  0.872  0.950  0.911  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.31e-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 1.04 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
## 1190 3042535216 4.284 2004      1702      3      3.095
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3194 -0.4178  0.0039  0.4257  3.0922
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2763     0.0956   13.34  <2e-16 ***
## LastAuthorFemale1 -0.0953     0.0503   -1.90    0.058 .
## Year1997         -0.1337     0.1503   -0.89    0.374
## Year1998          0.0431     0.1517    0.28    0.777
## Year1999         -0.3214     0.1357   -2.37    0.018 *
## Year2000         -0.2313     0.1175   -1.97    0.049 *
## Year2001          0.0841     0.1165    0.72    0.470
## Year2002         -0.0324     0.1284   -0.25    0.801
## Year2003         -0.1612     0.1202   -1.34    0.180
## Year2004         -0.0846     0.1175   -0.72    0.472
## Year2005         -0.1172     0.1210   -0.97    0.333
## Year2006         -0.0236     0.1248   -0.19    0.850
```

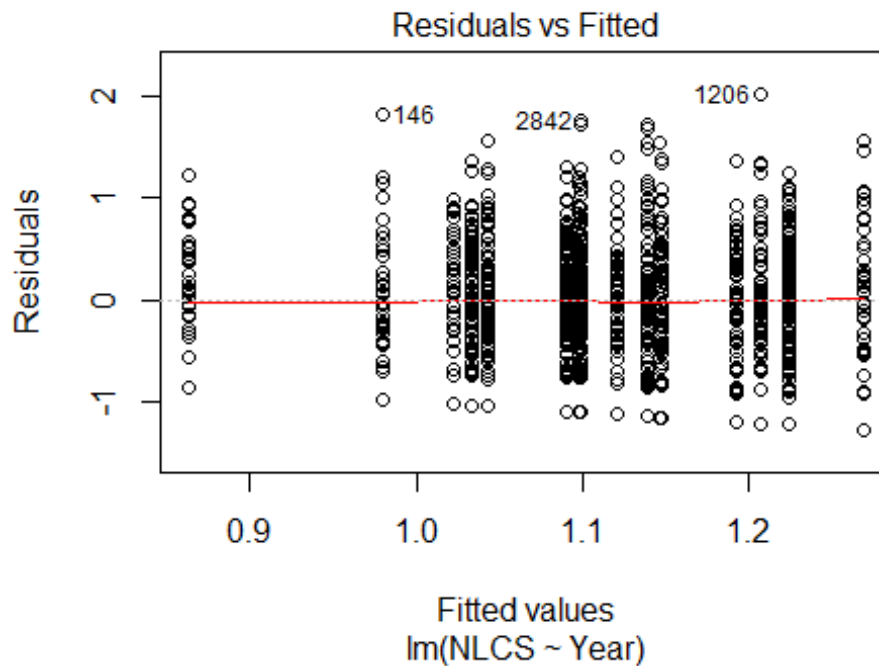
```

## Year2007          -0.2540      0.1147   -2.21    0.027 *
## Year2008          -0.0398      0.1155   -0.34    0.730
## Year2009          -0.1220      0.1163   -1.05    0.295
## Year2010          -0.0967      0.1128   -0.86    0.391
## Year2011          -0.1501      0.1113   -1.35    0.178
## Year2012          -0.1488      0.1141   -1.30    0.193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.622
## Multiple R-squared:  0.0266, Adjusted R-squared:  0.0126
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 460 is an outlier with |weight| = 0 ( < 8.3e-05);
## 102 weights are ~= 1. The remaining 1100 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.132  0.874  0.950  0.911  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.31e-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: 1203"
## [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
##   99   80   68   71   76  105   91   96  102  139  157  141  160  184  153
## 2011 2012
##  161  162
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   42   38   35   37   32   43   62   63   58   86   95   80   96  117   99

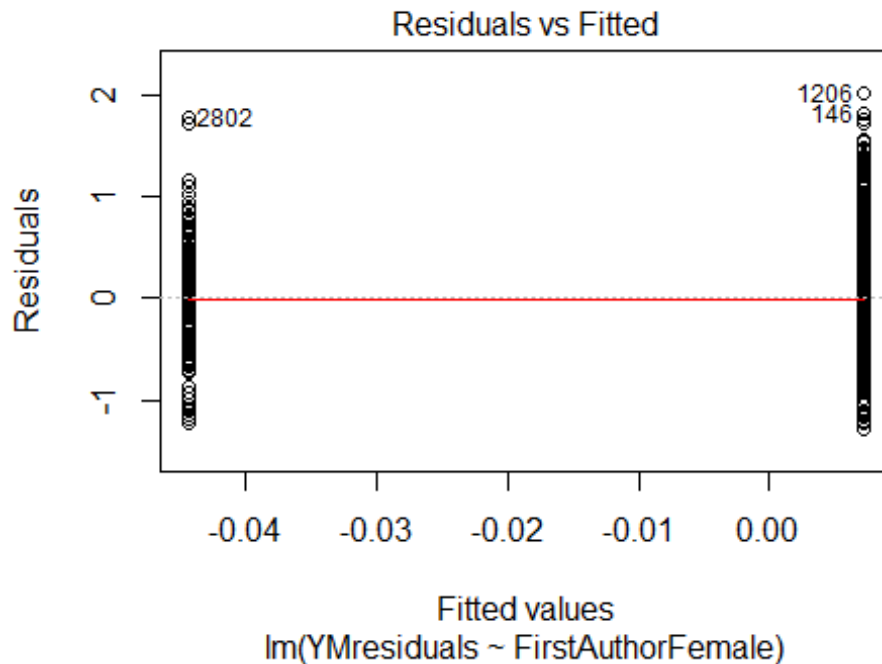
```



```
## 2011 2012
## 109 108
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 35 30 32 32 25 38 54 56 50 68 75 68 84 100 89
## 2011 2012
## 79 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 = 17, df = 16, p-value = 0.4
```

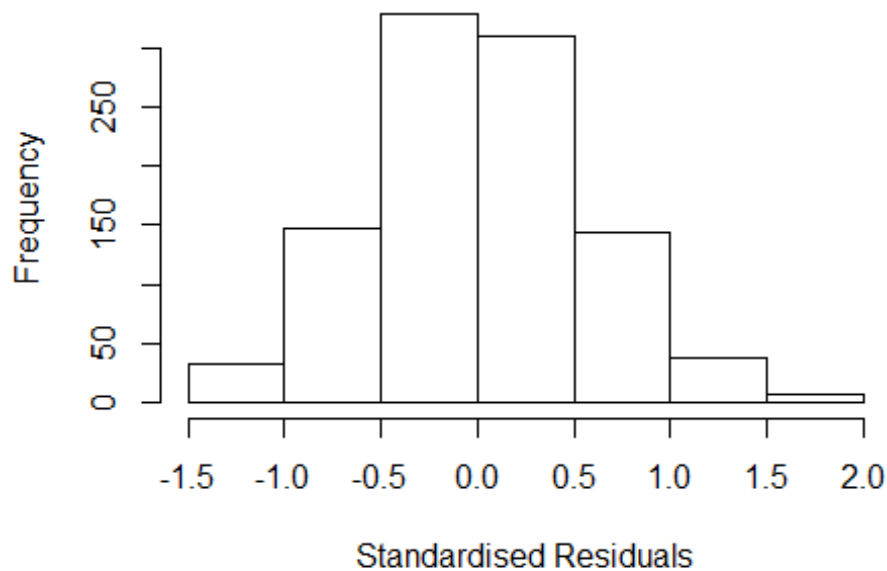


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.50823470647222"
## [1] "Male first author team size 2018 geometric mean: 2.15883095447275"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 390, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.5874010519682"
## [1] "Male last author team size 2018 geometric mean: 2.30480809821264"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, 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 1.280 1          1.131
## LastAuthorFemale  1.252 1          1.119
## UniqueAuthors     1.421 4          1.045
## Year               1.504 16         1.013
```

Residuals from first and last author and team size



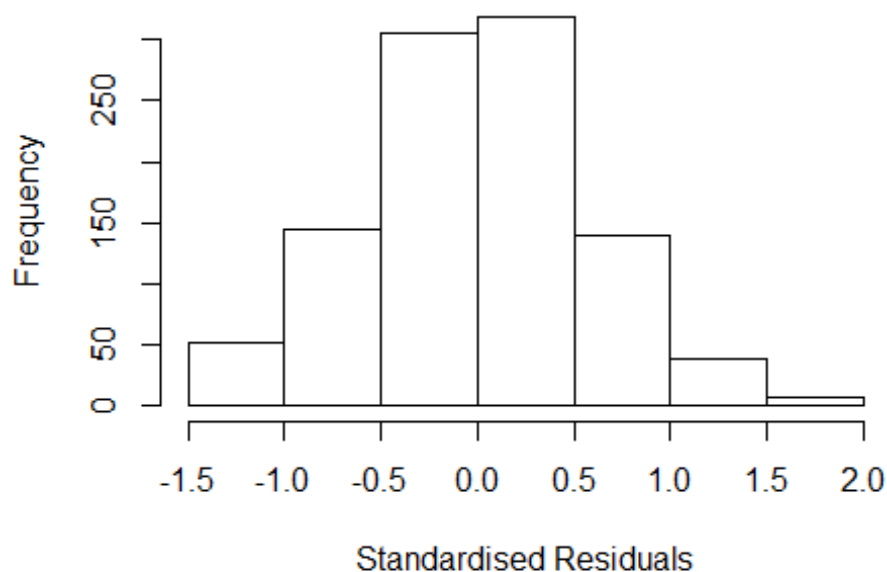
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4743 -0.3688 -0.0138 0.3715 1.9556
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2839 0.1346 9.54 < 2e-16 ***
## FirstAuthorFemale1 -0.0350 0.0477 -0.73 0.4627
## LastAuthorFemale1 -0.0901 0.0563 -1.60 0.1100
## UniqueAuthors2 0.1904 0.0458 4.16 3.5e-05 ***
## UniqueAuthors3 0.2584 0.0525 4.92 1.0e-06 ***
## UniqueAuthors4 0.3159 0.0778 4.06 5.2e-05 ***
## UniqueAuthors5 0.4581 0.0907 5.05 5.3e-07 ***
## Year1997 -0.4364 0.1797 -2.43 0.0153 *
## Year1998 -0.1303 0.1610 -0.81 0.4186
## Year1999 -0.5233 0.1807 -2.90 0.0039 **
```

```

## Year2000      -0.2598      0.1647      -1.58      0.1149
## Year2001      -0.3189      0.1651      -1.93      0.0537 .
## Year2002      -0.2928      0.1529      -1.91      0.0559 .
## Year2003      -0.2077      0.1538      -1.35      0.1771
## Year2004      -0.3037      0.1498      -2.03      0.0429 *
## Year2005      -0.2749      0.1478      -1.86      0.0632 .
## Year2006      -0.2154      0.1502      -1.43      0.1520
## Year2007      -0.3209      0.1506      -2.13      0.0334 *
## Year2008      -0.4094      0.1450      -2.82      0.0048 **
## Year2009      -0.3759      0.1455      -2.58      0.0099 **
## Year2010      -0.3334      0.1497      -2.23      0.0262 *
## Year2011      -0.3715      0.1471      -2.53      0.0117 *
## Year2012      -0.3950      0.1455      -2.72      0.0067 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.548
## Multiple R-squared:  0.0772, Adjusted R-squared:  0.0566
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 918 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.175  0.857  0.949  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          9.94e-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 1.207 1          1.099
## LastAuthorFemale  1.196 1          1.094
## Year              1.130 16          1.004

```

Residuals from first and last author



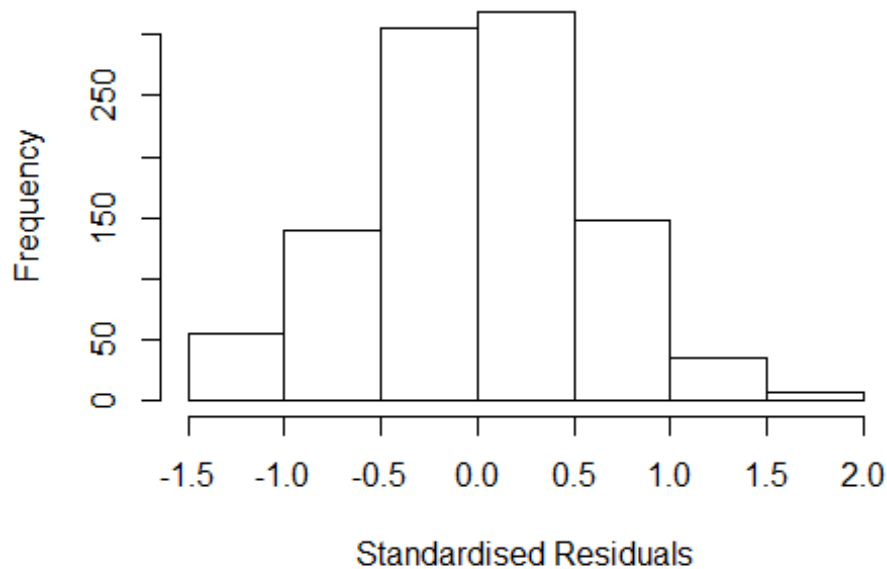
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.41725 -0.35514 0.00176 0.38278 1.84993
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4172 0.1383 10.24 <2e-16 ***
## FirstAuthorFemale1 -0.0158 0.0486 -0.32 0.7461
## LastAuthorFemale1 -0.0833 0.0588 -1.42 0.1573
## Year1997 -0.4642 0.1805 -2.57 0.0103 *
## Year1998 -0.1628 0.1669 -0.98 0.3296
## Year1999 -0.5694 0.1894 -3.01 0.0027 **
## Year2000 -0.2477 0.1733 -1.43 0.1533
## Year2001 -0.3198 0.1724 -1.86 0.0638 .
## Year2002 -0.2972 0.1577 -1.88 0.0598 .
## Year2003 -0.2400 0.1607 -1.49 0.1355
## Year2004 -0.3025 0.1572 -1.92 0.0547 .
## Year2005 -0.2469 0.1506 -1.64 0.1014
```

```

## Year2006          -0.1602      0.1573   -1.02   0.3089
## Year2007          -0.2726      0.1567   -1.74   0.0823 .
## Year2008          -0.4034      0.1516   -2.66   0.0079 **
## Year2009          -0.3361      0.1511   -2.22   0.0264 *
## Year2010          -0.3011      0.1562   -1.93   0.0542 .
## Year2011          -0.3128      0.1523   -2.05   0.0403 *
## Year2012          -0.3504      0.1529   -2.29   0.0221 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0316, Adjusted R-squared:  0.0139
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 100 weights are ~= 1. The remaining 906 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.251  0.859   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      9.94e-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 1.079 1      1.039
## Year              1.079 16      1.002

```

Residuals from first author



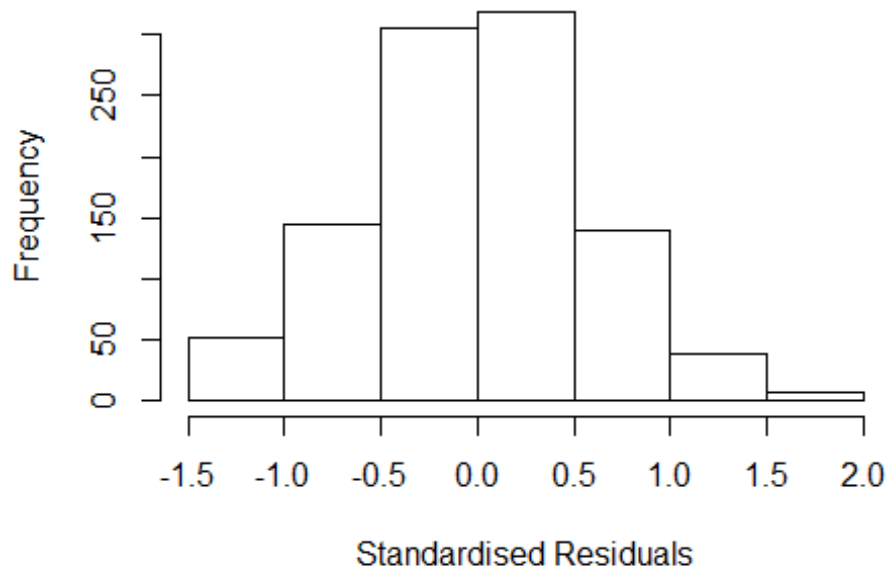
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41769 -0.35793  0.00219  0.38384  1.85053
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4177    0.1410   10.05  <2e-16 ***
## FirstAuthorFemale1 -0.0427    0.0469   -0.91   0.3622
## Year1997         -0.4652    0.1828   -2.54   0.0111 *
## Year1998         -0.1689    0.1694   -1.00   0.3190
## Year1999         -0.5775    0.1913   -3.02   0.0026 **
## Year2000         -0.2555    0.1753   -1.46   0.1452
## Year2001         -0.3276    0.1742   -1.88   0.0604 .
## Year2002         -0.3004    0.1599   -1.88   0.0605 .
## Year2003         -0.2454    0.1626   -1.51   0.1315
## Year2004         -0.3064    0.1595   -1.92   0.0551 .
## Year2005         -0.2600    0.1526   -1.70   0.0888 .
## Year2006         -0.1709    0.1594   -1.07   0.2837
```

```

## Year2007          -0.2790      0.1590   -1.75   0.0796 .
## Year2008          -0.4060      0.1539   -2.64   0.0085 **
## Year2009          -0.3492      0.1527   -2.29   0.0224 *
## Year2010          -0.3082      0.1588   -1.94   0.0526 .
## Year2011          -0.3200      0.1545   -2.07   0.0386 *
## Year2012          -0.3528      0.1553   -2.27   0.0233 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.0296, Adjusted R-squared:  0.0129
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 87 weights are ~= 1. The remaining 919 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.255  0.861  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      9.94e-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 1.066 1          1.032
## Year            1.066 16          1.002

```


Residuals from last author



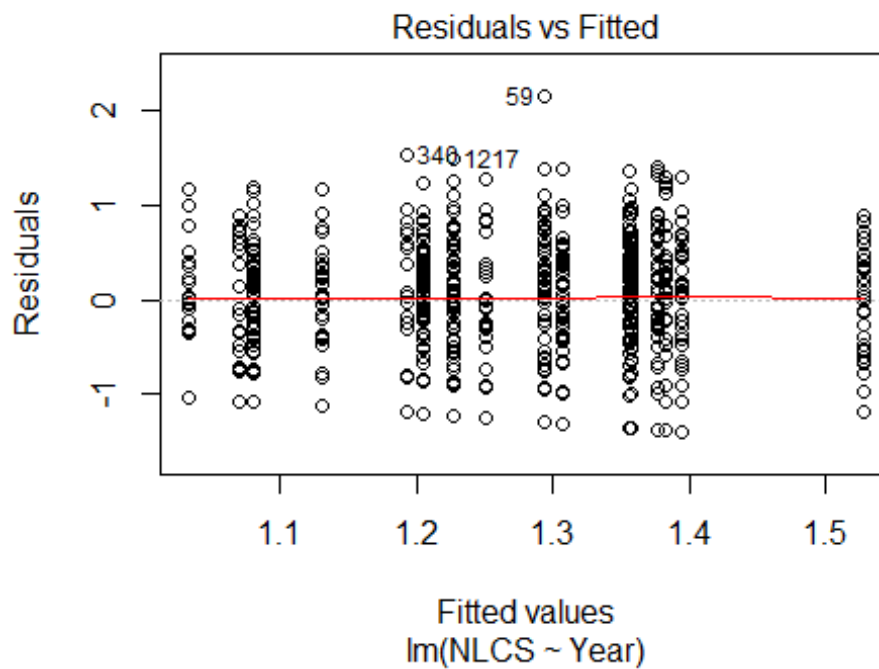
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.41538 -0.35423 0.00351 0.37730 1.85101
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4154 0.1381 10.25 <2e-16 ***
## LastAuthorFemale1 -0.0893 0.0557 -1.60 0.1090
## Year1997 -0.4634 0.1805 -2.57 0.0104 *
## Year1998 -0.1619 0.1667 -0.97 0.3317
## Year1999 -0.5682 0.1894 -3.00 0.0028 **
## Year2000 -0.2461 0.1731 -1.42 0.1555
## Year2001 -0.3208 0.1719 -1.87 0.0623 .
## Year2002 -0.2968 0.1574 -1.88 0.0597 .
## Year2003 -0.2392 0.1604 -1.49 0.1362
## Year2004 -0.3019 0.1571 -1.92 0.0549 .
## Year2005 -0.2471 0.1501 -1.65 0.1001
## Year2006 -0.1595 0.1571 -1.02 0.3102
```

```

## Year2007          -0.2723      0.1564    -1.74    0.0821 .
## Year2008          -0.4039      0.1512    -2.67    0.0077 **
## Year2009          -0.3351      0.1509    -2.22    0.0266 *
## Year2010          -0.3010      0.1559    -1.93    0.0538 .
## Year2011          -0.3136      0.1518    -2.07    0.0390 *
## Year2012          -0.3507      0.1525    -2.30    0.0217 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0316, Adjusted R-squared:  0.0149
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 95 weights are ~= 1. The remaining 911 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.251  0.861  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      9.94e-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: 1006"
## [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
##   55   62   48   55   55   50   51   55   55   84   87   83   85  100  105
## 2011 2012
##   86  101
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   29   33   21   29   29   22   25   31   30   45   45   38   44   58   75
## 2011 2012

```

```
## 63 67
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 26 26 18 20 18 19 22 23 25 32 33 25 35 44 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 = 27, df = 16, p-value = 0.04
```



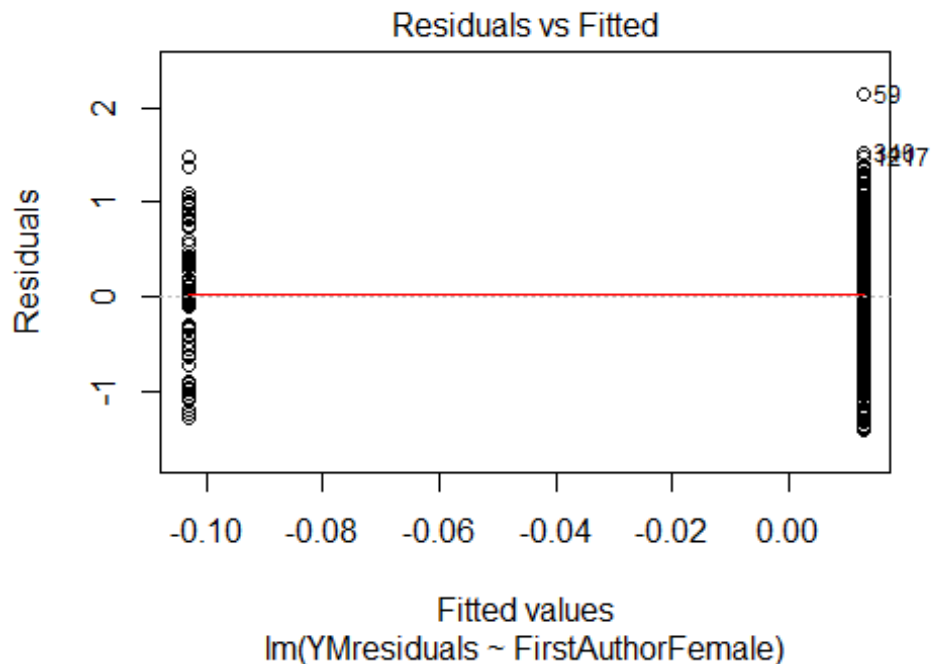
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.2, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 2.91295063024394"
## [1] "Male first author team size 2018 geometric mean: 3.20777844727166"

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

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 61, p-value = 0.6
## 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: 3.22078352880252"

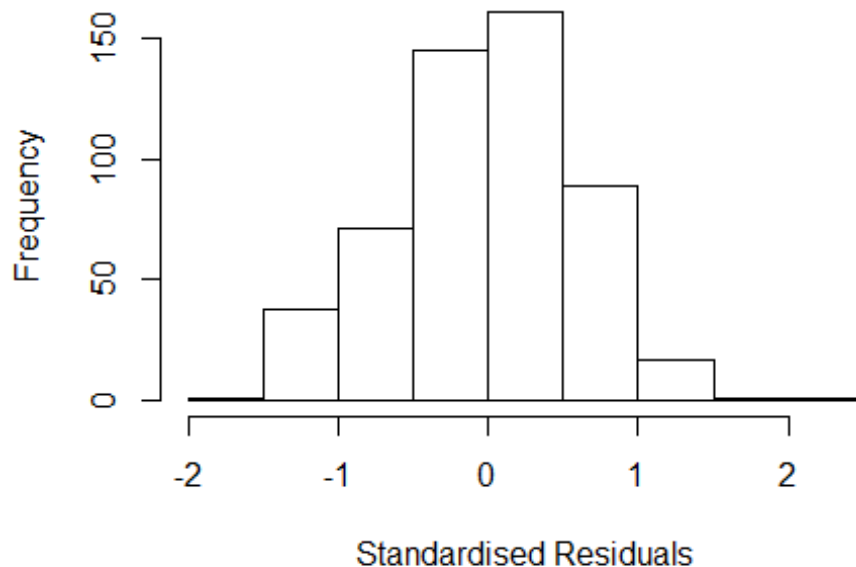
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 25, p-value = 0.4
## 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	1.222	1	1.105
LastAuthorFemale	1.200	1	1.096
UniqueAuthors	2.065	4	1.095
Year	2.425	16	1.028

Residuals from first and last author and team size



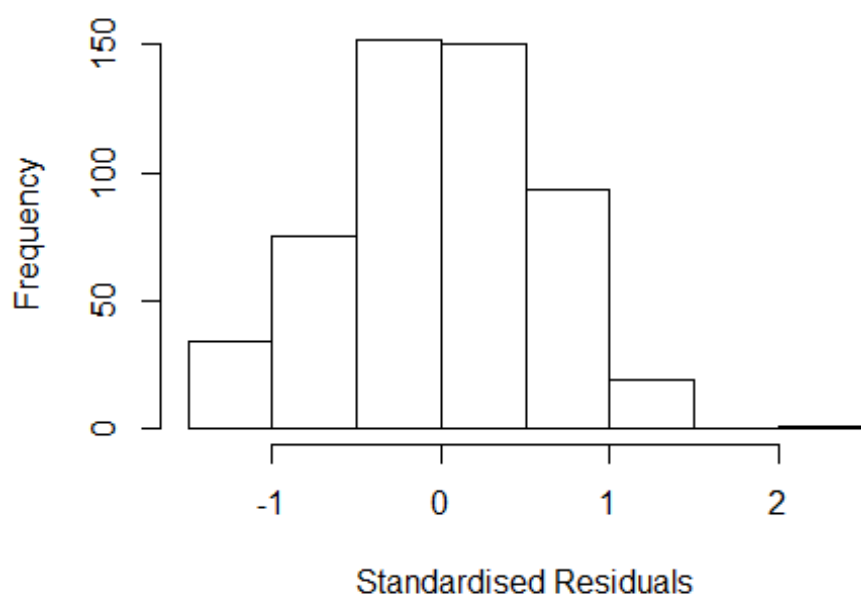
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5301 -0.4398  0.0282  0.4034  2.0403
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.226342   0.181120   6.77 3.6e-11 ***
## FirstAuthorFemale1 -0.034880   0.098480  -0.35  0.7234
## LastAuthorFemale1  0.011239   0.140393   0.08  0.9362
## UniqueAuthors2     0.184354   0.086649   2.13  0.0339 *
## UniqueAuthors3     0.257578   0.091549   2.81  0.0051 **
## UniqueAuthors4     0.309234   0.121867   2.54  0.0115 *
## UniqueAuthors5     0.109270   0.137697   0.79  0.4278
## Year1997          -0.019271   0.221624  -0.09  0.9307
## Year1998          -0.247308   0.224075  -1.10  0.2703
## Year1999          -0.153873   0.222579  -0.69  0.4897
```

```

## Year2000      -0.000907    0.271771    0.00    0.9973
## Year2001      -0.134268    0.271344   -0.49    0.6209
## Year2002      -0.377092    0.234929   -1.61    0.1091
## Year2003      -0.208332    0.219149   -0.95    0.3422
## Year2004       0.227922    0.210782    1.08    0.2801
## Year2005       0.060998    0.225842    0.27    0.7872
## Year2006       0.046167    0.214592    0.22    0.8297
## Year2007       0.066080    0.221226    0.30    0.7653
## Year2008      -0.061782    0.206999   -0.30    0.7655
## Year2009      -0.432548    0.195166   -2.22    0.0271 *
## Year2010      -0.155866    0.192647   -0.81    0.4189
## Year2011      -0.203151    0.193621   -1.05    0.2946
## Year2012      -0.103273    0.193524   -0.53    0.5938
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.627
## Multiple R-squared:  0.0825, Adjusted R-squared:  0.0422
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 476 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.269  0.865  0.952  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.91e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.137 1      1.067
## LastAuthorFemale  1.201 1      1.096
## Year              1.250 16      1.007

```

Residuals from first and last author



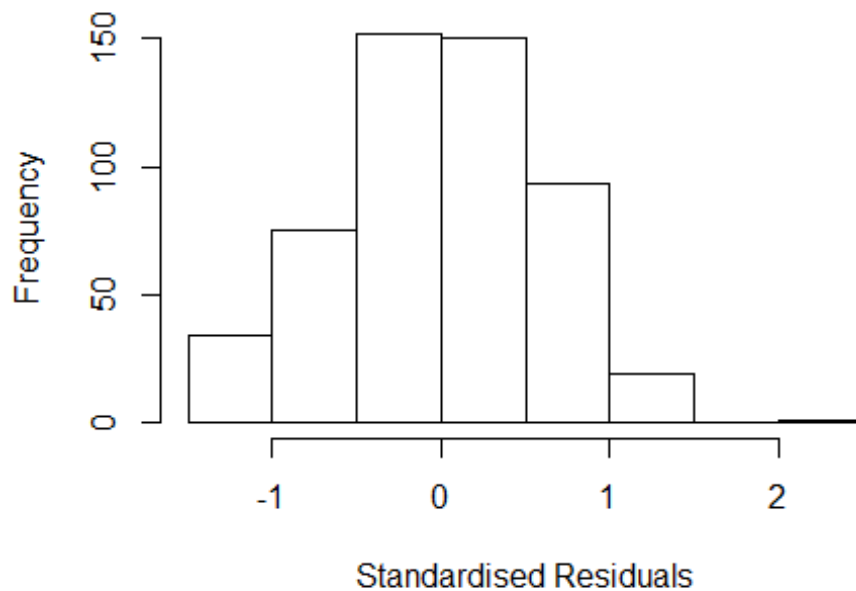
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47393 -0.39564  0.00296  0.40200  2.09088
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.36012    0.18910     7.19 2.3e-12 ***
## FirstAuthorFemale1 -0.02105    0.10022    -0.21  0.834
## LastAuthorFemale1  0.00607    0.14807     0.04  0.967
## Year1997        -0.02150    0.23284    -0.09  0.926
## Year1998        -0.28993    0.23290    -1.24  0.214
## Year1999        -0.19329    0.23143    -0.84  0.404
## Year2000         0.02722    0.27346     0.10  0.921
## Year2001        -0.10841    0.29041    -0.37  0.709
## Year2002        -0.37849    0.24819    -1.52  0.128
## Year2003        -0.19160    0.23318    -0.82  0.412
## Year2004         0.26732    0.22612     1.18  0.238
## Year2005         0.11381    0.23074     0.49  0.622
```

```

## Year2006          0.08906    0.22488    0.40    0.692
## Year2007          0.10061    0.22847    0.44    0.660
## Year2008         -0.04045    0.21943   -0.18    0.854
## Year2009         -0.36786    0.20719   -1.78    0.076 .
## Year2010         -0.09515    0.20600   -0.46    0.644
## Year2011         -0.14215    0.20249   -0.70    0.483
## Year2012         -0.02100    0.20668   -0.10    0.919
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.624
## Multiple R-squared:  0.0629, Adjusted R-squared:  0.0295
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 51 weights are ~= 1. The remaining 473 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.238  0.866  0.947  0.906  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.91e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



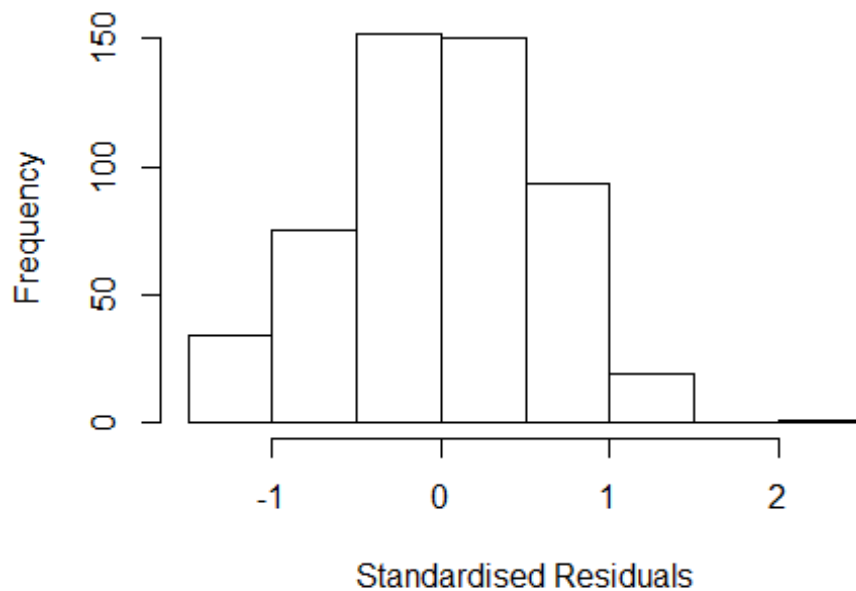
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.47399 -0.39565 0.00274 0.40214 2.09034
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3607 0.1884 7.22 1.9e-12 ***
## FirstAuthorFemale1 -0.0197 0.0989 -0.20 0.842
## Year1997 -0.0221 0.2322 -0.10 0.924
## Year1998 -0.2903 0.2322 -1.25 0.212
## Year1999 -0.1934 0.2307 -0.84 0.402
## Year2000 0.0263 0.2727 0.10 0.923
## Year2001 -0.1098 0.2896 -0.38 0.705
## Year2002 -0.3786 0.2472 -1.53 0.126
## Year2003 -0.1919 0.2324 -0.83 0.409
## Year2004 0.2676 0.2245 1.19 0.234
## Year2005 0.1133 0.2305 0.49 0.623
## Year2006 0.0880 0.2249 0.39 0.696
```

```

## Year2007          0.1000      0.2281      0.44      0.661
## Year2008          -0.0411     0.2188     -0.19     0.851
## Year2009          -0.3677     0.2065     -1.78     0.076 .
## Year2010          -0.0958     0.2051     -0.47     0.641
## Year2011          -0.1424     0.2020     -0.71     0.481
## Year2012          -0.0217     0.2062     -0.11     0.916
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.628
## Multiple R-squared:  0.0626, Adjusted R-squared:  0.0311
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 472 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.245  0.867  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      1.91e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



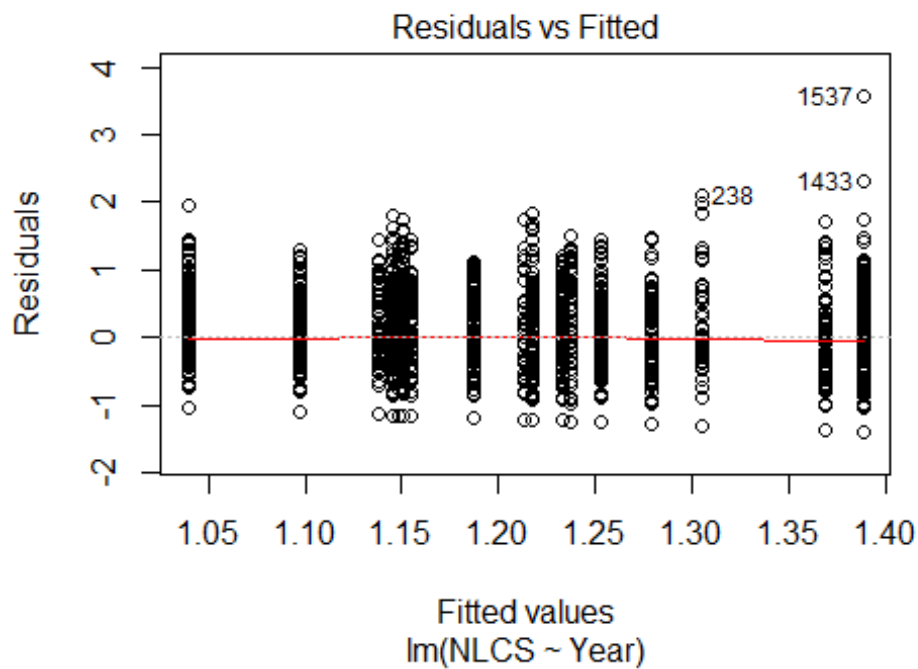
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.47259 -0.39304 0.00421 0.39548 2.09166
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.359344 0.189361 7.18 2.5e-12 ***
## LastAuthorFemale1 -0.000736 0.144779 -0.01 0.996
## Year1997 -0.022429 0.232964 -0.10 0.923
## Year1998 -0.289372 0.232965 -1.24 0.215
## Year1999 -0.195367 0.230449 -0.85 0.397
## Year2000 0.025385 0.273432 0.09 0.926
## Year2001 -0.108826 0.290724 -0.37 0.708
## Year2002 -0.381398 0.245357 -1.55 0.121
## Year2003 -0.192363 0.232660 -0.83 0.409
## Year2004 0.265604 0.225290 1.18 0.239
## Year2005 0.113243 0.231090 0.49 0.624
## Year2006 0.086967 0.225742 0.39 0.700
```

```

## Year2007      0.097339    0.227974    0.43    0.670
## Year2008     -0.041844    0.218955   -0.19    0.849
## Year2009     -0.366458    0.207439   -1.77    0.078 .
## Year2010     -0.096106    0.205707   -0.47    0.641
## Year2011     -0.143005    0.202404   -0.71    0.480
## Year2012     -0.023124    0.206179   -0.11    0.911
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.63
## Multiple R-squared:  0.0624, Adjusted R-squared:  0.0309
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 472 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.248  0.869  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      1.91e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 524"
## [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
##   99   95   72  106  119  139  159  112  146  225  288  268  255  246  226
## 2011 2012
##  226  218
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   52   58   38   58   62   71  103   70   89  124  154  158  161  140  135
## 2011 2012

```

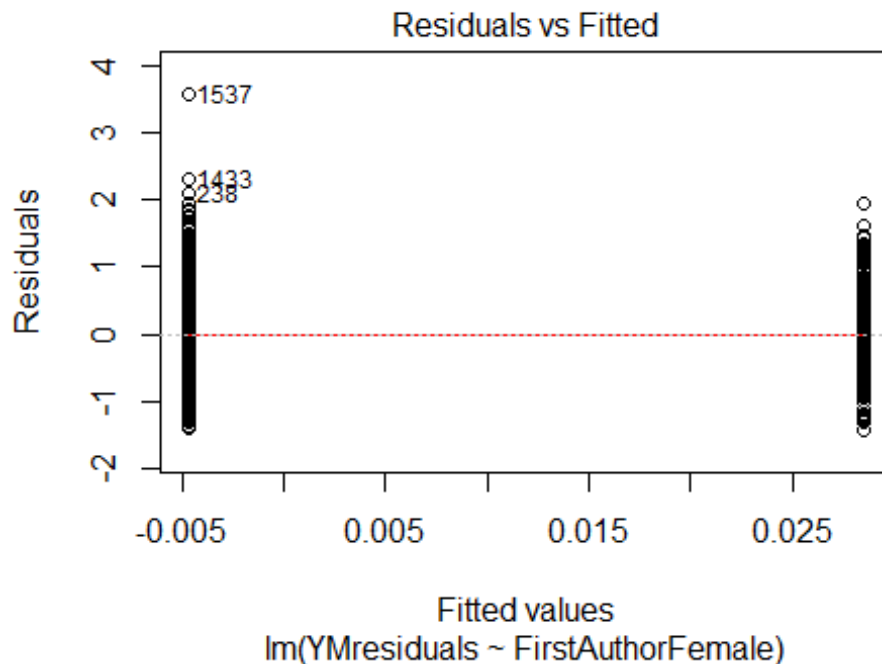
```
## 145 135
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 42 52 33 47 51 61 75 59 63 97 113 108 124 106 108
## 2011 2012
## 116 106
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 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: 2.62173937710938"
## [1] "Male first author team size 2018 geometric mean: 2.54838334237675"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

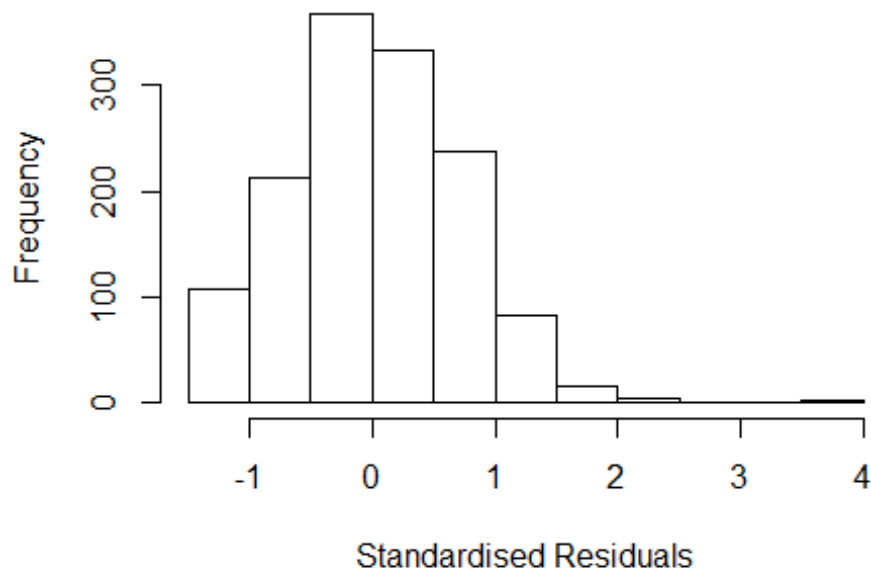
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 210, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.44948974278318"
## [1] "Male last author team size 2018 geometric mean: 2.57982256288372"

## Warning in wilcox.test.default(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 = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.173 1      1.083
## LastAuthorFemale  1.184 1      1.088
## UniqueAuthors    1.237 4      1.027
## Year              1.340 16     1.009
```

Residuals from first and last author and team size



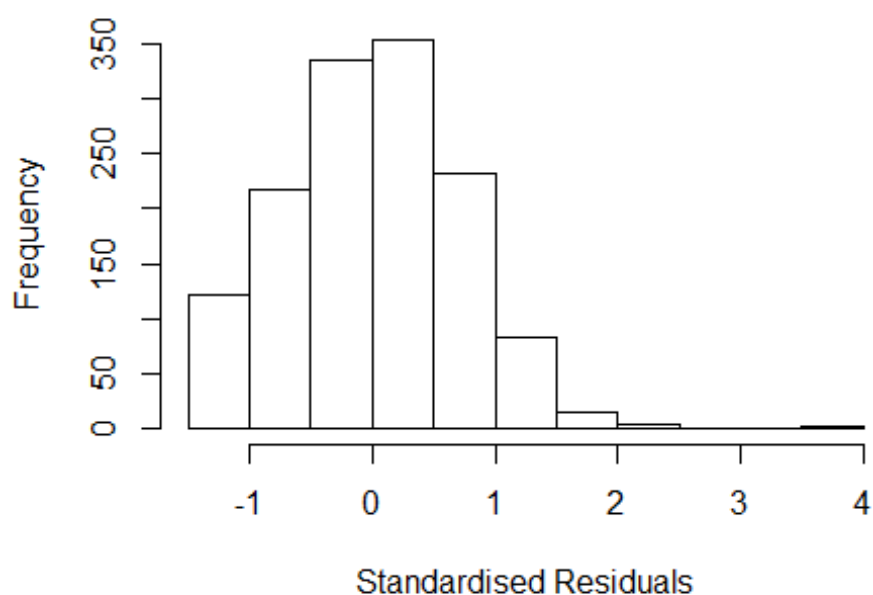
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1537 13844296408 4.951 2005      1705      2      3.767
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36725 -0.46626 -0.00669  0.49330  3.76738
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06e+00   1.17e-01   9.03  < 2e-16 ***
## FirstAuthorFemale1  4.86e-02   5.75e-02   0.84  0.39871
## LastAuthorFemale1 -7.30e-02   5.73e-02  -1.27  0.20354
## UniqueAuthors2    1.84e-01   5.21e-02   3.53  0.00044 ***
## UniqueAuthors3    2.80e-01   5.69e-02   4.92  9.6e-07 ***
## UniqueAuthors4    1.40e-01   8.84e-02   1.58  0.11356
## UniqueAuthors5    3.87e-01   9.12e-02   4.24  2.4e-05 ***
## Year1997          6.96e-03   1.52e-01   0.05  0.96359
## Year1998          6.77e-02   1.94e-01   0.35  0.72670
## Year1999         -1.83e-02   1.63e-01  -0.11  0.91104
```

```

## Year2000          7.21e-05    1.78e-01    0.00  0.99968
## Year2001          1.58e-01    1.49e-01    1.06  0.28825
## Year2002         -1.07e-03    1.43e-01   -0.01  0.99404
## Year2003         -8.41e-03    1.50e-01   -0.06  0.95537
## Year2004          1.24e-01    1.36e-01    0.91  0.36189
## Year2005          1.23e-01    1.42e-01    0.87  0.38448
## Year2006         -8.28e-02    1.36e-01   -0.61  0.54359
## Year2007         -2.58e-01    1.29e-01   -2.01  0.04500 *
## Year2008         -9.79e-02    1.27e-01   -0.77  0.44260
## Year2009         -1.65e-01    1.28e-01   -1.29  0.19657
## Year2010         -4.52e-02    1.30e-01   -0.35  0.72786
## Year2011         -3.84e-02    1.31e-01   -0.29  0.77011
## Year2012         -9.99e-02    1.34e-01   -0.75  0.45538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.711
## Multiple R-squared:  0.0429, Adjusted R-squared:  0.0272
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 558 is an outlier with |weight| = 0 ( < 7.3e-05);
## 107 weights are ~= 1. The remaining 1253 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.261  0.877  0.951  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           7.35e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.169 1 1.081
## LastAuthorFemale 1.161 1 1.078
## Year 1.108 16 1.003

```


Residuals from first and last author

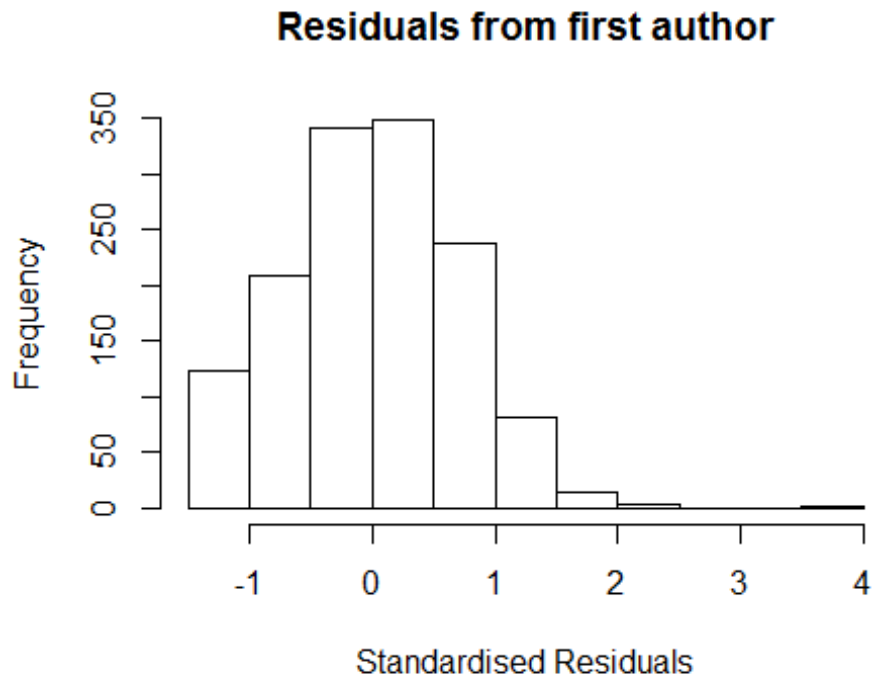


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1537 13844296408 4.951 2005      1705      2      3.587
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3928 -0.4954  0.0104  0.4902  3.5871
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.17708    0.11395   10.33  <2e-16 ***
## FirstAuthorFemale1  0.04309    0.05732    0.75    0.45
## LastAuthorFemale1 -0.08670    0.05707   -1.52    0.13
## Year1997          0.00613    0.15131    0.04    0.97
## Year1998          0.06993    0.19379    0.36    0.72
## Year1999          0.02247    0.16712    0.13    0.89
## Year2000          0.02486    0.17796    0.14    0.89
## Year2001          0.21569    0.14969    1.44    0.15
## Year2002          0.04864    0.14705    0.33    0.74
## Year2003          0.05044    0.14715    0.34    0.73
## Year2004          0.14766    0.13602    1.09    0.28
## Year2005          0.18687    0.14160    1.32    0.19
```

```

## Year2006      -0.01697    0.13697   -0.12    0.90
## Year2007      -0.20364    0.13026   -1.56    0.12
## Year2008      -0.04467    0.12751   -0.35    0.73
## Year2009      -0.09507    0.12775   -0.74    0.46
## Year2010       0.01985    0.12904    0.15    0.88
## Year2011       0.01673    0.13139    0.13    0.90
## Year2012      -0.00927    0.13462   -0.07    0.95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.713
## Multiple R-squared:  0.0216, Adjusted R-squared:  0.00851
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 558 is an outlier with |weight| = 0 ( < 7.3e-05);
## 125 weights are ~= 1. The remaining 1235 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.263  0.873   0.947   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      7.35e-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 1.053 1          1.026
## Year              1.053 16          1.002

```

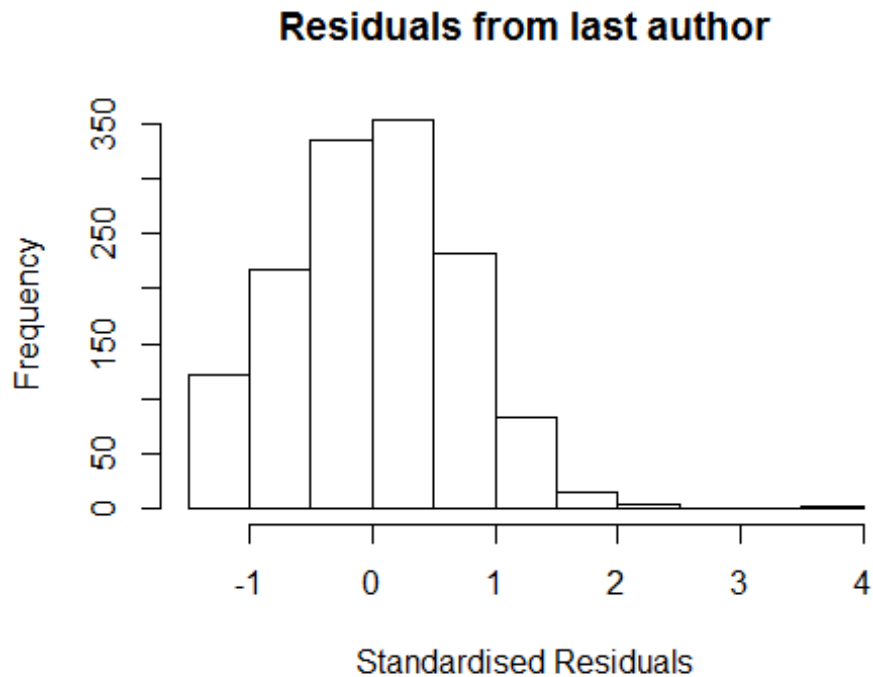


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1537 13844296408 4.951 2005      1705      2      3.587
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.38331 -0.48851  0.00518  0.49520  3.59149
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.17313    0.11448   10.25  <2e-16 ***
## FirstAuthorFemale1  0.01656    0.05438    0.30    0.76
## Year1997        -0.00311    0.15120   -0.02    0.98
## Year1998         0.06151    0.19356    0.32    0.75
## Year1999         0.02257    0.16705    0.14    0.89
## Year2000         0.02145    0.17815    0.12    0.90
## Year2001         0.21018    0.15048    1.40    0.16
## Year2002         0.04733    0.14753    0.32    0.75
## Year2003         0.04810    0.14707    0.33    0.74
## Year2004         0.13716    0.13594    1.01    0.31
## Year2005         0.18637    0.14227    1.31    0.19
## Year2006        -0.01970    0.13787   -0.14    0.89
```

```

## Year2007          -0.20530    0.13055   -1.57    0.12
## Year2008          -0.04841    0.12786   -0.38    0.71
## Year2009          -0.10029    0.12802   -0.78    0.43
## Year2010           0.01645    0.12951    0.13    0.90
## Year2011           0.00667    0.13186    0.05    0.96
## Year2012          -0.01732    0.13498   -0.13    0.90
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.713
## Multiple R-squared:  0.0201, Adjusted R-squared:  0.00765
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 558 is an outlier with |weight| = 0 ( < 7.3e-05);
## 122 weights are ~= 1. The remaining 1238 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.260  0.874  0.947  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      7.35e-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 1.046 1          1.023
## Year            1.046 16          1.001

```



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1537 13844296408 4.951 2005      1705      2      3.587
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3946 -0.4938  0.0144  0.4944  3.5817
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17961    0.11399   10.35  <2e-16 ***
## LastAuthorFemale1 -0.07335    0.05423   -1.35    0.18
## Year1997         0.00200    0.15110    0.01    0.99
## Year1998         0.06799    0.19423    0.35    0.73
## Year1999         0.02371    0.16729    0.14    0.89
## Year2000         0.02164    0.17755    0.12    0.90
## Year2001         0.21499    0.14968    1.44    0.15
## Year2002         0.05153    0.14716    0.35    0.73
## Year2003         0.05280    0.14698    0.36    0.72
## Year2004         0.14932    0.13597    1.10    0.27
## Year2005         0.18966    0.14167    1.34    0.18
## Year2006        -0.01369    0.13721   -0.10    0.92
```

```

## Year2007          -0.20267      0.13048      -1.55      0.12
## Year2008          -0.04168      0.12764      -0.33      0.74
## Year2009          -0.09380      0.12792      -0.73      0.46
## Year2010           0.02302      0.12920       0.18      0.86
## Year2011           0.02018      0.13141       0.15      0.88
## Year2012          -0.00413      0.13453      -0.03      0.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.716
## Multiple R-squared:  0.0212, Adjusted R-squared:  0.00884
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 558 is an outlier with |weight| = 0 ( < 7.3e-05);
## 127 weights are ~= 1. The remaining 1233 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.269  0.873   0.948   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      7.35e-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: 1361"
## [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
##   312   14  313  261  313  368  381  311  364  487  618  571  630  763  653
## 2011 2012
##   652  724
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   122    1  142  110  138  162  207  176  198  255  340  307  342  422  391

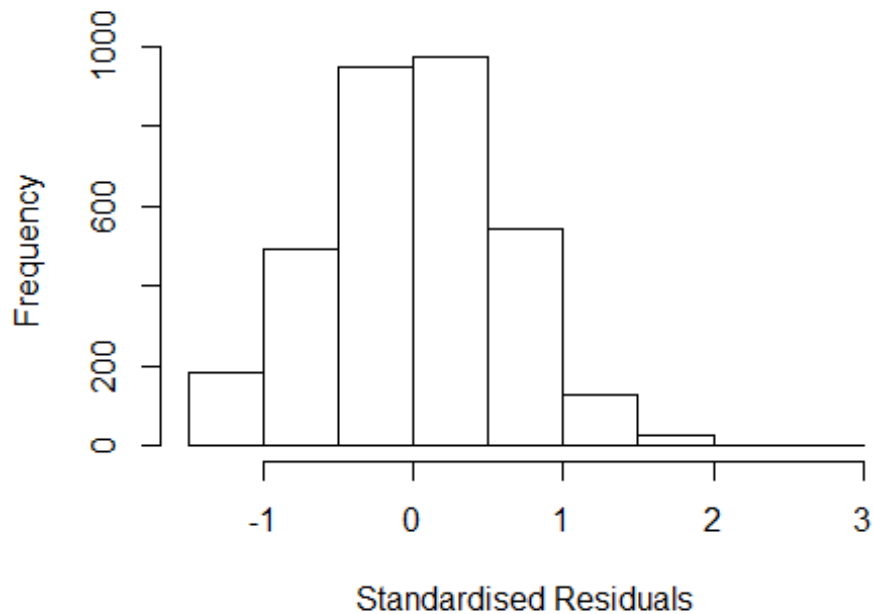
```

```

## 2011 2012
## 376 446
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 108 1 119 92 120 129 170 142 158 197 251 237 274 334 311
## 2011 2012
## 314 338
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.5974786082779"
## [1] "Male first author team size 2018 geometric mean: 2.59968261481002"
##
## 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] "Female last author team size 2018 geometric mean: 2.14167157788204"
## [1] "Male last author team size 2018 geometric mean: 2.74069891918195"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, 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 1.346 1 1.160
## LastAuthorFemale 9.942 1 3.153
## UniqueAuthors 13.245 4 1.381
## Year 24.720 16 1.105

```

Residuals from first and last author and team size



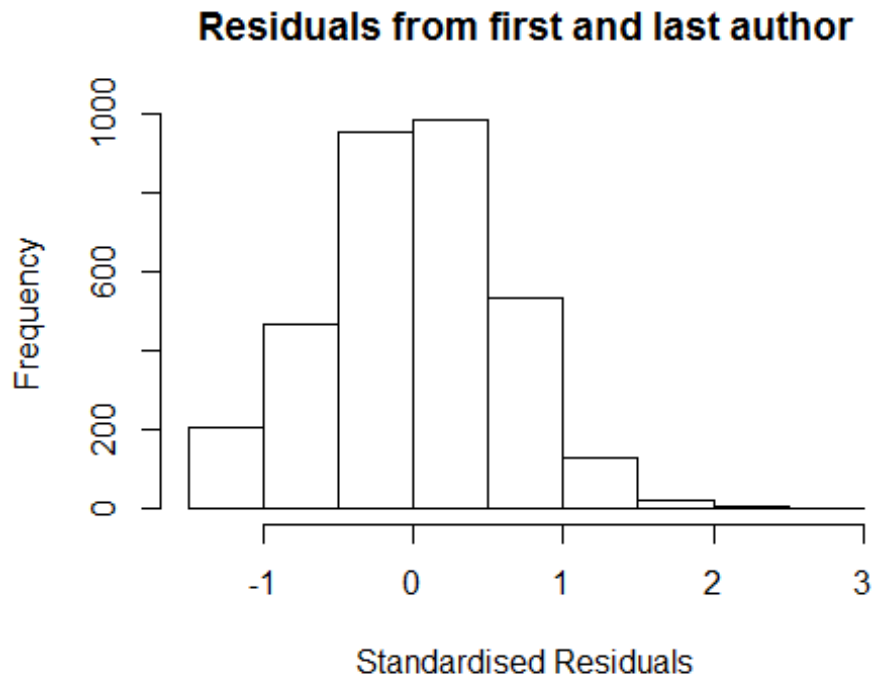
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 418 0031987382 3.783 1998    1706    4      2.54
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3491 -0.4088  0.0087  0.4156  2.5397
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.1443    0.0724   15.81 < 2e-16 ***
## FirstAuthorFemale1  0.0171    0.0322    0.53  0.5961
## LastAuthorFemale1 -0.0388    0.0355   -1.09  0.2745
## UniqueAuthors2    0.1867    0.0306    6.09 1.2e-09 ***
## UniqueAuthors3    0.2432    0.0315    7.71 1.6e-14 ***
## UniqueAuthors4    0.1960    0.0425    4.61 4.1e-06 ***
## UniqueAuthors5    0.3291    0.0445    7.40 1.7e-13 ***
## Year1997         -0.1816    0.0882   -2.06  0.0396 *
## Year1998         -0.1442    0.0932   -1.55  0.1219
## Year1999         -0.0383    0.1067   -0.36  0.7197
```



```

## Year2000          -0.1958      0.0863   -2.27   0.0233 *
## Year2001          -0.1976      0.1044   -1.89   0.0586 .
## Year2002          -0.2974      0.0910   -3.27   0.0011 **
## Year2003          -0.1049      0.0840   -1.25   0.2120
## Year2004          -0.1081      0.0871   -1.24   0.2144
## Year2005          -0.0808      0.0831   -0.97   0.3310
## Year2006          -0.1351      0.0810   -1.67   0.0954 .
## Year2007          -0.1148      0.0822   -1.40   0.1623
## Year2008          -0.1631      0.0805   -2.03   0.0428 *
## Year2009          -0.0770      0.0801   -0.96   0.3366
## Year2010          -0.1271      0.0793   -1.60   0.1091
## Year2011          -0.1568      0.0787   -1.99   0.0464 *
## Year2012          -0.1736      0.0788   -2.20   0.0277 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.0331, Adjusted R-squared:  0.0266
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 270 weights are ~= 1. The remaining 3025 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.038  0.872  0.950  0.909  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          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 1.349 1          1.161
## LastAuthorFemale  9.716 1          3.117
## Year             10.909 16          1.078

```

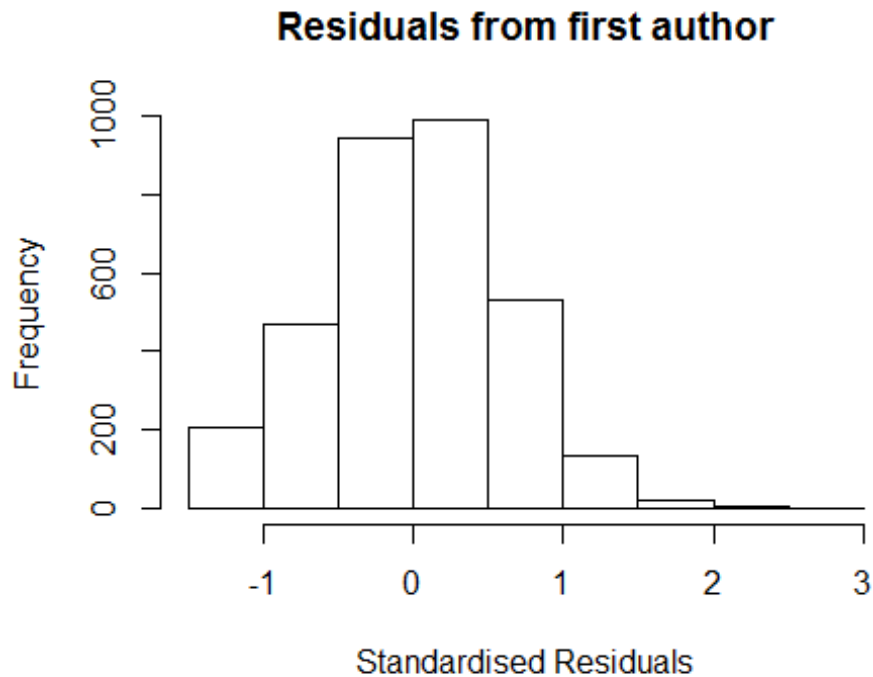


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 418 0031987382 3.783 1998      1706      4      2.645
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2595 -0.3999  0.0143  0.4087  2.6451
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2566     0.0716   17.55  <2e-16 ***
## FirstAuthorFemale1  0.0190     0.0324    0.59  0.5582
## LastAuthorFemale1 -0.0507     0.0359   -1.41  0.1572
## Year1997          0.0471     0.0794    0.59  0.5530
## Year1998         -0.1187     0.0939   -1.26  0.2063
## Year1999         -0.0289     0.1092   -0.26  0.7913
## Year2000         -0.1603     0.0870   -1.84  0.0655 .
## Year2001         -0.1242     0.1031   -1.21  0.2282
## Year2002         -0.2316     0.0898   -2.58  0.0099 **
## Year2003         -0.0747     0.0847   -0.88  0.3779
## Year2004         -0.0652     0.0877   -0.74  0.4572
## Year2005         -0.0264     0.0834   -0.32  0.7515
```

```

## Year2006          -0.0696      0.0806   -0.86   0.3881
## Year2007          -0.0510      0.0823   -0.62   0.5353
## Year2008          -0.0954      0.0805   -1.19   0.2359
## Year2009          -0.0160      0.0804   -0.20   0.8418
## Year2010          -0.0597      0.0791   -0.75   0.4508
## Year2011          -0.0877      0.0789   -1.11   0.2661
## Year2012          -0.0979      0.0786   -1.25   0.2128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.00743,    Adjusted R-squared:  0.00197
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 293 weights are ~= 1. The remaining 3002 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0192 0.8690 0.9510 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 1.208 1          1.099
## Year              1.208 16          1.006

```

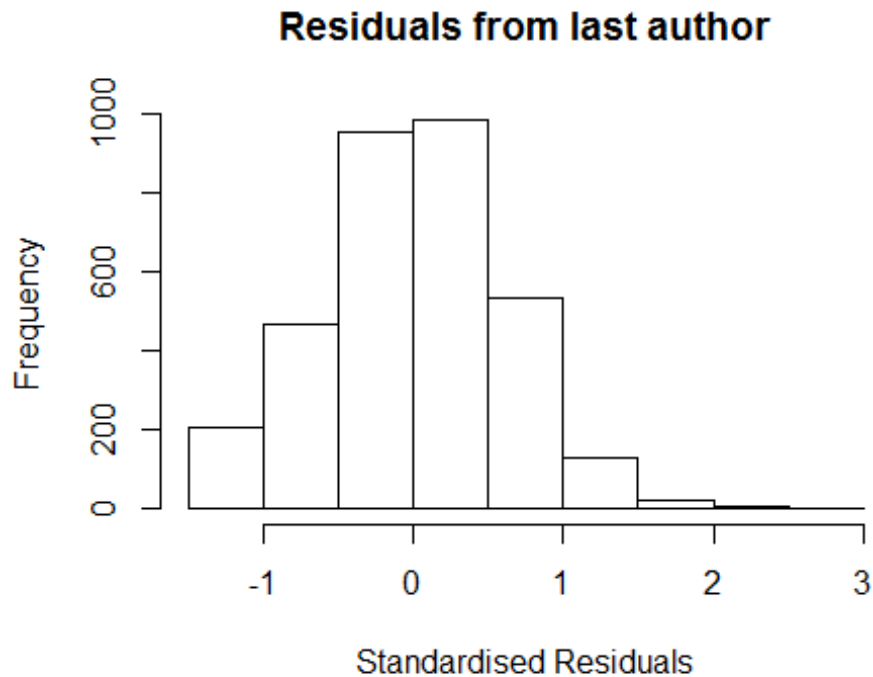


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 418 0031987382 3.783 1998    1706      4      2.645
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.256 -0.398  0.015  0.410  2.650
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.254740   0.071487   17.55  <2e-16 ***
## FirstAuthorFemale1 0.000995   0.030957    0.03   0.974
## Year1997      -0.001740   0.071487   -0.02   0.981
## Year1998      -0.121383   0.093671   -1.30   0.195
## Year1999      -0.031282   0.109305   -0.29   0.775
## Year2000      -0.162821   0.086899   -1.87   0.061 .
## Year2001      -0.125680   0.103168   -1.22   0.223
## Year2002      -0.231352   0.089758   -2.58   0.010 **
## Year2003      -0.075912   0.084583   -0.90   0.370
## Year2004      -0.067844   0.087717   -0.77   0.439
## Year2005      -0.029554   0.083385   -0.35   0.723
## Year2006      -0.071201   0.080497   -0.88   0.376
```

```

## Year2007          -0.051700    0.082138    -0.63    0.529
## Year2008          -0.098195    0.080378    -1.22    0.222
## Year2009          -0.019511    0.080289    -0.24    0.808
## Year2010          -0.061698    0.079003    -0.78    0.435
## Year2011          -0.090070    0.078792    -1.14    0.253
## Year2012          -0.099619    0.078475    -1.27    0.204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.00675,    Adjusted R-squared:  0.0016
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 294 weights are ~= 1. The remaining 3001 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0185 0.8700 0.9500 0.9080 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 8.692 1          2.948
## Year            8.692 16          1.070

```



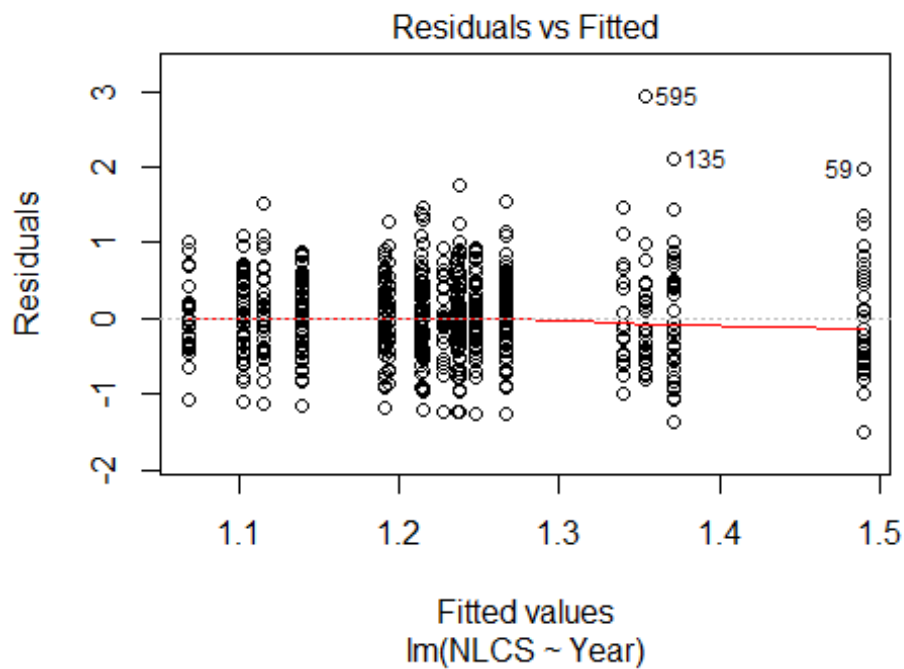
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 418 0031987382 3.783 1998    1706    4    2.645
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2577 -0.3991  0.0129  0.4082  2.6437
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2577     0.0716   17.57  <2e-16 ***
## LastAuthorFemale1 -0.0431     0.0339   -1.27    0.204
## Year1997          0.0384     0.0782    0.49    0.623
## Year1998         -0.1184     0.0938   -1.26    0.207
## Year1999         -0.0289     0.1092   -0.26    0.792
## Year2000         -0.1609     0.0870   -1.85    0.065 .
## Year2001         -0.1234     0.1030   -1.20    0.231
## Year2002         -0.2298     0.0897   -2.56    0.010 *
## Year2003         -0.0737     0.0847   -0.87    0.384
## Year2004         -0.0648     0.0877   -0.74    0.460
## Year2005         -0.0258     0.0834   -0.31    0.758
## Year2006         -0.0688     0.0806   -0.85    0.393
```

```

## Year2007          -0.0496      0.0822   -0.60      0.546
## Year2008          -0.0943      0.0804   -1.17      0.241
## Year2009          -0.0160      0.0804   -0.20      0.843
## Year2010          -0.0585      0.0791   -0.74      0.460
## Year2011          -0.0869      0.0789   -1.10      0.271
## Year2012          -0.0966      0.0785   -1.23      0.219
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.00732,    Adjusted R-squared:  0.00217
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 294 weights are ~= 1. The remaining 3001 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0193 0.8700 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      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: 3295"
## [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
##   55   63   30   62   62   60   62   57   47   74   77   78   86   86   87
## 2011 2012
##   87  111
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   35   18   27   32   32   41   35   30   37   49   44   55   59   54
## 2011 2012

```

```
##      62      76
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    27    32    14    20    22    26    29    27    25    29    37    36    41    48    46
## 2011 2012
##    45    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 = 35, df = 16, p-value = 0.004
```

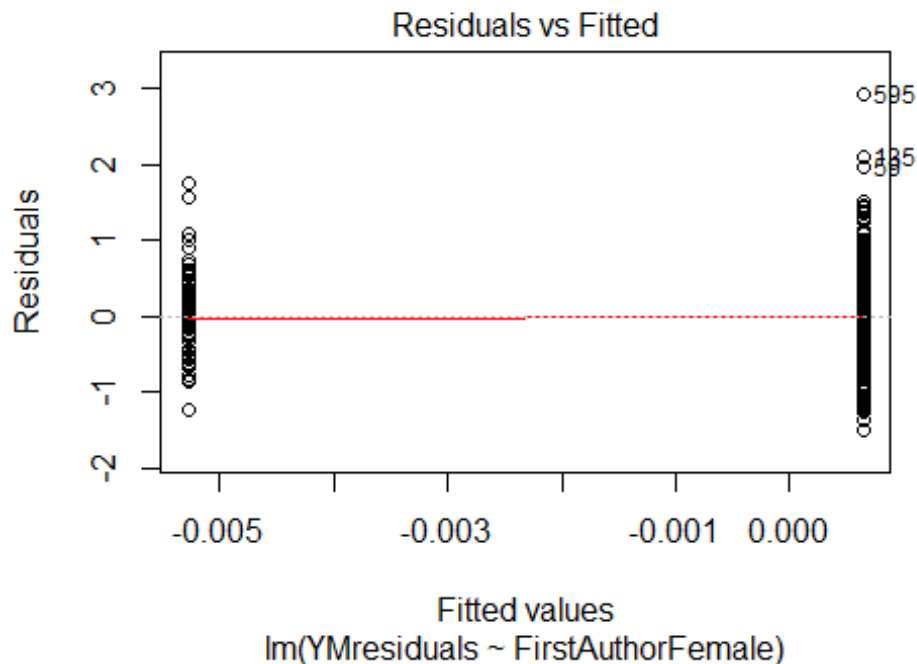


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2, df = 1, p-value = 0.2
## [1] "Female first author team size 2018 geometric mean: 2.58666719213051"
## [1] "Male first author team size 2018 geometric mean: 2.87253419458644"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



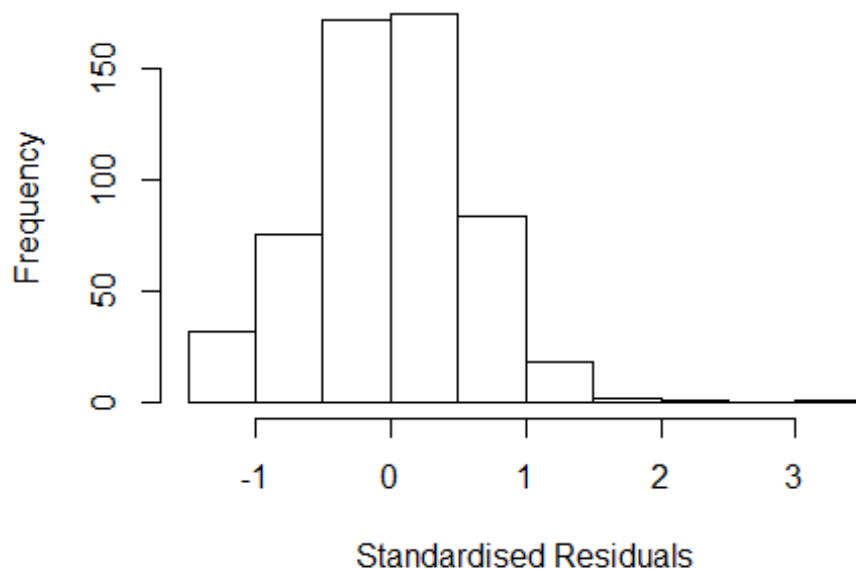
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 86, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.49146187923103"
## [1] "Male last author team size 2018 geometric mean: 2.84797909585204"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 55, 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 1.358 1      1.166
## LastAuthorFemale  1.333 1      1.155
## UniqueAuthors    1.899 4      1.083
## Year              2.172 16     1.025
```

Residuals from first and last author and team size



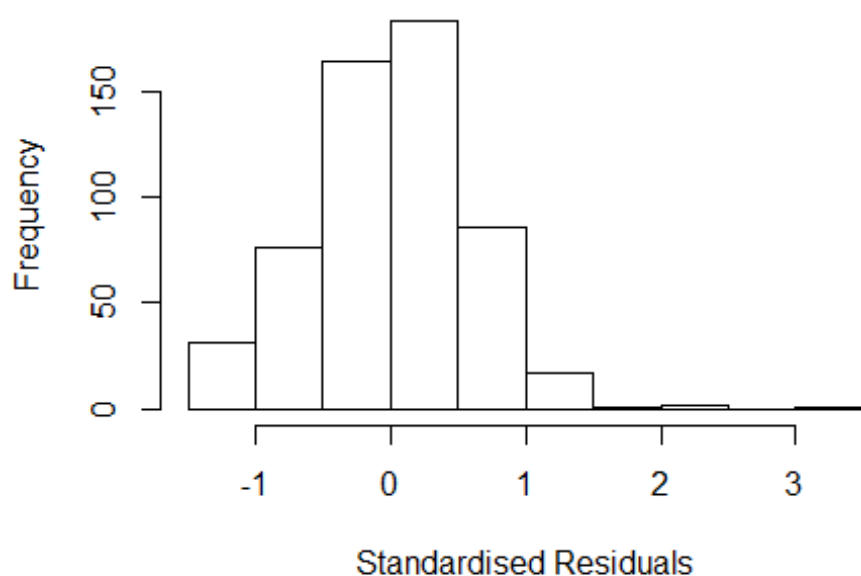
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 595 3042535216 4.284 2004      1702      3      3.095
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4862 -0.3583  0.0018  0.3865  3.0945
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3552    0.1632   8.30 8.3e-16 ***
## FirstAuthorFemale1 -0.1176    0.0767  -1.53  0.126
## LastAuthorFemale1  0.0967    0.0941   1.03  0.304
## UniqueAuthors2     0.1310    0.0794   1.65  0.099 .
## UniqueAuthors3     0.0496    0.0867   0.57  0.567
## UniqueAuthors4     0.2557    0.1030   2.48  0.013 *
## UniqueAuthors5     0.1594    0.1007   1.58  0.114
## Year1997          -0.1342    0.2134  -0.63  0.530
## Year1998          -0.0361    0.2405  -0.15  0.881
## Year1999          -0.1728    0.1900  -0.91  0.364
```

```

## Year2000          -0.3457      0.2230    -1.55      0.122
## Year2001          -0.2114      0.1816    -1.16      0.245
## Year2002          -0.3170      0.1965    -1.61      0.107
## Year2003          -0.4046      0.1843    -2.20      0.029 *
## Year2004          -0.1657      0.1768    -0.94      0.349
## Year2005          -0.3861      0.1791    -2.16      0.032 *
## Year2006          -0.3655      0.1823    -2.01      0.045 *
## Year2007          -0.2873      0.1663    -1.73      0.085 .
## Year2008          -0.1859      0.1700    -1.09      0.275
## Year2009          -0.3083      0.1756    -1.76      0.080 .
## Year2010          -0.1958      0.1759    -1.11      0.266
## Year2011          -0.2200      0.1824    -1.21      0.228
## Year2012          -0.2075      0.1798    -1.15      0.249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0463, Adjusted R-squared:  0.00734
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 203 is an outlier with |weight| = 0 ( < 0.00018);
## 54 weights are ~= 1. The remaining 506 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0698 0.8650 0.9470 0.9010 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          1.78e-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 1.294 1          1.138
## LastAuthorFemale 1.277 1          1.130
## Year          1.248 16          1.007

```

Residuals from first and last author



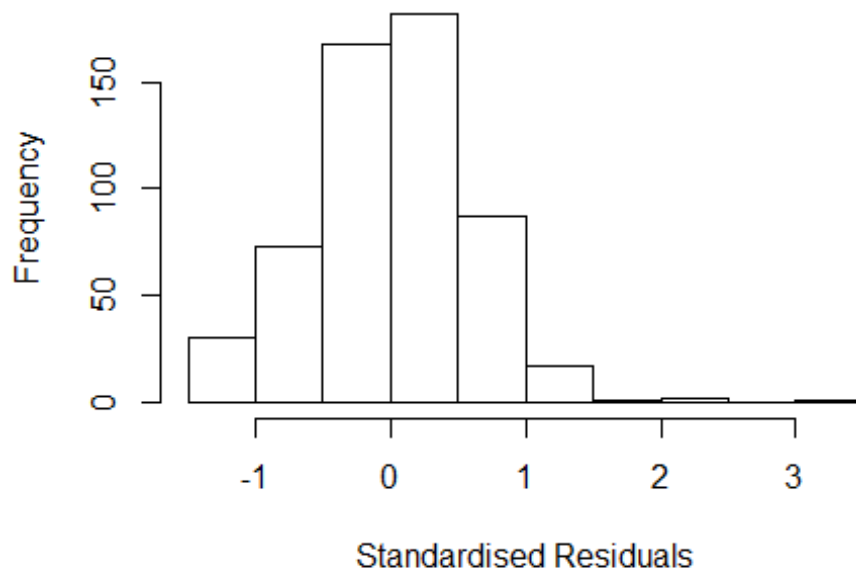
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 595 3042535216 4.284 2004      1702      3      3.013
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4388 -0.3746  0.0196  0.4077  3.0127
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.43883    0.15177     9.48  <2e-16 ***
## FirstAuthorFemale1 -0.12327    0.07602    -1.62   0.105
## LastAuthorFemale1  0.09247    0.09611     0.96   0.336
## Year1997        -0.13101    0.21298    -0.62   0.539
## Year1998         0.00922    0.24989     0.04   0.971
## Year1999        -0.17125    0.19141    -0.89   0.371
## Year2000        -0.32471    0.23266    -1.40   0.163
## Year2001        -0.19043    0.18220    -1.05   0.296
## Year2002        -0.26642    0.19117    -1.39   0.164
## Year2003        -0.37004    0.18492    -2.00   0.046 *
## Year2004        -0.16757    0.17924    -0.93   0.350
## Year2005        -0.36873    0.17913    -2.06   0.040 *
```

```

## Year2006      -0.34532    0.18198   -1.90    0.058 .
## Year2007      -0.24925    0.16557   -1.51    0.133
## Year2008      -0.19698    0.16864   -1.17    0.243
## Year2009      -0.30471    0.17755   -1.72    0.087 .
## Year2010      -0.17780    0.17604   -1.01    0.313
## Year2011      -0.18552    0.18001   -1.03    0.303
## Year2012      -0.18071    0.17713   -1.02    0.308
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.0323, Adjusted R-squared:  0.000211
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 203 is an outlier with |weight| = 0 ( < 0.00018);
## 44 weights are ~= 1. The remaining 516 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.110  0.867   0.949   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      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 1.119 1          1.058
## Year              1.119 16          1.004

```

Residuals from first author

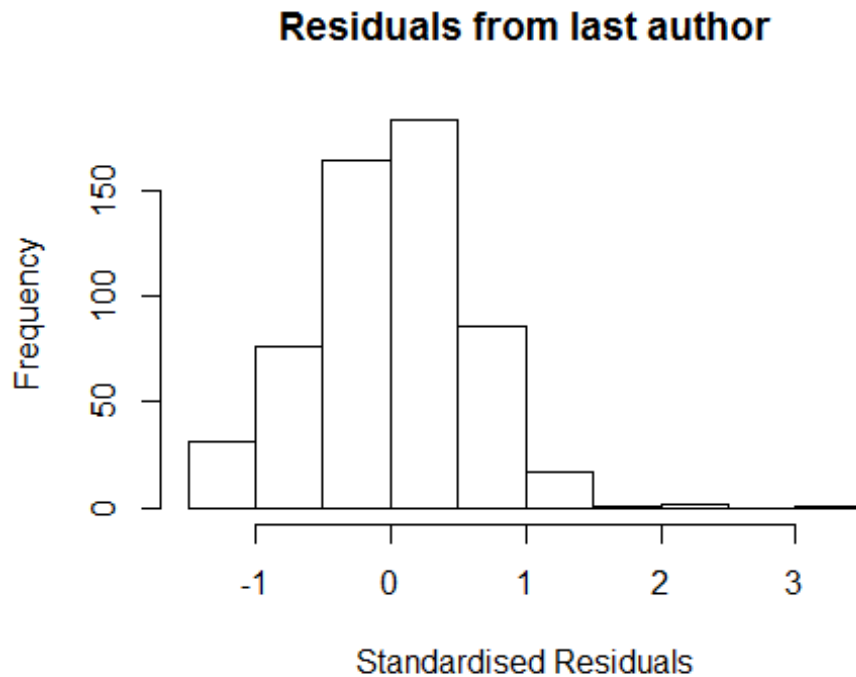


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 595 3042535216 4.284 2004      1702      3      3.013
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4481 -0.3747  0.0208  0.4079  3.0077
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4481     0.1531   9.46  <2e-16 ***
## FirstAuthorFemale1 -0.1028     0.0711  -1.45    0.149
## Year1997         -0.1280     0.2154  -0.59    0.553
## Year1998          0.0129     0.2460   0.05    0.958
## Year1999         -0.1634     0.1927  -0.85    0.397
## Year2000         -0.3340     0.2336  -1.43    0.153
## Year2001         -0.2009     0.1836  -1.09    0.274
## Year2002         -0.2788     0.1917  -1.45    0.146
## Year2003         -0.3781     0.1854  -2.04    0.042 *
## Year2004         -0.1718     0.1817  -0.95    0.345
## Year2005         -0.3786     0.1794  -2.11    0.035 *
## Year2006         -0.3434     0.1843  -1.86    0.063 .
```

```

## Year2007          -0.2495      0.1676   -1.49    0.137
## Year2008          -0.1965      0.1702   -1.15    0.249
## Year2009          -0.3034      0.1797   -1.69    0.092 .
## Year2010          -0.1854      0.1776   -1.04    0.297
## Year2011          -0.1908      0.1816   -1.05    0.294
## Year2012          -0.1876      0.1786   -1.05    0.294
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.0308, Adjusted R-squared:  0.000504
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 203 is an outlier with |weight| = 0 ( < 0.00018);
## 44 weights are ~= 1. The remaining 516 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.114  0.867  0.948  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      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 1.099 1          1.048
## Year          1.099 16          1.003

```



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 595 3042535216 4.284 2004      1702      3      3.013
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4434 -0.3761  0.0258  0.4031  3.0216
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.443363   0.152273   9.48  <2e-16 ***
## LastAuthorFemale1 0.057713   0.090939   0.63   0.526
## Year1997       -0.133716   0.214012  -0.62   0.532
## Year1998       -0.000234   0.244461   0.00   0.999
## Year1999       -0.189202   0.192979  -0.98   0.327
## Year2000       -0.329232   0.232693  -1.41   0.158
## Year2001       -0.201118   0.184677  -1.09   0.277
## Year2002       -0.290005   0.190468  -1.52   0.128
## Year2003       -0.394259   0.184776  -2.13   0.033 *
## Year2004       -0.180994   0.180387  -1.00   0.316
## Year2005       -0.390149   0.179332  -2.18   0.030 *
## Year2006       -0.351480   0.182040  -1.93   0.054 .
```

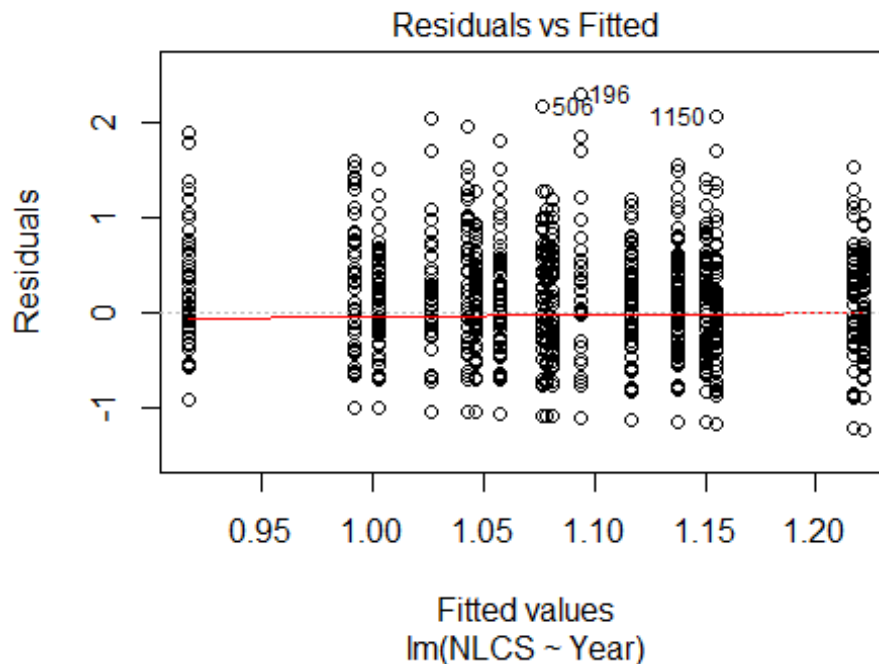


```

## Year2007          -0.261574    0.166040    -1.58     0.116
## Year2008          -0.212910    0.168069    -1.27     0.206
## Year2009          -0.314729    0.177978    -1.77     0.078 .
## Year2010          -0.196699    0.175405    -1.12     0.263
## Year2011          -0.208017    0.179398    -1.16     0.247
## Year2012          -0.207161    0.175362    -1.18     0.238
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.567
## Multiple R-squared:  0.0286, Adjusted R-squared:  -0.00182
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 203 is an outlier with |weight| = 0 ( < 0.00018);
## 43 weights are ~= 1. The remaining 517 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.115  0.869  0.949  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      1.78e-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 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
##   92   75  105   75  121  113  135   78   89  114  166  140  141  142  121
## 2011 2012
##  141  139
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   53   38   58   41   47   46   68   52   58   73   95   76   86   77   72

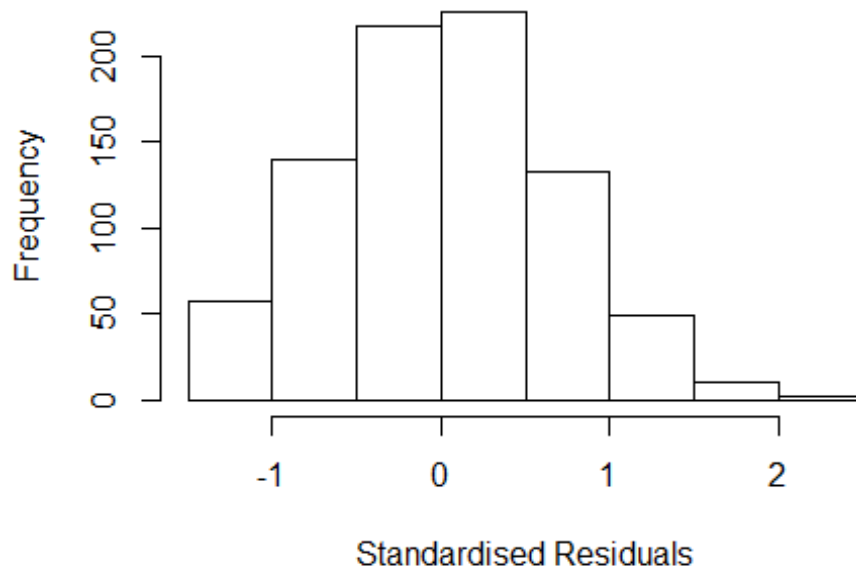
```

```
## 2011 2012
## 80 72
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 43 31 48 35 38 34 50 46 44 57 65 65 61 64 50
## 2011 2012
## 47 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 = 50, df = 16, p-value = 2e-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: 3.23774081372113"
## [1] "Male first author team size 2018 geometric mean: 2.6108073834666"
## 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



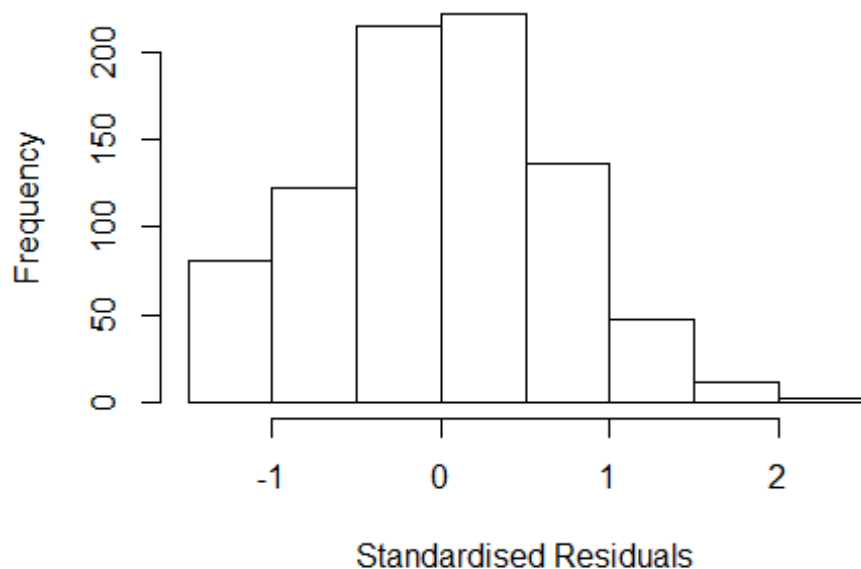
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.37653 -0.47192 0.00249 0.45201 2.04606
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.93093 0.15588 5.97 3.5e-09 ***
## FirstAuthorFemale1 -0.09021 0.07636 -1.18 0.23780
## LastAuthorFemale1 0.05897 0.07707 0.77 0.44440
## UniqueAuthors2 0.11804 0.06947 1.70 0.08966 .
## UniqueAuthors3 0.27093 0.07788 3.48 0.00053 ***
## UniqueAuthors4 0.16753 0.11730 1.43 0.15362
## UniqueAuthors5 0.29302 0.11748 2.49 0.01282 *
## Year1997 0.15008 0.22436 0.67 0.50375
## Year1998 -0.01426 0.20092 -0.07 0.94345
## Year1999 0.08927 0.18402 0.49 0.62773
```

```

## Year2000      0.04194      0.19767      0.21  0.83201
## Year2001     -0.01104      0.20230     -0.05  0.95648
## Year2002     -0.24442      0.17821     -1.37  0.17060
## Year2003      0.02900      0.18029      0.16  0.87226
## Year2004      0.23152      0.18042      1.28  0.19978
## Year2005      0.01605      0.17430      0.09  0.92665
## Year2006      0.11568      0.17529      0.66  0.50948
## Year2007      0.04748      0.17219      0.28  0.78279
## Year2008      0.08297      0.16720      0.50  0.61986
## Year2009      0.00759      0.17227      0.04  0.96486
## Year2010     -0.14603      0.17516     -0.83  0.40471
## Year2011      0.07071      0.16963      0.42  0.67691
## Year2012     -0.07019      0.17647     -0.40  0.69090
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.697
## Multiple R-squared:  0.0444, Adjusted R-squared:  0.0185
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 761 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.369  0.874  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      1.20e-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 1.124 1      1.060
## LastAuthorFemale  1.147 1      1.071
## Year              1.133 16      1.004

```

Residuals from first and last author



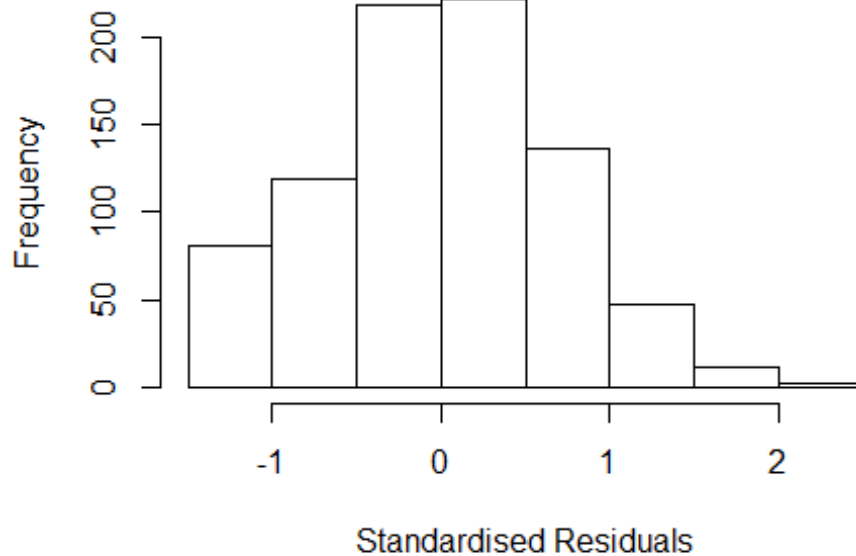
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.307090 -0.479523 0.000127 0.478247 2.237990
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03263 0.15434 6.69 4.1e-11 ***
## FirstAuthorFemale1 -0.09451 0.07766 -1.22 0.22
## LastAuthorFemale1 0.06105 0.07752 0.79 0.43
## Year1997 0.12738 0.22547 0.56 0.57
## Year1998 -0.03136 0.20761 -0.15 0.88
## Year1999 0.09221 0.19169 0.48 0.63
## Year2000 0.07780 0.20115 0.39 0.70
## Year2001 0.00503 0.20237 0.02 0.98
## Year2002 -0.20175 0.18217 -1.11 0.27
## Year2003 0.05526 0.18464 0.30 0.76
## Year2004 0.27446 0.18129 1.51 0.13
## Year2005 0.06489 0.17662 0.37 0.71
```

```

## Year2006          0.14670      0.17987      0.82      0.41
## Year2007          0.09947      0.17344      0.57      0.57
## Year2008          0.17069      0.16779      1.02      0.31
## Year2009          0.07902      0.17577      0.45      0.65
## Year2010         -0.11732      0.17777     -0.66      0.51
## Year2011          0.13221      0.17225      0.77      0.44
## Year2012         -0.00157      0.17954     -0.01      0.99
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.696
## Multiple R-squared:  0.0258, Adjusted R-squared:  0.00436
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 769 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.280  0.874  0.948  0.914  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.20e-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 1.05 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
## -1.311005 -0.473838 -0.000334 0.475725 2.226611
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03967 0.15446 6.73 3.2e-11 ***
## FirstAuthorFemale1 -0.07778 0.07537 -1.03 0.30
## Year1997 0.13172 0.22490 0.59 0.56
## Year1998 -0.03051 0.20848 -0.15 0.88
## Year1999 0.08983 0.19218 0.47 0.64
## Year2000 0.07868 0.20159 0.39 0.70
## Year2001 0.00719 0.20195 0.04 0.97
## Year2002 -0.20425 0.18248 -1.12 0.26
## Year2003 0.05120 0.18440 0.28 0.78
## Year2004 0.27133 0.18144 1.50 0.14
## Year2005 0.06236 0.17660 0.35 0.72
## Year2006 0.14335 0.18010 0.80 0.43
```

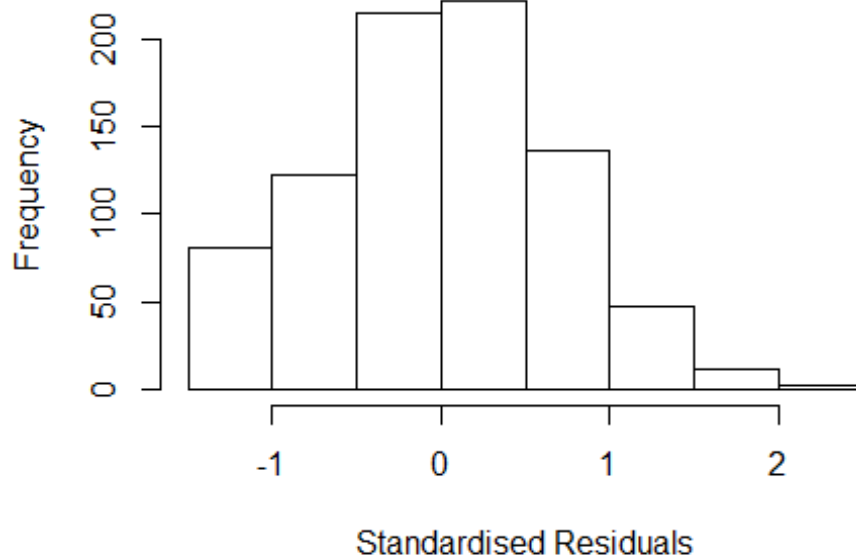


```

## Year2007          0.09498    0.17339    0.55    0.58
## Year2008          0.16580    0.16796    0.99    0.32
## Year2009          0.07680    0.17590    0.44    0.66
## Year2010         -0.12046    0.17809   -0.68    0.50
## Year2011          0.13088    0.17247    0.76    0.45
## Year2012         -0.00265    0.17923   -0.01    0.99
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.695
## Multiple R-squared:  0.0251, Adjusted R-squared:  0.00486
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 771 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.284  0.873  0.948  0.914  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.20e-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 1.071 1          1.035
## Year            1.071 16          1.002

```

Residuals from last author



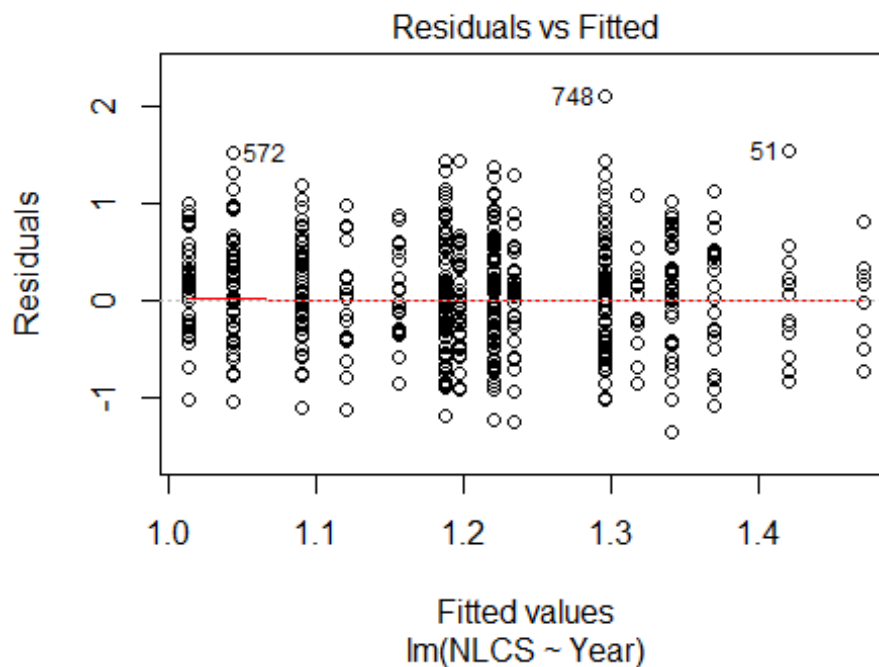
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29745 -0.47651 -0.00184 0.47372 2.23571
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02336 0.15508 6.60 7.4e-11 ***
## LastAuthorFemale1 0.03412 0.07493 0.46 0.65
## Year1997 0.13893 0.22522 0.62 0.54
## Year1998 -0.03000 0.20868 -0.14 0.89
## Year1999 0.09062 0.19281 0.47 0.64
## Year2000 0.07776 0.20167 0.39 0.70
## Year2001 0.00232 0.20289 0.01 0.99
## Year2002 -0.19915 0.18327 -1.09 0.28
## Year2003 0.05830 0.18574 0.31 0.75
## Year2004 0.27409 0.18253 1.50 0.13
## Year2005 0.07115 0.17735 0.40 0.69
## Year2006 0.14610 0.18079 0.81 0.42
```

```

## Year2007      0.09966      0.17452      0.57      0.57
## Year2008      0.16781      0.16903      0.99      0.32
## Year2009      0.08102      0.17632      0.46      0.65
## Year2010     -0.11272      0.17867     -0.63      0.53
## Year2011      0.12710      0.17316      0.73      0.46
## Year2012      0.00118      0.18052      0.01      0.99
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.695
## Multiple R-squared:  0.0241, Adjusted R-squared:  0.00377
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 762 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.280  0.876  0.945  0.913  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.20e-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: 836"
## [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   26   35   31   41   29   48   41   44   55   73  108   82   99   98
## 2011 2012
##  102  102
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8   12   21   18   27   15   32   28   29   43   47   66   39   67   63
## 2011 2012

```

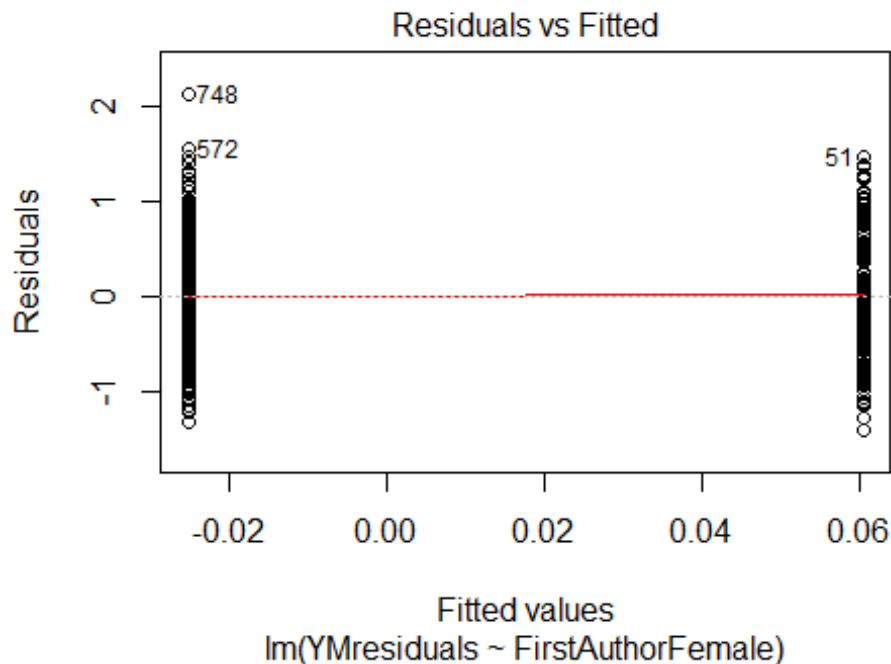
```
## 68 74
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 10 19 16 21 13 26 26 22 35 38 54 34 57 49
## 2011 2012
## 59 57
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6
```



```
##
## 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: 2.09574075705412"
## [1] "Male first author team size 2018 geometric mean: 2.16192013278378"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

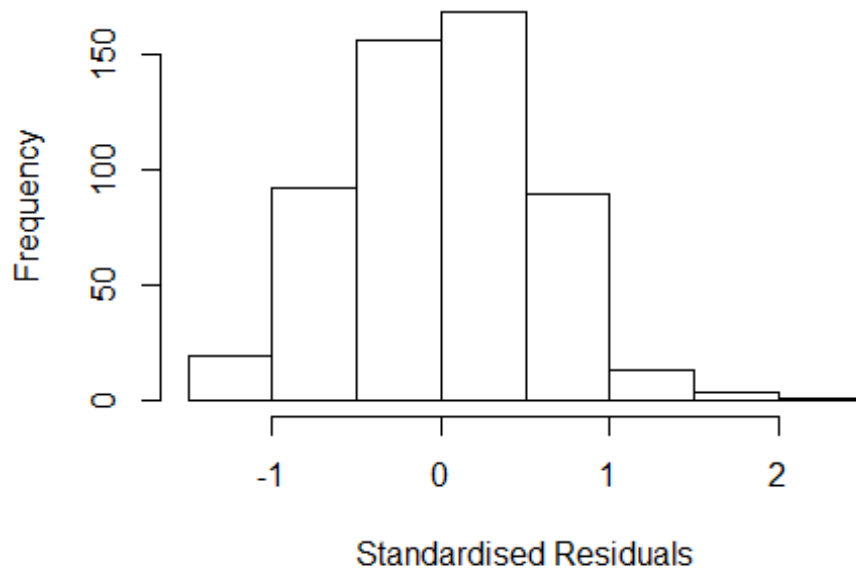
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 430, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.164812151717"
## [1] "Male last author team size 2018 geometric mean: 2.12435407201994"

## Warning in wilcox.test.default(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 = 0.9
## 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 1.461  1      1.209
## LastAuthorFemale  1.520  1      1.233
## UniqueAuthors    1.935  4      1.086
## Year              2.187 16      1.025
```

Residuals from first and last author and team size



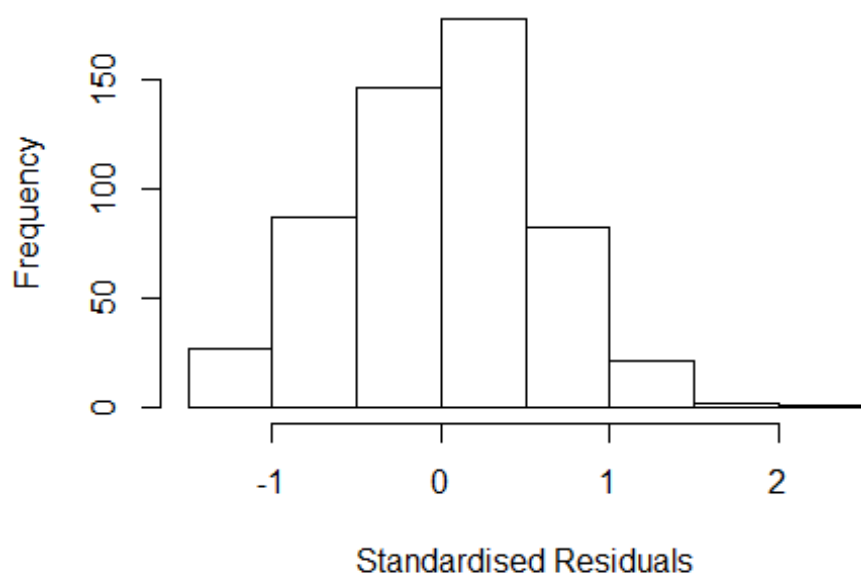
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4479 -0.3881 0.0174 0.3705 2.3770
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4218 0.1517 9.37 < 2e-16 ***
## FirstAuthorFemale1 0.0743 0.0678 1.10 0.2737
## LastAuthorFemale1 0.0345 0.0719 0.48 0.6311
## UniqueAuthors2 0.1316 0.0704 1.87 0.0621 .
## UniqueAuthors3 0.3721 0.0746 4.98 8.5e-07 ***
## UniqueAuthors4 0.2600 0.1150 2.26 0.0242 *
## UniqueAuthors5 0.3004 0.1021 2.94 0.0034 **
## Year1997 -0.2622 0.2613 -1.00 0.3161
## Year1998 -0.2798 0.2182 -1.28 0.2004
## Year1999 -0.4329 0.2204 -1.96 0.0500 *
```

```

## Year2000          -0.4090      0.1774    -2.31    0.0216 *
## Year2001          -0.3111      0.2166    -1.44    0.1515
## Year2002          -0.3537      0.1858    -1.90    0.0576 .
## Year2003          -0.3824      0.1863    -2.05    0.0405 *
## Year2004          -0.4682      0.1805    -2.59    0.0098 **
## Year2005          -0.6581      0.1808    -3.64    0.0003 ***
## Year2006          -0.2339      0.1918    -1.22    0.2234
## Year2007          -0.5566      0.1819    -3.06    0.0023 **
## Year2008          -0.4230      0.1796    -2.36    0.0189 *
## Year2009          -0.4108      0.1726    -2.38    0.0177 *
## Year2010          -0.4484      0.1777    -2.52    0.0119 *
## Year2011          -0.5096      0.1786    -2.85    0.0045 **
## Year2012          -0.4431      0.1773    -2.50    0.0128 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.581
## Multiple R-squared:  0.0938, Adjusted R-squared:  0.0556
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 500 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0564 0.8710 0.9500 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          1.84e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.438 1          1.199
## LastAuthorFemale 1.517 1          1.231
## Year 1.236 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
## -1.4688 -0.3915 0.0386 0.4031 2.2027
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4501 0.1676 8.65 <2e-16 ***
## FirstAuthorFemale1 0.0896 0.0695 1.29 0.198
## LastAuthorFemale1 0.0269 0.0740 0.36 0.716
## Year1997 -0.2015 0.2638 -0.76 0.445
## Year1998 -0.0933 0.2287 -0.41 0.683
## Year1999 -0.3599 0.2288 -1.57 0.116
## Year2000 -0.3557 0.1882 -1.89 0.059 .
## Year2001 -0.1843 0.2180 -0.85 0.398
## Year2002 -0.2318 0.1936 -1.20 0.232
## Year2003 -0.2323 0.1947 -1.19 0.233
## Year2004 -0.3932 0.1927 -2.04 0.042 *
## Year2005 -0.5033 0.1974 -2.55 0.011 *
```

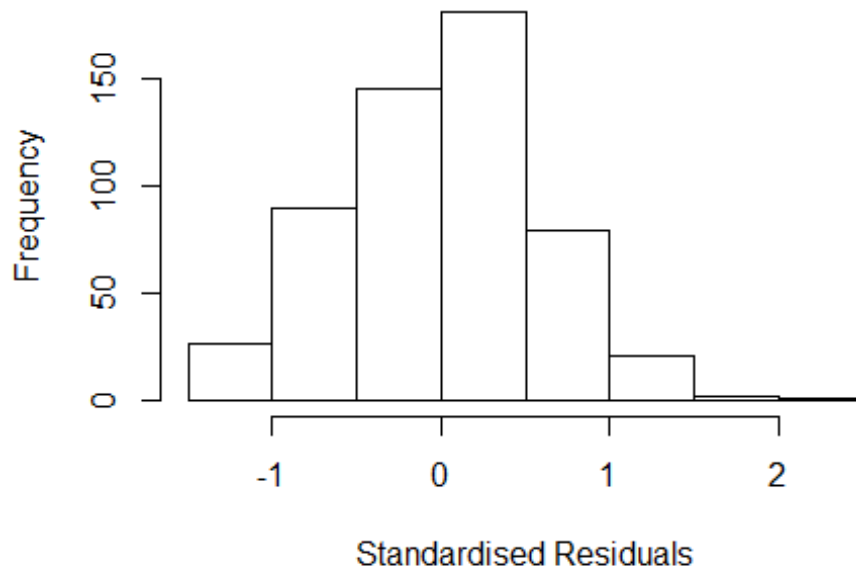


```

## Year2006          -0.0979      0.2002   -0.49    0.625
## Year2007          -0.4562      0.1964   -2.32    0.021 *
## Year2008          -0.2670      0.1899   -1.41    0.160
## Year2009          -0.2648      0.1869   -1.42    0.157
## Year2010          -0.3055      0.1901   -1.61    0.109
## Year2011          -0.3582      0.1876   -1.91    0.057 .
## Year2012          -0.2886      0.1871   -1.54    0.124
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0444, Adjusted R-squared:  0.0116
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 492 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.127  0.862  0.946  0.906  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.84e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.107 1      1.052
## Year              1.107 16      1.003

```

Residuals from first author



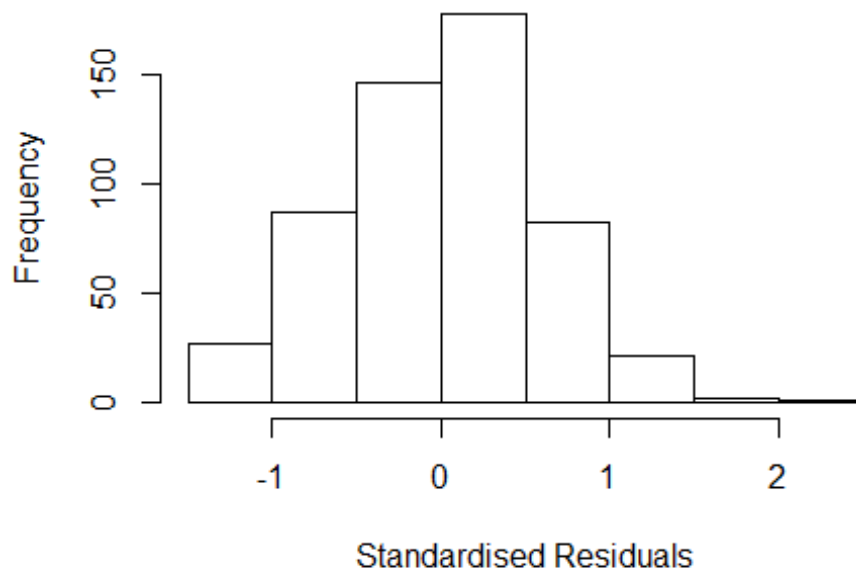
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4553 -0.3948 0.0431 0.3987 2.1984
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4543 0.1702 8.55 <2e-16 ***
## FirstAuthorFemale1 0.1013 0.0612 1.65 0.099 .
## Year1997 -0.1972 0.2653 -0.74 0.458
## Year1998 -0.0928 0.2309 -0.40 0.688
## Year1999 -0.3612 0.2308 -1.56 0.118
## Year2000 -0.3573 0.1904 -1.88 0.061 .
## Year2001 -0.1896 0.2208 -0.86 0.391
## Year2002 -0.2329 0.1958 -1.19 0.235
## Year2003 -0.2306 0.1967 -1.17 0.242
## Year2004 -0.3917 0.1947 -2.01 0.045 *
## Year2005 -0.5036 0.1992 -2.53 0.012 *
## Year2006 -0.1004 0.2019 -0.50 0.619
```

```

## Year2007          -0.4571      0.1984   -2.30    0.022 *
## Year2008          -0.2670      0.1920   -1.39    0.165
## Year2009          -0.2647      0.1890   -1.40    0.162
## Year2010          -0.3077      0.1925   -1.60    0.111
## Year2011          -0.3604      0.1897   -1.90    0.058 .
## Year2012          -0.2890      0.1894   -1.53    0.128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0442, Adjusted R-squared:  0.0133
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 488 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.129  0.861  0.945  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      1.84e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.175 1      1.084
## Year              1.175 16      1.005

```

Residuals from last author



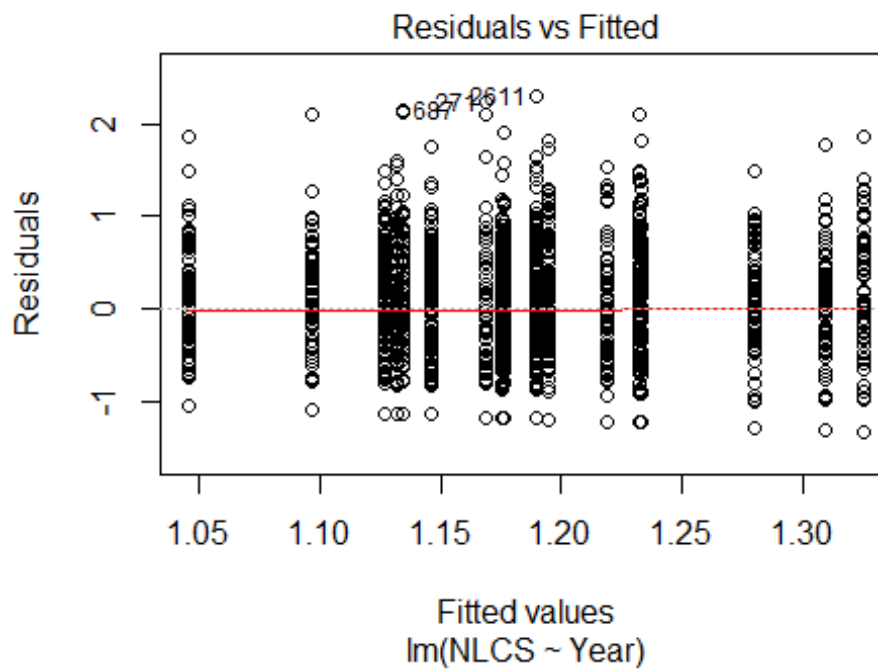
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4435 -0.3876 0.0425 0.3912 2.1886
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4536 0.1658 8.77 <2e-16 ***
## LastAuthorFemale1 0.0689 0.0654 1.05 0.293
## Year1997 -0.1979 0.2596 -0.76 0.446
## Year1998 -0.0853 0.2249 -0.38 0.705
## Year1999 -0.3391 0.2264 -1.50 0.135
## Year2000 -0.3438 0.1879 -1.83 0.068 .
## Year2001 -0.1640 0.2141 -0.77 0.444
## Year2002 -0.2166 0.1930 -1.12 0.262
## Year2003 -0.2130 0.1939 -1.10 0.273
## Year2004 -0.3849 0.1904 -2.02 0.044 *
## Year2005 -0.5039 0.1949 -2.59 0.010 *
## Year2006 -0.0790 0.1981 -0.40 0.690
```

```

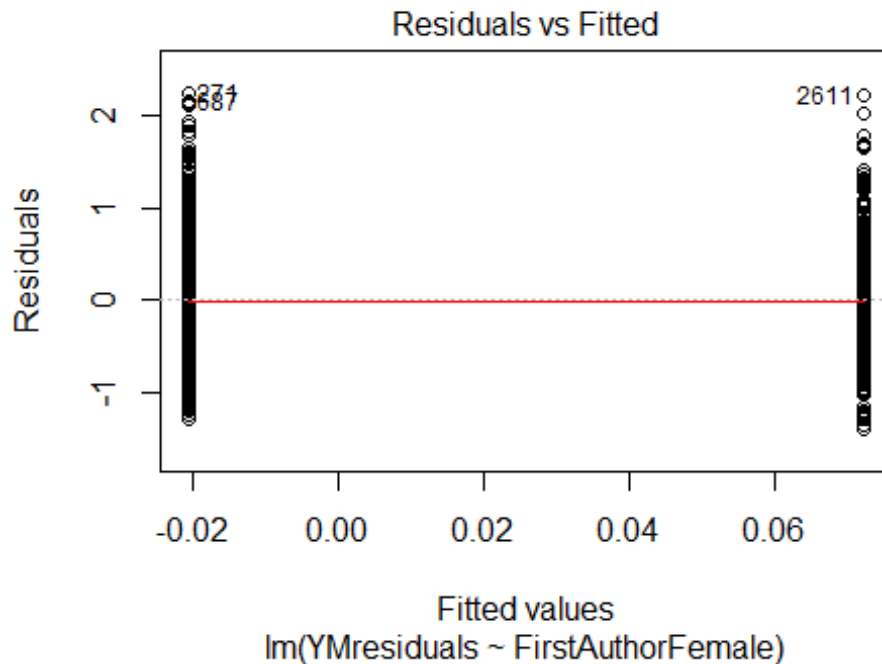
## Year2007          -0.4547      0.1952   -2.33    0.020 *
## Year2008          -0.2560      0.1883   -1.36    0.175
## Year2009          -0.2543      0.1860   -1.37    0.172
## Year2010          -0.2864      0.1897   -1.51    0.132
## Year2011          -0.3491      0.1861   -1.88    0.061 .
## Year2012          -0.2724      0.1862   -1.46    0.144
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0409, Adjusted R-squared:  0.00988
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 49 weights are ~= 1. The remaining 495 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.133  0.865  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      1.84e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 544"
## [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
## 125 125 115 111 122 124 138 123 131 147 212 189 214 257 212
## 2011 2012
## 209 195
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 78 59 72 70 76 78 98 86 86 90 125 126 149 162 131
## 2011 2012

```

```
## 131 136
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 68 46 62 64 73 66 85 73 71 73 102 89 131 131 102
## 2011 2012
## 115 107
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.03
```

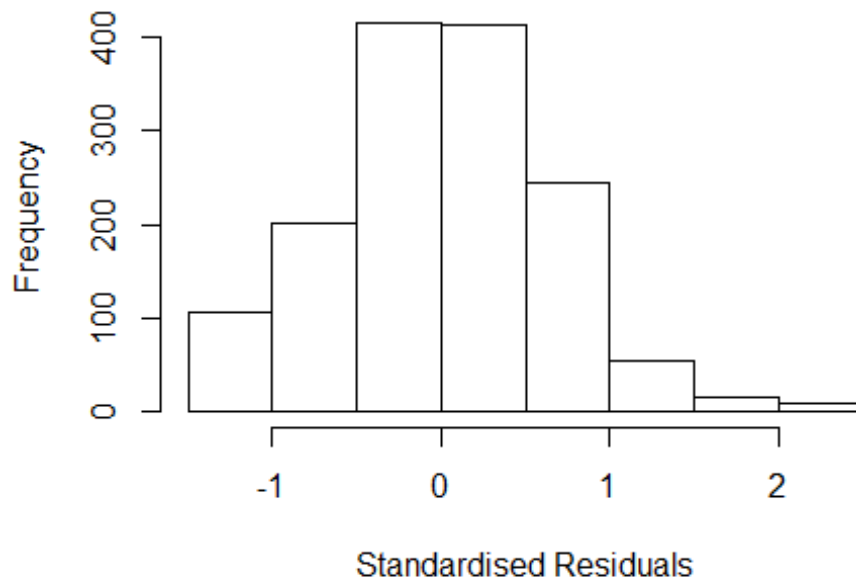


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.6, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.28570791237748"
## [1] "Male first author team size 2018 geometric mean: 2.38829298879865"
##
## 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] "Female last author team size 2018 geometric mean: 1.94125884787121"
## [1] "Male last author team size 2018 geometric mean: 2.50785348994575"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 690, 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 1.442 1          1.201
## LastAuthorFemale  1.456 1          1.206
## UniqueAuthors    1.209 4          1.024
## Year              1.280 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
## -1.46746 -0.43227 0.00782 0.44702 2.32812
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1663 0.0983 11.87 < 2e-16 ***
## FirstAuthorFemale1 0.0849 0.0513 1.65 0.0981 .
## LastAuthorFemale1 0.0270 0.0533 0.51 0.6118
## UniqueAuthors2 0.2060 0.0417 4.93 9.0e-07 ***
## UniqueAuthors3 0.2304 0.0509 4.52 6.6e-06 ***
## UniqueAuthors4 0.2521 0.0847 2.97 0.0030 **
## UniqueAuthors5 0.2743 0.0756 3.63 0.0003 ***
## Year1997 -0.2337 0.1325 -1.76 0.0780 .
## Year1998 -0.0790 0.1304 -0.61 0.5446
## Year1999 -0.1809 0.1233 -1.47 0.1427
```

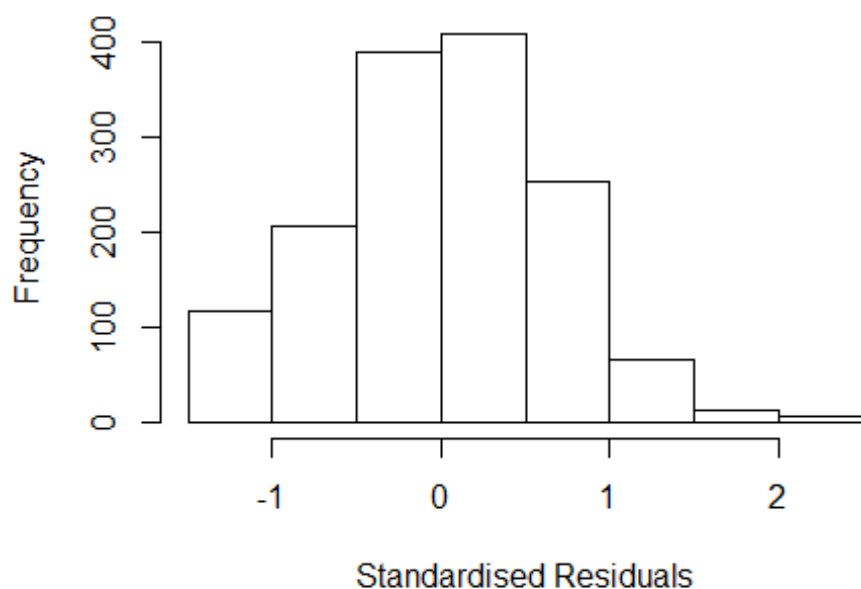


```

## Year2000          -0.2134      0.1281    -1.67    0.0958 .
## Year2001          -0.0148      0.1231    -0.12    0.9043
## Year2002          -0.2304      0.1220    -1.89    0.0590 .
## Year2003          -0.1426      0.1140    -1.25    0.2112
## Year2004          -0.0141      0.1159    -0.12    0.9034
## Year2005          -0.3670      0.1212    -3.03    0.0025 **
## Year2006          -0.0357      0.1203    -0.30    0.7668
## Year2007          -0.1543      0.1196    -1.29    0.1975
## Year2008          -0.1543      0.1046    -1.48    0.1404
## Year2009          -0.1624      0.1128    -1.44    0.1504
## Year2010          -0.1875      0.1141    -1.64    0.1006
## Year2011          -0.1590      0.1155    -1.38    0.1689
## Year2012          -0.2236      0.1138    -1.96    0.0497 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.65
## Multiple R-squared:  0.0437, Adjusted R-squared:  0.0291
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1322 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.173  0.871  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          6.86e-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 1.404 1          1.185
## LastAuthorFemale 1.409 1          1.187
## Year              1.092 16          1.003

```

Residuals from first and last author



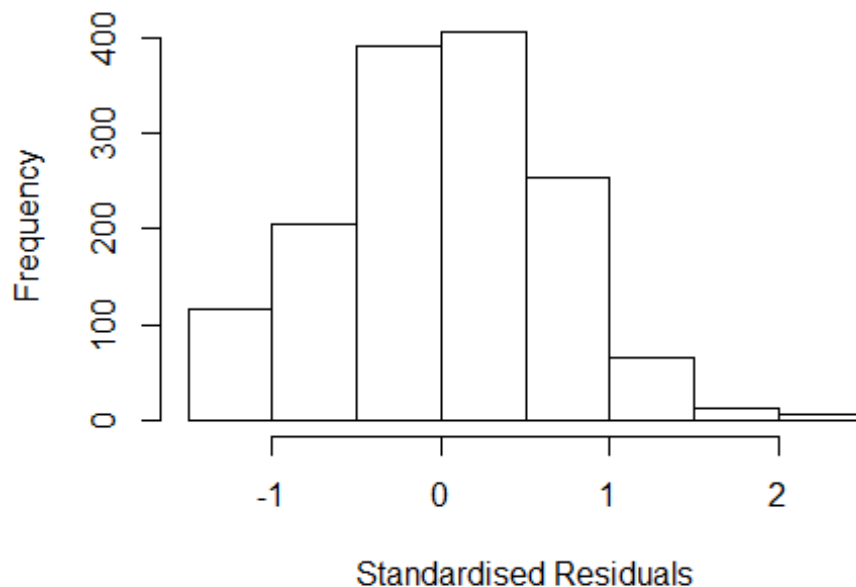
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3840 -0.4393 0.0182 0.4513 2.3708
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29299 0.09755 13.26 <2e-16 ***
## FirstAuthorFemale1 0.07838 0.05094 1.54 0.1241
## LastAuthorFemale1 0.01264 0.05316 0.24 0.8121
## Year1997 -0.26581 0.13956 -1.90 0.0570 .
## Year1998 -0.10594 0.13343 -0.79 0.4274
## Year1999 -0.20380 0.12852 -1.59 0.1130
## Year2000 -0.22205 0.13129 -1.69 0.0910 .
## Year2001 -0.01411 0.12682 -0.11 0.9114
## Year2002 -0.22205 0.12447 -1.78 0.0746 .
## Year2003 -0.13091 0.11758 -1.11 0.2657
## Year2004 -0.00321 0.12058 -0.03 0.9788
## Year2005 -0.33373 0.12479 -2.67 0.0076 **
```

```

## Year2006      -0.02821    0.12280   -0.23    0.8184
## Year2007      -0.10820    0.12307   -0.88    0.3794
## Year2008      -0.12937    0.10803   -1.20    0.2313
## Year2009      -0.13934    0.11673   -1.19    0.2328
## Year2010      -0.15534    0.11740   -1.32    0.1860
## Year2011      -0.11111    0.11884   -0.93    0.3500
## Year2012      -0.19240    0.11790   -1.63    0.1029
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.661
## Multiple R-squared:  0.0187, Adjusted R-squared:  0.00639
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1324 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.171  0.875   0.949   0.908   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.86e-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 1.04 1      1.020
## Year              1.04 16      1.001

```

Residuals from first author



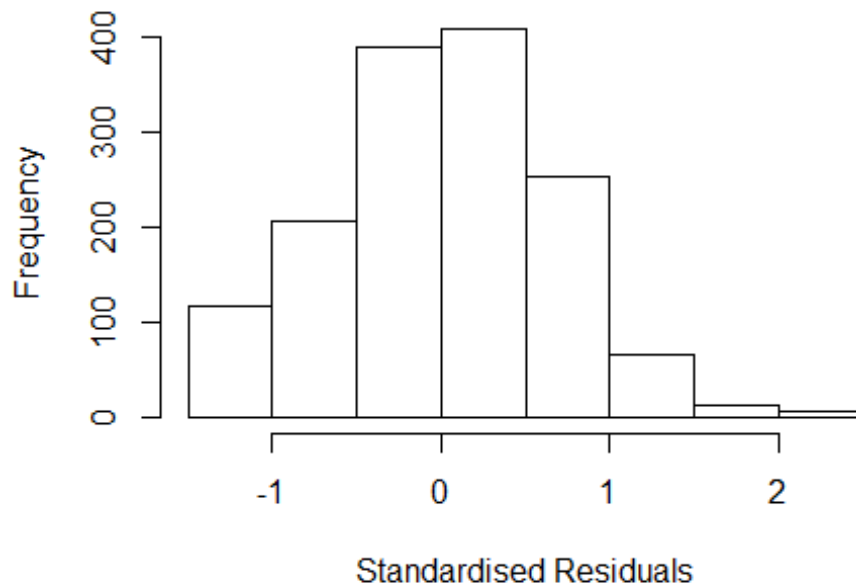
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.379 -0.442 0.018 0.455 2.368
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29389 0.09749 13.27 <2e-16 ***
## FirstAuthorFemale1 0.08483 0.04386 1.93 0.0533 .
## Year1997 -0.26393 0.13925 -1.90 0.0582 .
## Year1998 -0.10543 0.13355 -0.79 0.4300
## Year1999 -0.20268 0.12836 -1.58 0.1146
## Year2000 -0.22084 0.13120 -1.68 0.0926 .
## Year2001 -0.01370 0.12680 -0.11 0.9140
## Year2002 -0.22170 0.12454 -1.78 0.0753 .
## Year2003 -0.13079 0.11758 -1.11 0.2662
## Year2004 -0.00273 0.12048 -0.02 0.9820
## Year2005 -0.33366 0.12477 -2.67 0.0076 **
## Year2006 -0.02852 0.12273 -0.23 0.8163
```

```

## Year2007          -0.10758    0.12300   -0.87    0.3819
## Year2008          -0.12990    0.10801   -1.20    0.2293
## Year2009          -0.13955    0.11677   -1.20    0.2323
## Year2010          -0.15403    0.11710   -1.32    0.1886
## Year2011          -0.11104    0.11886   -0.93    0.3504
## Year2012          -0.19269    0.11790   -1.63    0.1024
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.662
## Multiple R-squared:  0.0186, Adjusted R-squared:  0.007
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1324 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.174  0.875  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      6.86e-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 1.041 1          1.020
## Year            1.041 16          1.001

```

Residuals from last author



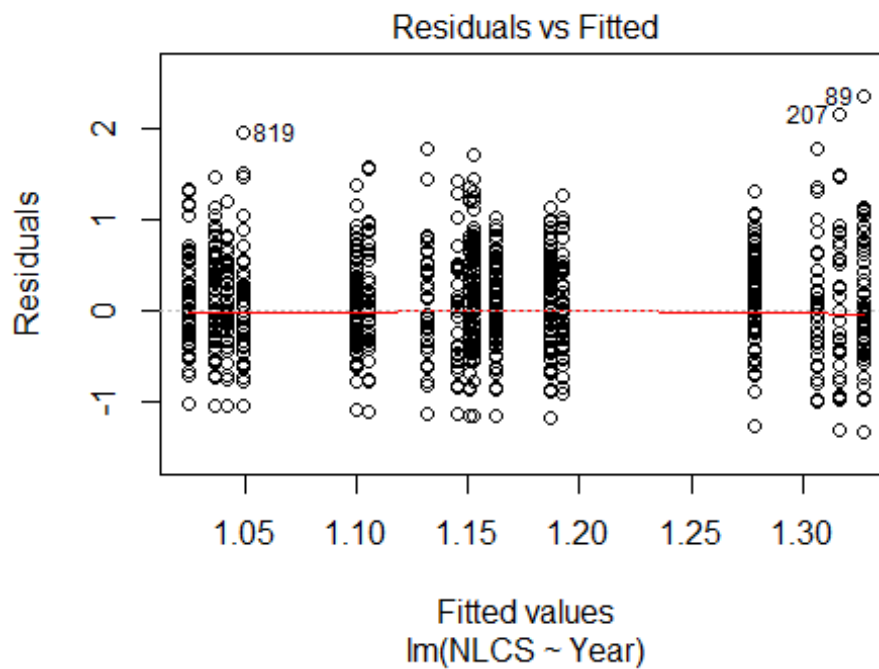
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3519 -0.4450 0.0105 0.4463 2.3715
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29556 0.09734 13.31 <2e-16 ***
## LastAuthorFemale1 0.05499 0.04575 1.20 0.2296
## Year1997 -0.26902 0.13941 -1.93 0.0538 .
## Year1998 -0.09694 0.13232 -0.73 0.4639
## Year1999 -0.20282 0.12800 -1.58 0.1133
## Year2000 -0.21807 0.13101 -1.66 0.0962 .
## Year2001 -0.01144 0.12655 -0.09 0.9280
## Year2002 -0.21229 0.12343 -1.72 0.0857 .
## Year2003 -0.12511 0.11721 -1.07 0.2860
## Year2004 0.00134 0.12061 0.01 0.9911
## Year2005 -0.32744 0.12444 -2.63 0.0086 **
## Year2006 -0.01812 0.12228 -0.15 0.8822
```

```

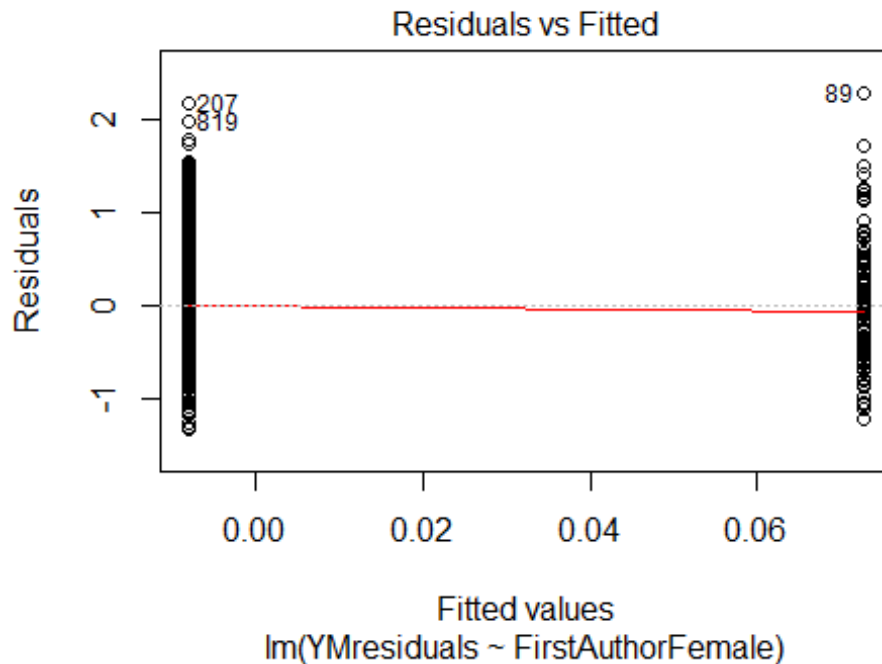
## Year2007          -0.10592      0.12300    -0.86    0.3893
## Year2008          -0.12193      0.10756    -1.13    0.2571
## Year2009          -0.13446      0.11663    -1.15    0.2492
## Year2010          -0.15067      0.11747    -1.28    0.1998
## Year2011          -0.09856      0.11770    -0.84    0.4025
## Year2012          -0.18126      0.11747    -1.54    0.1230
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.661
## Multiple R-squared:  0.017, Adjusted R-squared:  0.0054
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 127 weights are ~= 1. The remaining 1331 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.171  0.876  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      6.86e-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: 1458"
## [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
##  109  92  91  73  76  80  89  91  97  103  146  153  173  156  187
## 2011 2012
##  159  173
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   57  37  46  31  40  38  59  52  53  49  69  88  100  86  93
## 2011 2012

```

```
## 93 100
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 48 26 31 20 25 29 43 39 42 30 44 63 75 70 72
## 2011 2012
## 68 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 = 41, df = 16, p-value = 6e-04
```

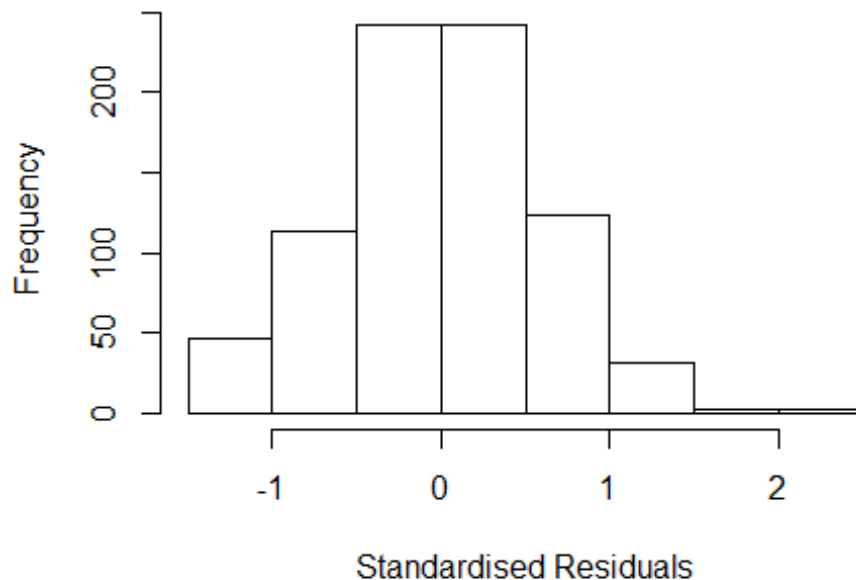


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.1, df = 1, p-value = 0.08
```

```
## [1] "Female first author team size 2018 geometric mean: 2.72774773317531"
## [1] "Male first author team size 2018 geometric mean: 2.47237660638662"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 340, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.74945927399721"
## [1] "Male last author team size 2018 geometric mean: 2.49014210396401"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 200, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.250  1          1.118
## LastAuthorFemale  1.158  1          1.076
## UniqueAuthors    1.648  4          1.064
## Year              1.935 16          1.021
```

Residuals from first and last author and team size



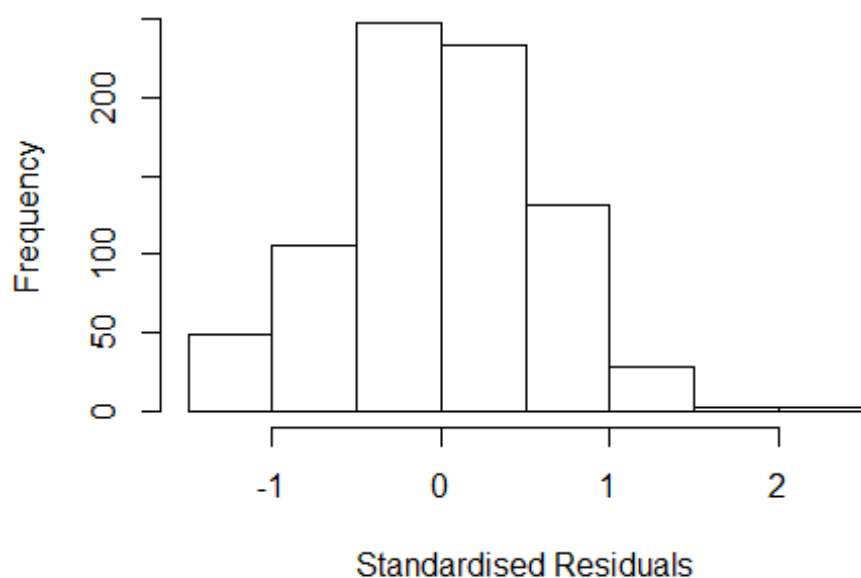
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.43843 -0.38152 -0.00133 0.40413 2.34026
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19602 0.11157 10.72 <2e-16 ***
## FirstAuthorFemale1 -0.00259 0.07898 -0.03 0.974
## LastAuthorFemale1 0.08980 0.10857 0.83 0.408
## UniqueAuthors2 0.13930 0.05868 2.37 0.018 *
## UniqueAuthors3 0.08986 0.06189 1.45 0.147
## UniqueAuthors4 0.31437 0.10153 3.10 0.002 **
## UniqueAuthors5 0.19768 0.15145 1.31 0.192
## Year1997 0.15255 0.21856 0.70 0.485
## Year1998 -0.04350 0.15575 -0.28 0.780
## Year1999 -0.24952 0.14994 -1.66 0.097 .
```

```

## Year2000      -0.13058    0.19316   -0.68    0.499
## Year2001      -0.04543    0.14553   -0.31    0.755
## Year2002      -0.27156    0.13847   -1.96    0.050 .
## Year2003      -0.33673    0.13561   -2.48    0.013 *
## Year2004      -0.15621    0.13509   -1.16    0.248
## Year2005      -0.08019    0.13824   -0.58    0.562
## Year2006      -0.25555    0.13780   -1.85    0.064 .
## Year2007      -0.30916    0.12537   -2.47    0.014 *
## Year2008      -0.19581    0.12077   -1.62    0.105
## Year2009      -0.12387    0.12448   -1.00    0.320
## Year2010      -0.12548    0.12541   -1.00    0.317
## Year2011      -0.06980    0.13311   -0.52    0.600
## Year2012      -0.25352    0.13579   -1.87    0.062 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.588
## Multiple R-squared:  0.0503, Adjusted R-squared:  0.0235
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 738 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0777 0.8800 0.9520 0.9090 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.25e-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 1.195 1 1.093
## LastAuthorFemale 1.126 1 1.061
## Year 1.265 16 1.007

```

Residuals from first and last author



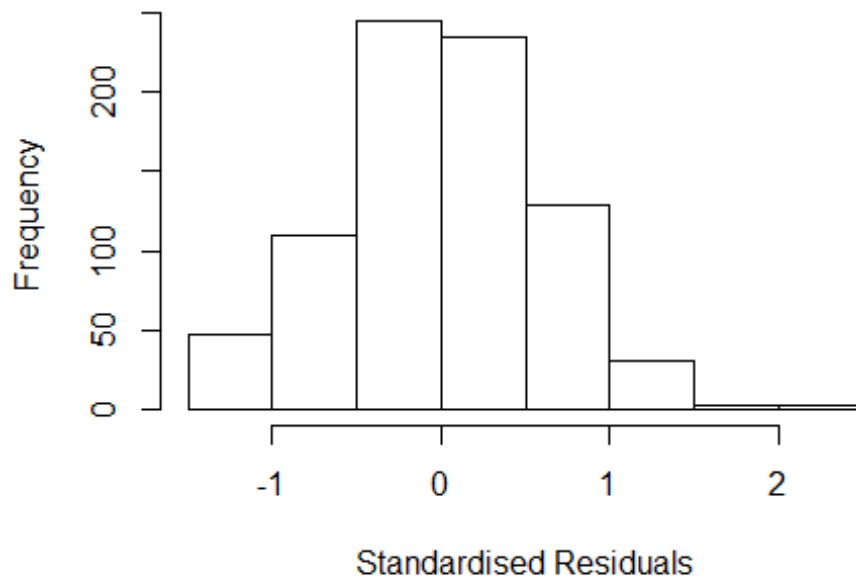
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.44603 -0.38653 -0.00848 0.40318 2.40136
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27448 0.10610 12.01 <2e-16 ***
## FirstAuthorFemale1 -0.00283 0.07792 -0.04 0.971
## LastAuthorFemale1 0.10608 0.11038 0.96 0.337
## Year1997 0.17155 0.22476 0.76 0.446
## Year1998 -0.02079 0.15871 -0.13 0.896
## Year1999 -0.25998 0.15894 -1.64 0.102
## Year2000 -0.13188 0.19348 -0.68 0.496
## Year2001 -0.01759 0.14958 -0.12 0.906
## Year2002 -0.22032 0.13508 -1.63 0.103
## Year2003 -0.31658 0.13848 -2.29 0.023 *
## Year2004 -0.15222 0.13520 -1.13 0.261
## Year2005 -0.05618 0.13941 -0.40 0.687
```

```

## Year2006      -0.24566    0.13989   -1.76    0.079 .
## Year2007      -0.27658    0.12682   -2.18    0.029 *
## Year2008      -0.20095    0.12191   -1.65    0.100 .
## Year2009      -0.08055    0.12571   -0.64    0.522
## Year2010      -0.09085    0.12802   -0.71    0.478
## Year2011      -0.01895    0.13335   -0.14    0.887
## Year2012      -0.19851    0.13684   -1.45    0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.591
## Multiple R-squared:  0.0356, Adjusted R-squared:  0.0134
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 59 weights are ~= 1. The remaining 742 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0614 0.8800 0.9540 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.25e-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 1.163 1      1.079
## Year      1.163 16      1.005

```

Residuals from first author



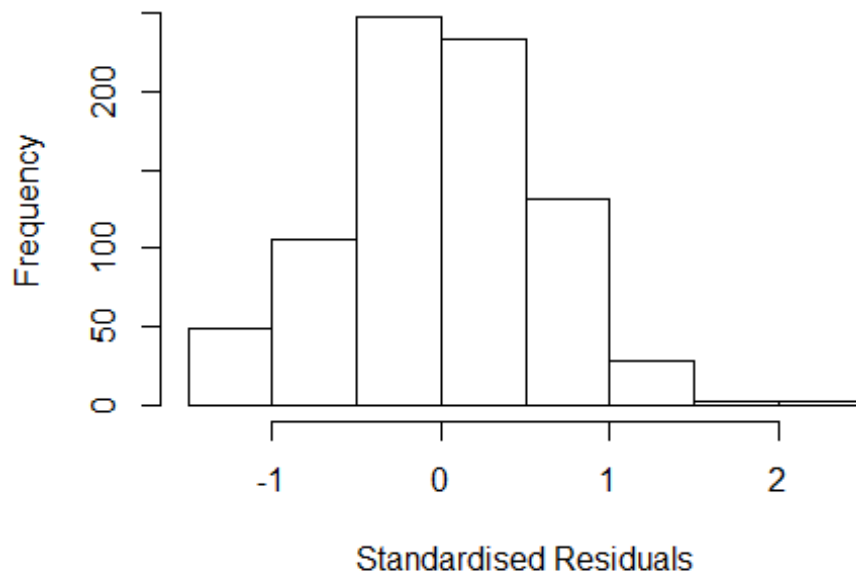
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.45241 -0.38186 -0.00438 0.39932 2.37984
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2820 0.1079 11.88 <2e-16 ***
## FirstAuthorFemale1 0.0112 0.0769 0.15 0.884
## Year1997 0.1705 0.2220 0.77 0.443
## Year1998 -0.0248 0.1579 -0.16 0.875
## Year1999 -0.2525 0.1576 -1.60 0.109
## Year2000 -0.1390 0.1928 -0.72 0.471
## Year2001 -0.0231 0.1510 -0.15 0.879
## Year2002 -0.2289 0.1365 -1.68 0.094 .
## Year2003 -0.3214 0.1392 -2.31 0.021 *
## Year2004 -0.1553 0.1362 -1.14 0.255
## Year2005 -0.0650 0.1407 -0.46 0.644
## Year2006 -0.2493 0.1415 -1.76 0.079 .
```

```

## Year2007          -0.2763      0.1285    -2.15      0.032 *
## Year2008          -0.1995      0.1231    -1.62      0.105
## Year2009          -0.0843      0.1277    -0.66      0.509
## Year2010          -0.0938      0.1296    -0.72      0.469
## Year2011          -0.0186      0.1348    -0.14      0.890
## Year2012          -0.1985      0.1392    -1.43      0.154
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.591
## Multiple R-squared:  0.0337, Adjusted R-squared:  0.0127
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 734 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0685 0.8750 0.9510 0.9080 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      1.25e-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 1.096 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
## -1.44620 -0.38635 -0.00875 0.40320 2.39825
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2747 0.1060 12.02 <2e-16 ***
## LastAuthorFemale1 0.1054 0.1089 0.97 0.333
## Year1997 0.1715 0.2244 0.76 0.445
## Year1998 -0.0212 0.1587 -0.13 0.894
## Year1999 -0.2603 0.1588 -1.64 0.102
## Year2000 -0.1322 0.1933 -0.68 0.494
## Year2001 -0.0182 0.1503 -0.12 0.903
## Year2002 -0.2209 0.1341 -1.65 0.100 .
## Year2003 -0.3170 0.1376 -2.30 0.021 *
## Year2004 -0.1524 0.1354 -1.13 0.261
## Year2005 -0.0567 0.1395 -0.41 0.685
## Year2006 -0.2459 0.1399 -1.76 0.079 .
```

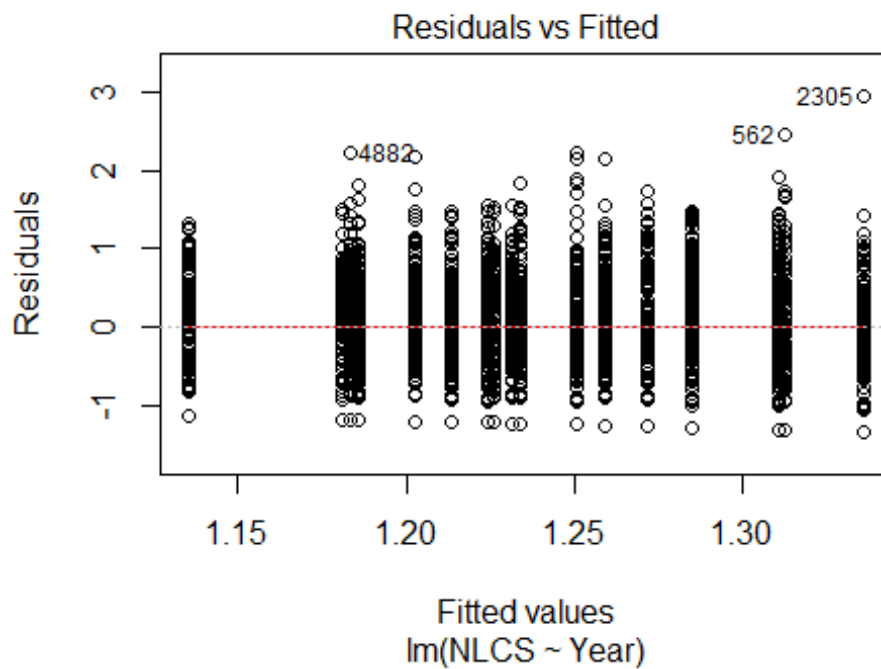


```

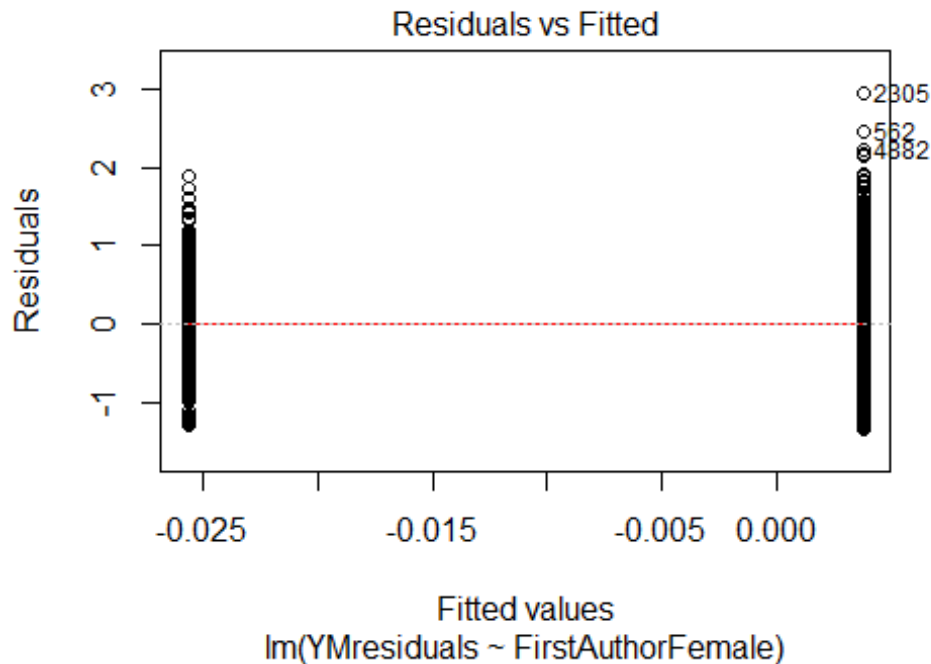
## Year2007          -0.2772      0.1256    -2.21      0.028 *
## Year2008          -0.2014      0.1216    -1.66      0.098 .
## Year2009          -0.0812      0.1247    -0.65      0.515
## Year2010          -0.0915      0.1272    -0.72      0.472
## Year2011          -0.0197      0.1322    -0.15      0.882
## Year2012          -0.1992      0.1359    -1.47      0.143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.593
## Multiple R-squared:  0.0356, Adjusted R-squared:  0.0147
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 59 weights are ~= 1. The remaining 742 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0647 0.8800 0.9540 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      1.25e-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 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
## 224 211 188 202 218 234 258 225 201 294 374 346 382 392 364
## 2011 2012
## 366 397
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 115 110 102 117 114 133 160 127 129 177 226 196 220 234 221
## 2011 2012

```

```
## 239 246
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 96 91 86 95 94 102 133 98 98 135 170 149 186 184 166
## 2011 2012
## 171 189
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 56, df = 16, p-value = 3e-06
```

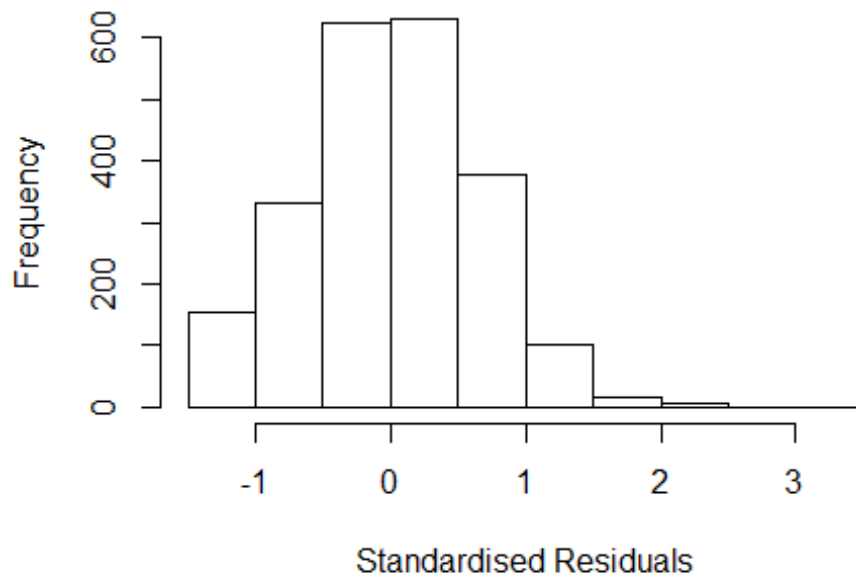


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.011, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 2.87564792441209"
## [1] "Male first author team size 2018 geometric mean: 2.68911945580678"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.61802978029759"
## [1] "Male last author team size 2018 geometric mean: 2.72883162763753"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 680, 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 1.128 1      1.062
## LastAuthorFemale  1.137 1      1.066
## UniqueAuthors    1.270 4      1.030
## Year              1.269 16     1.007
```

Residuals from first and last author and team size



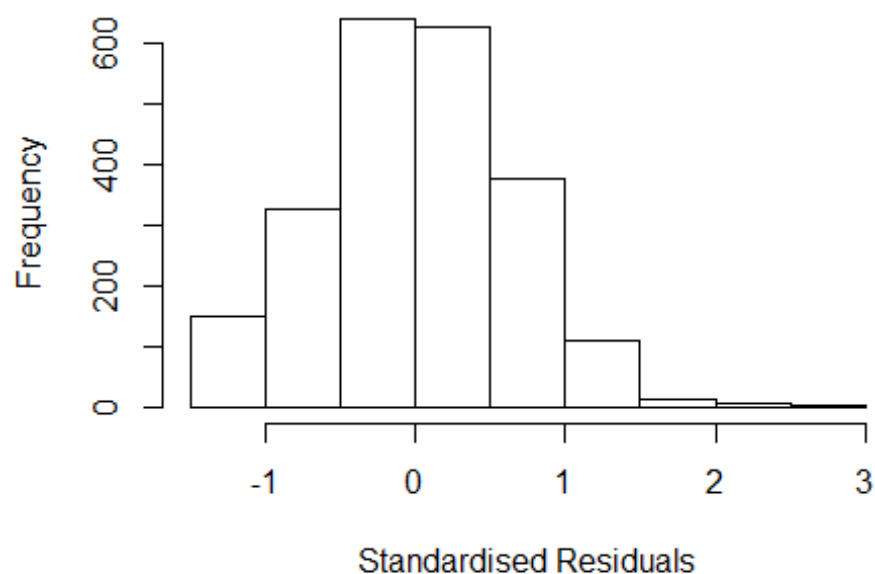
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2305 3042535216 4.284 2004      1702      3      3.051
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4227 -0.4243  0.0072  0.4307  3.0510
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18064    0.08161   14.47  < 2e-16 ***
## FirstAuthorFemale1 -0.07322    0.04301   -1.70  0.0888 .
## LastAuthorFemale1  0.04435    0.04630    0.96  0.3382
## UniqueAuthors2    0.15337    0.03764    4.08 4.8e-05 ***
## UniqueAuthors3    0.24465    0.04152    5.89 4.4e-09 ***
## UniqueAuthors4    0.19365    0.06440    3.01  0.0027 **
## UniqueAuthors5    0.29946    0.06381    4.69 2.9e-06 ***
## Year1997         -0.09626    0.11515   -0.84  0.4033
## Year1998         -0.03468    0.11295   -0.31  0.7589
## Year1999         -0.13018    0.10720   -1.21  0.2247
```

```

## Year2000      -0.12902    0.10517   -1.23    0.2200
## Year2001      -0.08197    0.10134   -0.81    0.4187
## Year2002      -0.14318    0.09991   -1.43    0.1520
## Year2003      -0.14543    0.10311   -1.41    0.1586
## Year2004       0.05235    0.10251    0.51    0.6097
## Year2005      -0.04399    0.10011   -0.44    0.6604
## Year2006      -0.00262    0.09669   -0.03    0.9784
## Year2007      -0.11190    0.09511   -1.18    0.2395
## Year2008      -0.13892    0.09304   -1.49    0.1355
## Year2009      -0.16678    0.09544   -1.75    0.0807 .
## Year2010      -0.13341    0.09413   -1.42    0.1566
## Year2011      -0.04822    0.09068   -0.53    0.5950
## Year2012      -0.13076    0.09101   -1.44    0.1509
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.64
## Multiple R-squared:  0.029, Adjusted R-squared:  0.0194
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 814 is an outlier with |weight| = 0 ( < 4.5e-05);
## 179 weights are ~= 1. The remaining 2063 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.132  0.873  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           4.46e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.103 1 1.050
## LastAuthorFemale 1.118 1 1.057
## Year 1.027 16 1.001

```

Residuals from first and last author



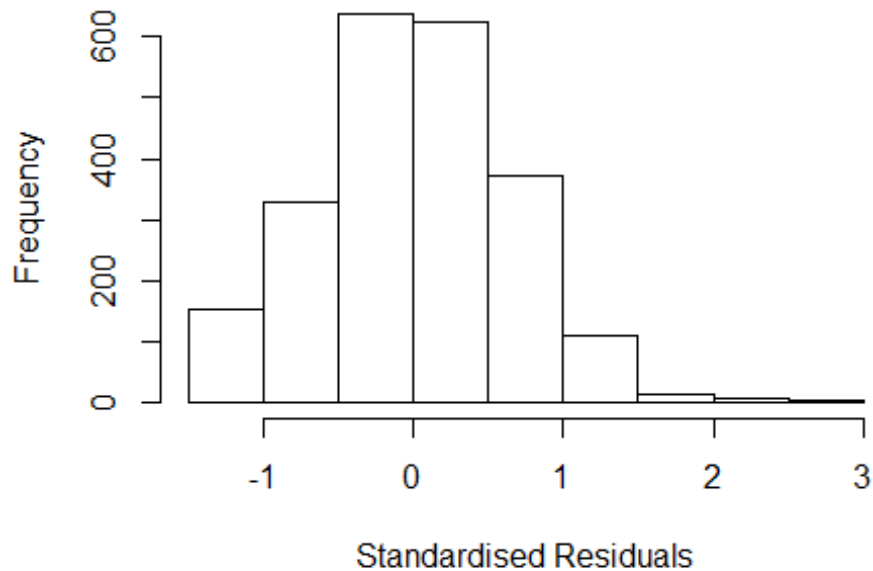
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 562  0031987382 3.783 1998    1706     4    2.522
## 2305 3042535216 4.284 2004    1702     3    2.929
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.35527 -0.42384  0.00256  0.43913  2.92873
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.28629    0.08258   15.58  <2e-16 ***
## FirstAuthorFemale1 -0.06730    0.04267   -1.58    0.11
## LastAuthorFemale1  0.03697    0.04648    0.80    0.43
## Year1997        -0.09088    0.11708   -0.78    0.44
## Year1998        -0.02555    0.11665   -0.22    0.83
## Year1999        -0.13836    0.11051   -1.25    0.21
## Year2000        -0.10469    0.10604   -0.99    0.32
## Year2001        -0.04988    0.10246   -0.49    0.63
## Year2002        -0.10954    0.10221   -1.07    0.28
## Year2003        -0.11575    0.10578   -1.09    0.27
## Year2004         0.06898    0.10567    0.65    0.51
```

```

## Year2005          0.00717    0.10090    0.07    0.94
## Year2006          0.03785    0.09871    0.38    0.70
## Year2007         -0.06866    0.09689   -0.71    0.48
## Year2008         -0.08785    0.09437   -0.93    0.35
## Year2009         -0.10958    0.09719   -1.13    0.26
## Year2010         -0.07659    0.09594   -0.80    0.42
## Year2011          0.00906    0.09196    0.10    0.92
## Year2012         -0.06735    0.09245   -0.73    0.47
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.639
## Multiple R-squared:  0.00887,    Adjusted R-squared:  0.000847
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 184 weights are ~= 1. The remaining 2059 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.002  0.868  0.951  0.908  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.006 1          1.003
## Year              1.006 16          1.000

```

Residuals from first author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 562  0031987382 3.783 1998    1706     4    2.522
## 2305 3042535216 4.284 2004    1702     3    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
## -1.35928 -0.42296  0.00203  0.43900  2.92472
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.28904    0.08264   15.60  <2e-16 ***
## FirstAuthorFemale1 -0.05618    0.04086   -1.38    0.17
## Year1997       -0.09148    0.11731   -0.78    0.44
## Year1998       -0.02432    0.11688   -0.21    0.84
## Year1999       -0.13856    0.11057   -1.25    0.21
## Year2000       -0.10569    0.10623   -0.99    0.32
## Year2001       -0.04978    0.10252   -0.49    0.63
## Year2002       -0.11020    0.10243   -1.08    0.28
## Year2003       -0.11605    0.10591   -1.10    0.27
## Year2004        0.07025    0.10582    0.66    0.51
## Year2005        0.00734    0.10099    0.07    0.94
```

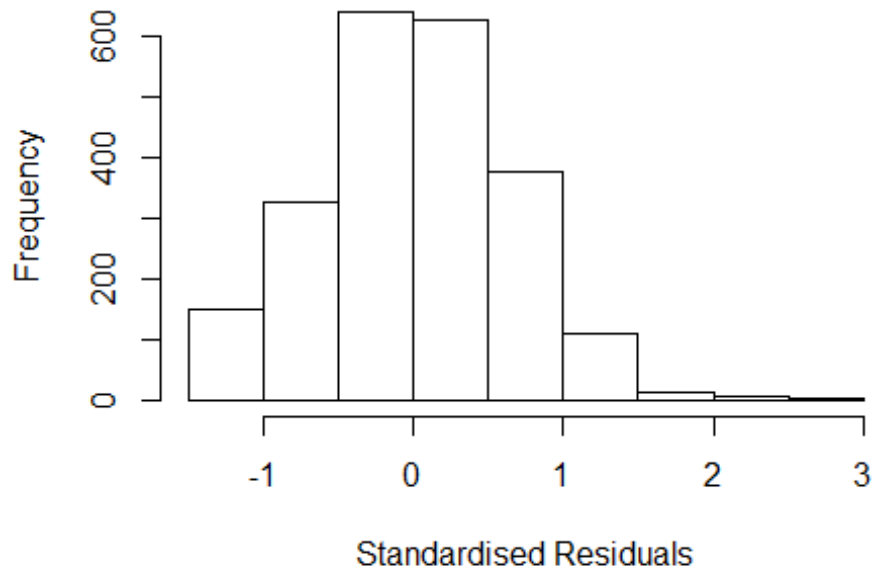


```

## Year2006          0.03791    0.09878    0.38    0.70
## Year2007          -0.06878    0.09700   -0.71    0.48
## Year2008          -0.08707    0.09452   -0.92    0.36
## Year2009          -0.10998    0.09731   -1.13    0.26
## Year2010          -0.07736    0.09601   -0.81    0.42
## Year2011           0.00892    0.09208    0.10    0.92
## Year2012          -0.06848    0.09254   -0.74    0.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.639
## Multiple R-squared:  0.00861,    Adjusted R-squared:  0.00103
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 181 weights are ~= 1. The remaining 2062 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0021 0.8690 0.9520 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      4.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.018 1          1.009
## Year              1.018 16          1.001

```

Residuals from last author



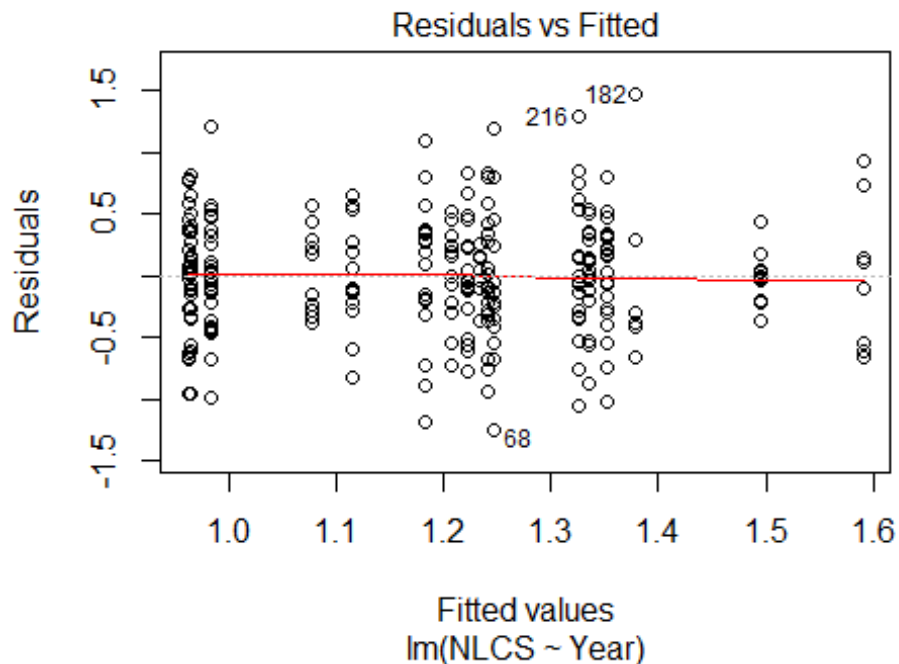
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 562  0031987382 3.783 1998    1706     4    2.522
## 2305 3042535216 4.284 2004    1702     3    2.929
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.34915 -0.42283  0.00272  0.44296  2.93485
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.28175    0.08306   15.43  <2e-16 ***
## LastAuthorFemale1 0.01341    0.04426    0.30   0.76
## Year1997      -0.08937    0.11777   -0.76   0.45
## Year1998      -0.02351    0.11750   -0.20   0.84
## Year1999      -0.14078    0.11135   -1.26   0.21
## Year2000      -0.10581    0.10644   -0.99   0.32
## Year2001      -0.05072    0.10305   -0.49   0.62
## Year2002      -0.11202    0.10257   -1.09   0.27
## Year2003      -0.11546    0.10614   -1.09   0.28
## Year2004       0.06740    0.10593    0.64   0.52
## Year2005       0.00653    0.10133    0.06   0.95
```

```

## Year2006      0.03431      0.09907      0.35      0.73
## Year2007     -0.07169      0.09722     -0.74      0.46
## Year2008     -0.08948      0.09498     -0.94      0.35
## Year2009     -0.11002      0.09776     -1.13      0.26
## Year2010     -0.07908      0.09637     -0.82      0.41
## Year2011      0.00711      0.09247      0.08      0.94
## Year2012     -0.07165      0.09280     -0.77      0.44
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.639
## Multiple R-squared:  0.0078, Adjusted R-squared:  0.000223
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 186 weights are ~= 1. The remaining 2057 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0016 0.8680 0.9510 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      4.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2243"
## [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
##   25   23   21   12   10   17   15   11   11   17   22   28   30   31   29
## 2011 2012
##   39   38
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   21   18   14   11    8   13   12    8    6   11   18   20   17   21   19

```

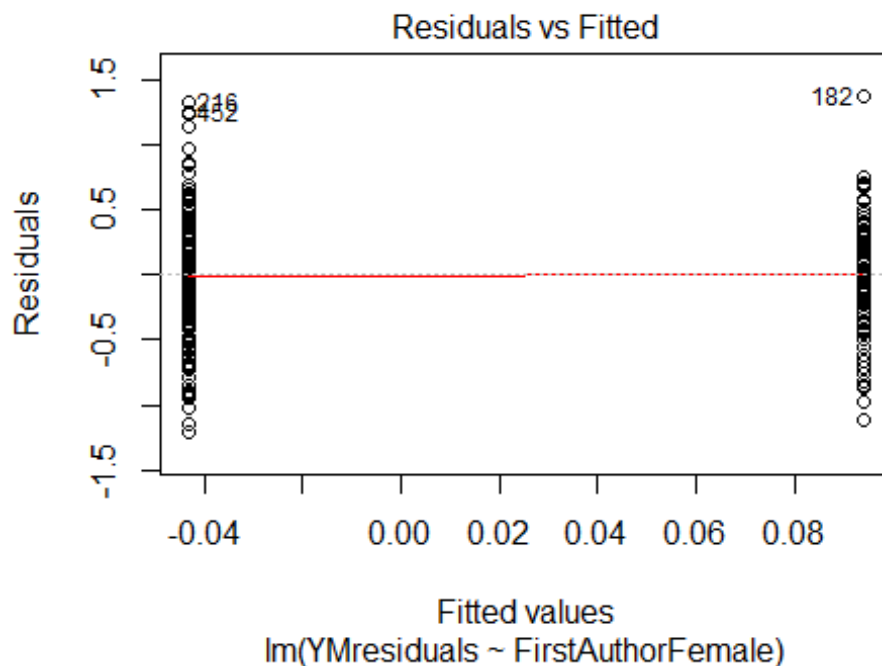
```
## 2011 2012
## 31 31
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 18 18 11 10 8 11 11 7 6 8 14 19 15 19 19
## 2011 2012
## 27 29
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 34, df = 16, p-value = 0.006
```



```
##
## 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: 2.64111740110372"
## [1] "Male first author team size 2018 geometric mean: 1.97747460401501"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

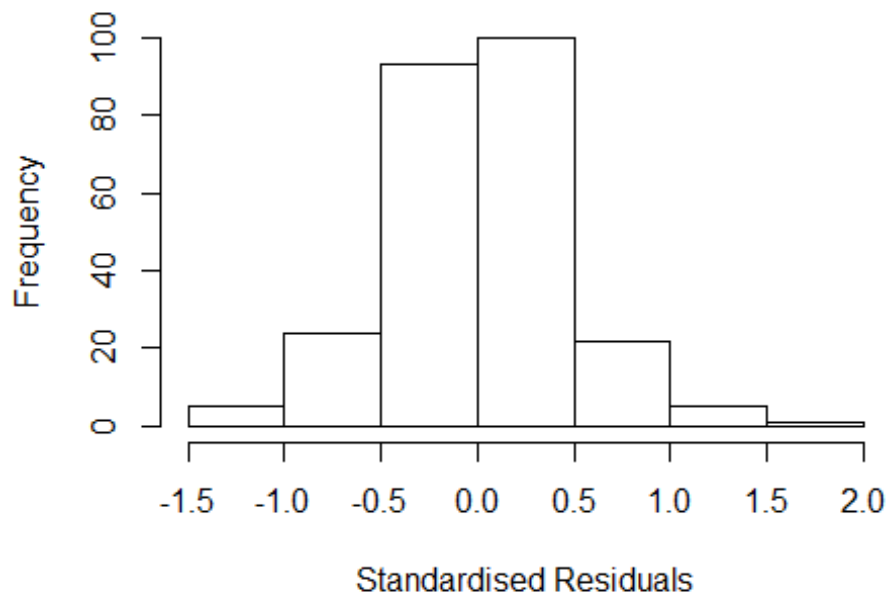
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.39246239770263"
## [1] "Male last author team size 2018 geometric mean: 2.16848670726769"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 90, 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 1.618  1      1.272
## LastAuthorFemale  1.609  1      1.268
## UniqueAuthors    3.256  4      1.159
## Year              4.362 16      1.047
```

Residuals from first and last author and team size



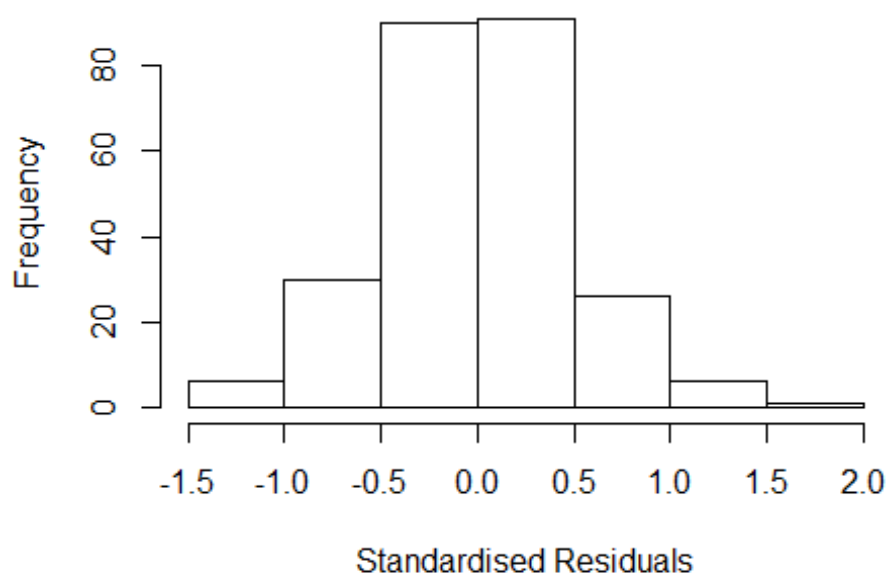
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.44553 -0.25051 0.00404 0.30400 1.57304
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18510 0.14011 8.46 3.3e-15 ***
## FirstAuthorFemale1 0.02843 0.07245 0.39 0.695
## LastAuthorFemale1 0.16727 0.08179 2.05 0.042 *
## UniqueAuthors2 0.09316 0.07889 1.18 0.239
## UniqueAuthors3 0.20087 0.07883 2.55 0.011 *
## UniqueAuthors4 0.39945 0.16266 2.46 0.015 *
## UniqueAuthors5 0.25780 0.12532 2.06 0.041 *
## Year1997 -0.34314 0.19364 -1.77 0.078 .
## Year1998 -0.17162 0.27321 -0.63 0.531
## Year1999 -0.17505 0.18517 -0.95 0.345
```

```

## Year2000      0.27610      0.33045      0.84      0.404
## Year2001     -0.13769      0.20138     -0.68      0.495
## Year2002     -0.04631      0.17029     -0.27      0.786
## Year2003     -0.10656      0.17052     -0.62      0.533
## Year2004     -0.10484      0.28346     -0.37      0.712
## Year2005      0.09817      0.17655      0.56      0.579
## Year2006      0.04134      0.22451      0.18      0.854
## Year2007      0.04438      0.17918      0.25      0.805
## Year2008      0.00498      0.17304      0.03      0.977
## Year2009     -0.15862      0.17515     -0.91      0.366
## Year2010     -0.13105      0.17301     -0.76      0.450
## Year2011     -0.32359      0.15923     -2.03      0.043 *
## Year2012     -0.39542      0.16761     -2.36      0.019 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.453
## Multiple R-squared:  0.179, Adjusted R-squared:  0.0992
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 220 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.203  0.888  0.948  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      4.00e-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 1.341 1      1.158
## LastAuthorFemale  1.643 1      1.282
## Year              1.627 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
## -1.37437 -0.28412 -0.00579 0.29628 1.50040
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2303 0.1475 8.34 6.8e-15 ***
## FirstAuthorFemale1 0.0482 0.0703 0.69 0.493
## LastAuthorFemale1 0.1440 0.0854 1.69 0.093 .
## Year1997 -0.3117 0.1978 -1.58 0.116
## Year1998 -0.1747 0.2682 -0.65 0.515
## Year1999 -0.1504 0.1929 -0.78 0.436
## Year2000 0.2396 0.3219 0.74 0.457
## Year2001 -0.1009 0.2073 -0.49 0.627
## Year2002 0.0268 0.1740 0.15 0.878
## Year2003 -0.0162 0.1626 -0.10 0.921
## Year2004 -0.0740 0.2745 -0.27 0.788
## Year2005 0.2464 0.1658 1.49 0.139
```

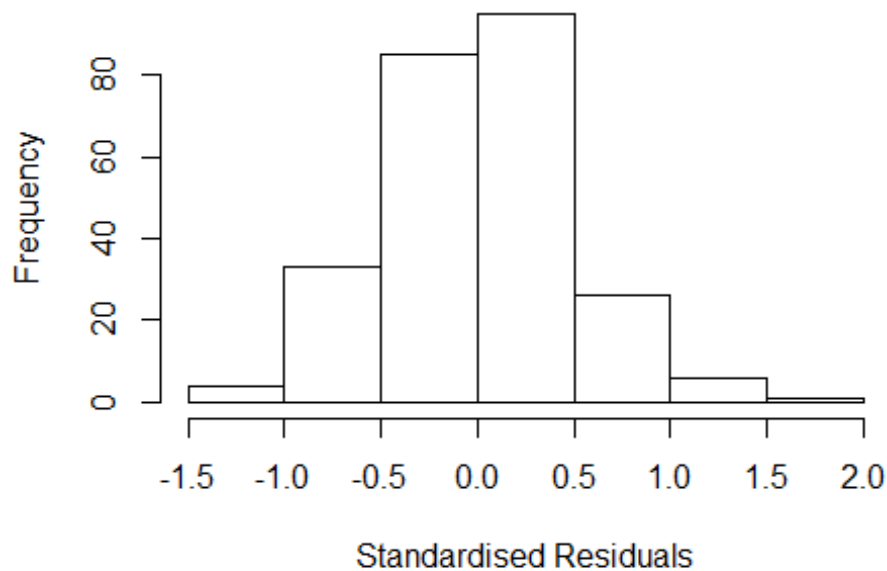


```

## Year2006          0.1586      0.2515      0.63      0.529
## Year2007          0.1097      0.1826      0.60      0.549
## Year2008          0.0360      0.1795      0.20      0.841
## Year2009         -0.0563      0.1808     -0.31      0.756
## Year2010         -0.0715      0.1819     -0.39      0.694
## Year2011         -0.2804      0.1701     -1.65      0.101
## Year2012         -0.3037      0.1727     -1.76      0.080 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.452
## Multiple R-squared:  0.144, Adjusted R-squared:  0.0776
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 229 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.247  0.880  0.956  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      4.00e-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 1.182 1      1.087
## Year              1.182 16      1.005

```

Residuals from first author



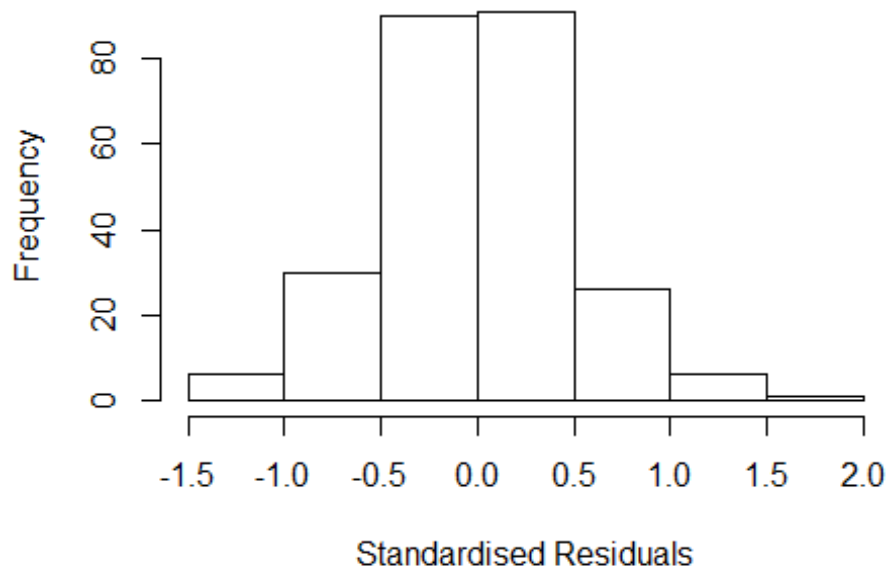
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2364 -0.3014 0.0148 0.3053 1.5891
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2364 0.1413 8.75 4.5e-16 ***
## FirstAuthorFemale1 0.1046 0.0668 1.57 0.119
## Year1997 -0.3000 0.1929 -1.56 0.121
## Year1998 -0.1425 0.2628 -0.54 0.588
## Year1999 -0.1604 0.1873 -0.86 0.393
## Year2000 0.2741 0.2983 0.92 0.359
## Year2001 -0.0868 0.2071 -0.42 0.675
## Year2002 0.0216 0.1715 0.13 0.900
## Year2003 -0.0177 0.1554 -0.11 0.909
## Year2004 -0.0811 0.2695 -0.30 0.764
## Year2005 0.2499 0.1559 1.60 0.110
## Year2006 0.1587 0.2344 0.68 0.499
```

```

## Year2007          0.1113      0.1777      0.63      0.531
## Year2008          0.0698      0.1701      0.41      0.682
## Year2009         -0.0398      0.1735     -0.23      0.819
## Year2010         -0.0592      0.1755     -0.34      0.736
## Year2011         -0.2686      0.1635     -1.64      0.102
## Year2012         -0.2958      0.1663     -1.78      0.077 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.47
## Multiple R-squared:  0.129, Adjusted R-squared:  0.0648
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 227 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.230  0.886  0.954  0.906  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.00e-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 1.421 1      1.192
## Year              1.421 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
## -1.40502 -0.27221 0.00305 0.28515 1.53356
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2383 0.1444 8.58 1.4e-15 ***
## LastAuthorFemale1 0.1667 0.0798 2.09 0.038 *
## Year1997 -0.3049 0.1972 -1.55 0.123
## Year1998 -0.1884 0.2655 -0.71 0.479
## Year1999 -0.1462 0.1888 -0.77 0.440
## Year2000 0.2412 0.3162 0.76 0.446
## Year2001 -0.1077 0.2036 -0.53 0.597
## Year2002 0.0264 0.1715 0.15 0.878
## Year2003 -0.0133 0.1609 -0.08 0.934
## Year2004 -0.0896 0.2674 -0.33 0.738
## Year2005 0.2563 0.1646 1.56 0.121
## Year2006 0.1598 0.2501 0.64 0.524
```

```

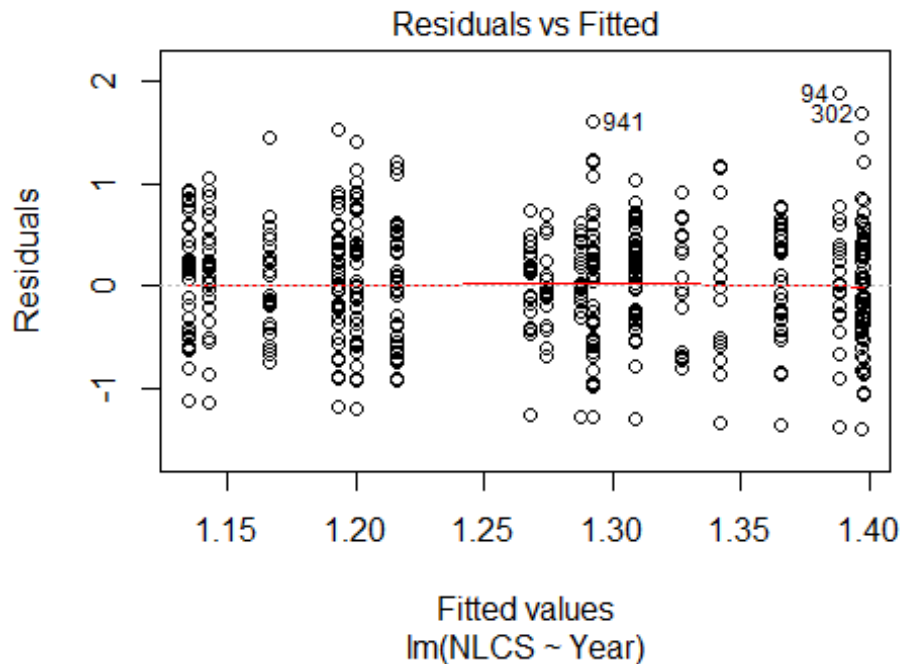
## Year2007          0.1143      0.1789      0.64      0.523
## Year2008          0.0444      0.1790      0.25      0.804
## Year2009         -0.0581      0.1805     -0.32      0.748
## Year2010         -0.0655      0.1787     -0.37      0.714
## Year2011         -0.2758      0.1680     -1.64      0.102
## Year2012         -0.2994      0.1704     -1.76      0.080 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.143, Adjusted R-squared:  0.0803
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 224 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.221  0.870   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      4.00e-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: 250"
## [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 2007 2009 2011
##    1    1    1    2
##
## 2006 2007 2009 2011
##    0    0    1    2
##
## 2006 2007 2009 2011
##    0    0    1    1

```

```

## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.44948974278318"
## [1] "Male first author team size 2018 geometric mean: NaN"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
## [1] "Regression 3: First author gender, Year as factors"
## [1] "Regression 4: Last author gender, Year as factors"
## [1] "Sample size for the above analysis: 2"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
## 34 38 51 33 33 39 37 25 40 31 79 76 68 67 63
## 2011 2012
## 62 50
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 25 22 30 20 17 31 23 17 31 22 58 48 44 44 44
## 2011 2012
## 45 38
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 22 19 29 16 13 30 22 14 30 20 47 42 37 37 40
## 2011 2012
## 39 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 = 12, df = 16, p-value = 0.7

```



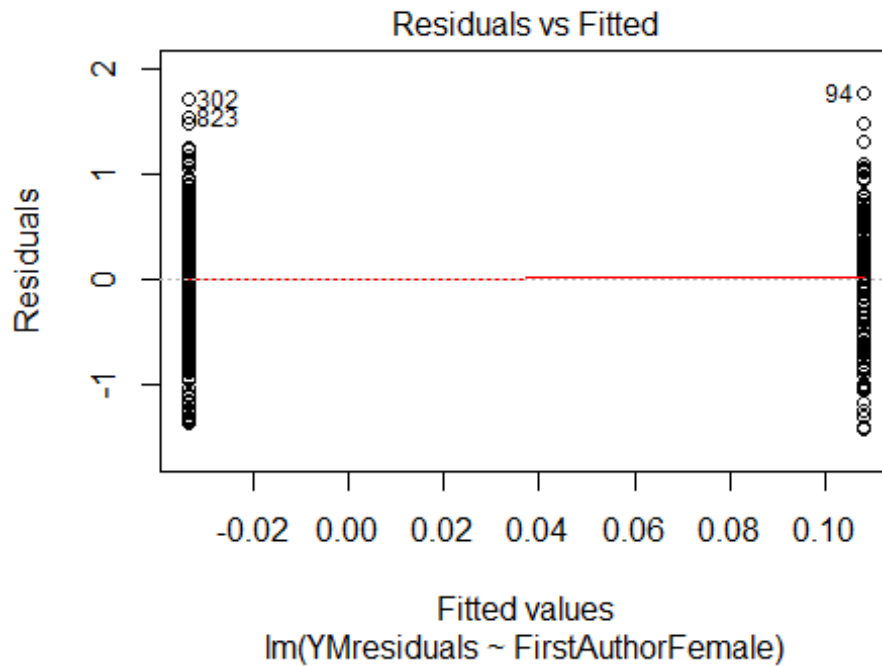
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.2, df = 1, p-value = 0.02

## [1] "Female first author team size 2018 geometric mean: 3.17397468794681"
## [1] "Male first author team size 2018 geometric mean: 1.69673094542681"

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

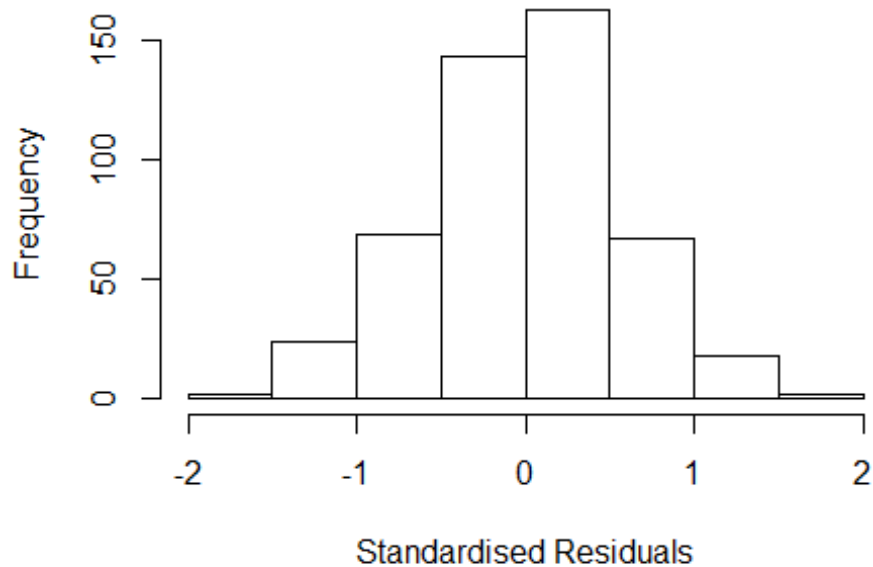
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 270, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.73053814226843"
## [1] "Male last author team size 2018 geometric mean: 1.90671685891466"

## Warning in wilcox.test.default(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 = 0.1
## 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 1.298 1          1.139
## LastAuthorFemale  1.253 1          1.119
## UniqueAuthors    1.839 4          1.079
## Year              2.284 16         1.026
```


Residuals from first and last author and team size



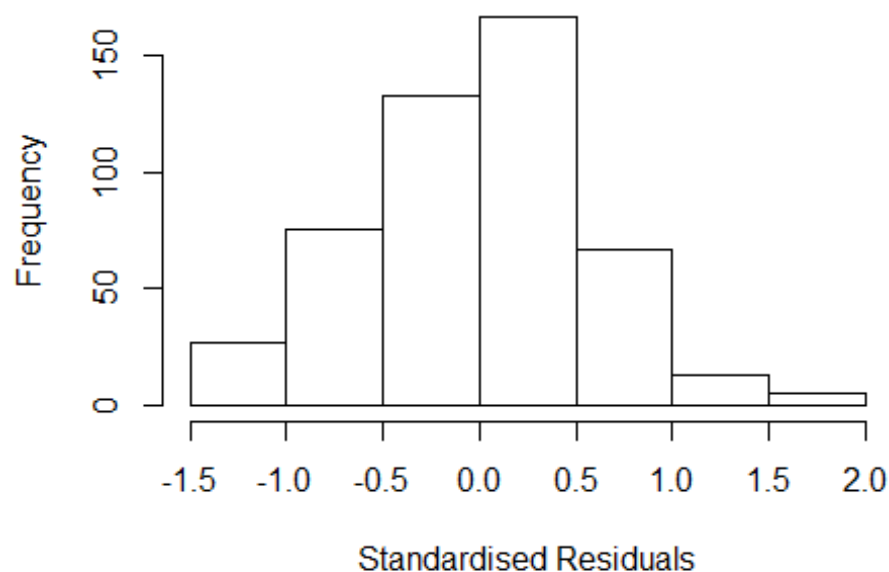
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5511 -0.3820 0.0176 0.3450 1.6380
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99760 0.12764 7.82 3.7e-14 ***
## FirstAuthorFemale1 0.10269 0.07609 1.35 0.178
## LastAuthorFemale1 -0.00160 0.08521 -0.02 0.985
## UniqueAuthors2 0.28444 0.06706 4.24 2.7e-05 ***
## UniqueAuthors3 0.39100 0.07716 5.07 5.8e-07 ***
## UniqueAuthors4 0.30792 0.12200 2.52 0.012 *
## UniqueAuthors5 -0.33346 0.21761 -1.53 0.126
## Year1997 0.22577 0.17908 1.26 0.208
## Year1998 0.15096 0.16029 0.94 0.347
## Year1999 0.12045 0.16533 0.73 0.467
```

```

## Year2000      0.09650    0.22782    0.42    0.672
## Year2001      0.12888    0.16069    0.80    0.423
## Year2002      0.14580    0.15508    0.94    0.348
## Year2003      0.07552    0.23262    0.32    0.746
## Year2004     -0.01894    0.15803   -0.12    0.905
## Year2005     -0.01190    0.15199   -0.08    0.938
## Year2006     -0.00332    0.15805   -0.02    0.983
## Year2007      0.03101    0.15466    0.20    0.841
## Year2008     -0.03130    0.16404   -0.19    0.849
## Year2009     -0.04824    0.16239   -0.30    0.767
## Year2010      0.05979    0.16646    0.36    0.720
## Year2011     -0.05591    0.15252   -0.37    0.714
## Year2012      0.16978    0.16400    1.04    0.301
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.541
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0592
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 442 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.340  0.864  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      2.05e-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 1.331 1      1.154
## LastAuthorFemale  1.350 1      1.162
## Year              1.407 16      1.011

```

Residuals from first and last author



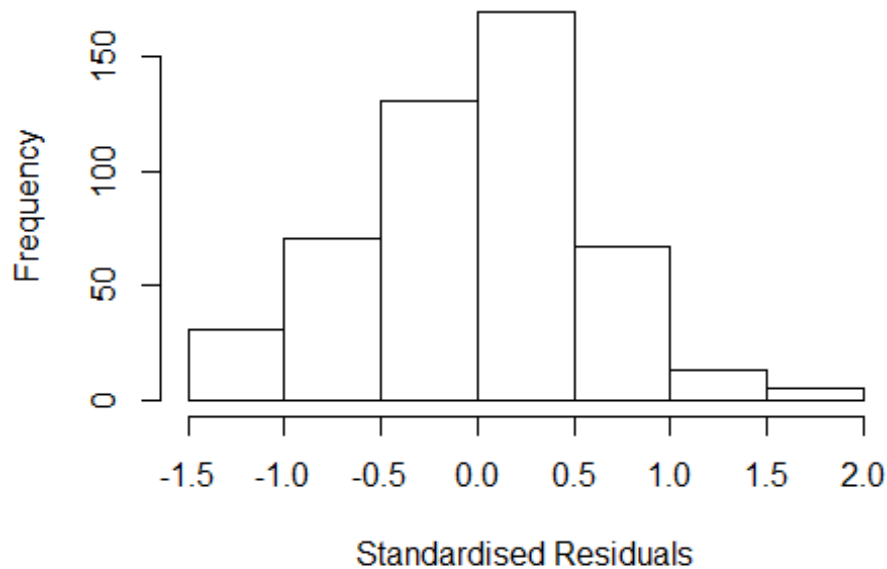
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.383 -0.409 0.029 0.385 1.739
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1642 0.1157 10.06 <2e-16 ***
## FirstAuthorFemale1 0.1414 0.0791 1.79 0.074 .
## LastAuthorFemale1 -0.0548 0.0942 -0.58 0.561
## Year1997 0.2416 0.1790 1.35 0.178
## Year1998 0.1775 0.1544 1.15 0.251
## Year1999 0.0498 0.1572 0.32 0.752
## Year2000 0.1750 0.1897 0.92 0.357
## Year2001 0.1837 0.1597 1.15 0.250
## Year2002 0.1604 0.1505 1.07 0.287
## Year2003 0.0557 0.2365 0.24 0.814
## Year2004 -0.0162 0.1658 -0.10 0.922
## Year2005 0.1248 0.1438 0.87 0.386
```

```

## Year2006          0.0543      0.1497      0.36      0.717
## Year2007          0.1324      0.1477      0.90      0.371
## Year2008          0.0152      0.1555      0.10      0.922
## Year2009          0.0341      0.1582      0.22      0.829
## Year2010          0.0773      0.1664      0.46      0.642
## Year2011         -0.0124      0.1507     -0.08      0.934
## Year2012          0.2537      0.1598      1.59      0.113
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.58
## Multiple R-squared:  0.0298, Adjusted R-squared:  -0.00747
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 445 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.348  0.876  0.949  0.904  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.05e-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 1.152 1      1.073
## Year              1.152 16      1.004

```

Residuals from first author



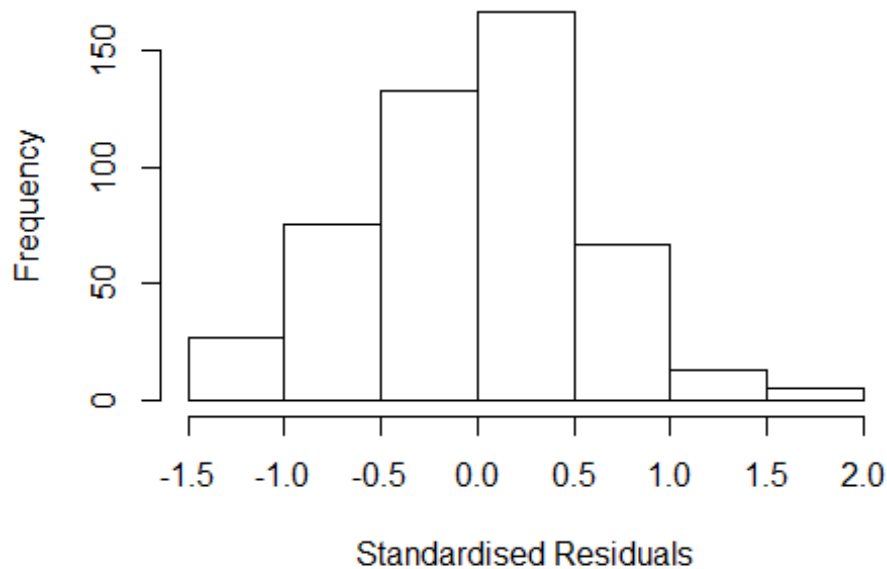
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4147 -0.4008 0.0304 0.3862 1.7466
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1622 0.1162 10.00 <2e-16 ***
## FirstAuthorFemale1 0.1234 0.0743 1.66 0.097 .
## Year1997 0.2397 0.1775 1.35 0.177
## Year1998 0.1674 0.1526 1.10 0.273
## Year1999 0.0495 0.1581 0.31 0.755
## Year2000 0.1676 0.1873 0.89 0.371
## Year2001 0.1848 0.1601 1.15 0.249
## Year2002 0.1532 0.1498 1.02 0.307
## Year2003 0.0445 0.2344 0.19 0.849
## Year2004 -0.0196 0.1652 -0.12 0.906
## Year2005 0.1263 0.1439 0.88 0.380
## Year2006 0.0503 0.1494 0.34 0.737
```

```

## Year2007          0.1290      0.1475      0.87      0.382
## Year2008          0.0127      0.1562      0.08      0.935
## Year2009          0.0339      0.1575      0.22      0.830
## Year2010          0.0686      0.1671      0.41      0.681
## Year2011         -0.0169      0.1508     -0.11      0.911
## Year2012          0.2530      0.1601      1.58      0.115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.592
## Multiple R-squared:  0.0285, Adjusted R-squared:  -0.00662
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 446 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.364  0.883  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      2.05e-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 1.165 1      1.079
## Year              1.165 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
## -1.3681 -0.4159 0.0455 0.3840 1.8410
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19010 0.11150 10.67 <2e-16 ***
## LastAuthorFemale1 0.00354 0.08425 0.04 0.966
## Year1997 0.24090 0.17416 1.38 0.167
## Year1998 0.16830 0.15127 1.11 0.266
## Year1999 0.04741 0.15991 0.30 0.767
## Year2000 0.16990 0.19231 0.88 0.377
## Year2001 0.17801 0.15621 1.14 0.255
## Year2002 0.14503 0.14850 0.98 0.329
## Year2003 0.02903 0.23916 0.12 0.903
## Year2004 -0.02300 0.16283 -0.14 0.888
## Year2005 0.10898 0.14327 0.76 0.447
## Year2006 0.05120 0.14862 0.34 0.731
```

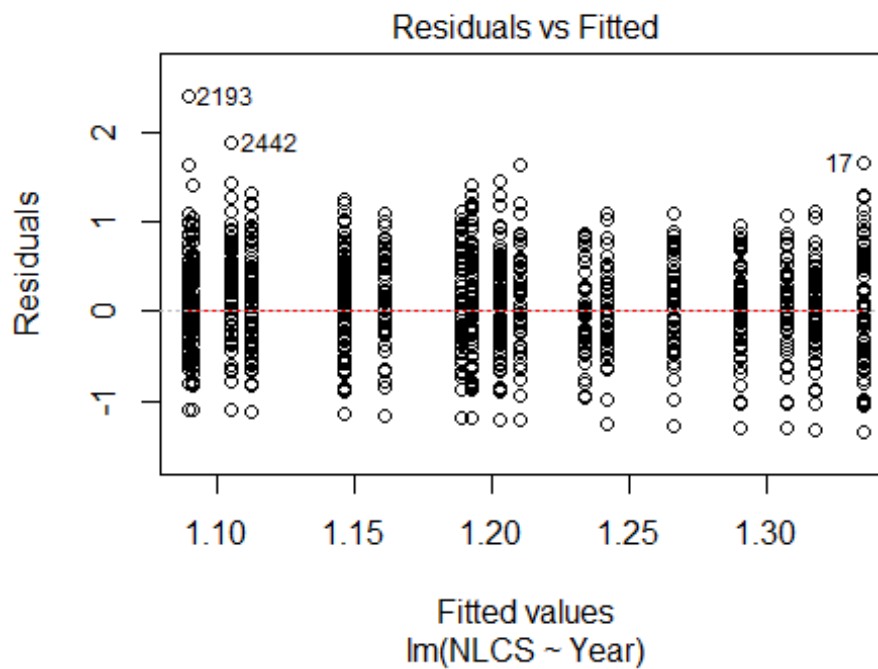
```

## Year2007      0.15063      0.14440      1.04      0.297
## Year2008      0.00789      0.15360      0.05      0.959
## Year2009      0.02752      0.15313      0.18      0.857
## Year2010      0.06264      0.16413      0.38      0.703
## Year2011     -0.02480      0.14867     -0.17      0.868
## Year2012      0.26724      0.15859      1.69      0.093 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.578
## Multiple R-squared:  0.0219, Adjusted R-squared:  -0.0135
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 34 weights are ~= 1. The remaining 454 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.290  0.881  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      2.05e-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: 488"
## [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
## 119 125 120 101 98 101 101 82 87 84 130 150 176 172 158
## 2011 2012
## 151 145
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 66 51 63 54 59 52 54 45 59 48 87 95 101 92 102
## 2011 2012

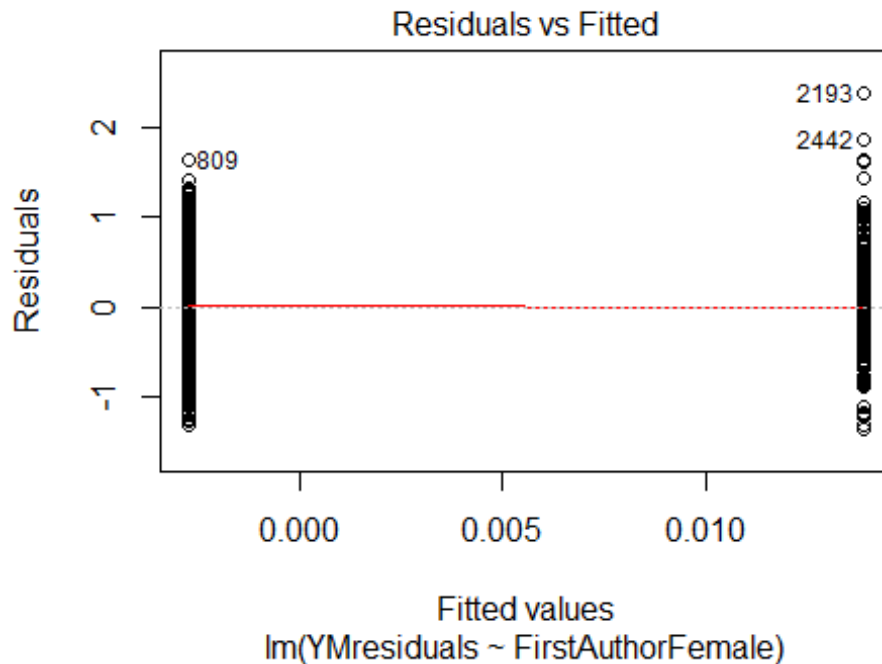
```



```
## 96 100
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 58 44 60 48 50 48 51 36 47 43 66 82 87 77 77
## 2011 2012
## 78 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 = 27, df = 16, p-value = 0.05
```

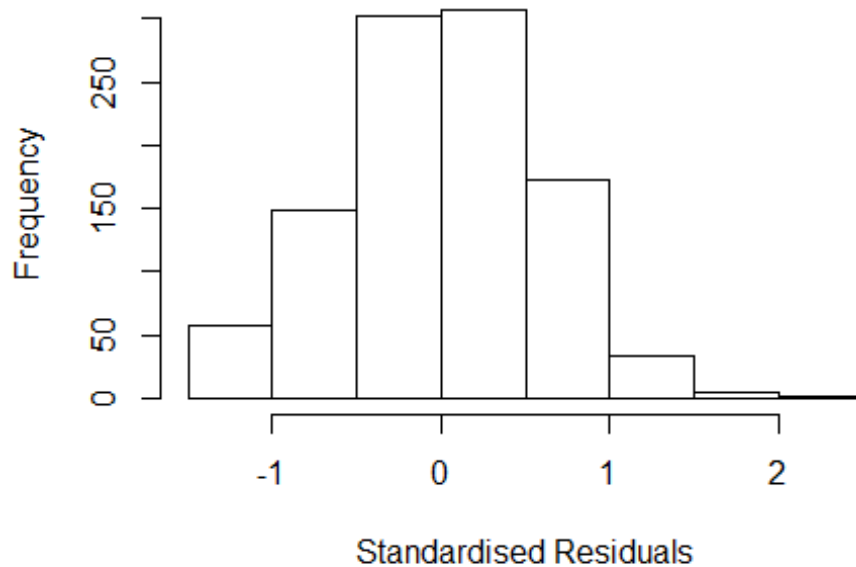


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.2, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.16217851836141"
## [1] "Male first author team size 2018 geometric mean: 2.1706289401368"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.97066953654195"
## [1] "Male last author team size 2018 geometric mean: 2.22991655734963"
##
## 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 1.200 1          1.095
## LastAuthorFemale  1.158 1          1.076
## UniqueAuthors    1.327 4          1.036
## Year              1.535 16         1.013
```

Residuals from first and last author and team size



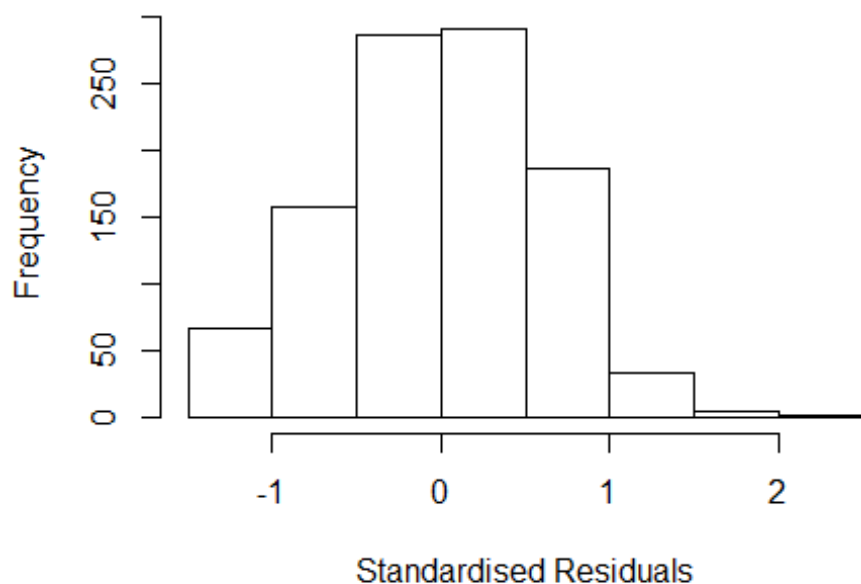
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.42817 -0.37768 0.00825 0.40989 2.48180
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14415 0.09956 11.49 < 2e-16 ***
## FirstAuthorFemale1 -0.05112 0.05424 -0.94 0.34620
## LastAuthorFemale1 0.04575 0.05899 0.78 0.43822
## UniqueAuthors2 0.20753 0.05058 4.10 4.4e-05 ***
## UniqueAuthors3 0.36318 0.05246 6.92 7.9e-12 ***
## UniqueAuthors4 0.20304 0.09734 2.09 0.03725 *
## UniqueAuthors5 0.44325 0.12004 3.69 0.00023 ***
## Year1997 -0.16064 0.14018 -1.15 0.25211
## Year1998 -0.13036 0.12666 -1.03 0.30363
## Year1999 0.04114 0.12698 0.32 0.74599
```

```

## Year2000      -0.10019    0.12126   -0.83   0.40886
## Year2001      -0.06492    0.12592   -0.52   0.60630
## Year2002      -0.06111    0.12212   -0.50   0.61689
## Year2003      -0.05416    0.12931   -0.42   0.67544
## Year2004      -0.00638    0.11244   -0.06   0.95474
## Year2005      -0.12816    0.12104   -1.06   0.28995
## Year2006      -0.25160    0.11446   -2.20   0.02817 *
## Year2007      -0.12490    0.11717   -1.07   0.28668
## Year2008      -0.23172    0.11330   -2.05   0.04110 *
## Year2009      -0.29637    0.11800   -2.51   0.01218 *
## Year2010      -0.37134    0.13342   -2.78   0.00548 **
## Year2011      -0.16150    0.11269   -1.43   0.15214
## Year2012      -0.20557    0.12671   -1.62   0.10503
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.595
## Multiple R-squared:  0.0759, Adjusted R-squared:  0.0556
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 96 weights are ~= 1. The remaining 930 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0426 0.8600 0.9490 0.9070 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.75e-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 1.137 1 1.066
## LastAuthorFemale 1.125 1 1.061
## Year 1.151 16 1.004

```

Residuals from first and last author



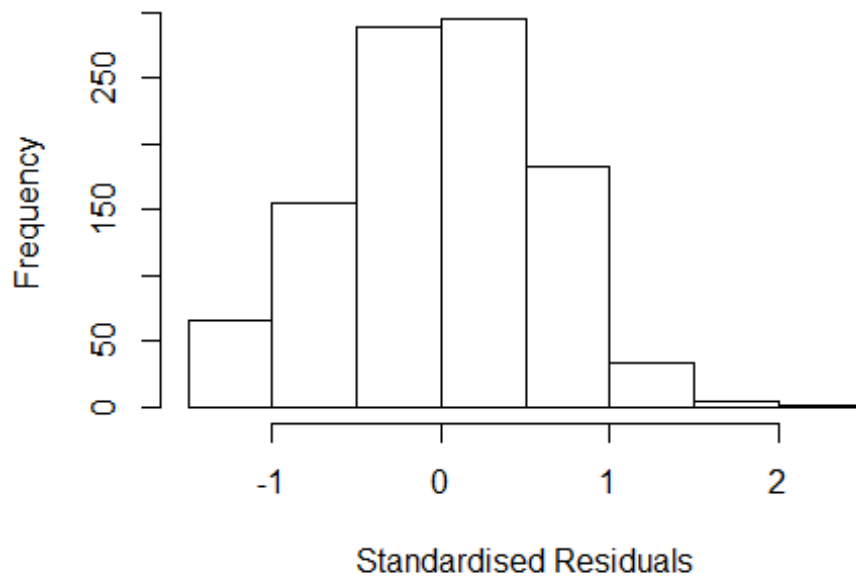
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34777 -0.39558 0.00289 0.43679 2.46373
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31072 0.09836 13.33 <2e-16 ***
## FirstAuthorFemale1 -0.02351 0.05418 -0.43 0.664
## LastAuthorFemale1 0.05098 0.05852 0.87 0.384
## Year1997 -0.16961 0.14248 -1.19 0.234
## Year1998 -0.14693 0.12840 -1.14 0.253
## Year1999 0.00984 0.13222 0.07 0.941
## Year2000 -0.04843 0.12339 -0.39 0.695
## Year2001 -0.08757 0.13218 -0.66 0.508
## Year2002 -0.04933 0.12734 -0.39 0.699
## Year2003 -0.01812 0.12900 -0.14 0.888
## Year2004 0.03705 0.11800 0.31 0.754
## Year2005 -0.07086 0.12530 -0.57 0.572
```

```

## Year2006          -0.18790      0.12209      -1.54      0.124
## Year2007          -0.13098      0.12365      -1.06      0.290
## Year2008          -0.21934      0.11865      -1.85      0.065 .
## Year2009          -0.26494      0.12511      -2.12      0.034 *
## Year2010          -0.33464      0.13469      -2.48      0.013 *
## Year2011          -0.11208      0.11910      -0.94      0.347
## Year2012          -0.13473      0.13276      -1.01      0.310
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.62
## Multiple R-squared:  0.0264, Adjusted R-squared:  0.00905
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 105 weights are ~= 1. The remaining 921 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0791 0.8640 0.9460 0.9090 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.75e-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 1.081 1      1.040
## Year              1.081 16      1.002

```

Residuals from first author



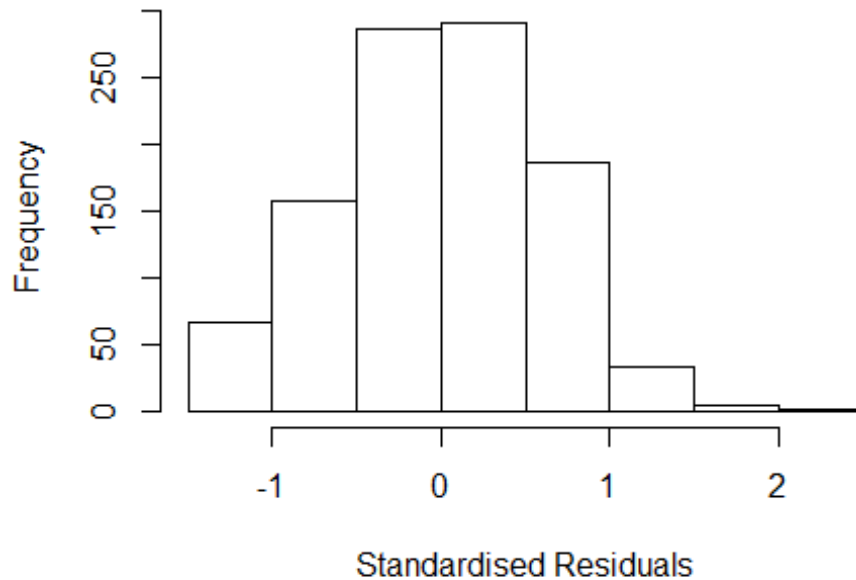
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35108 -0.39916 0.00706 0.43675 2.44621
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31773 0.09756 13.51 <2e-16 ***
## FirstAuthorFemale1 -0.01008 0.05278 -0.19 0.849
## Year1997 -0.16773 0.14252 -1.18 0.240
## Year1998 -0.14816 0.12823 -1.16 0.248
## Year1999 0.00776 0.13193 0.06 0.953
## Year2000 -0.04329 0.12349 -0.35 0.726
## Year2001 -0.09115 0.13205 -0.69 0.490
## Year2002 -0.05144 0.12725 -0.40 0.686
## Year2003 -0.02377 0.12848 -0.19 0.853
## Year2004 0.03335 0.11742 0.28 0.776
## Year2005 -0.07659 0.12482 -0.61 0.540
## Year2006 -0.19059 0.12178 -1.57 0.118
```

```

## Year2007          -0.13543    0.12310   -1.10    0.271
## Year2008          -0.22005    0.11842   -1.86    0.063 .
## Year2009          -0.26786    0.12460   -2.15    0.032 *
## Year2010          -0.33670    0.13536   -2.49    0.013 *
## Year2011          -0.11405    0.11870   -0.96    0.337
## Year2012          -0.13599    0.13250   -1.03    0.305
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.62
## Multiple R-squared:  0.0257, Adjusted R-squared:  0.00922
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 918 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.085  0.860  0.945  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.75e-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 1.065 1          1.032
## Year              1.065 16          1.002

```


Residuals from last author



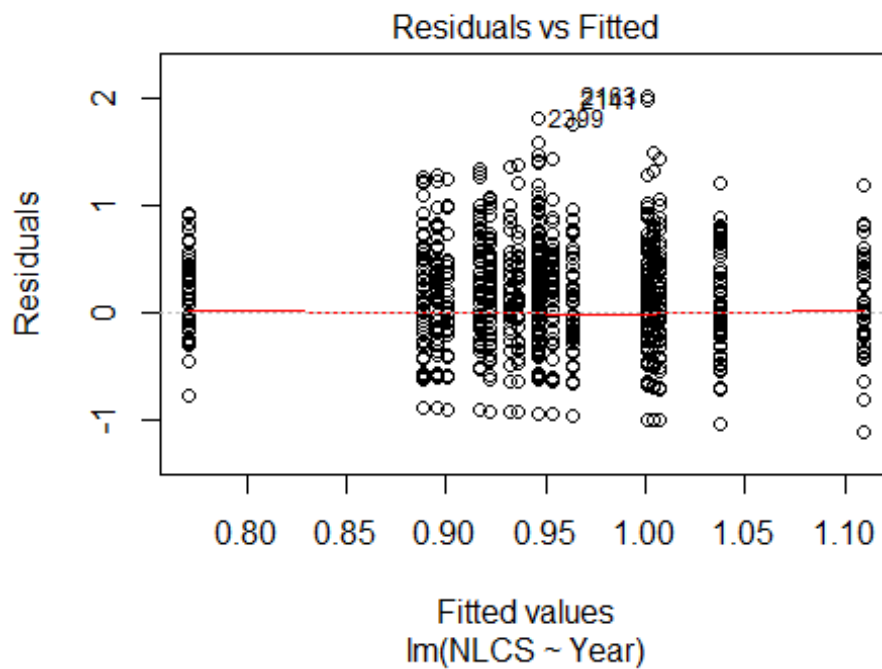
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35352 -0.39562 0.00394 0.43710 2.44339
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30964 0.09871 13.27 <2e-16 ***
## LastAuthorFemale1 0.04388 0.05715 0.77 0.443
## Year1997 -0.17192 0.14260 -1.21 0.228
## Year1998 -0.14928 0.12834 -1.16 0.245
## Year1999 0.00907 0.13216 0.07 0.945
## Year2000 -0.04830 0.12367 -0.39 0.696
## Year2001 -0.08819 0.13243 -0.67 0.506
## Year2002 -0.04851 0.12768 -0.38 0.704
## Year2003 -0.01924 0.12917 -0.15 0.882
## Year2004 0.03457 0.11766 0.29 0.769
## Year2005 -0.07346 0.12529 -0.59 0.558
## Year2006 -0.19072 0.12185 -1.57 0.118
```

```

## Year2007          -0.13440      0.12292    -1.09      0.274
## Year2008          -0.22091      0.11883    -1.86      0.063 .
## Year2009          -0.26704      0.12536    -2.13      0.033 *
## Year2010          -0.33723      0.13568    -2.49      0.013 *
## Year2011          -0.11537      0.11869    -0.97      0.331
## Year2012          -0.13835      0.13228    -1.05      0.296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.619
## Multiple R-squared:  0.0262, Adjusted R-squared:  0.00978
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 110 weights are ~= 1. The remaining 916 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0848 0.8630 0.9460 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.75e-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: 1026"
## [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
## 100 115 98 84 117 106 105 81 74 85 87 131 143 165 163
## 2011 2012
## 142 160
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 57 65 55 42 73 62 62 51 43 47 47 72 84 102 101
## 2011 2012

```

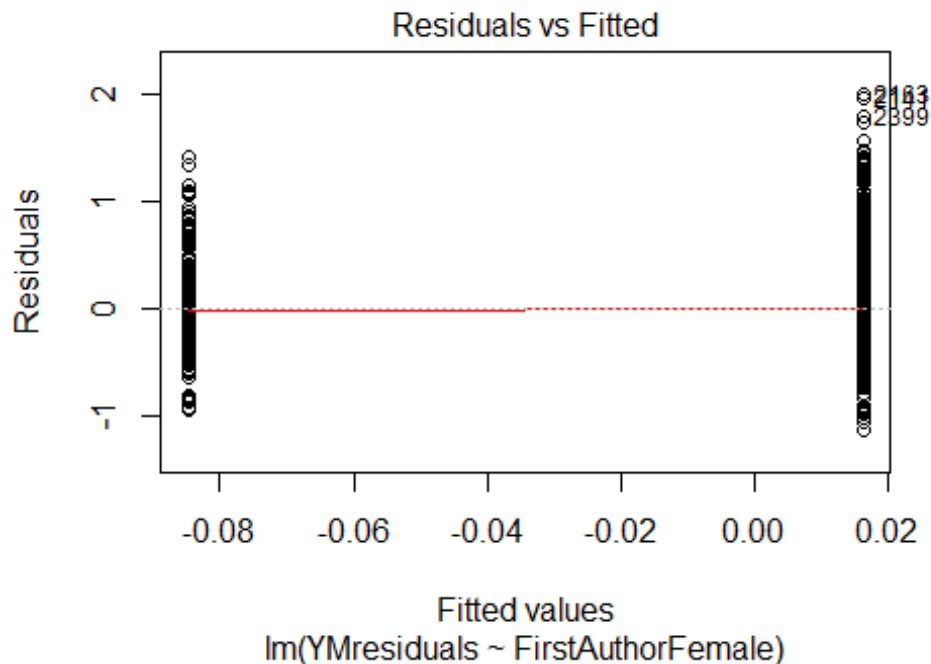
```
## 90 89
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 58 50 40 67 57 57 43 36 39 39 62 75 86 80
## 2011 2012
## 73 66
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.5
```



```
##
## 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: 1.96630722163743"
## [1] "Male first author team size 2018 geometric mean: 1.60796666725725"
## Warning in wilcox.test.default(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 = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.62238960361098"
## [1] "Male last author team size 2018 geometric mean: 1.68219037802369"

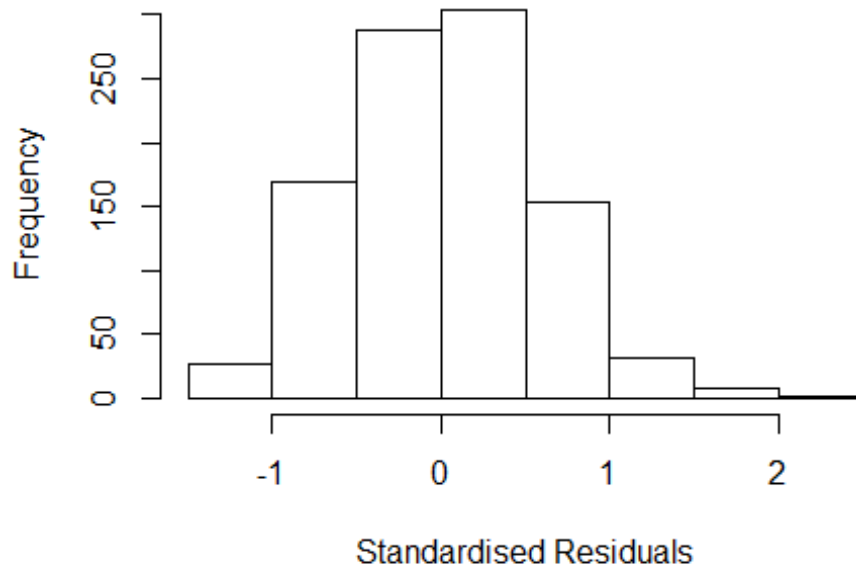
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	1.242	1	1.114
LastAuthorFemale	1.263	1	1.124
UniqueAuthors	1.340	4	1.037
Year	1.375	16	1.010

Residuals from first and last author and team size



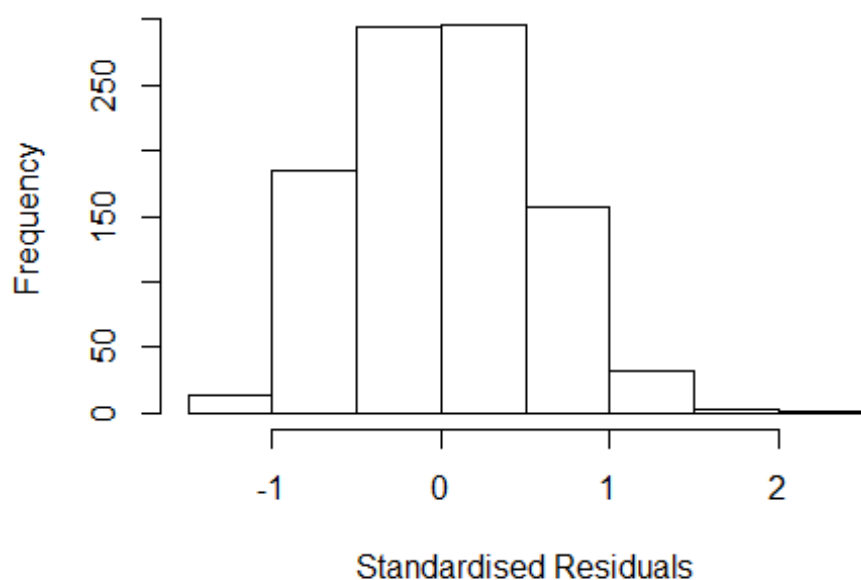
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23701 -0.39517 0.00776 0.39245 2.07611
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.91925 0.08243 11.15 < 2e-16 ***
## FirstAuthorFemale1 -0.16636 0.05693 -2.92 0.00356 **
## LastAuthorFemale1 0.08300 0.05931 1.40 0.16200
## UniqueAuthors2 0.16335 0.04304 3.79 0.00016 ***
## UniqueAuthors3 0.17938 0.06344 2.83 0.00479 **
## UniqueAuthors4 0.08836 0.13797 0.64 0.52202
## UniqueAuthors5 0.03083 0.13753 0.22 0.82265
## Year1997 -0.18439 0.10814 -1.71 0.08850 .
## Year1998 0.00355 0.12858 0.03 0.97797
## Year1999 -0.05109 0.12211 -0.42 0.67577
```

```

## Year2000      -0.08303    0.10890   -0.76  0.44598
## Year2001      -0.13467    0.11382   -1.18  0.23703
## Year2002       0.01035    0.10818    0.10  0.92382
## Year2003      -0.07245    0.11789   -0.61  0.53898
## Year2004      -0.03955    0.13186   -0.30  0.76431
## Year2005      -0.00679    0.11395   -0.06  0.95250
## Year2006       0.15441    0.12130    1.27  0.20334
## Year2007      -0.07299    0.10528   -0.69  0.48831
## Year2008       0.07999    0.10221    0.78  0.43407
## Year2009      -0.01036    0.10436   -0.10  0.92095
## Year2010      -0.09237    0.11806   -0.78  0.43417
## Year2011      -0.02406    0.11484   -0.21  0.83407
## Year2012      -0.06635    0.10933   -0.61  0.54403
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.0424, Adjusted R-squared:  0.0204
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 907 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.213  0.872  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      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 1.214 1      1.102
## LastAuthorFemale  1.238 1      1.112
## Year              1.057 16      1.002

```

Residuals from first and last author



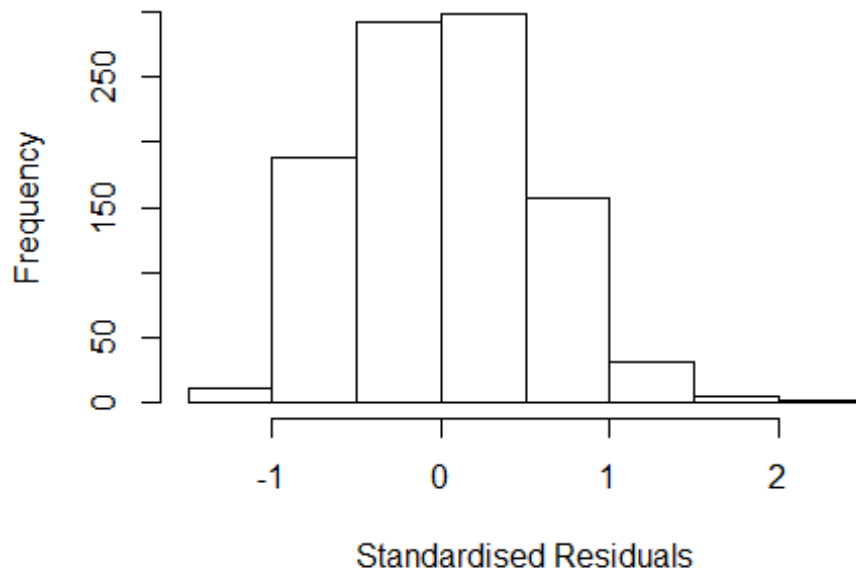
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18205 -0.41546 -0.00125 0.40936 2.02758
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9651 0.0789 12.23 <2e-16 ***
## FirstAuthorFemale1 -0.1374 0.0570 -2.41 0.016 *
## LastAuthorFemale1 0.0807 0.0590 1.37 0.172
## Year1997 -0.1858 0.1065 -1.74 0.082 .
## Year1998 0.0223 0.1300 0.17 0.864
## Year1999 -0.0441 0.1225 -0.36 0.719
## Year2000 -0.0802 0.1073 -0.75 0.455
## Year2001 -0.1097 0.1129 -0.97 0.331
## Year2002 0.0491 0.1063 0.46 0.644
## Year2003 -0.0101 0.1176 -0.09 0.931
## Year2004 0.0205 0.1309 0.16 0.876
## Year2005 0.0374 0.1137 0.33 0.742
```

```

## Year2006          0.2169      0.1234      1.76      0.079 .
## Year2007         -0.0354      0.1040     -0.34      0.734
## Year2008          0.1089      0.1002      1.09      0.277
## Year2009          0.0193      0.1015      0.19      0.849
## Year2010         -0.0409      0.1126     -0.36      0.716
## Year2011          0.0199      0.1131      0.18      0.860
## Year2012         -0.0312      0.1074     -0.29      0.772
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.609
## Multiple R-squared:  0.0247, Adjusted R-squared:  0.00647
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 894 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.244  0.871  0.950  0.915  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.016 1      1.008
## Year              1.016 16      1.000

```


Residuals from first author



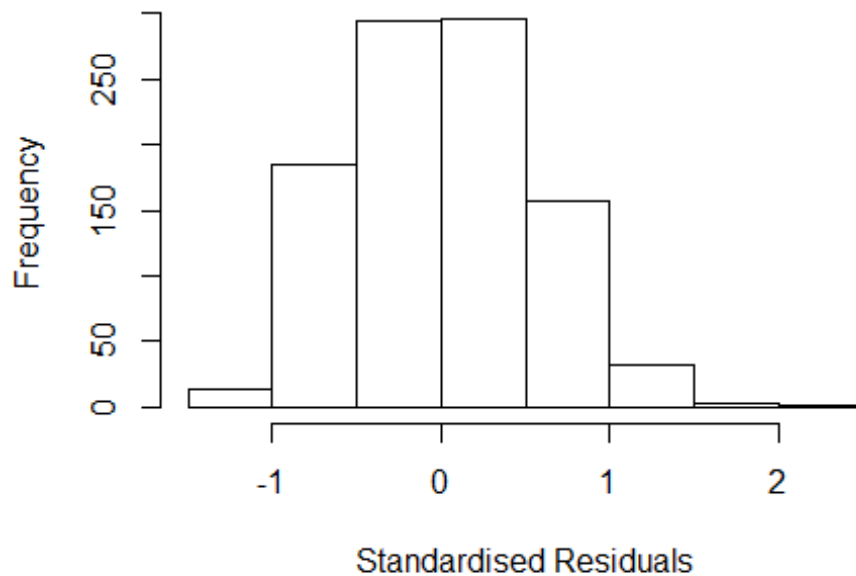
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19054 -0.42459 0.00182 0.41176 2.01810
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96785 0.07914 12.23 <2e-16 ***
## FirstAuthorFemale1 -0.10413 0.05197 -2.00 0.045 *
## Year1997 -0.18047 0.10720 -1.68 0.093 .
## Year1998 0.02198 0.12978 0.17 0.866
## Year1999 -0.04219 0.12303 -0.34 0.732
## Year2000 -0.07535 0.10761 -0.70 0.484
## Year2001 -0.10524 0.11370 -0.93 0.355
## Year2002 0.04974 0.10686 0.47 0.642
## Year2003 -0.00912 0.11756 -0.08 0.938
## Year2004 0.02363 0.13138 0.18 0.857
## Year2005 0.03816 0.11410 0.33 0.738
## Year2006 0.22269 0.12471 1.79 0.074 .
```

```

## Year2007      -0.03103    0.10415   -0.30    0.766
## Year2008      0.11112    0.10024    1.11    0.268
## Year2009      0.02605    0.10104    0.26    0.797
## Year2010     -0.04073    0.11309   -0.36    0.719
## Year2011      0.02392    0.11295    0.21    0.832
## Year2012     -0.03229    0.10743   -0.30    0.764
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.611
## Multiple R-squared:  0.023, Adjusted R-squared:  0.00575
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 902 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.253  0.876  0.951  0.916  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 1.037 1          1.018
## Year            1.037 16          1.001

```

Residuals from last author



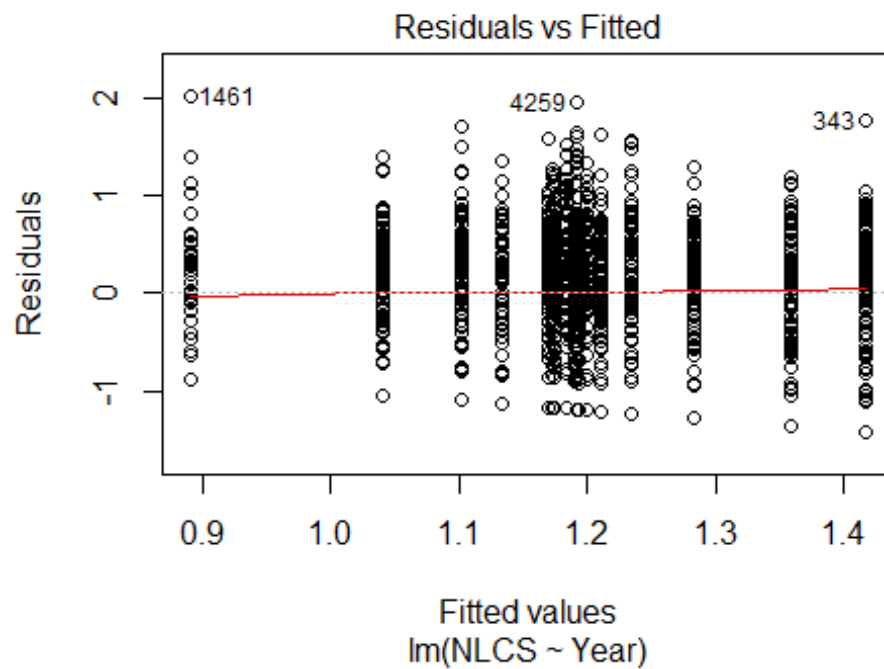
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17259 -0.42321 -0.00488 0.40721 2.03792
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95331 0.07819 12.19 <2e-16 ***
## LastAuthorFemale1 0.01433 0.05499 0.26 0.794
## Year1997 -0.17960 0.10660 -1.68 0.092 .
## Year1998 0.02944 0.13036 0.23 0.821
## Year1999 -0.04483 0.12193 -0.37 0.713
## Year2000 -0.08309 0.10641 -0.78 0.435
## Year2001 -0.10951 0.11282 -0.97 0.332
## Year2002 0.05149 0.10613 0.49 0.628
## Year2003 -0.00564 0.11614 -0.05 0.961
## Year2004 0.01684 0.13000 0.13 0.897
## Year2005 0.03866 0.11397 0.34 0.735
## Year2006 0.21927 0.12376 1.77 0.077 .
```

```

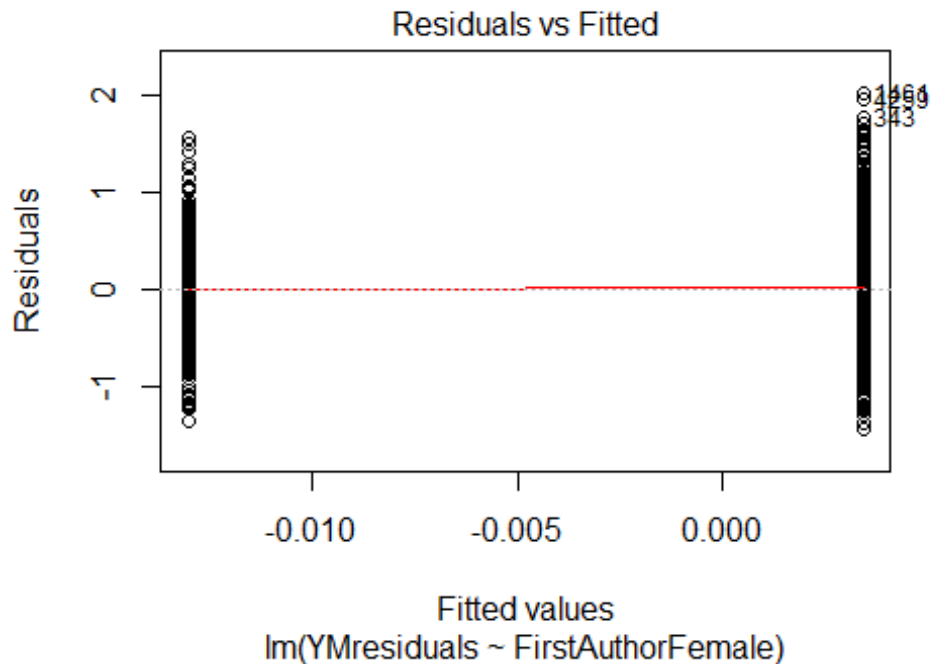
## Year2007          -0.04289      0.10362    -0.41      0.679
## Year2008           0.10645      0.09941      1.07      0.285
## Year2009           0.02077      0.10063      0.21      0.837
## Year2010          -0.05197      0.11484     -0.45      0.651
## Year2011           0.01459      0.11262      0.13      0.897
## Year2012          -0.04643      0.10828     -0.43      0.668
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.612
## Multiple R-squared:  0.0188, Adjusted R-squared:  0.00153
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 906 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.244  0.873   0.952   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      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 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
##  309  260  189  184  194  258  194  186  222  175  191  213  213  193  210
## 2011 2012
##  212  201
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  145  106   74   94   76   50  103  129  120  110  108  118  122  127  133
## 2011 2012

```

```
## 127 122
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 130 91 66 83 70 44 90 113 106 88 89 105 106 116 119
## 2011 2012
## 112 105
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 49, df = 16, p-value = 3e-05
```

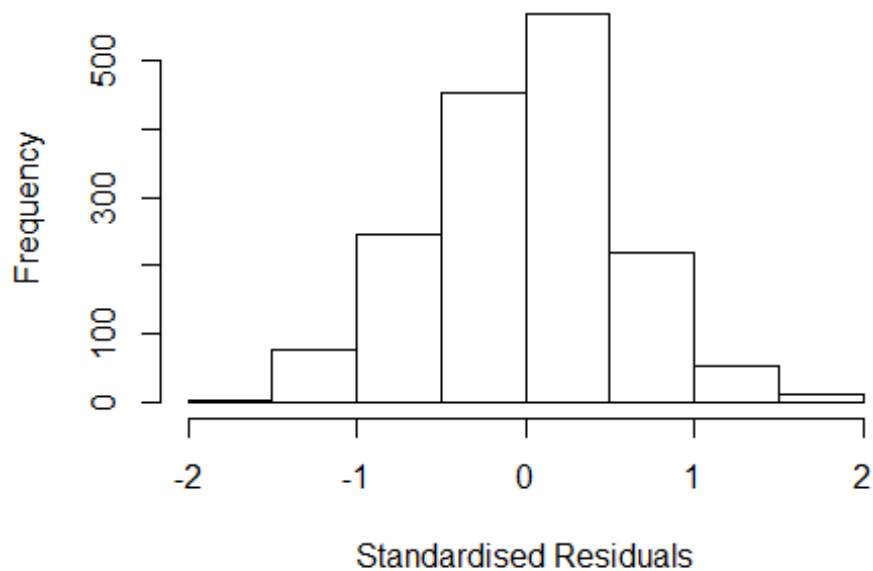


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.5, df = 1, p-value = 0.01
```



```
## [1] "Female first author team size 2018 geometric mean: 3.04698360600842"
## [1] "Male first author team size 2018 geometric mean: 2.76774665673149"
##
## 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: 2.69016427014372"
## [1] "Male last author team size 2018 geometric mean: 2.86219375743547"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.133 1      1.064
## LastAuthorFemale  1.127 1      1.062
## UniqueAuthors     1.381 4      1.041
## Year              1.470 16     1.012
```

Residuals from first and last author and team size



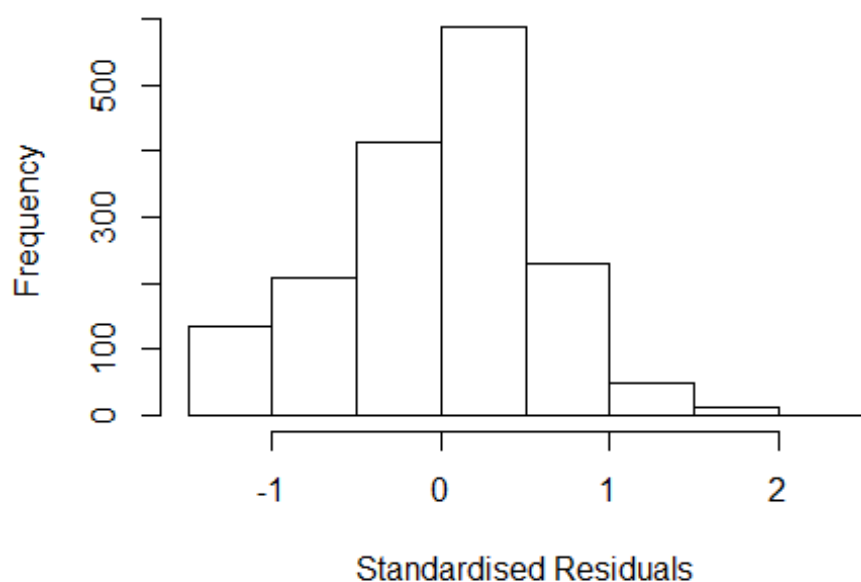
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5541 -0.3714 0.0262 0.3784 1.9564
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2760 0.0645 19.77 < 2e-16 ***
## FirstAuthorFemale1 -0.0176 0.0350 -0.50 0.61429
## LastAuthorFemale1 0.0155 0.0431 0.36 0.71861
## UniqueAuthors2 0.2782 0.0440 6.32 3.3e-10 ***
## UniqueAuthors3 0.3162 0.0442 7.16 1.2e-12 ***
## UniqueAuthors4 0.3578 0.0527 6.79 1.5e-11 ***
## UniqueAuthors5 0.4545 0.0564 8.06 1.5e-15 ***
## Year1997 -0.0553 0.0867 -0.64 0.52385
## Year1998 -0.2012 0.1175 -1.71 0.08698 .
## Year1999 -0.3568 0.0908 -3.93 8.9e-05 ***
```

```

## Year2000      -0.3808      0.0973      -3.91      9.5e-05 ***
## Year2001      -0.6055      0.1037      -5.84      6.4e-09 ***
## Year2002      -0.3245      0.0897      -3.62      0.00031 ***
## Year2003      -0.3059      0.0834      -3.67      0.00025 ***
## Year2004      -0.4782      0.0787      -6.08      1.5e-09 ***
## Year2005      -0.3541      0.0757      -4.68      3.1e-06 ***
## Year2006      -0.3402      0.0765      -4.45      9.2e-06 ***
## Year2007      -0.3242      0.0804      -4.03      5.8e-05 ***
## Year2008      -0.2512      0.0779      -3.23      0.00128 **
## Year2009      -0.4181      0.0808      -5.17      2.6e-07 ***
## Year2010      -0.2942      0.0863      -3.41      0.00067 ***
## Year2011      -0.3446      0.0892      -3.86      0.00012 ***
## Year2012      -0.3528      0.0835      -4.23      2.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.548
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0897
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 1465 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.175  0.861  0.948   0.894  0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.12e-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 1.100 1          1.049
## LastAuthorFemale  1.086 1          1.042
## Year              1.118 16          1.003

```


Residuals from first and last author



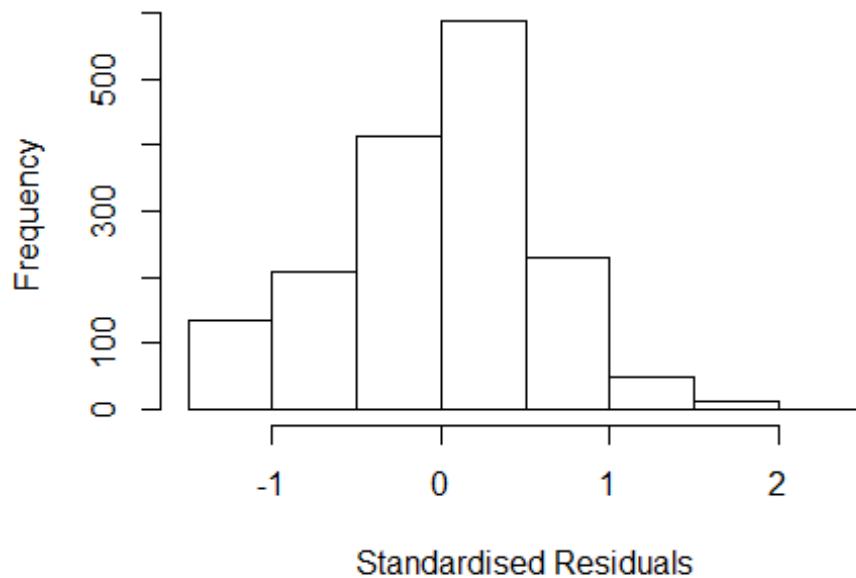
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4570 -0.3750 0.0507 0.3857 2.0526
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.457008 0.060594 24.05 < 2e-16 ***
## FirstAuthorFemale1 -0.003985 0.035374 -0.11 0.91031
## LastAuthorFemale1 -0.000319 0.042690 -0.01 0.99404
## Year1997 -0.038002 0.090806 -0.42 0.67564
## Year1998 -0.225965 0.119948 -1.88 0.05976 .
## Year1999 -0.274260 0.089243 -3.07 0.00215 **
## Year2000 -0.316245 0.105141 -3.01 0.00267 **
## Year2001 -0.595655 0.112090 -5.31 1.2e-07 ***
## Year2002 -0.259383 0.088526 -2.93 0.00344 **
## Year2003 -0.275526 0.083832 -3.29 0.00104 **
## Year2004 -0.430338 0.081756 -5.26 1.6e-07 ***
## Year2005 -0.258404 0.076086 -3.40 0.00070 ***
```

```

## Year2006          -0.275677    0.078245    -3.52    0.00044 ***
## Year2007          -0.227156    0.079782    -2.85    0.00447 **
## Year2008          -0.151137    0.081114    -1.86    0.06261 .
## Year2009          -0.349742    0.083210    -4.20    2.8e-05 ***
## Year2010          -0.240276    0.084463    -2.84    0.00450 **
## Year2011          -0.285274    0.091241    -3.13    0.00180 **
## Year2012          -0.292865    0.084460    -3.47    0.00054 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared:  0.0406, Adjusted R-squared:  0.0299
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1495 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.139  0.848   0.946   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      6.12e-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 1.068 1      1.033
## Year              1.068 16      1.002

```

Residuals from first author



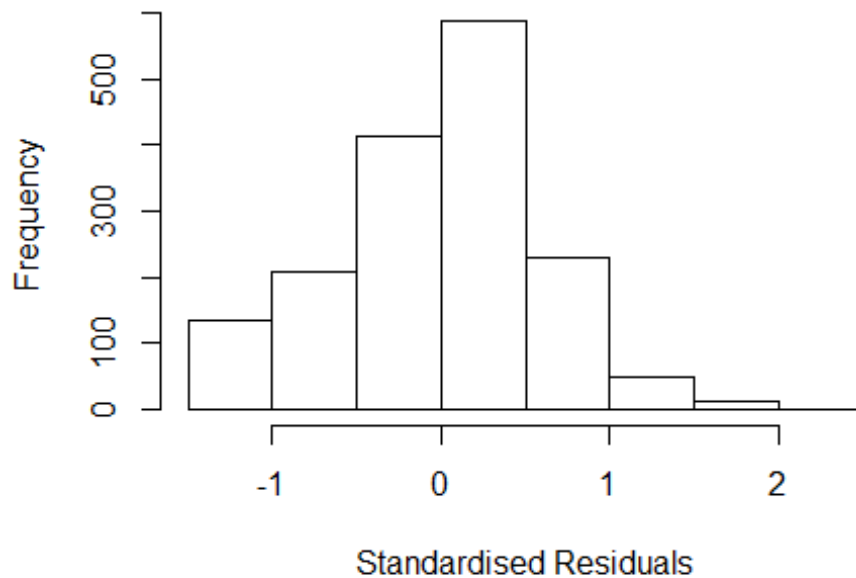
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4567 -0.3755 0.0508 0.3858 2.0525
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.45669 0.06029 24.16 < 2e-16 ***
## FirstAuthorFemale1 -0.00406 0.03497 -0.12 0.90762
## Year1997 -0.03799 0.09076 -0.42 0.67559
## Year1998 -0.22600 0.11970 -1.89 0.05919 .
## Year1999 -0.27407 0.08913 -3.08 0.00214 **
## Year2000 -0.31622 0.10499 -3.01 0.00264 **
## Year2001 -0.59523 0.11199 -5.31 1.2e-07 ***
## Year2002 -0.25924 0.08848 -2.93 0.00344 **
## Year2003 -0.27539 0.08377 -3.29 0.00103 **
## Year2004 -0.43010 0.08171 -5.26 1.6e-07 ***
## Year2005 -0.25825 0.07610 -3.39 0.00071 ***
## Year2006 -0.27542 0.07821 -3.52 0.00044 ***
```

```

## Year2007          -0.22698    0.07974   -2.85  0.00447 **
## Year2008          -0.15103    0.08109   -1.86  0.06272 .
## Year2009          -0.34951    0.08317   -4.20  2.8e-05 ***
## Year2010          -0.24007    0.08431   -2.85  0.00446 **
## Year2011          -0.28483    0.09117   -3.12  0.00181 **
## Year2012          -0.29254    0.08447   -3.46  0.00055 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.0405, Adjusted R-squared:  0.0304
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1495 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.142  0.849  0.947  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      6.12e-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 1.054 1          1.027
## Year            1.054 16          1.002

```

Residuals from last author



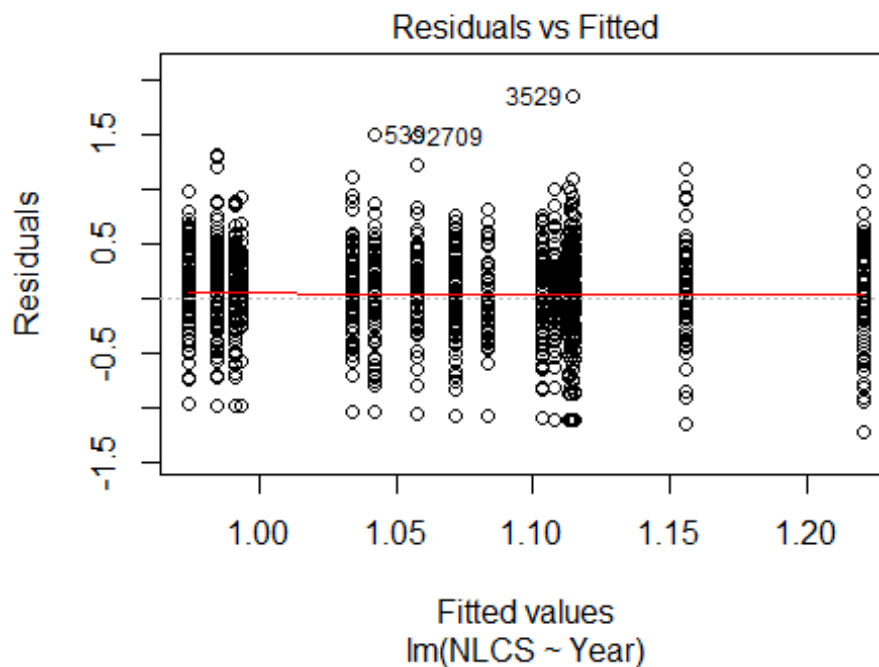
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4566 -0.3740 0.0513 0.3848 2.0533
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.45660 0.06035 24.13 < 2e-16 ***
## LastAuthorFemale1 -0.00154 0.04216 -0.04 0.97085
## Year1997 -0.03815 0.09077 -0.42 0.67435
## Year1998 -0.22592 0.11994 -1.88 0.05981 .
## Year1999 -0.27471 0.08922 -3.08 0.00211 **
## Year2000 -0.31627 0.10511 -3.01 0.00266 **
## Year2001 -0.59594 0.11199 -5.32 1.2e-07 ***
## Year2002 -0.25970 0.08844 -2.94 0.00337 **
## Year2003 -0.27616 0.08319 -3.32 0.00092 ***
## Year2004 -0.43041 0.08176 -5.26 1.6e-07 ***
## Year2005 -0.25860 0.07609 -3.40 0.00069 ***
## Year2006 -0.27579 0.07825 -3.52 0.00044 ***
```

```

## Year2007          -0.22754      0.07973      -2.85   0.00437 **
## Year2008          -0.15140      0.08112      -1.87   0.06215 .
## Year2009          -0.34992      0.08326      -4.20   2.8e-05 ***
## Year2010          -0.24057      0.08451      -2.85   0.00448 **
## Year2011          -0.28603      0.09111      -3.14   0.00172 **
## Year2012          -0.29355      0.08419      -3.49   0.00050 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared:  0.0406, Adjusted R-squared:  0.0305
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1497 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.138  0.848  0.946  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      6.12e-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: 1633"
## [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
## 151 165 124 129 131 144 146 161 144 145 168 169 156 143 157
## 2011 2012
## 151 145
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 81 70 68 51 44 55 74 93 85 83 92 92 88 68 77
## 2011 2012

```

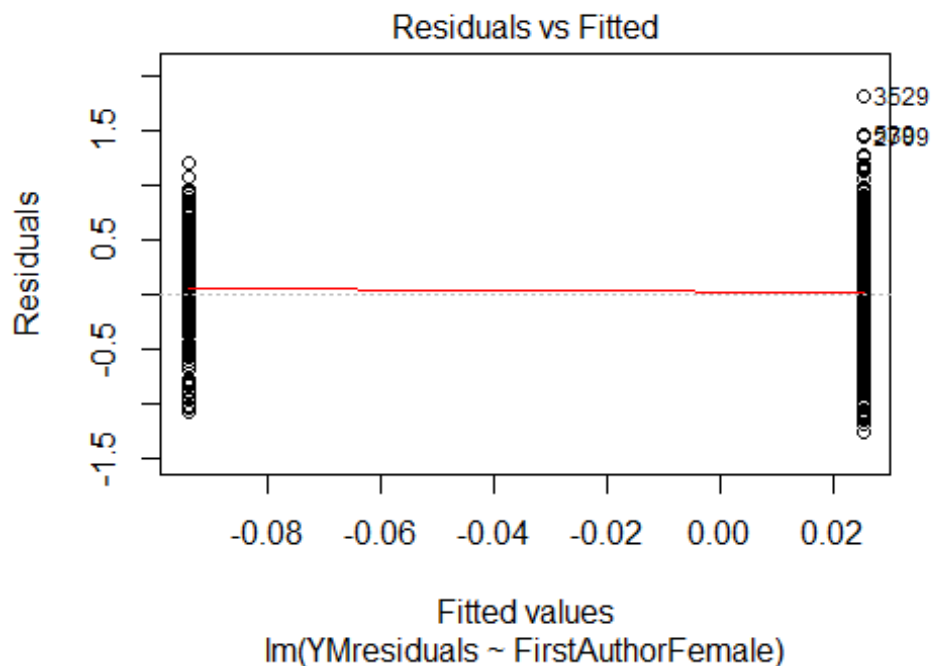
```
## 72 84
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 69 62 61 46 35 47 62 73 77 69 74 71 72 55 63
## 2011 2012
## 58 72
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 13, df = 16, p-value = 0.6
```



```
##
## 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: 2.99343057517387"
## [1] "Male first author team size 2018 geometric mean: 2.76110811154142"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

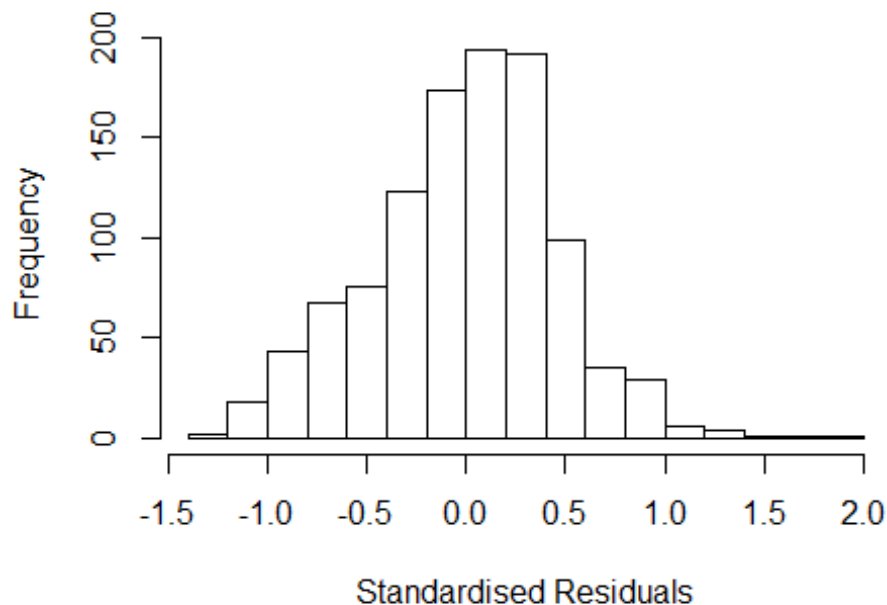
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 340, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.51002580517068"
## [1] "Male last author team size 2018 geometric mean: 2.72849698489624"

## Warning in wilcox.test.default(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 = 0.2
## 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  1.097 1      1.047
## UniqueAuthors    1.430 4      1.046
## Year              1.582 16     1.014
```


Residuals from first and last author and team size



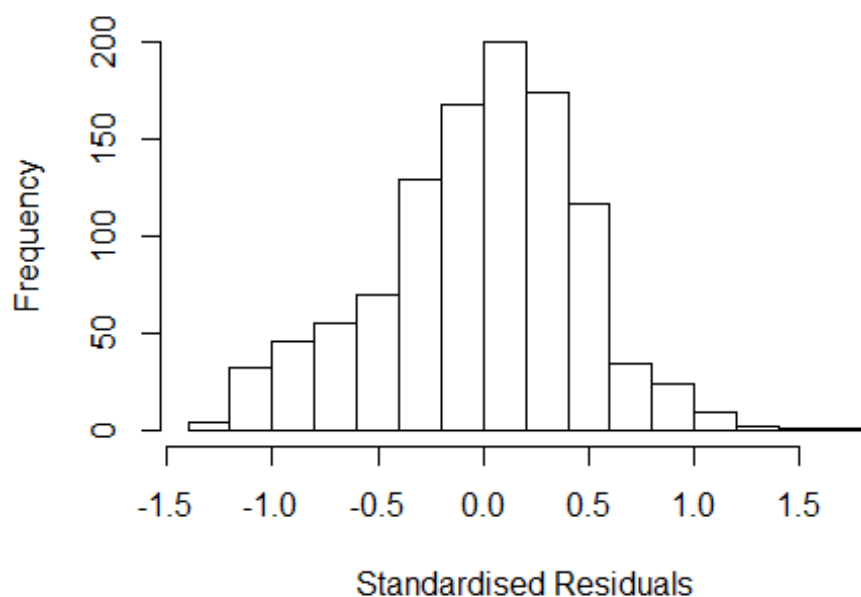
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3381 -0.2810  0.0329  0.2932  1.8053
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1334    0.0648   17.49 < 2e-16 ***
## FirstAuthorFemale1 -0.0711    0.0381   -1.87  0.06228 .
## LastAuthorFemale1 -0.1894    0.0490   -3.86  0.00012 ***
## UniqueAuthors2     0.2047    0.0407    5.02  6.0e-07 ***
## UniqueAuthors3     0.1884    0.0443    4.25  2.3e-05 ***
## UniqueAuthors4     0.2475    0.0573    4.32  1.7e-05 ***
## UniqueAuthors5     0.3392    0.0584    5.81  8.4e-09 ***
## Year1997          -0.1435    0.0850   -1.69  0.09172 .
## Year1998          -0.2377    0.0855   -2.78  0.00552 **
## Year1999          -0.1018    0.0874   -1.17  0.24406
```

```

## Year2000          -0.1320      0.0907    -1.46    0.14589
## Year2001          -0.1203      0.0807    -1.49    0.13626
## Year2002          -0.0396      0.0884    -0.45    0.65442
## Year2003          -0.2688      0.0798    -3.37    0.00079 ***
## Year2004          -0.1338      0.0789    -1.70    0.09020 .
## Year2005          -0.1103      0.0805    -1.37    0.17127
## Year2006          -0.2766      0.0806    -3.43    0.00062 ***
## Year2007          -0.3040      0.0854    -3.56    0.00039 ***
## Year2008          -0.2210      0.0761    -2.91    0.00374 **
## Year2009          -0.2124      0.0883    -2.41    0.01630 *
## Year2010          -0.2164      0.0869    -2.49    0.01294 *
## Year2011          -0.1891      0.0860    -2.20    0.02819 *
## Year2012          -0.1701      0.0827    -2.06    0.04007 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.111, Adjusted R-squared:  0.0921
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 988 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0416 0.8550 0.9520 0.8970 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.38e-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 1.098 1          1.048
## LastAuthorFemale 1.122 1          1.059
## Year          1.215 16          1.006

```

Residuals from first and last author



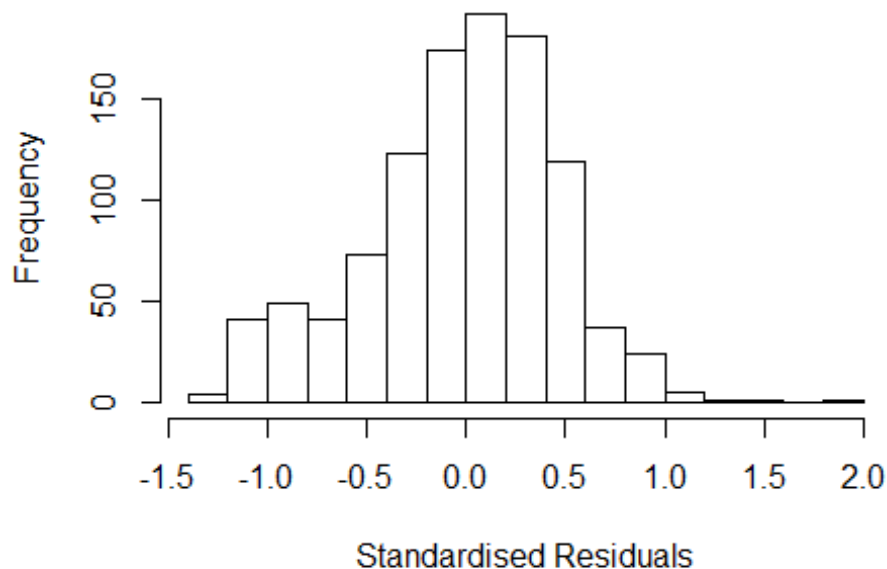
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2778 -0.2999 0.0396 0.2878 1.7933
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2778 0.0599 21.32 < 2e-16 ***
## FirstAuthorFemale1 -0.0542 0.0401 -1.35 0.17684
## LastAuthorFemale1 -0.2073 0.0539 -3.85 0.00013 ***
## Year1997 -0.1581 0.0848 -1.86 0.06253 .
## Year1998 -0.2660 0.0843 -3.15 0.00165 **
## Year1999 -0.1033 0.0877 -1.18 0.23889
## Year2000 -0.1382 0.0893 -1.55 0.12212
## Year2001 -0.1047 0.0816 -1.28 0.19947
## Year2002 -0.0455 0.0915 -0.50 0.61916
## Year2003 -0.2525 0.0839 -3.01 0.00269 **
## Year2004 -0.1197 0.0802 -1.49 0.13597
## Year2005 -0.1069 0.0805 -1.33 0.18446
```

```

## Year2006          -0.2605      0.0811   -3.21  0.00136 **
## Year2007          -0.2867      0.0917   -3.13  0.00181 **
## Year2008          -0.1990      0.0784   -2.54  0.01127 *
## Year2009          -0.1968      0.0892   -2.21  0.02760 *
## Year2010          -0.1853      0.0883   -2.10  0.03608 *
## Year2011          -0.1198      0.0856   -1.40  0.16201
## Year2012          -0.1141      0.0825   -1.38  0.16697
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0607, Adjusted R-squared:  0.0445
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 981 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.049  0.858  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      9.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.089 1          1.044
## Year              1.089 16          1.003

```

Residuals from first author



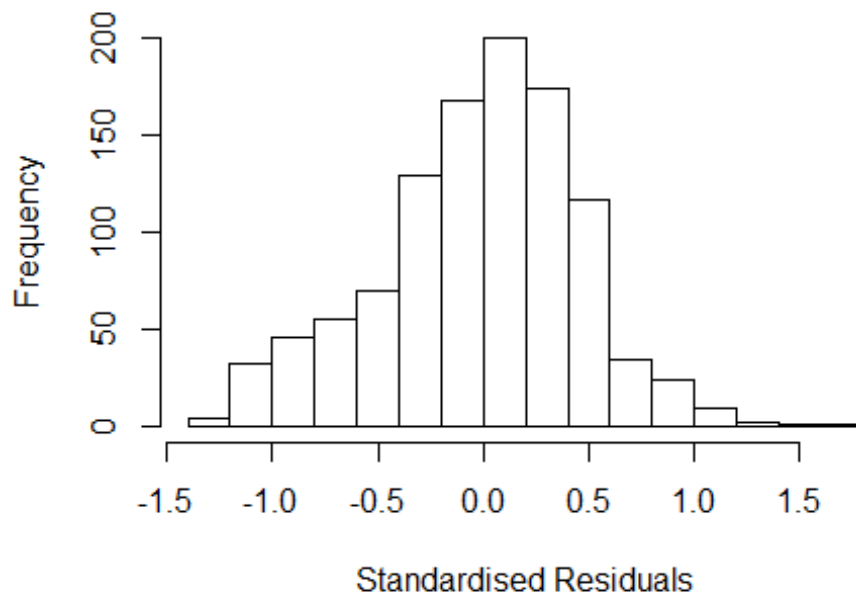
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2615 -0.2925 0.0365 0.2933 1.8151
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2615 0.0605 20.84 <2e-16 ***
## FirstAuthorFemale1 -0.0961 0.0448 -2.15 0.0320 *
## Year1997 -0.1485 0.0845 -1.76 0.0790 .
## Year1998 -0.2552 0.0848 -3.01 0.0027 **
## Year1999 -0.1104 0.0872 -1.27 0.2056
## Year2000 -0.1581 0.0911 -1.73 0.0832 .
## Year2001 -0.1227 0.0832 -1.47 0.1408
## Year2002 -0.0553 0.0927 -0.60 0.5505
## Year2003 -0.2531 0.0855 -2.96 0.0031 **
## Year2004 -0.1164 0.0799 -1.46 0.1452
## Year2005 -0.1155 0.0815 -1.42 0.1568
## Year2006 -0.2604 0.0836 -3.12 0.0019 **
```

```

## Year2007          -0.2890      0.0939   -3.08   0.0021 **
## Year2008          -0.1845      0.0787   -2.34   0.0192 *
## Year2009          -0.2089      0.0940   -2.22   0.0265 *
## Year2010          -0.1889      0.0902   -2.10   0.0364 *
## Year2011          -0.1227      0.0849   -1.45   0.1487
## Year2012          -0.1196      0.0832   -1.44   0.1511
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.435
## Multiple R-squared:  0.0357, Adjusted R-squared:  0.02
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 990 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0427 0.8620 0.9540 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      9.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.126 1          1.061
## Year            1.126 16          1.004

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2758 -0.2966  0.0436  0.2907  1.8069
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2758     0.0600  21.28 < 2e-16 ***
## LastAuthorFemale1 -0.2233     0.0571  -3.91  9.9e-05 ***
## Year1997         -0.1588     0.0850  -1.87  0.06182 .
## Year1998         -0.2727     0.0843  -3.23  0.00126 **
## Year1999         -0.1065     0.0877  -1.21  0.22500
## Year2000         -0.1383     0.0885  -1.56  0.11818
## Year2001         -0.1044     0.0822  -1.27  0.20412
## Year2002         -0.0529     0.0912  -0.58  0.56175
## Year2003         -0.2576     0.0841  -3.06  0.00225 **
## Year2004         -0.1253     0.0807  -1.55  0.12065
## Year2005         -0.1146     0.0802  -1.43  0.15309
## Year2006         -0.2697     0.0808  -3.34  0.00087 ***
```

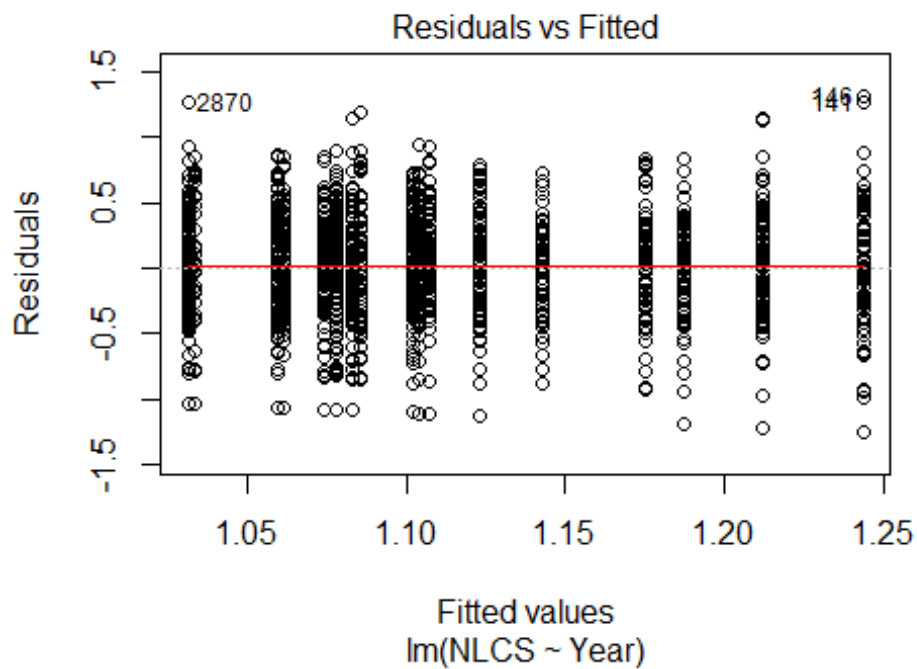
```

## Year2007          -0.2941      0.0921   -3.20   0.00144 **
## Year2008          -0.2107      0.0785   -2.68   0.00740 **
## Year2009          -0.2029      0.0898   -2.26   0.02404 *
## Year2010          -0.1956      0.0875   -2.24   0.02561 *
## Year2011          -0.1334      0.0844   -1.58   0.11444
## Year2012          -0.1257      0.0821   -1.53   0.12604
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.0575, Adjusted R-squared:  0.0422
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 989 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0431 0.8610 0.9520 0.8930 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1066"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
## 199 188 190 172 188 191 203 200 164 148 232 238 238 245 229
## 2011 2012
## 263 252
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 79 78 78 84 70 40 96 108 80 80 125 140 134 131 102
## 2011 2012

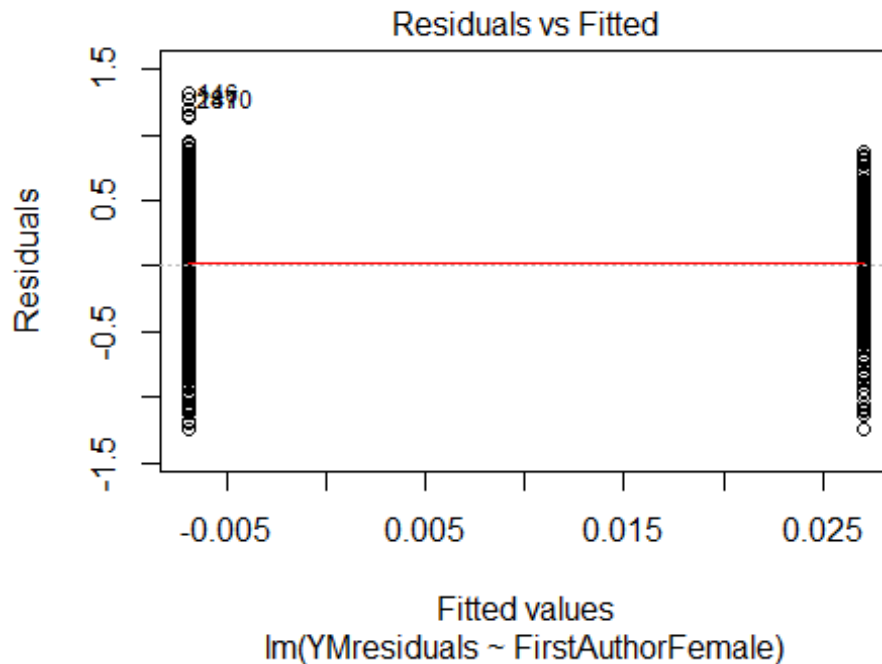
```



```
## 133 155
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 70 65 68 61 56 31 73 89 65 57 87 106 100 98 80
## 2011 2012
## 111 131
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 35, df = 16, p-value = 0.005
```

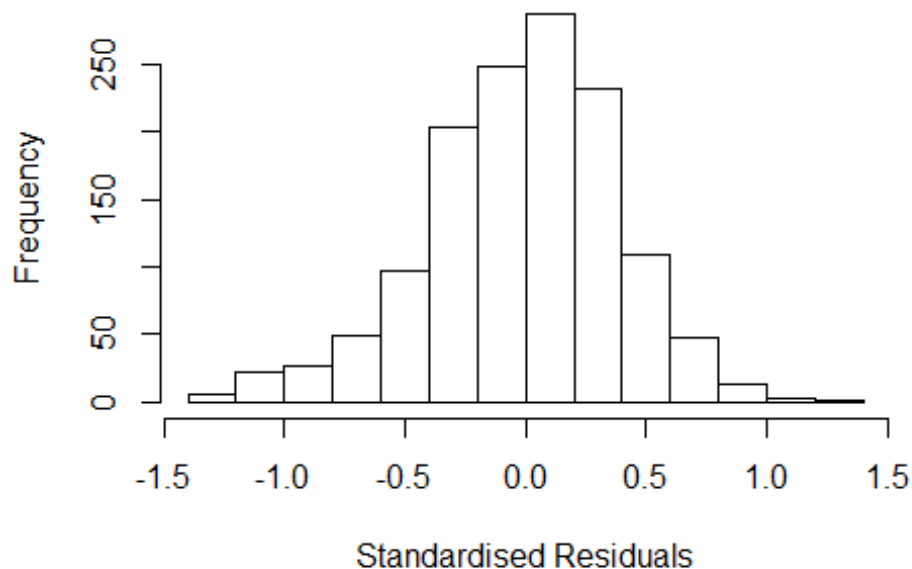


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.7, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 2.84779018119064"
## [1] "Male first author team size 2018 geometric mean: 2.98335633119354"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.84310494069997"
## [1] "Male last author team size 2018 geometric mean: 2.96064671810011"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, 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 1.123 1      1.060
## LastAuthorFemale  1.082 1      1.040
## UniqueAuthors    1.364 4      1.040
## Year              1.487 16     1.012
```

Residuals from first and last author and team size



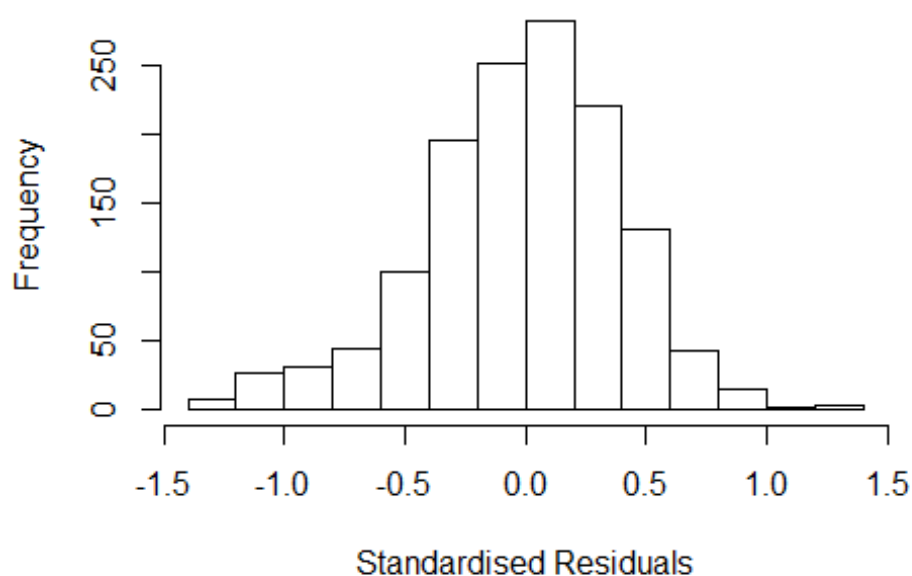
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2964 -0.2646  0.0157  0.2560  1.3082
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22826    0.06698   18.34 < 2e-16 ***
## FirstAuthorFemale1  0.04243    0.02688    1.58  0.1147
## LastAuthorFemale1  0.03422    0.03466    0.99  0.3237
## UniqueAuthors2   -0.02343    0.03309   -0.71  0.4791
## UniqueAuthors3   -0.00203    0.03734   -0.05  0.9566
## UniqueAuthors4    0.09592    0.03937    2.44  0.0150 *
## UniqueAuthors5    0.18519    0.04392    4.22 2.6e-05 ***
## Year1997         -0.10083    0.08563   -1.18  0.2392
## Year1998         -0.16397    0.08442   -1.94  0.0523 .
## Year1999          0.00390    0.07664    0.05  0.9594
```

```

## Year2000      -0.13813      0.07918      -1.74      0.0813 .
## Year2001      -0.26886      0.14377      -1.87      0.0617 .
## Year2002      -0.00851      0.08055      -0.11      0.9159
## Year2003      -0.17369      0.07350      -2.36      0.0183 *
## Year2004      -0.12683      0.08119      -1.56      0.1185
## Year2005      -0.07980      0.08137      -0.98      0.3269
## Year2006      -0.18480      0.07417      -2.49      0.0128 *
## Year2007      -0.22919      0.07394      -3.10      0.0020 **
## Year2008      -0.17967      0.07110      -2.53      0.0116 *
## Year2009      -0.15373      0.07356      -2.09      0.0368 *
## Year2010      -0.23471      0.07645      -3.07      0.0022 **
## Year2011      -0.22174      0.07333      -3.02      0.0025 **
## Year2012      -0.17244      0.07246      -2.38      0.0175 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0577, Adjusted R-squared:  0.0421
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 110 weights are ~= 1. The remaining 1238 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.213  0.871  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           7.42e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.089 1           1.043
## LastAuthorFemale  1.064 1           1.031
## Year              1.144 16           1.004

```

Residuals from first and last author



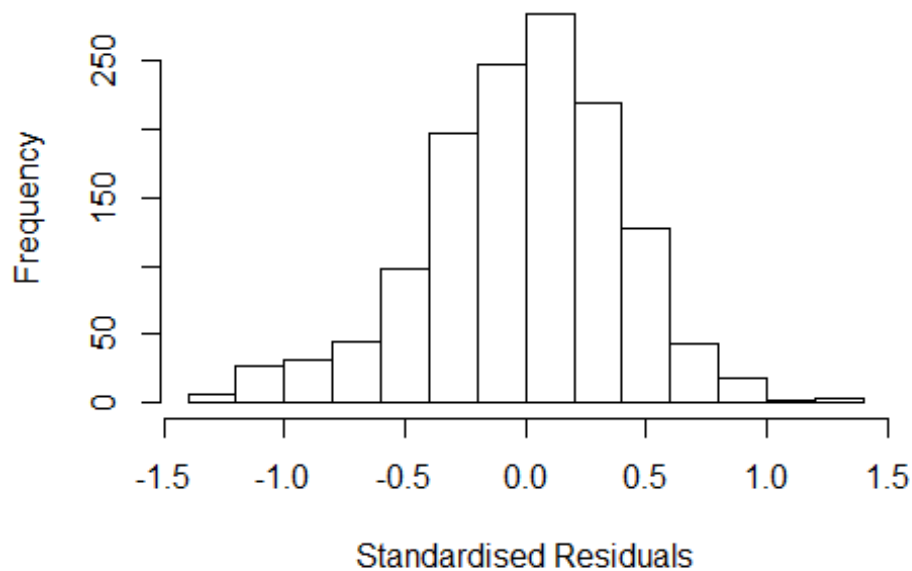
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3252 -0.2667  0.0127  0.2582  1.2898
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.23863    0.05872   21.09  <2e-16 ***
## FirstAuthorFemale1  0.04343    0.02737    1.59   0.1128
## LastAuthorFemale1  0.05378    0.03465    1.55   0.1209
## Year1997        -0.09117    0.08234   -1.11   0.2684
## Year1998        -0.18058    0.08150   -2.22   0.0269 *
## Year1999        -0.00628    0.07297   -0.09   0.9314
## Year2000        -0.13665    0.07576   -1.80   0.0715 .
## Year2001        -0.28221    0.14179   -1.99   0.0468 *
## Year2002        -0.01063    0.07727   -0.14   0.8906
## Year2003        -0.17023    0.07161   -2.38   0.0176 *
## Year2004        -0.12883    0.07809   -1.65   0.0992 .
## Year2005        -0.08960    0.07881   -1.14   0.2557
```

```

## Year2006          -0.16119      0.07149    -2.25    0.0243 *
## Year2007          -0.22648      0.07065    -3.21    0.0014 **
## Year2008          -0.16187      0.06858    -2.36    0.0184 *
## Year2009          -0.15396      0.07068    -2.18    0.0296 *
## Year2010          -0.21759      0.07419    -2.93    0.0034 **
## Year2011          -0.19799      0.07045    -2.81    0.0050 **
## Year2012          -0.13866      0.06911    -2.01    0.0450 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.0326, Adjusted R-squared:  0.0195
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1227 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.219  0.870   0.949   0.898   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      7.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.088 1      1.043
## Year              1.088 16      1.003

```

Residuals from first author



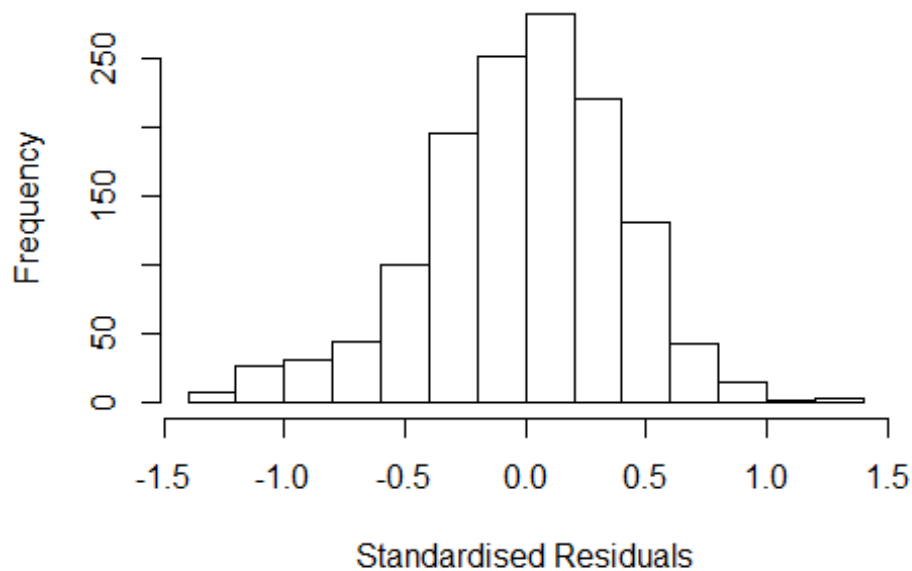
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2825 -0.2649  0.0114  0.2615  1.2894
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.24001    0.05852   21.19  <2e-16 ***
## FirstAuthorFemale1  0.05142    0.02762    1.86   0.0629 .
## Year1997        -0.09096    0.08231   -1.11   0.2693
## Year1998        -0.17796    0.08135   -2.19   0.0289 *
## Year1999        -0.00294    0.07291   -0.04   0.9679
## Year2000        -0.12510    0.07516   -1.66   0.0963 .
## Year2001        -0.27548    0.14034   -1.96   0.0499 *
## Year2002        -0.00897    0.07754   -0.12   0.9079
## Year2003        -0.16619    0.07138   -2.33   0.0201 *
## Year2004        -0.12778    0.07822   -1.63   0.1026
## Year2005        -0.08853    0.07853   -1.13   0.2598
## Year2006        -0.16033    0.07142   -2.24   0.0249 *
```

```

## Year2007          -0.22741    0.07053   -3.22    0.0013 **
## Year2008          -0.15679    0.06838   -2.29    0.0220 *
## Year2009          -0.14990    0.07056   -2.12    0.0338 *
## Year2010          -0.21550    0.07396   -2.91    0.0036 **
## Year2011          -0.19234    0.07058   -2.73    0.0065 **
## Year2012          -0.13324    0.06870   -1.94    0.0526 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.0305, Adjusted R-squared:  0.0181
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1232 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.256  0.874  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      7.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.06 1          1.030
## Year              1.06 16          1.002

```


Residuals from last author



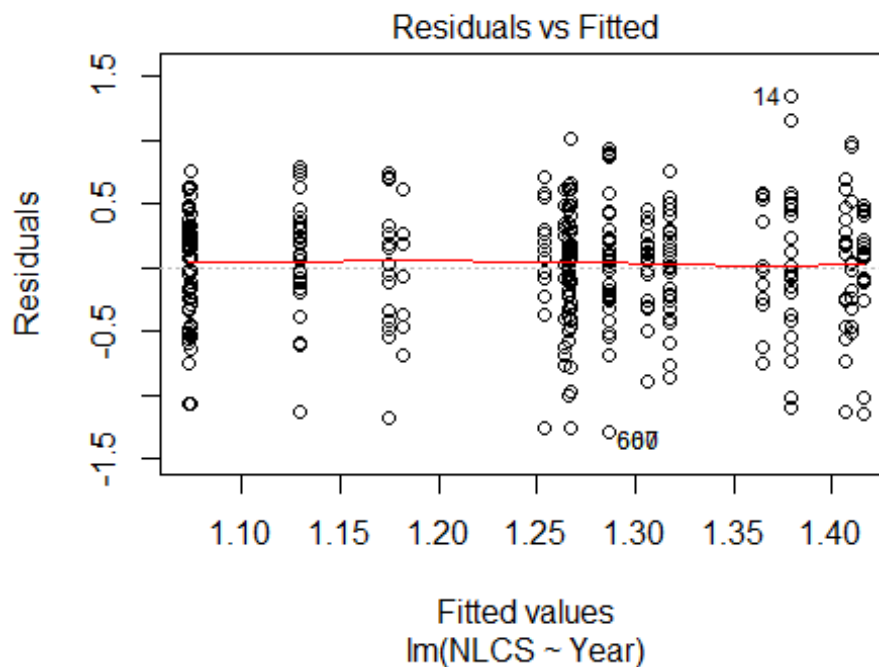
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29789 -0.26367  0.00857  0.25674  1.28123
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.24084    0.05859   21.18  <2e-16 ***
## LastAuthorFemale1 0.06367    0.03494    1.82  0.0686 .
## Year1997       -0.08986    0.08213   -1.09  0.2741
## Year1998       -0.17735    0.08159   -2.17  0.0299 *
## Year1999       -0.00603    0.07304   -0.08  0.9342
## Year2000       -0.13760    0.07593   -1.81  0.0702 .
## Year2001       -0.28478    0.14109   -2.02  0.0437 *
## Year2002       -0.00662    0.07724   -0.09  0.9317
## Year2003       -0.16610    0.07158   -2.32  0.0205 *
## Year2004       -0.12422    0.07778   -1.60  0.1105
## Year2005       -0.07881    0.07778   -1.01  0.3111
## Year2006       -0.15126    0.07095   -2.13  0.0332 *
```

```

## Year2007          -0.22007      0.07046    -3.12    0.0018 **
## Year2008          -0.15404      0.06816    -2.26    0.0240 *
## Year2009          -0.14599      0.07025    -2.08    0.0379 *
## Year2010          -0.21167      0.07398    -2.86    0.0043 **
## Year2011          -0.18824      0.07009    -2.69    0.0073 **
## Year2012          -0.12932      0.06843    -1.89    0.0590 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.0307, Adjusted R-squared:  0.0183
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 117 weights are ~= 1. The remaining 1231 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.240  0.871  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      7.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1348"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
##   49   25   26   31   36   34   28   26   37   34   47   49   45   54   43
## 2011 2012
##   46   64
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   25   11   13    9   16   14   15   16   21   17   28   29   29   32   24
## 2011 2012

```

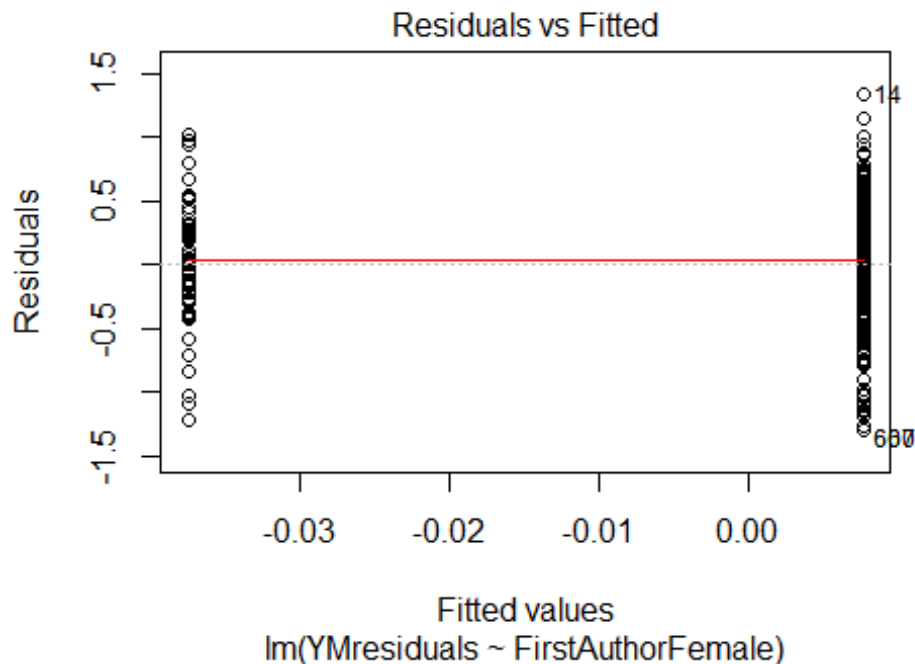
```
## 30 40
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 19 6 13 8 14 14 12 12 14 14 24 25 26 27 22
## 2011 2012
## 28 33
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 20, df = 16, p-value = 0.2
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.43, df = 1, p-value = 0.5
## [1] "Female first author team size 2018 geometric mean: 2.24416353346015"
## [1] "Male first author team size 2018 geometric mean: 2.85178850191061"
## Warning in wilcox.test.default(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 = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.98570940619587"
## [1] "Male last author team size 2018 geometric mean: 2.92961560642889"

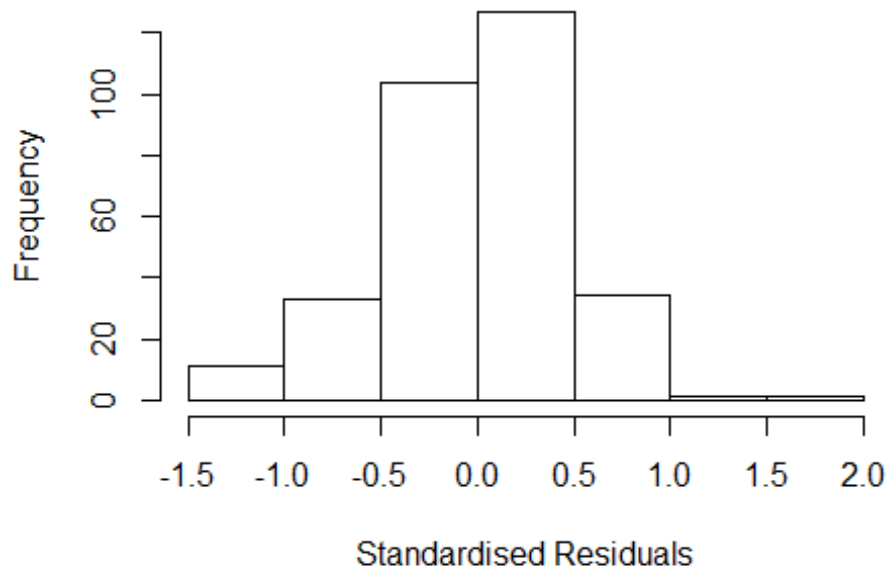
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 92, 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	1.435	1	1.198
LastAuthorFemale	1.366	1	1.169
UniqueAuthors	3.355	4	1.163
Year	5.202	16	1.053

Residuals from first and last author and team size



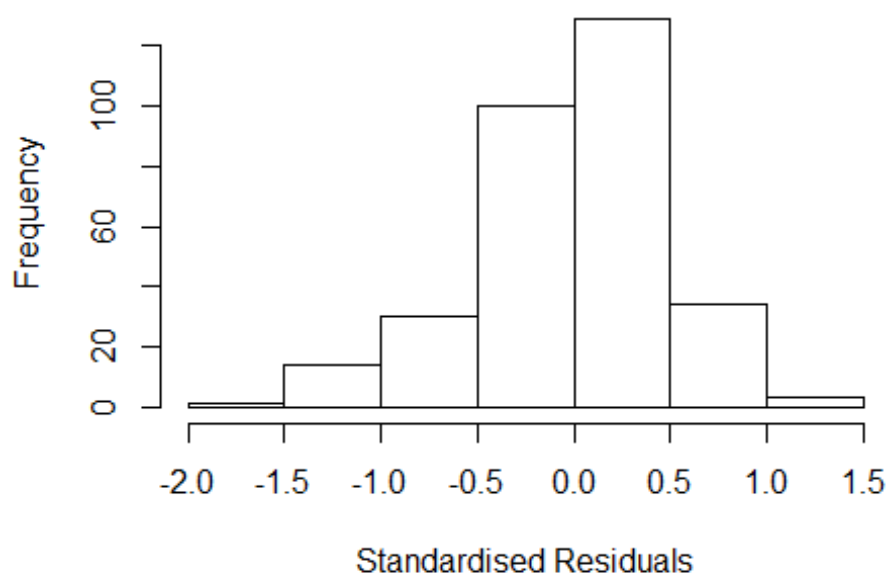
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32369 -0.28823  0.00924  0.28327  1.53360
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19040    0.19287   6.17 2.3e-09 ***
## FirstAuthorFemale1  0.00265    0.08090   0.03  0.9739
## LastAuthorFemale1  0.06953    0.07624   0.91  0.3626
## UniqueAuthors2    0.12211    0.09897   1.23  0.2183
## UniqueAuthors3    0.20826    0.08849   2.35  0.0193 *
## UniqueAuthors4    0.23436    0.09446   2.48  0.0137 *
## UniqueAuthors5    0.30003    0.11505   2.61  0.0096 **
## Year1997          0.07447    0.23720   0.31  0.7538
## Year1998          0.05376    0.22855   0.24  0.8142
## Year1999         -0.14861    0.24565  -0.60  0.5457
```

```

## Year2000      0.14837    0.22115    0.67    0.5028
## Year2001     -0.06838    0.21746   -0.31    0.7534
## Year2002     -0.24647    0.24286   -1.01    0.3110
## Year2003      0.01117    0.24153    0.05    0.9632
## Year2004     -0.03356    0.20168   -0.17    0.8680
## Year2005      0.05540    0.20321    0.27    0.7853
## Year2006     -0.08711    0.19025   -0.46    0.6474
## Year2007     -0.32418    0.20458   -1.58    0.1142
## Year2008     -0.05442    0.19467   -0.28    0.7800
## Year2009     -0.03558    0.20780   -0.17    0.8642
## Year2010     -0.00118    0.20378   -0.01    0.9954
## Year2011     -0.14964    0.19928   -0.75    0.4533
## Year2012     -0.28313    0.19214   -1.47    0.1417
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.105, Adjusted R-squared:  0.0365
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 279 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.181  0.860  0.952  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      3.22e-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 1.342 1      1.159
## LastAuthorFemale  1.273 1      1.128
## Year              1.666 16      1.016

```

Residuals from first and last author



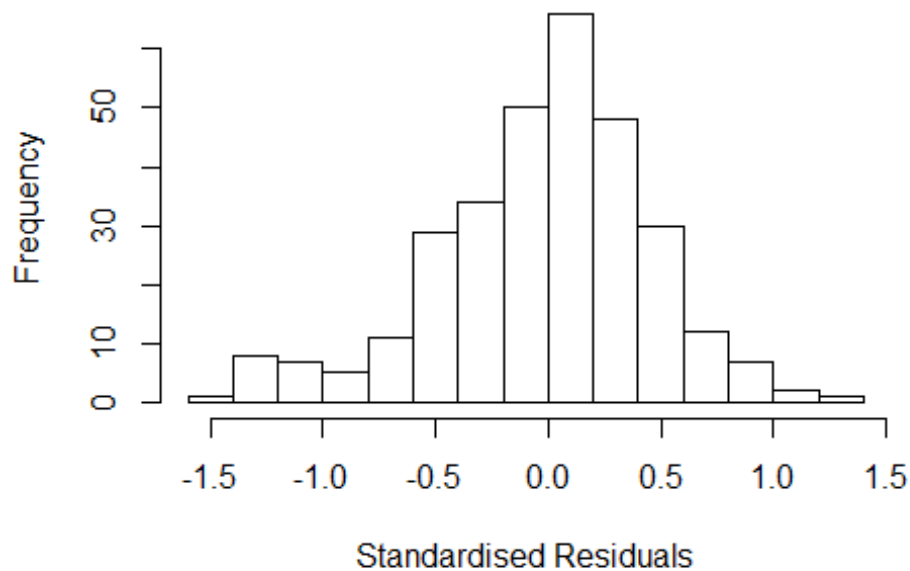
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.502 -0.318 0.027 0.275 1.394
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.329700 0.155568 8.55 7.1e-16 ***
## FirstAuthorFemale1 0.030634 0.081638 0.38 0.708
## LastAuthorFemale1 0.091033 0.076851 1.18 0.237
## Year1997 0.064292 0.206569 0.31 0.756
## Year1998 -0.000477 0.210183 0.00 0.998
## Year1999 -0.190101 0.229774 -0.83 0.409
## Year2000 0.164947 0.209511 0.79 0.432
## Year2001 -0.046658 0.203535 -0.23 0.819
## Year2002 -0.248695 0.226925 -1.10 0.274
## Year2003 0.050680 0.221136 0.23 0.819
## Year2004 -0.034758 0.182940 -0.19 0.849
## Year2005 0.129636 0.185112 0.70 0.484
```

```

## Year2006      -0.081050    0.168386   -0.48    0.631
## Year2007      -0.303540    0.182524   -1.66    0.097 .
## Year2008      -0.071860    0.175787   -0.41    0.683
## Year2009      -0.022038    0.187404   -0.12    0.906
## Year2010      -0.020896    0.189309   -0.11    0.912
## Year2011      -0.150071    0.181896   -0.83    0.410
## Year2012      -0.262815    0.173842   -1.51    0.132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0739, Adjusted R-squared:  0.0168
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 274 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.216  0.868  0.946  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      3.22e-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 1.319 1      1.148
## Year      1.319 16      1.009

```


Residuals from first author



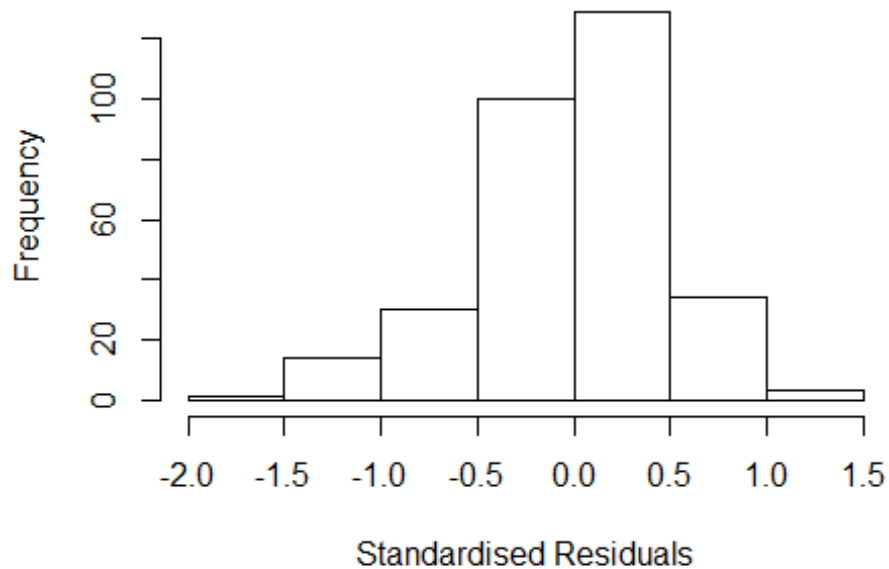
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4041 -0.3072  0.0255  0.2720  1.3945
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.32951    0.15570     8.54 7.5e-16 ***
## FirstAuthorFemale1 0.03261    0.08212     0.40  0.69
## Year1997         0.08119    0.20370     0.40  0.69
## Year1998         0.00787    0.20855     0.04  0.97
## Year1999        -0.19081    0.22983    -0.83  0.41
## Year2000         0.16551    0.20963     0.79  0.43
## Year2001        -0.04353    0.20163    -0.22  0.83
## Year2002        -0.23722    0.22752    -1.04  0.30
## Year2003         0.04201    0.22262     0.19  0.85
## Year2004        -0.01902    0.18189    -0.10  0.92
## Year2005         0.13859    0.18406     0.75  0.45
## Year2006        -0.06491    0.16815    -0.39  0.70
```

```

## Year2007      -0.28078    0.17873   -1.57    0.12
## Year2008      -0.05954    0.17444   -0.34    0.73
## Year2009      -0.01112    0.18668   -0.06    0.95
## Year2010      -0.00606    0.18983   -0.03    0.97
## Year2011      -0.13158    0.18203   -0.72    0.47
## Year2012      -0.25427    0.17431   -1.46    0.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.0699, Adjusted R-squared:  0.0159
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 283 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.280  0.862  0.948  0.892  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.22e-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 1.247 1          1.117
## Year            1.247 16          1.007

```

Residuals from last author



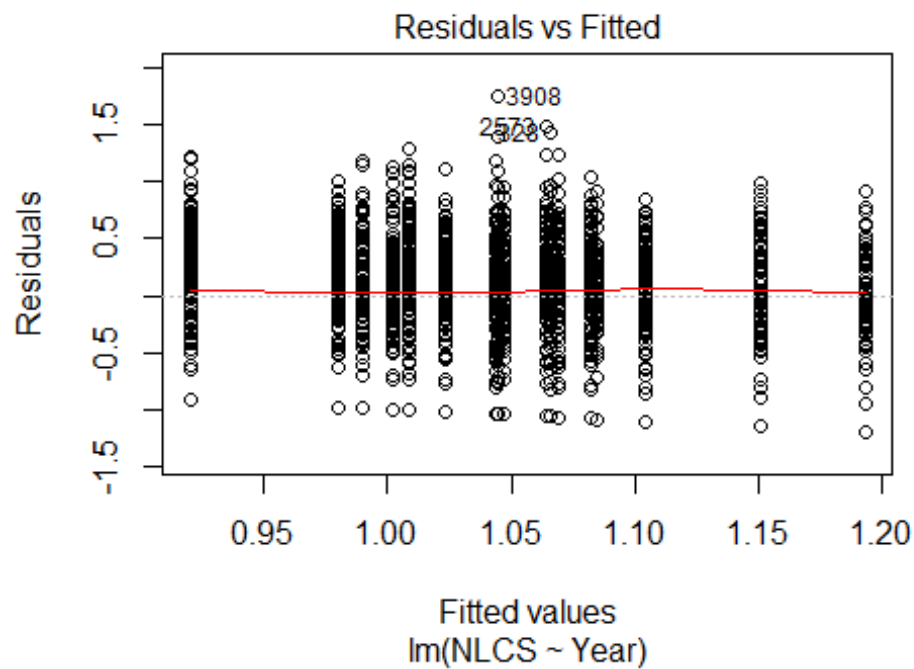
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4747 -0.3163 0.0228 0.2791 1.3918
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.33218 0.15604 8.54 7.6e-16 ***
## LastAuthorFemale1 0.09170 0.07712 1.19 0.24
## Year1997 0.06748 0.20679 0.33 0.74
## Year1998 -0.00146 0.21054 -0.01 0.99
## Year1999 -0.17846 0.22795 -0.78 0.43
## Year2000 0.16304 0.20995 0.78 0.44
## Year2001 -0.04587 0.20441 -0.22 0.82
## Year2002 -0.24382 0.22548 -1.08 0.28
## Year2003 0.05077 0.22202 0.23 0.82
## Year2004 -0.03495 0.18319 -0.19 0.85
## Year2005 0.13302 0.18636 0.71 0.48
## Year2006 -0.08207 0.16913 -0.49 0.63
```

```

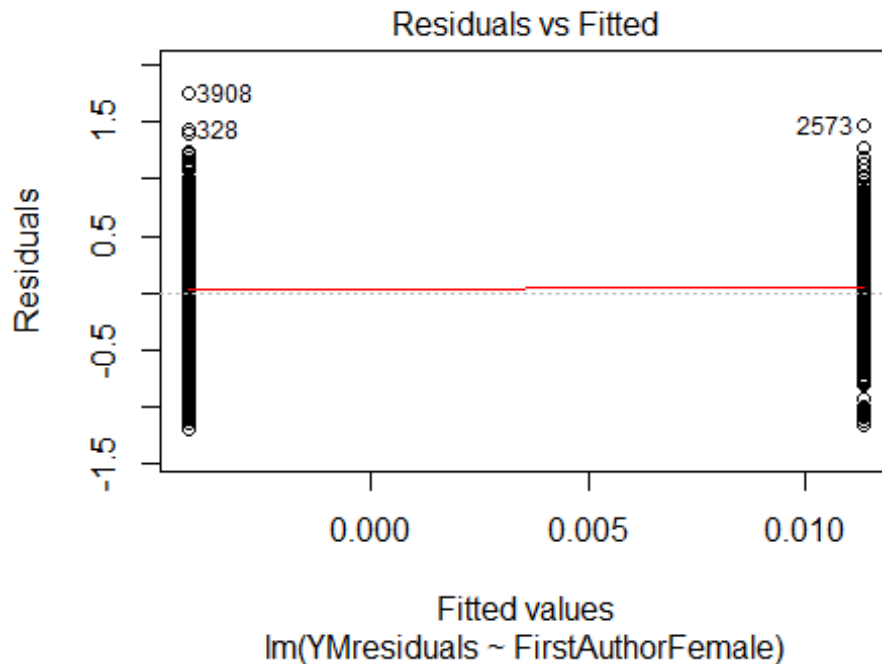
## Year2007          -0.29902      0.18281    -1.64      0.10
## Year2008          -0.06647      0.17636    -0.38      0.71
## Year2009          -0.02128      0.18837    -0.11      0.91
## Year2010          -0.01675      0.18934    -0.09      0.93
## Year2011          -0.14868      0.18325    -0.81      0.42
## Year2012          -0.25814      0.17499    -1.48      0.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.0733, Adjusted R-squared:  0.0196
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 35 weights are ~= 1. The remaining 276 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.230  0.864  0.944  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      3.22e-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: 311"
## [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
## 159 168 183 139 161 203 179 162 152 153 203 198 237 251 219
## 2011 2012
## 211 226
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 87 81 90 62 70 88 112 112 99 116 123 119 161 168 159
## 2011 2012

```

```
## 136 156
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 76 75 78 55 62 81 105 101 83 103 107 106 144 148 144
## 2011 2012
## 119 146
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.5
```

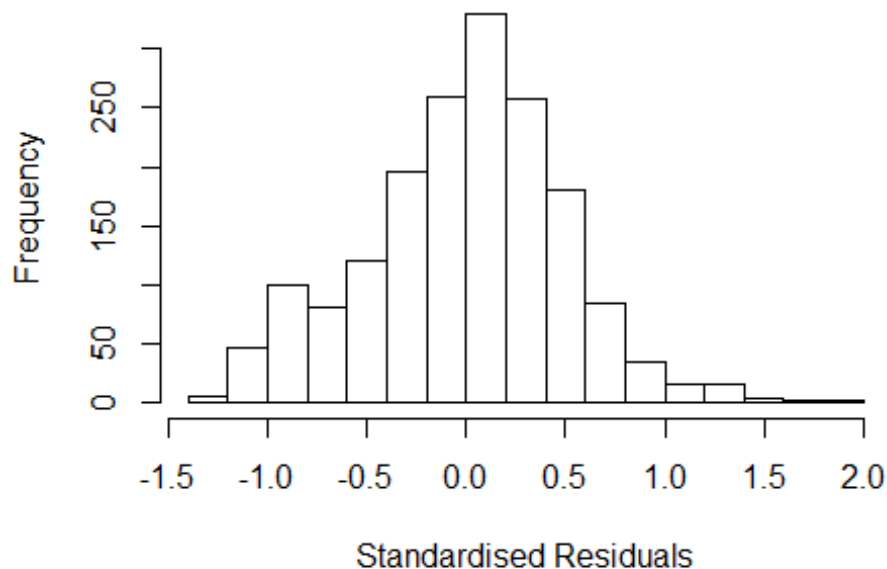


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.3, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 2.2869936228051"
## [1] "Male first author team size 2018 geometric mean: 2.37698367778376"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2400, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.19014121661784"
## [1] "Male last author team size 2018 geometric mean: 2.39797666404305"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, 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 1.126 1      1.061
## LastAuthorFemale  1.063 1      1.031
## UniqueAuthors    1.295 4      1.033
## Year              1.323 16     1.009
```

Residuals from first and last author and team size



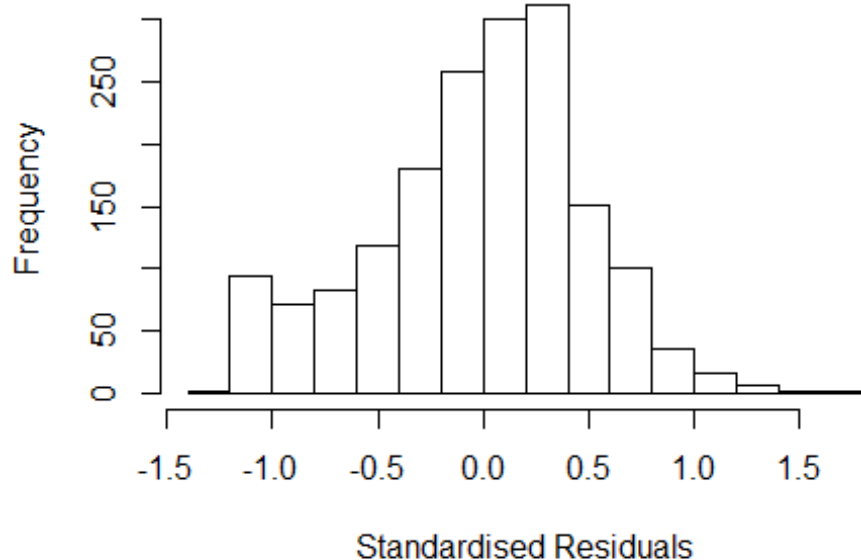
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2903 -0.3121 0.0362 0.3083 1.8870
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02350 0.06328 16.17 < 2e-16 ***
## FirstAuthorFemale1 0.03126 0.02658 1.18 0.23970
## LastAuthorFemale1 -0.11964 0.03536 -3.38 0.00073 ***
## UniqueAuthors2 0.21768 0.03459 6.29 3.9e-10 ***
## UniqueAuthors3 0.18664 0.03889 4.80 1.7e-06 ***
## UniqueAuthors4 0.17104 0.04297 3.98 7.2e-05 ***
## UniqueAuthors5 0.26798 0.04583 5.85 6.0e-09 ***
## Year1997 -0.09144 0.08140 -1.12 0.26144
## Year1998 -0.05851 0.08378 -0.70 0.48500
## Year1999 0.04620 0.08147 0.57 0.57079
```

```

## Year2000      -0.06358      0.07992      -0.80      0.42635
## Year2001      -0.10203      0.07703      -1.32      0.18549
## Year2002      -0.11167      0.07753      -1.44      0.14998
## Year2003      -0.10386      0.08190      -1.27      0.20491
## Year2004      -0.07655      0.08442      -0.91      0.36464
## Year2005      -0.00119      0.07382      -0.02      0.98718
## Year2006      -0.06198      0.07941      -0.78      0.43521
## Year2007      -0.07851      0.07935      -0.99      0.32260
## Year2008      -0.14988      0.07078      -2.12      0.03435 *
## Year2009      -0.20728      0.07813      -2.65      0.00805 **
## Year2010      -0.14934      0.07466      -2.00      0.04562 *
## Year2011      -0.11850      0.07538      -1.57      0.11612
## Year2012      -0.18428      0.07203      -2.56      0.01061 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.0678, Adjusted R-squared:  0.0558
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1597 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0491 0.8570 0.9490 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          5.77e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.095 1      1.046
## LastAuthorFemale  1.051 1      1.025
## Year              1.129 16      1.004

```


Residuals from first and last author



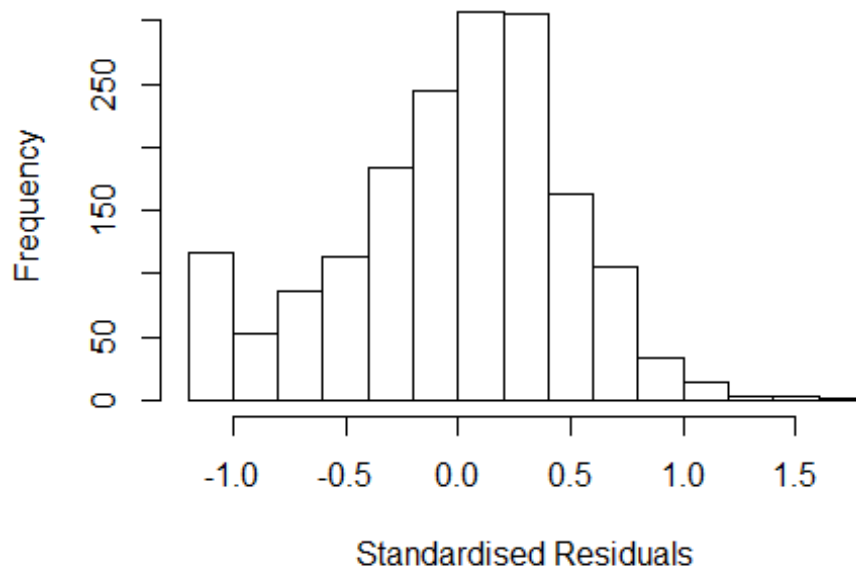
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2112 -0.3025 0.0424 0.2958 1.7328
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1629 0.0595 19.54 < 2e-16 ***
## FirstAuthorFemale1 0.0506 0.0269 1.88 0.061 .
## LastAuthorFemale1 -0.1473 0.0366 -4.02 6.1e-05 ***
## Year1997 -0.0818 0.0826 -0.99 0.322
## Year1998 -0.0954 0.0830 -1.15 0.251
## Year1999 0.0483 0.0822 0.59 0.557
## Year2000 -0.0567 0.0811 -0.70 0.484
## Year2001 -0.0703 0.0781 -0.90 0.368
## Year2002 -0.1200 0.0783 -1.53 0.125
## Year2003 -0.1252 0.0844 -1.48 0.138
## Year2004 -0.0607 0.0869 -0.70 0.485
## Year2005 0.0137 0.0751 0.18 0.856
```

```

## Year2006          -0.0487      0.0788   -0.62    0.536
## Year2007          -0.0671      0.0819   -0.82    0.412
## Year2008          -0.1357      0.0717   -1.89    0.059 .
## Year2009          -0.2001      0.0804   -2.49    0.013 *
## Year2010          -0.1448      0.0762   -1.90    0.057 .
## Year2011          -0.1038      0.0759   -1.37    0.172
## Year2012          -0.1496      0.0724   -2.07    0.039 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.0275, Adjusted R-squared:  0.0173
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1592 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.120  0.846  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      5.77e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.104 1      1.051
## Year              1.104 16      1.003

```

Residuals from first author



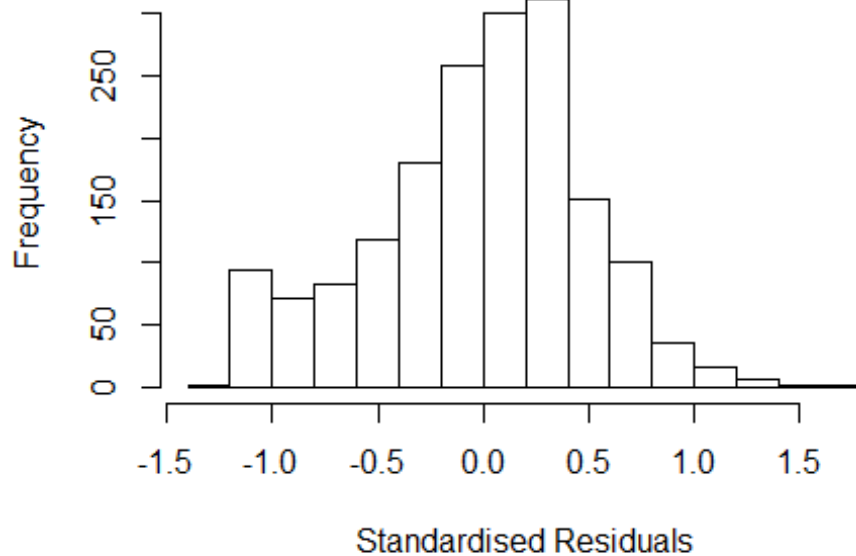
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1955 -0.3093  0.0427  0.3074  1.7586
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.16020    0.05978   19.41  <2e-16 ***
## FirstAuthorFemale1 0.01721    0.02966    0.58  0.5618
## Year1997       -0.09423    0.08255   -1.14  0.2538
## Year1998       -0.09848    0.08323   -1.18  0.2369
## Year1999        0.03532    0.08139    0.43  0.6643
## Year2000       -0.06198    0.08078   -0.77  0.4431
## Year2001       -0.08753    0.07935   -1.10  0.2702
## Year2002       -0.12647    0.07812   -1.62  0.1056
## Year2003       -0.15279    0.08472   -1.80  0.0715 .
## Year2004       -0.07070    0.08728   -0.81  0.4180
## Year2005       -0.00409    0.07582   -0.05  0.9570
## Year2006       -0.05890    0.07906   -0.75  0.4563
```

```

## Year2007          -0.07957    0.08209   -0.97    0.3325
## Year2008          -0.15109    0.07218   -2.09    0.0365 *
## Year2009          -0.22132    0.08120   -2.73    0.0065 **
## Year2010          -0.15786    0.07685   -2.05    0.0401 *
## Year2011          -0.12679    0.07557   -1.68    0.0936 .
## Year2012          -0.16290    0.07237   -2.25    0.0245 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.46
## Multiple R-squared:  0.0161, Adjusted R-squared:  0.00634
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1604 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.111  0.846  0.952  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      5.77e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.045 1          1.022
## Year            1.045 16          1.001

```

Residuals from last author



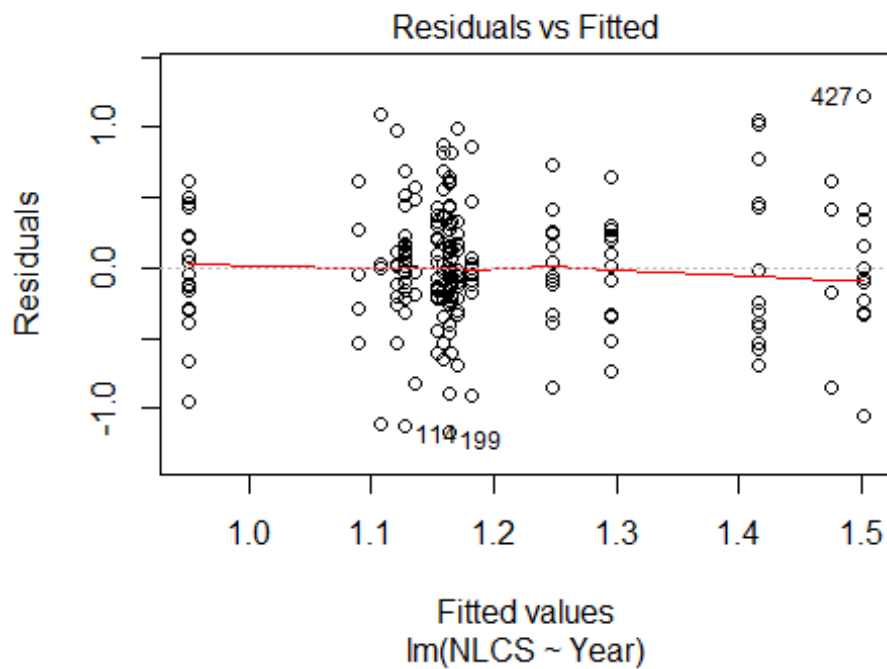
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2245 -0.3098 0.0462 0.2992 1.7212
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1672 0.0595 19.61 < 2e-16 ***
## LastAuthorFemale1 -0.1313 0.0383 -3.43 0.00062 ***
## Year1997 -0.0799 0.0827 -0.97 0.33420
## Year1998 -0.0917 0.0828 -1.11 0.26843
## Year1999 0.0573 0.0820 0.70 0.48477
## Year2000 -0.0455 0.0809 -0.56 0.57397
## Year2001 -0.0673 0.0784 -0.86 0.39094
## Year2002 -0.1126 0.0782 -1.44 0.15018
## Year2003 -0.1217 0.0845 -1.44 0.14996
## Year2004 -0.0606 0.0866 -0.70 0.48413
## Year2005 0.0197 0.0752 0.26 0.79336
## Year2006 -0.0411 0.0788 -0.52 0.60225
```

```

## Year2007          -0.0577      0.0816   -0.71  0.47965
## Year2008          -0.1294      0.0716   -1.81  0.07079 .
## Year2009          -0.1919      0.0801   -2.40  0.01671 *
## Year2010          -0.1374      0.0760   -1.81  0.07073 .
## Year2011          -0.0964      0.0757   -1.27  0.20291
## Year2012          -0.1374      0.0717   -1.92  0.05557 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.458
## Multiple R-squared:  0.0265, Adjusted R-squared:  0.0169
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 130 weights are ~= 1. The remaining 1603 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.126  0.851  0.952  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      5.77e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1733"
## [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
##   38   21   48   20   18   23   35    8   10   17   41   24   21   16   18
## 2011 2012
##   15   29
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   16   12   23    5    4   13   22    4    5   11   21   14   10   11    8
## 2011 2012

```

```
## 13 21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 9 17 5 4 10 21 4 4 11 18 10 8 9 8
## 2011 2012
## 13 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 = 18, df = 16, p-value = 0.3
```



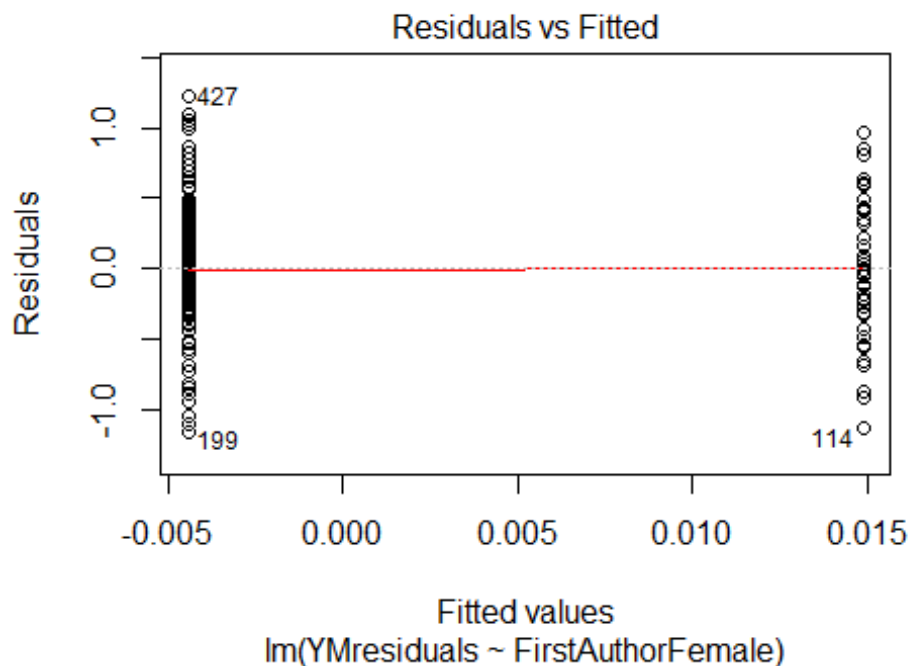
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.2, df = 1, p-value = 0.3

## [1] "Female first author team size 2018 geometric mean: 2.9653098171939"
## [1] "Male first author team size 2018 geometric mean: 2.3284355309218"

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

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 72, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.63214802590498"
## [1] "Male last author team size 2018 geometric mean: 2.48050943289873"

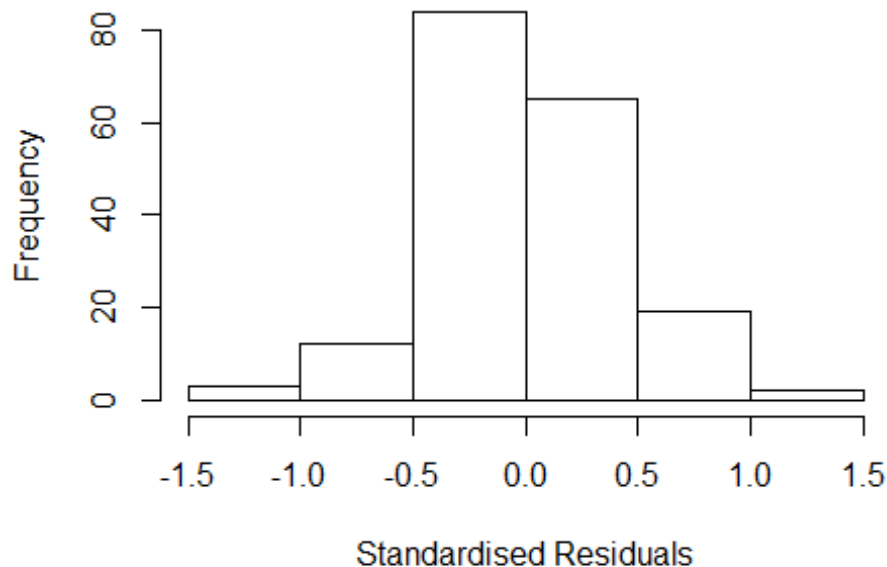
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



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

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	1.546	1	1.243
LastAuthorFemale	1.697	1	1.303
UniqueAuthors	15.683	4	1.411
Year	25.526	16	1.107

Residuals from first and last author and team size



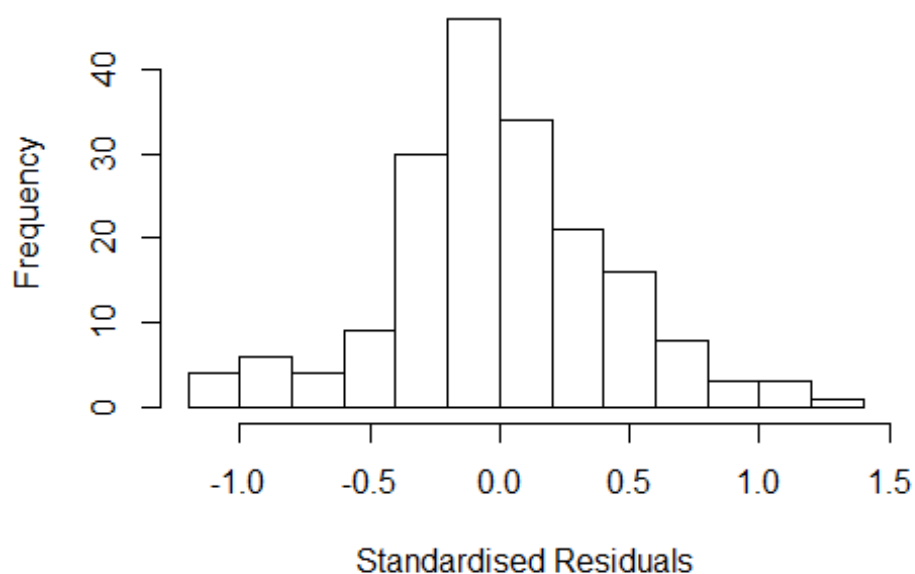
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4094 -0.2208 -0.0254 0.2120 1.2895
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26919 0.12927 9.82 <2e-16 ***
## FirstAuthorFemale1 0.01199 0.08606 0.14 0.889
## LastAuthorFemale1 0.00807 0.08817 0.09 0.927
## UniqueAuthors2 -0.14153 0.13675 -1.03 0.302
## UniqueAuthors3 -0.08768 0.12953 -0.68 0.499
## UniqueAuthors4 0.01743 0.11914 0.15 0.884
## UniqueAuthors5 0.02823 0.12773 0.22 0.825
## Year1997 0.12196 0.16177 0.75 0.452
## Year1998 -0.07776 0.10875 -0.71 0.476
## Year1999 0.01074 0.26073 0.04 0.967
```

```

## Year2000      0.14018      0.67418      0.21      0.836
## Year2001      0.02940      0.15706      0.19      0.852
## Year2002      0.02718      0.12858      0.21      0.833
## Year2003      0.33055      0.36304      0.91      0.364
## Year2004     -0.19449      0.26651     -0.73      0.467
## Year2005     -0.06393      0.16179     -0.40      0.693
## Year2006     -0.27188      0.13373     -2.03      0.044 *
## Year2007      0.01586      0.19582      0.08      0.936
## Year2008     -0.06495      0.16586     -0.39      0.696
## Year2009      0.20058      0.16325      1.23      0.221
## Year2010     -0.17588      0.14496     -1.21      0.227
## Year2011     -0.08728      0.12706     -0.69      0.493
## Year2012     -0.21345      0.13334     -1.60      0.111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.126, Adjusted R-squared:  0.00777
## Convergence in 38 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 170 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0867 0.8640 0.9540 0.8770 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      5.41e-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 1.391 1      1.179
## LastAuthorFemale 1.812 1      1.346
## Year      2.514 16      1.029

```

Residuals from first and last author



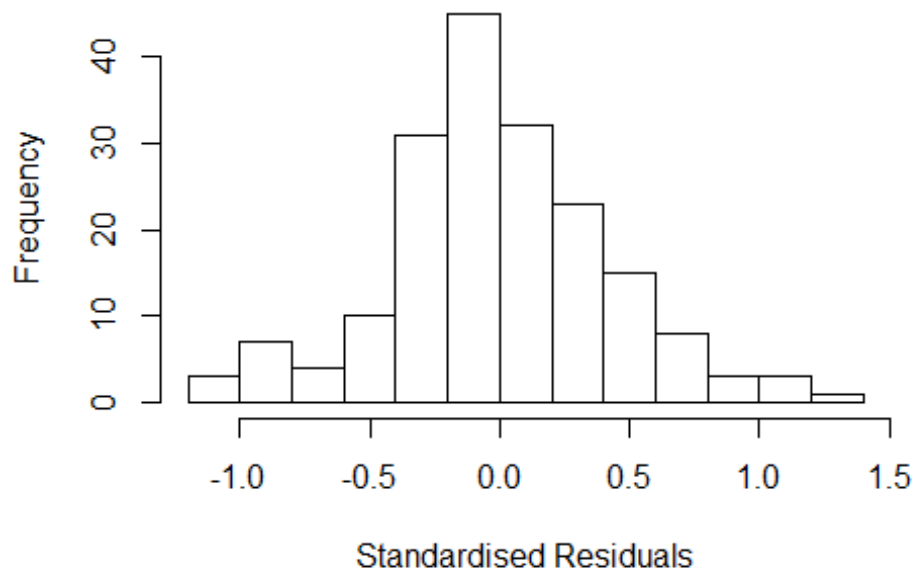
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1840 -0.2367 -0.0253 0.2349 1.2292
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17469 0.08320 14.12 <2e-16 ***
## FirstAuthorFemale1 0.04058 0.08708 0.47 0.642
## LastAuthorFemale1 -0.02798 0.09776 -0.29 0.775
## Year1997 0.16339 0.16396 1.00 0.320
## Year1998 -0.01893 0.10536 -0.18 0.858
## Year1999 0.02137 0.30039 0.07 0.943
## Year2000 -0.03877 0.81168 -0.05 0.962
## Year2001 0.05500 0.14948 0.37 0.713
## Year2002 0.00928 0.12256 0.08 0.940
## Year2003 0.42170 0.43837 0.96 0.337
## Year2004 -0.22584 0.26422 -0.85 0.394
## Year2005 0.01572 0.13360 0.12 0.906
```

```

## Year2006      -0.22856    0.13376   -1.71    0.089 .
## Year2007      0.02914    0.20359    0.14    0.886
## Year2008      0.01374    0.14280    0.10    0.923
## Year2009      0.24682    0.15196    1.62    0.106
## Year2010     -0.13616    0.13974   -0.97    0.331
## Year2011     -0.04342    0.12153   -0.36    0.721
## Year2012     -0.13740    0.12193   -1.13    0.261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.357
## Multiple R-squared:  0.101, Adjusted R-squared:  0.00397
## Convergence in 46 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 171 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.212  0.856  0.954  0.876  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      5.41e-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 1.327 1      1.152
## Year              1.327 16      1.009

```

Residuals from first author



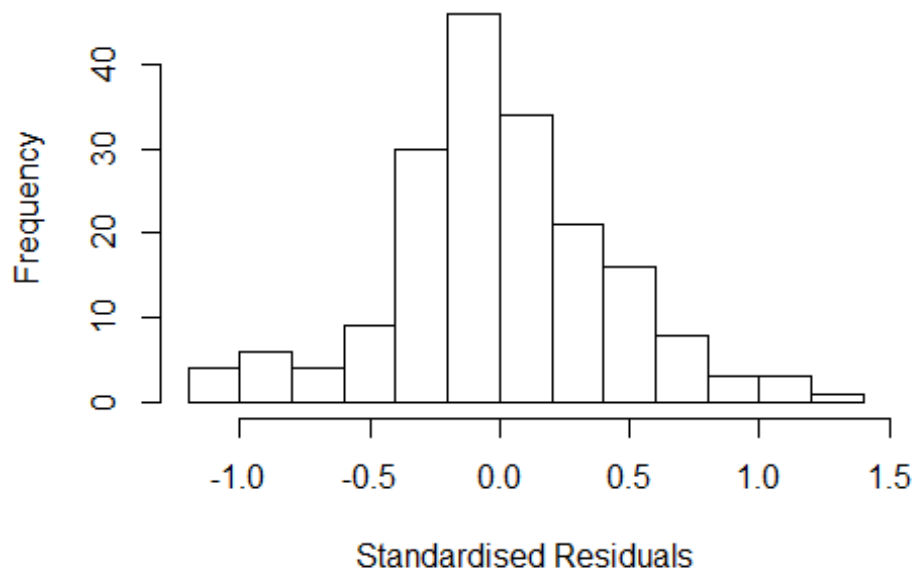
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1793 -0.2344 -0.0361 0.2429 1.2100
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16963 0.08005 14.61 <2e-16 ***
## FirstAuthorFemale1 0.03687 0.08590 0.43 0.67
## Year1997 0.15889 0.16557 0.96 0.34
## Year1998 -0.02724 0.10940 -0.25 0.80
## Year1999 0.01608 0.28683 0.06 0.96
## Year2000 -0.02379 0.67139 -0.04 0.97
## Year2001 0.05431 0.14887 0.36 0.72
## Year2002 0.00425 0.12280 0.03 0.97
## Year2003 0.39139 0.40672 0.96 0.34
## Year2004 -0.21144 0.25746 -0.82 0.41
## Year2005 0.01155 0.13568 0.09 0.93
## Year2006 -0.23550 0.13350 -1.76 0.08 .
```

```

## Year2007          0.05332    0.19730    0.27    0.79
## Year2008          0.02451    0.13999    0.18    0.86
## Year2009          0.24351    0.15134    1.61    0.11
## Year2010         -0.12307    0.14124   -0.87    0.38
## Year2011         -0.04037    0.12263   -0.33    0.74
## Year2012         -0.13204    0.12199   -1.08    0.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0913, Adjusted R-squared:  -0.00119
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 172 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.293  0.872  0.960  0.890  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      5.41e-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 1.777 1          1.333
## Year            1.777 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
## -1.1988 -0.2385 -0.0363 0.2221 1.2245
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18013 0.08361 14.11 <2e-16 ***
## LastAuthorFemale1 -0.02480 0.09886 -0.25 0.80
## Year1997 0.15798 0.16409 0.96 0.34
## Year1998 -0.00879 0.10417 -0.08 0.93
## Year1999 0.01698 0.30143 0.06 0.96
## Year2000 -0.04037 0.81517 -0.05 0.96
## Year2001 0.04944 0.14974 0.33 0.74
## Year2002 0.01872 0.12470 0.15 0.88
## Year2003 0.41776 0.43992 0.95 0.34
## Year2004 -0.20694 0.26274 -0.79 0.43
## Year2005 0.00724 0.13049 0.06 0.96
## Year2006 -0.22128 0.13382 -1.65 0.10
```

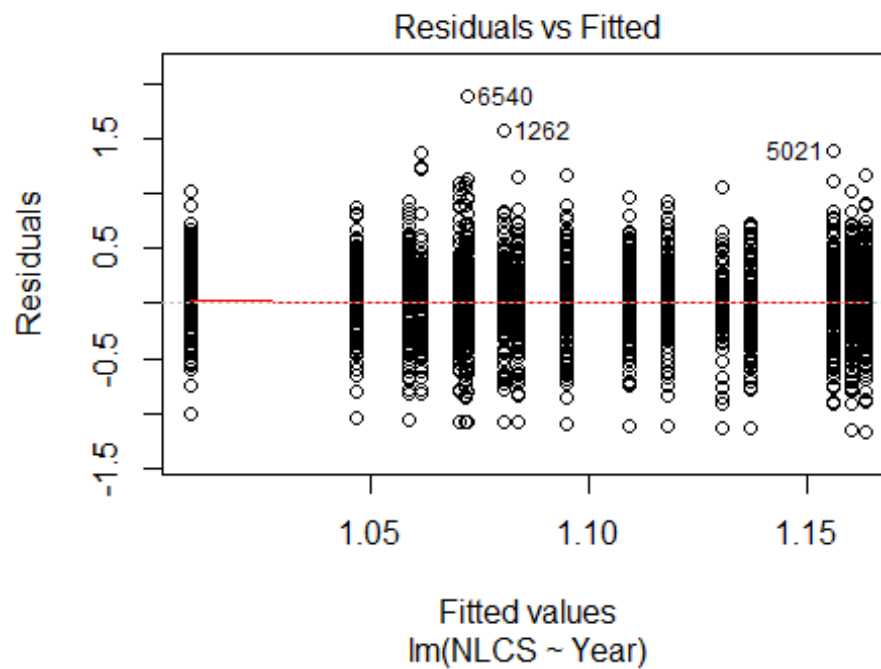
```

## Year2007      0.02836      0.20390      0.14      0.89
## Year2008      0.01340      0.14231      0.09      0.93
## Year2009      0.25176      0.15452      1.63      0.11
## Year2010     -0.13339      0.14012     -0.95      0.34
## Year2011     -0.03928      0.12418     -0.32      0.75
## Year2012     -0.13404      0.12299     -1.09      0.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.355
## Multiple R-squared:  0.1, Adjusted R-squared:  0.00842
## Convergence in 47 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 172 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.210  0.850  0.952  0.875  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+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-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 185"
## [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
## 338 301 268 261 250 284 285 253 251 232 303 258 283 296 249
## 2011 2012
## 253 258
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 184 161 133 113 75 96 159 135 149 138 183 144 159 178 143
## 2011 2012

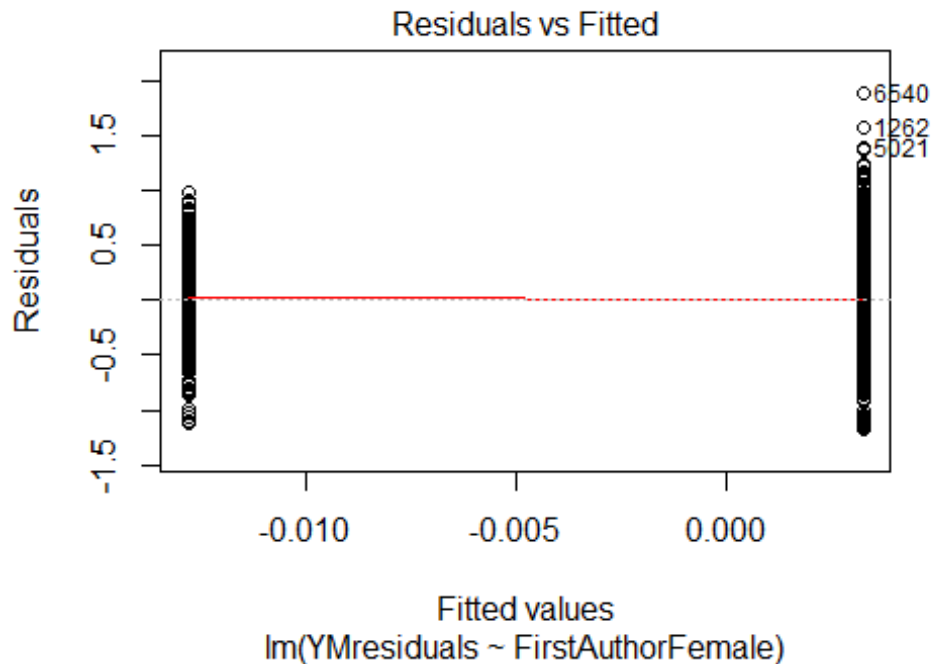
```



```
## 157 158
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 160 138 120 103 62 81 138 102 129 113 158 112 130 148 123
## 2011 2012
## 133 133
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 10, df = 16, p-value = 0.8
```

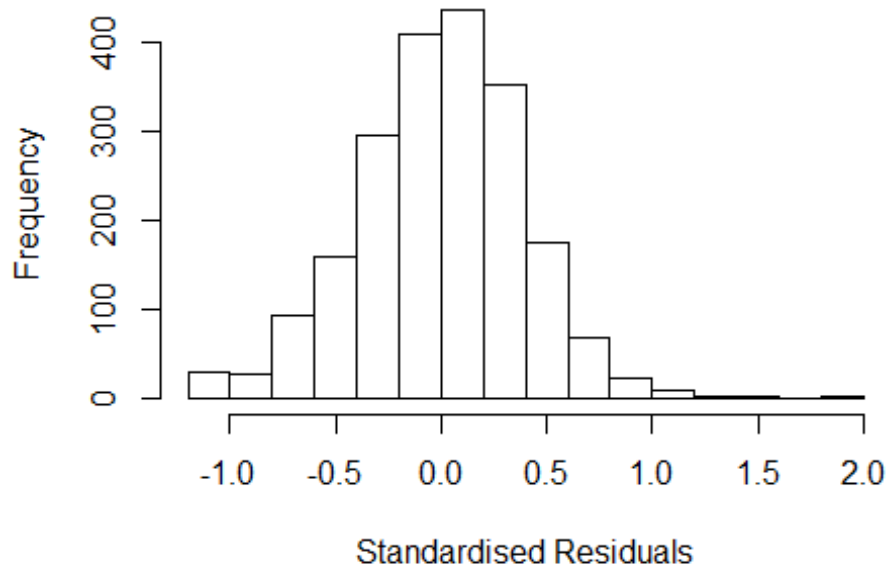


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.8, df = 1, p-value = 0.09
```



```
## [1] "Female first author team size 2018 geometric mean: 2.9379470432908"
## [1] "Male first author team size 2018 geometric mean: 2.97040532194174"
##
## 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: 2.9628866053996"
## [1] "Male last author team size 2018 geometric mean: 2.9577534428385"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 700, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "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  1.076 1      1.037
## UniqueAuthors    1.213 4      1.024
## Year              1.292 16     1.008
```

Residuals from first and last author and team size



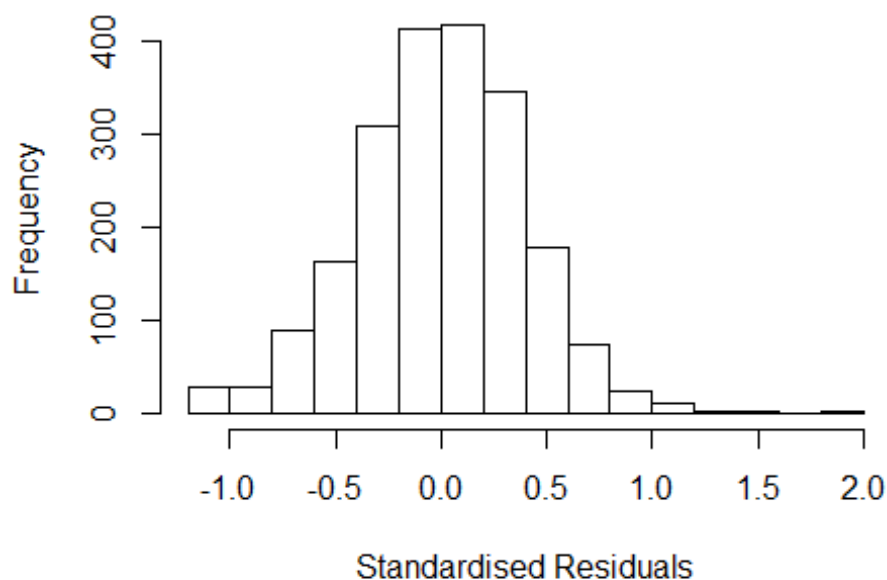
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17189 -0.24555  0.00982  0.25345  1.91578
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19217    0.03984   29.92 < 2e-16 ***
## FirstAuthorFemale1 -0.00852    0.02099   -0.41  0.68478
## LastAuthorFemale1  0.01767    0.02569    0.69  0.49170
## UniqueAuthors2   -0.03163    0.02636   -1.20  0.23029
## UniqueAuthors3   -0.00861    0.02880   -0.30  0.76501
## UniqueAuthors4    0.06419    0.03179    2.02  0.04358 *
## UniqueAuthors5    0.05757    0.03789    1.52  0.12878
## Year1997         -0.08355    0.04904   -1.70  0.08860 .
## Year1998         -0.12431    0.04909   -2.53  0.01141 *
## Year1999         -0.12349    0.05226   -2.36  0.01821 *
```

```

## Year2000      -0.07360      0.05346      -1.38      0.16876
## Year2001      -0.05051      0.05925      -0.85      0.39408
## Year2002      -0.02028      0.05141      -0.39      0.69329
## Year2003      -0.09813      0.05116      -1.92      0.05522 .
## Year2004      -0.07717      0.04708      -1.64      0.10138
## Year2005      -0.05044      0.05176      -0.97      0.32987
## Year2006      -0.14048      0.04650      -3.02      0.00255 **
## Year2007      -0.15170      0.04766      -3.18      0.00148 **
## Year2008      -0.05262      0.04753      -1.11      0.26833
## Year2009      -0.12803      0.04737      -2.70      0.00693 **
## Year2010      -0.19554      0.05145      -3.80      0.00015 ***
## Year2011      -0.14543      0.04646      -3.13      0.00177 **
## Year2012      -0.14234      0.04849      -2.94      0.00337 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0248, Adjusted R-squared:  0.0143
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2022 is an outlier with |weight| = 0 ( < 4.8e-05);
## 168 weights are ~= 1. The remaining 1914 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0383 0.8700 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      4.80e-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 1.060 1      1.030
## LastAuthorFemale 1.062 1      1.031
## Year      1.103 16      1.003

```

Residuals from first and last author



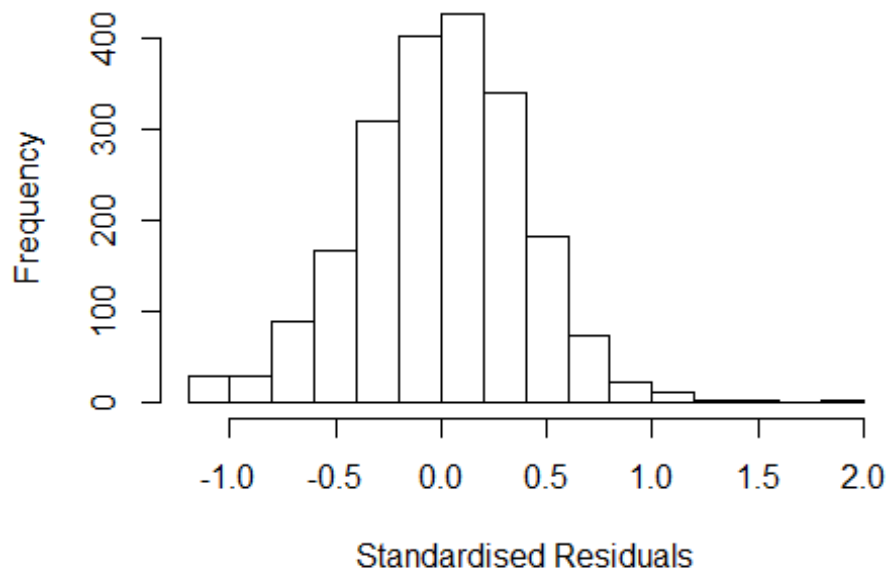
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17787 -0.24927  0.00462  0.25877  1.90458
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.17787    0.03454   34.10 < 2e-16 ***
## FirstAuthorFemale1 -0.01118    0.02097   -0.53  0.59394
## LastAuthorFemale1  0.02250    0.02583    0.87  0.38376
## Year1997          -0.07746    0.04931   -1.57  0.11633
## Year1998          -0.11160    0.04873   -2.29  0.02212 *
## Year1999          -0.11949    0.05214   -2.29  0.02203 *
## Year2000          -0.06001    0.05302   -1.13  0.25782
## Year2001          -0.04484    0.05910   -0.76  0.44810
## Year2002          -0.00744    0.05107   -0.15  0.88417
## Year2003          -0.08647    0.05117   -1.69  0.09121 .
## Year2004          -0.06661    0.04704   -1.42  0.15693
## Year2005          -0.03568    0.05145   -0.69  0.48807
```

```

## Year2006          -0.12431      0.04624      -2.69   0.00724 **
## Year2007          -0.14033      0.04697      -2.99   0.00284 **
## Year2008          -0.03584      0.04700      -0.76   0.44575
## Year2009          -0.10652      0.04632      -2.30   0.02157 *
## Year2010          -0.17829      0.05145      -3.47   0.00054 ***
## Year2011          -0.12596      0.04641      -2.71   0.00670 **
## Year2012          -0.12545      0.04799      -2.61   0.00901 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0173, Adjusted R-squared:  0.00868
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2022 is an outlier with |weight| = 0 ( < 4.8e-05);
## 173 weights are ~ = 1. The remaining 1909 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0377 0.8710 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          4.80e-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 1.047 1          1.023
## Year              1.047 16          1.001

```

Residuals from first author



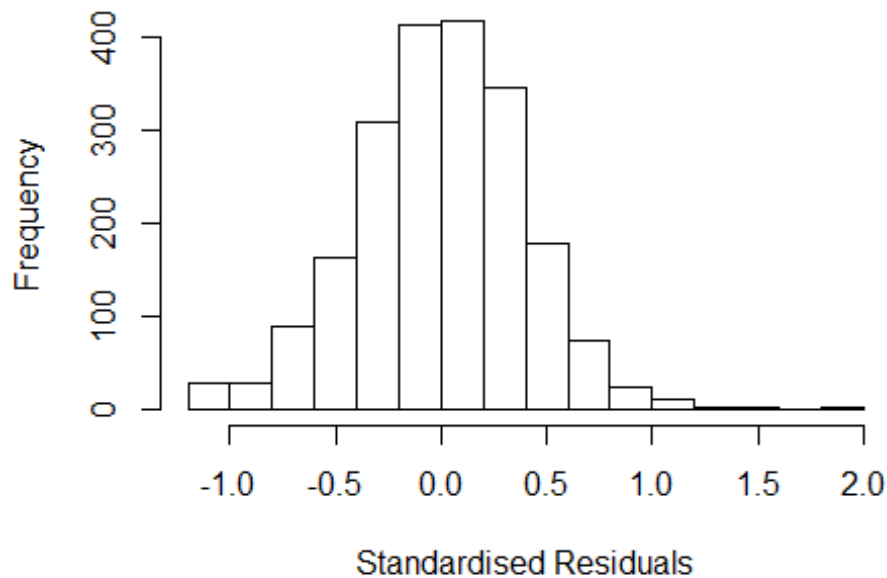
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1803 -0.2493 0.0067 0.2578 1.9024
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18030 0.03419 34.53 < 2e-16 ***
## FirstAuthorFemale1 -0.00803 0.02096 -0.38 0.70154
## Year1997 -0.07812 0.04926 -1.59 0.11294
## Year1998 -0.11217 0.04868 -2.30 0.02130 *
## Year1999 -0.11931 0.05203 -2.29 0.02194 *
## Year2000 -0.05796 0.05315 -1.09 0.27560
## Year2001 -0.04357 0.05903 -0.74 0.46056
## Year2002 -0.00754 0.05107 -0.15 0.88258
## Year2003 -0.08721 0.05121 -1.70 0.08870 .
## Year2004 -0.06742 0.04702 -1.43 0.15180
## Year2005 -0.03391 0.05139 -0.66 0.50937
## Year2006 -0.12545 0.04610 -2.72 0.00655 **
```

```

## Year2007          -0.13950      0.04702      -2.97      0.00305 **
## Year2008          -0.03560      0.04695      -0.76      0.44843
## Year2009          -0.10825      0.04603      -2.35      0.01878 *
## Year2010          -0.17893      0.05123      -3.49      0.00049 ***
## Year2011          -0.12585      0.04645      -2.71      0.00680 **
## Year2012          -0.12574      0.04787      -2.63      0.00869 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0169, Adjusted R-squared:  0.00882
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2022 is an outlier with |weight| = 0 ( < 4.8e-05);
## 172 weights are ~= 1. The remaining 1910 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0387 0.8720 0.9510 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          4.80e-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 1.052 1          1.026
## Year            1.052 16          1.002

```


Residuals from last author



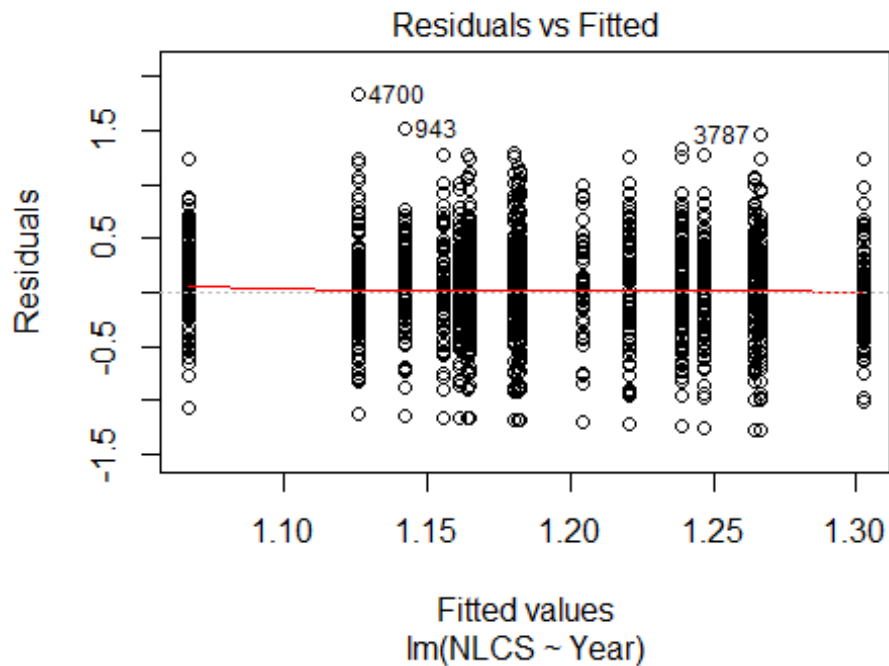
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17655 -0.24807  0.00511  0.25632  1.90823
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1765     0.0344   34.24  <2e-16 ***
## LastAuthorFemale1  0.0203     0.0257    0.79   0.4307
## Year1997        -0.0772     0.0493   -1.57   0.1173
## Year1998        -0.1121     0.0487   -2.30   0.0215 *
## Year1999        -0.1193     0.0522   -2.29   0.0223 *
## Year2000        -0.0601     0.0529   -1.14   0.2562
## Year2001        -0.0443     0.0592   -0.75   0.4541
## Year2002        -0.0083     0.0511   -0.16   0.8709
## Year2003        -0.0868     0.0512   -1.69   0.0903 .
## Year2004        -0.0668     0.0471   -1.42   0.1558
## Year2005        -0.0366     0.0513   -0.71   0.4760
## Year2006        -0.1257     0.0462   -2.72   0.0066 **
```

```

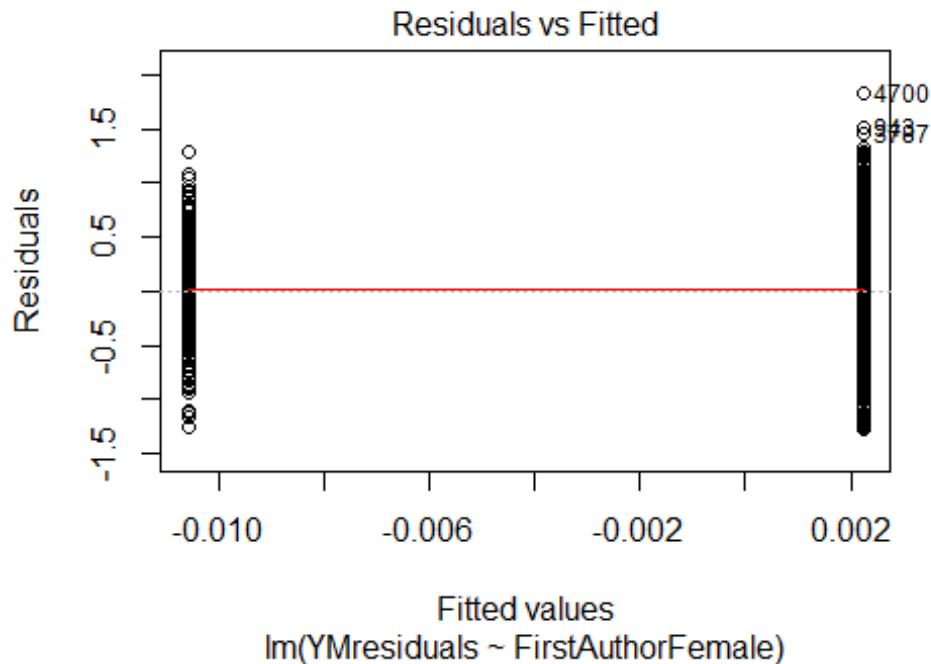
## Year2007          -0.1412      0.0469   -3.01   0.0027 **
## Year2008          -0.0375      0.0467   -0.80   0.4218
## Year2009          -0.1070      0.0463   -2.31   0.0210 *
## Year2010          -0.1792      0.0514   -3.48   0.0005 ***
## Year2011          -0.1275      0.0464   -2.75   0.0061 **
## Year2012          -0.1278      0.0478   -2.67   0.0076 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0171, Adjusted R-squared:  0.00903
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2022 is an outlier with |weight| = 0 ( < 4.8e-05);
## 176 weights are ~= 1. The remaining 1906 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0368 0.8700 0.9510 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      4.80e-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: 2083"
## [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
## 246 214 212 233 213 261 238 145 193 159 221 218 186 195 200
## 2011 2012
## 184 197
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 132 118 107 98 48 104 141 80 118 99 120 124 107 118 115

```

```
## 2011 2012
## 116 140
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 108 104 92 86 45 84 128 68 104 80 98 102 86 107 97
## 2011 2012
## 100 114
## [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
```

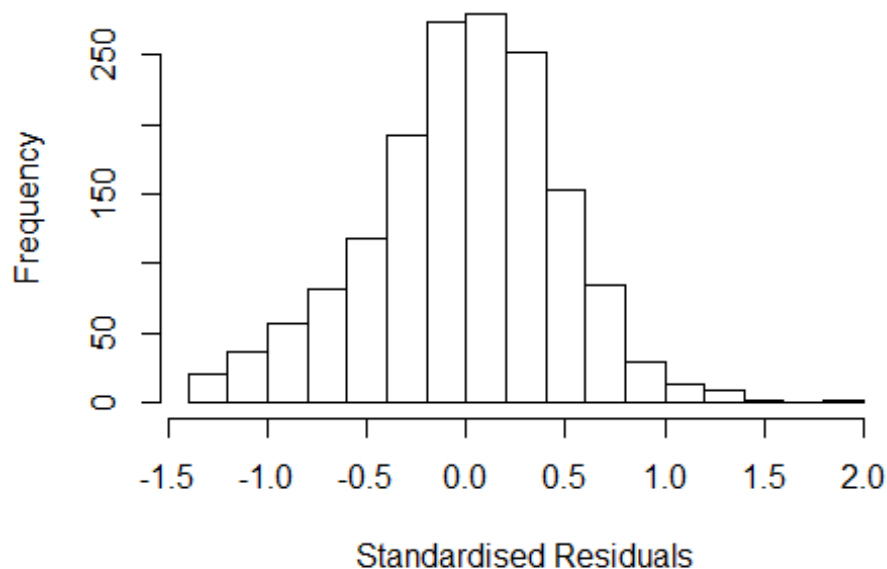


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.2, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 2.6396188965155"
## [1] "Male first author team size 2018 geometric mean: 2.78956471206162"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.71284103417964"
## [1] "Male last author team size 2018 geometric mean: 2.75769556644462"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, 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 1.156 1      1.075
## LastAuthorFemale  1.094 1      1.046
## UniqueAuthors     1.250 4      1.028
## Year               1.298 16     1.008
```

Residuals from first and last author and team size



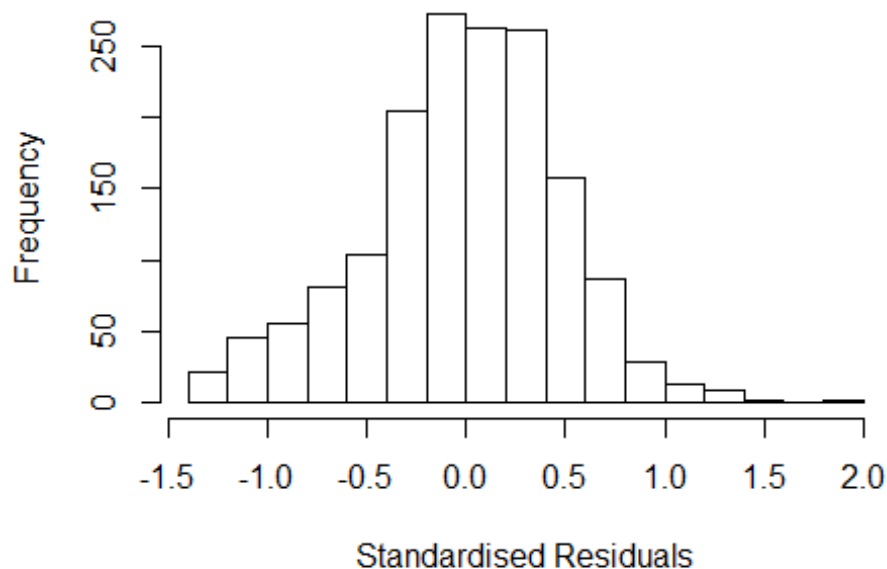
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3238 -0.3061 0.0184 0.3035 1.8232
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18706 0.05550 21.39 < 2e-16 ***
## FirstAuthorFemale1 -0.02204 0.02973 -0.74 0.45853
## LastAuthorFemale1 -0.02909 0.03408 -0.85 0.39355
## UniqueAuthors2 0.08620 0.03645 2.37 0.01815 *
## UniqueAuthors3 0.13673 0.03948 3.46 0.00055 ***
## UniqueAuthors4 0.20184 0.04244 4.76 2.2e-06 ***
## UniqueAuthors5 0.22799 0.04900 4.65 3.5e-06 ***
## Year1997 -0.07245 0.07067 -1.03 0.30545
## Year1998 -0.10817 0.07886 -1.37 0.17035
## Year1999 -0.08030 0.07383 -1.09 0.27696
```

```

## Year2000      -0.00476    0.09331   -0.05   0.95934
## Year2001      -0.02199    0.07027   -0.31   0.75435
## Year2002      -0.02827    0.06609   -0.43   0.66895
## Year2003      -0.16158    0.07688   -2.10   0.03575 *
## Year2004      -0.20247    0.07468   -2.71   0.00678 **
## Year2005      -0.02790    0.07664   -0.36   0.71584
## Year2006      -0.13968    0.06661   -2.10   0.03614 *
## Year2007      -0.09710    0.07036   -1.38   0.16779
## Year2008      -0.01155    0.06307   -0.18   0.85473
## Year2009      -0.02562    0.06564   -0.39   0.69637
## Year2010      -0.14104    0.07088   -1.99   0.04677 *
## Year2011      -0.10767    0.06383   -1.69   0.09184 .
## Year2012      -0.18995    0.06600   -2.88   0.00405 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.451
## Multiple R-squared:  0.0383, Adjusted R-squared:  0.025
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1469 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0646 0.8650 0.9480 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.24e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.113 1 1.055
## LastAuthorFemale 1.087 1 1.042
## Year 1.070 16 1.002

```

Residuals from first and last author



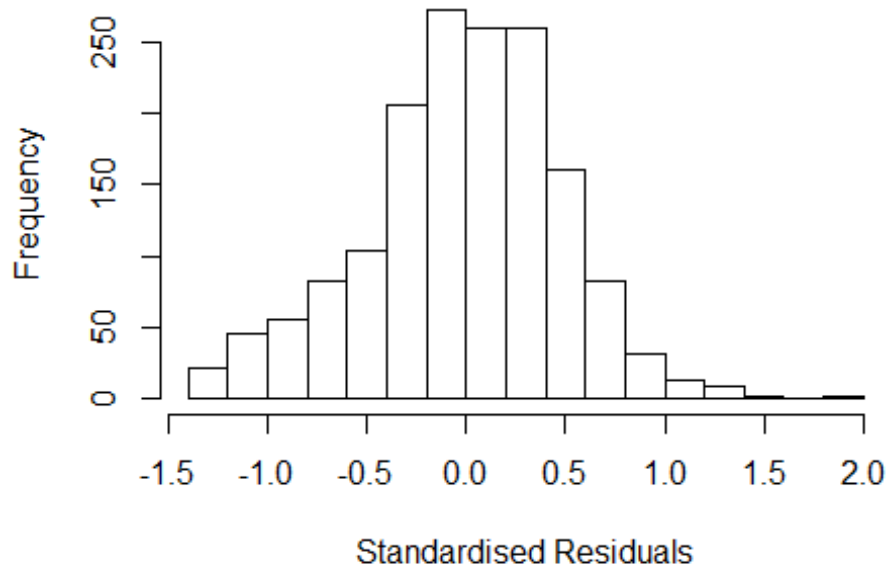
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2795 -0.3054 0.0115 0.3118 1.8393
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.269496 0.050462 25.16 <2e-16 ***
## FirstAuthorFemale1 -0.013068 0.029851 -0.44 0.662
## LastAuthorFemale1 -0.016403 0.034481 -0.48 0.634
## Year1997 -0.073034 0.071594 -1.02 0.308
## Year1998 -0.108479 0.078951 -1.37 0.170
## Year1999 -0.100340 0.073747 -1.36 0.174
## Year2000 0.000225 0.096638 0.00 0.998
## Year2001 -0.022464 0.071329 -0.31 0.753
## Year2002 -0.014983 0.067547 -0.22 0.824
## Year2003 -0.144621 0.076841 -1.88 0.060 .
## Year2004 -0.182642 0.077013 -2.37 0.018 *
## Year2005 -0.014006 0.076737 -0.18 0.855
```

```

## Year2006          -0.133146    0.067337    -1.98    0.048 *
## Year2007          -0.087166    0.070850    -1.23    0.219
## Year2008           0.012035    0.062768     0.19    0.848
## Year2009           0.010029    0.065523     0.15    0.878
## Year2010          -0.099599    0.070985    -1.40    0.161
## Year2011          -0.073762    0.064793    -1.14    0.255
## Year2012          -0.151786    0.065420    -2.32    0.020 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.454
## Multiple R-squared:  0.0167, Adjusted R-squared:  0.00553
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 123 weights are ~= 1. The remaining 1480 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0635 0.8620 0.9530 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.24e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.049 1      1.024
## Year              1.049 16      1.002

```


Residuals from first author



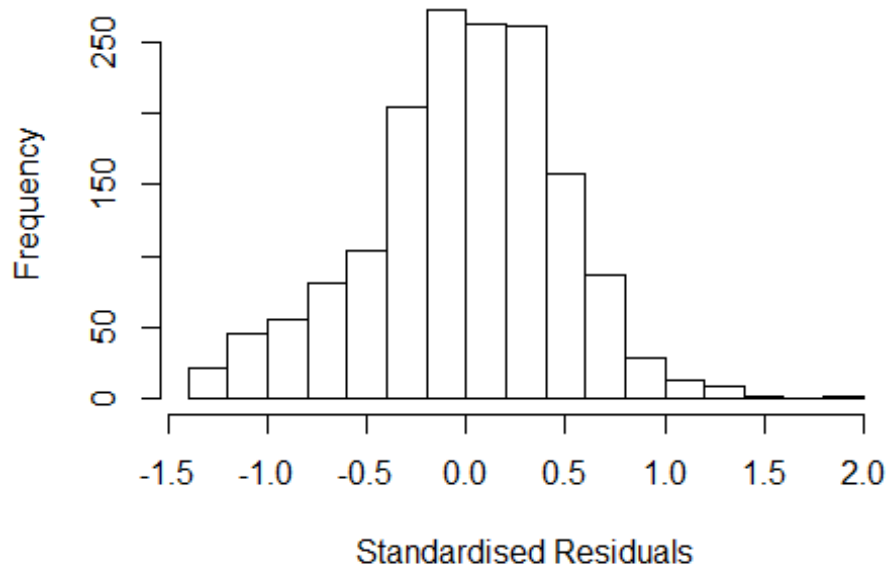
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27748 -0.30447  0.00903  0.31362  1.84088
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.267668   0.050218   25.24  <2e-16 ***
## FirstAuthorFemale1 -0.015989   0.029053   -0.55   0.582
## Year1997        -0.072697   0.071618   -1.02   0.310
## Year1998        -0.107788   0.078958   -1.37   0.172
## Year1999        -0.100142   0.073902   -1.36   0.176
## Year2000         -0.000684   0.096343   -0.01   0.994
## Year2001        -0.022364   0.071404   -0.31   0.754
## Year2002        -0.015780   0.067646   -0.23   0.816
## Year2003        -0.143616   0.076749   -1.87   0.061 .
## Year2004        -0.183580   0.077154   -2.38   0.017 *
## Year2005        -0.014060   0.076870   -0.18   0.855
## Year2006        -0.132946   0.067409   -1.97   0.049 *
```

```

## Year2007      -0.087812   0.071025   -1.24   0.217
## Year2008      0.012176   0.062834    0.19   0.846
## Year2009      0.009811   0.065673    0.15   0.881
## Year2010     -0.100215   0.070957   -1.41   0.158
## Year2011     -0.073515   0.064794   -1.13   0.257
## Year2012     -0.151548   0.065619   -2.31   0.021 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.455
## Multiple R-squared:  0.0165, Adjusted R-squared:  0.00598
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 123 weights are ~= 1. The remaining 1480 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0648 0.8640 0.9530 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      6.24e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.024 1          1.012
## Year            1.024 16          1.001

```

Residuals from last author



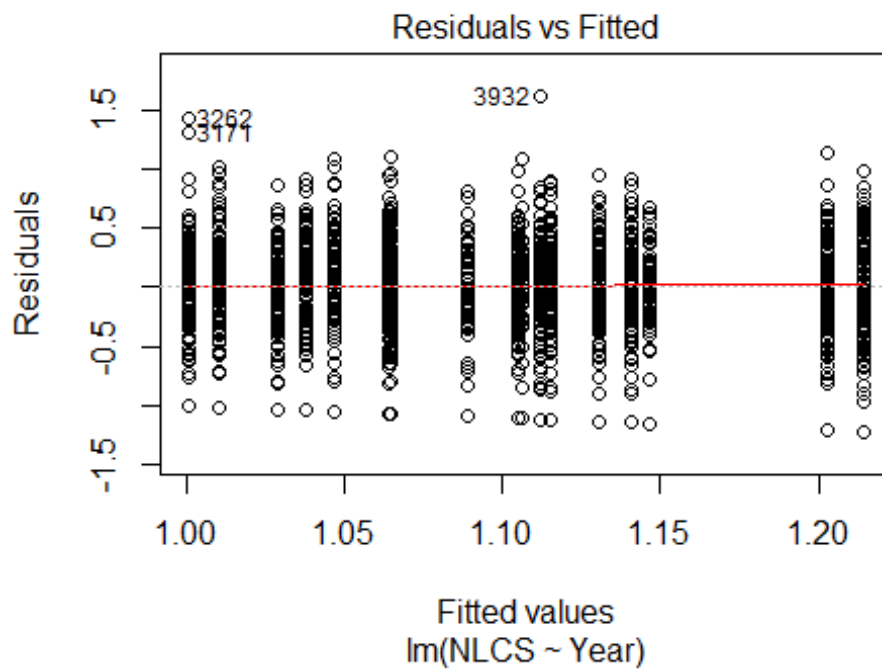
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27868 -0.30528 0.00893 0.31020 1.84250
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.267749 0.050100 25.30 <2e-16 ***
## LastAuthorFemale1 -0.019391 0.033558 -0.58 0.563
## Year1997 -0.072683 0.071546 -1.02 0.310
## Year1998 -0.108316 0.078962 -1.37 0.170
## Year1999 -0.099579 0.073696 -1.35 0.177
## Year2000 0.000941 0.096617 0.01 0.992
## Year2001 -0.022314 0.071316 -0.31 0.754
## Year2002 -0.015284 0.067473 -0.23 0.821
## Year2003 -0.144826 0.076849 -1.88 0.060 .
## Year2004 -0.182342 0.076921 -2.37 0.018 *
## Year2005 -0.014662 0.076658 -0.19 0.848
## Year2006 -0.133690 0.067374 -1.98 0.047 *
```

```

## Year2007          -0.087364    0.070819    -1.23    0.218
## Year2008           0.010110    0.062521     0.16    0.872
## Year2009           0.010931    0.065368     0.17    0.867
## Year2010          -0.100618    0.070920    -1.42    0.156
## Year2011          -0.075217    0.064866    -1.16    0.246
## Year2012          -0.153252    0.065360    -2.34    0.019 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.454
## Multiple R-squared:  0.0166, Adjusted R-squared:  0.00604
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1481 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0617 0.8640 0.9530 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.24e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1603"
## [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
## 257 198 200 175 168 175 213 179 204 170 203 228 219 195 178
## 2011 2012
## 205 240
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 118 90 100 68 50 57 118 100 118 97 109 136 127 117 98
## 2011 2012

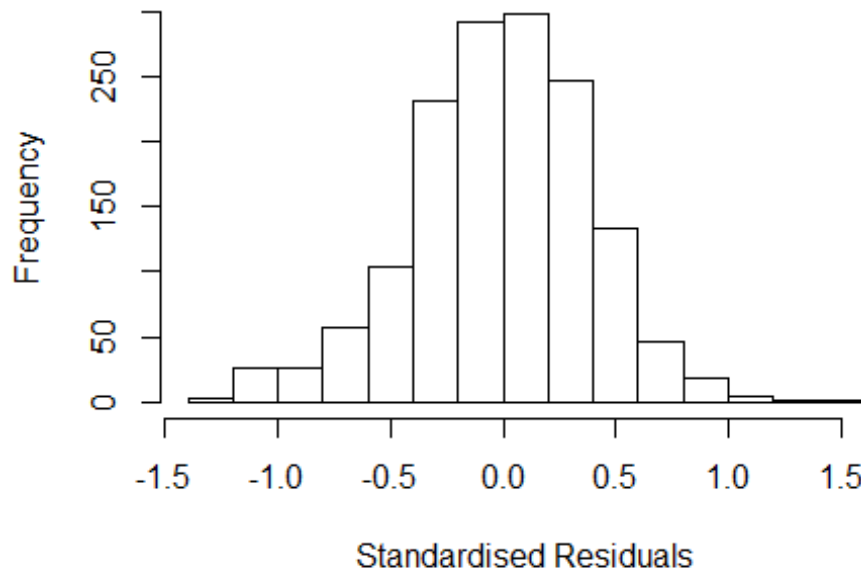
```

```
## 119 143
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 99 76 88 61 43 53 102 78 99 77 90 104 103 100 86
## 2011 2012
## 102 129
## [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 = 0.38, df = 1, p-value = 0.5
```


Residuals from first and last author and team size



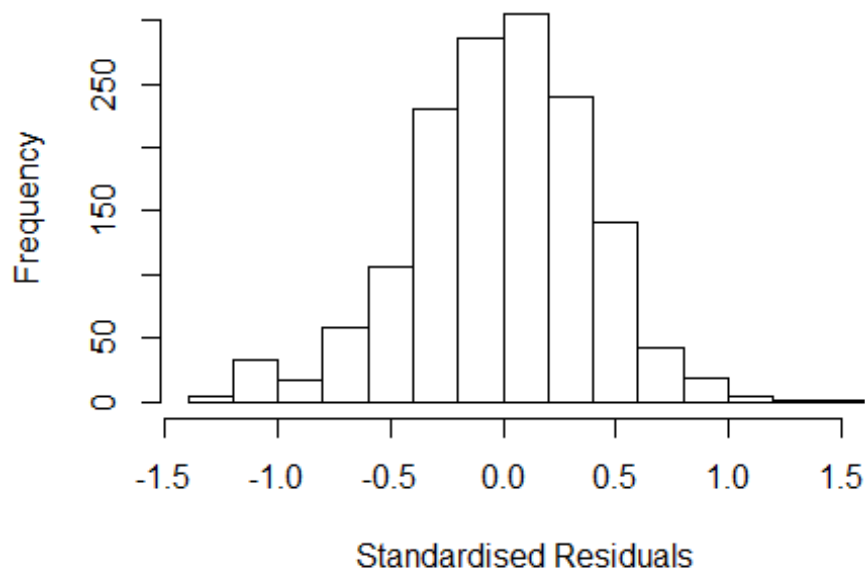
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29252 -0.26127 0.00514 0.25276 1.43548
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2586 0.0500 25.16 < 2e-16 ***
## FirstAuthorFemale1 0.0213 0.0260 0.82 0.41386
## LastAuthorFemale1 -0.0366 0.0310 -1.18 0.23868
## UniqueAuthors2 0.0339 0.0299 1.14 0.25579
## UniqueAuthors3 0.0334 0.0328 1.02 0.30894
## UniqueAuthors4 0.1394 0.0380 3.66 0.00026 ***
## UniqueAuthors5 0.1356 0.0432 3.14 0.00172 **
## Year1997 -0.1544 0.0710 -2.18 0.02978 *
## Year1998 -0.2368 0.0627 -3.78 0.00017 ***
## Year1999 -0.1792 0.0672 -2.67 0.00772 **
```

```

## Year2000          -0.1832      0.0661   -2.77  0.00562 **
## Year2001          -0.1559      0.0701   -2.22  0.02624 *
## Year2002          -0.0853      0.0635   -1.34  0.17955
## Year2003          -0.2500      0.0616   -4.06  5.1e-05 ***
## Year2004          -0.2107      0.0602   -3.50  0.00048 ***
## Year2005          -0.1562      0.0666   -2.35  0.01905 *
## Year2006          -0.2788      0.0623   -4.47  8.3e-06 ***
## Year2007          -0.2950      0.0609   -4.84  1.4e-06 ***
## Year2008          -0.1771      0.0602   -2.94  0.00329 **
## Year2009          -0.1761      0.0612   -2.88  0.00404 **
## Year2010          -0.2527      0.0707   -3.57  0.00036 ***
## Year2011          -0.2631      0.0583   -4.51  6.9e-06 ***
## Year2012          -0.3237      0.0570   -5.68  1.6e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.051, Adjusted R-squared:  0.0368
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1369 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.119  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          6.71e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.117 1 1.057
## LastAuthorFemale 1.059 1 1.029
## Year 1.147 16 1.004

```


Residuals from first and last author



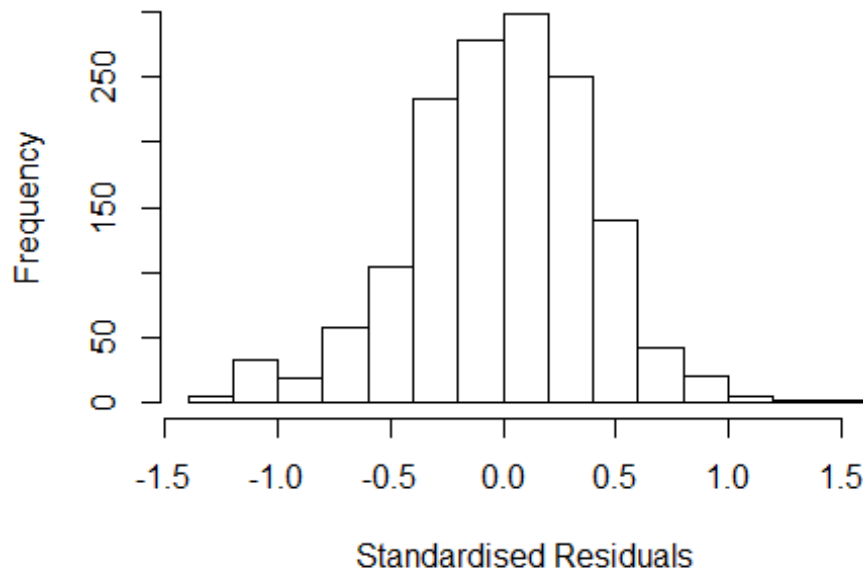
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2835 -0.2563  0.0038  0.2564  1.4159
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2835     0.0472   27.20 < 2e-16 ***
## FirstAuthorFemale1  0.0241     0.0261    0.92  0.35735
## LastAuthorFemale1 -0.0345     0.0305   -1.13  0.25817
## Year1997          -0.1375     0.0714   -1.93  0.05411 .
## Year1998          -0.2263     0.0626   -3.61  0.00031 ***
## Year1999          -0.1739     0.0678   -2.56  0.01043 *
## Year2000          -0.1610     0.0668   -2.41  0.01604 *
## Year2001          -0.1418     0.0695   -2.04  0.04135 *
## Year2002          -0.0692     0.0638   -1.08  0.27845
## Year2003          -0.2297     0.0616   -3.73  0.00020 ***
## Year2004          -0.1931     0.0606   -3.18  0.00148 **
## Year2005          -0.1396     0.0664   -2.10  0.03580 *
```

```

## Year2006          -0.2595      0.0627   -4.14  3.7e-05 ***
## Year2007          -0.2665      0.0614   -4.34  1.5e-05 ***
## Year2008          -0.1446      0.0593   -2.44  0.01486 *
## Year2009          -0.1517      0.0618   -2.45  0.01428 *
## Year2010          -0.2319      0.0713   -3.25  0.00118 **
## Year2011          -0.2328      0.0584   -3.99  7.1e-05 ***
## Year2012          -0.2928      0.0570   -5.13  3.2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.379
## Multiple R-squared:  0.0384, Adjusted R-squared:  0.0266
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1374 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.132  0.875   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      6.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.105 1      1.051
## Year              1.105 16      1.003

```

Residuals from first author



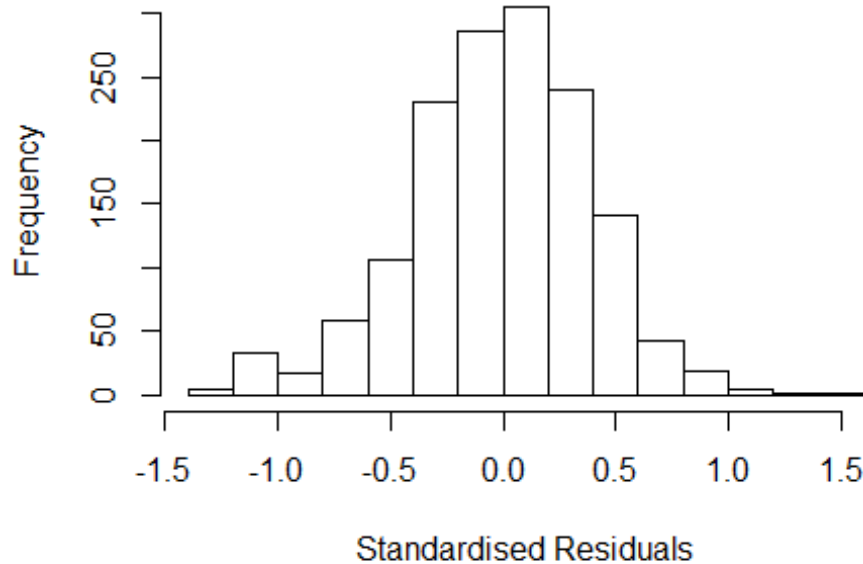
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28054 -0.25676  0.00514  0.25580  1.41952
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2805     0.0471   27.18 < 2e-16 ***
## FirstAuthorFemale1  0.0184     0.0260    0.71  0.48062
## Year1997        -0.1370     0.0712   -1.92  0.05460 .
## Year1998        -0.2258     0.0628   -3.60  0.00033 ***
## Year1999        -0.1750     0.0681   -2.57  0.01027 *
## Year2000        -0.1660     0.0666   -2.49  0.01273 *
## Year2001        -0.1418     0.0699   -2.03  0.04288 *
## Year2002        -0.0702     0.0638   -1.10  0.27108
## Year2003        -0.2304     0.0615   -3.74  0.00019 ***
## Year2004        -0.1937     0.0607   -3.19  0.00145 **
## Year2005        -0.1423     0.0668   -2.13  0.03326 *
## Year2006        -0.2588     0.0630   -4.11  4.2e-05 ***
```

```

## Year2007          -0.2671      0.0614   -4.35  1.5e-05 ***
## Year2008          -0.1451      0.0594   -2.44  0.01474 *
## Year2009          -0.1495      0.0619   -2.42  0.01585 *
## Year2010          -0.2306      0.0717   -3.22  0.00132 **
## Year2011          -0.2327      0.0584   -3.98  7.2e-05 ***
## Year2012          -0.2934      0.0572   -5.13  3.3e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.379
## Multiple R-squared:  0.0375, Adjusted R-squared:  0.0264
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1370 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.130  0.875   0.951   0.898   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      6.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.046 1          1.023
## Year            1.046 16          1.001

```

Residuals from last author



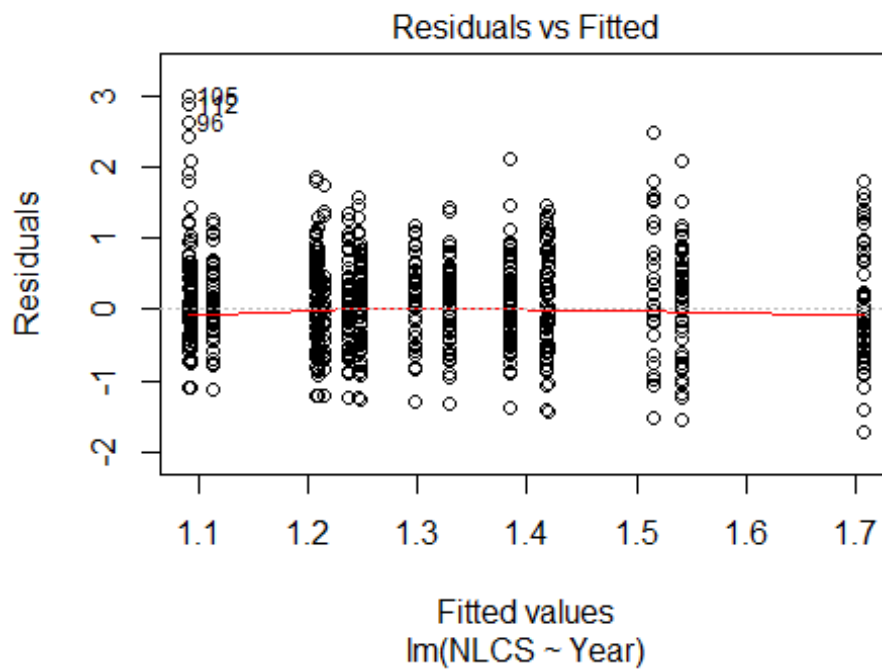
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28702 -0.25874 0.00275 0.25296 1.41134
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2870 0.0470 27.41 < 2e-16 ***
## LastAuthorFemale1 -0.0289 0.0305 -0.95 0.34348
## Year1997 -0.1409 0.0709 -1.99 0.04692 *
## Year1998 -0.2260 0.0626 -3.61 0.00032 ***
## Year1999 -0.1753 0.0677 -2.59 0.00968 **
## Year2000 -0.1630 0.0668 -2.44 0.01480 *
## Year2001 -0.1433 0.0693 -2.07 0.03894 *
## Year2002 -0.0676 0.0637 -1.06 0.28931
## Year2003 -0.2296 0.0614 -3.74 0.00019 ***
## Year2004 -0.1936 0.0606 -3.19 0.00144 **
## Year2005 -0.1391 0.0663 -2.10 0.03602 *
## Year2006 -0.2590 0.0629 -4.12 4.0e-05 ***
```

```

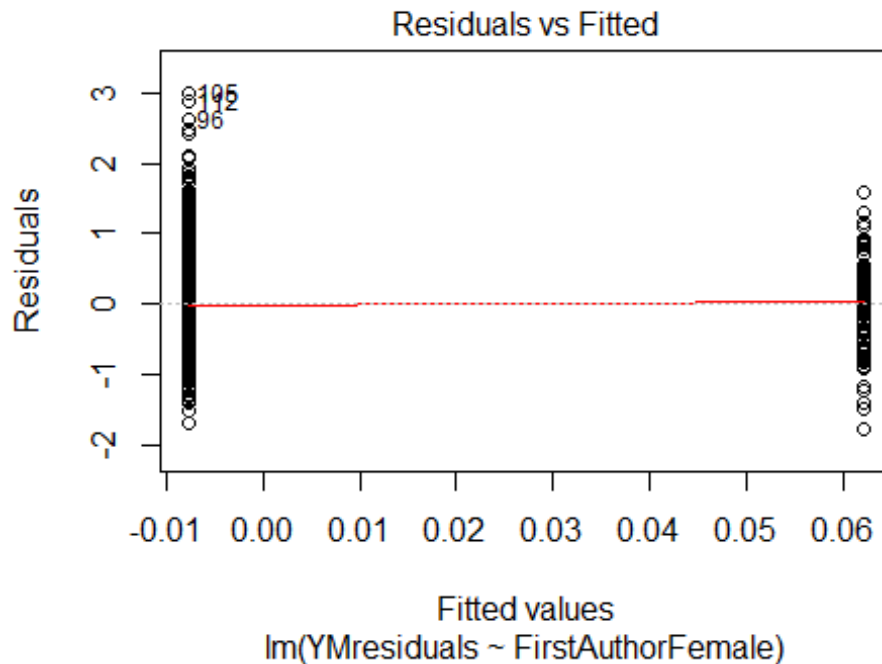
## Year2007          -0.2654      0.0614    -4.32  1.6e-05 ***
## Year2008          -0.1425      0.0591    -2.41  0.01607 *
## Year2009          -0.1517      0.0620    -2.45  0.01447 *
## Year2010          -0.2296      0.0712    -3.22  0.00129 **
## Year2011          -0.2312      0.0584    -3.96  8.0e-05 ***
## Year2012          -0.2891      0.0566    -5.11  3.7e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0379, Adjusted R-squared:  0.0268
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~ = 1. The remaining 1372 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.134  0.875  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      6.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1490"
## [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
## 160 131 139 148 123 112 124 141 123 159 149 150 195 211 172
## 2011 2012
## 171 163
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 34 37 29 27 31 51 57 51 72 62 72 82 93 65
## 2011 2012

```

```
##      74      74
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    47   28   31   27   24   24   37   46   49   50   45   55   60   79   51
## 2011 2012
##    63   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 = 120, df = 16, p-value <2e-16
```

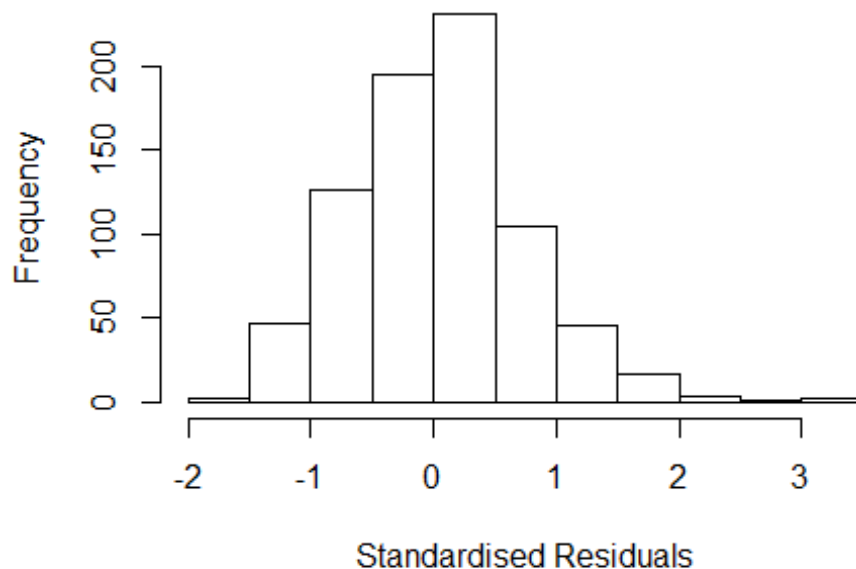


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.7, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 2.64674138967313"
## [1] "Male first author team size 2018 geometric mean: 2.83348826632123"
##
## 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] "Female last author team size 2018 geometric mean: 2.46567293663078"
## [1] "Male last author team size 2018 geometric mean: 2.84017507698443"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 760, 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 1.105 1      1.051
## LastAuthorFemale  1.107 1      1.052
## UniqueAuthors     1.573 4      1.058
## Year              1.729 16     1.017
```


Residuals from first and last author and team size



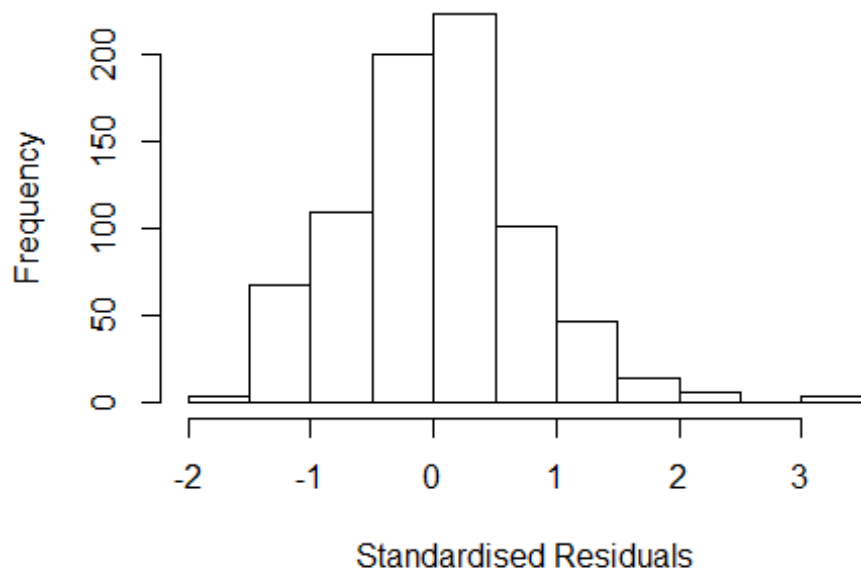
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 96  0030153193 3.713 1996      1909      1      2.882
## 105 0030209407 4.078 1996      1909      1      3.247
## 112 0030426687 3.967 1996      1909      1      3.440
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7143 -0.4447  0.0256  0.4300  3.4395
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5275     0.1234   4.27 2.2e-05 ***
## FirstAuthorFemale1  0.0542     0.0695   0.78 0.43618
## LastAuthorFemale1 -0.0887     0.0860  -1.03 0.30291
## UniqueAuthors2    0.3582     0.0757   4.73 2.7e-06 ***
## UniqueAuthors3    0.3037     0.0796   3.82 0.00015 ***
## UniqueAuthors4    0.2664     0.1043   2.55 0.01083 *
## UniqueAuthors5    0.2224     0.1191   1.87 0.06233 .
## Year1997         0.3379     0.2600   1.30 0.19413
```

```

## Year1998          0.4471      0.2078      2.15  0.03170 *
## Year1999          0.8286      0.2401      3.45  0.00059 ***
## Year2000          0.3251      0.2866      1.13  0.25709
## Year2001          0.4813      0.1836      2.62  0.00894 **
## Year2002          0.7858      0.1995      3.94  8.9e-05 ***
## Year2003          0.7601      0.1958      3.88  0.00011 ***
## Year2004          0.6091      0.1569      3.88  0.00011 ***
## Year2005          0.4352      0.1668      2.61  0.00928 **
## Year2006          0.5814      0.1695      3.43  0.00064 ***
## Year2007          0.5203      0.1554      3.35  0.00086 ***
## Year2008          0.4758      0.1474      3.23  0.00130 **
## Year2009          0.4691      0.1384      3.39  0.00074 ***
## Year2010          0.4001      0.1510      2.65  0.00824 **
## Year2011          0.3020      0.1503      2.01  0.04494 *
## Year2012          0.3131      0.1462      2.14  0.03262 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.656
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0815
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(32,34) are outliers with |weight| = 0 ( < 0.00013);
## 63 weights are ~= 1. The remaining 709 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0145 0.8580 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          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 1.066 1          1.032
## LastAuthorFemale 1.099 1          1.048
## Year 1.171 16          1.005

```

Residuals from first and last author



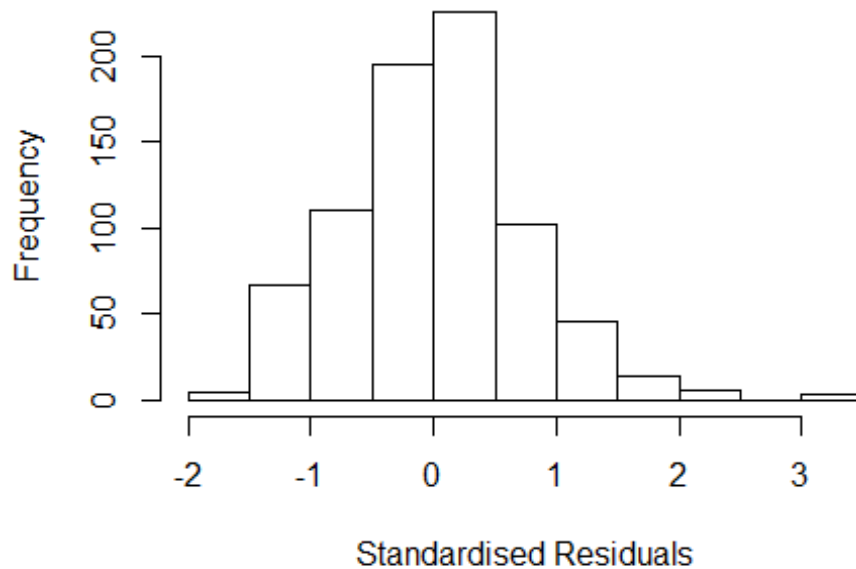
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 96  0030153193 3.713 1996      1909      1      3.030
## 105 0030209407 4.078 1996      1909      1      3.395
## 112 0030426687 3.967 1996      1909      1      3.284
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5516 -0.4668  0.0335  0.4325  3.3951
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6829    0.1405   4.86 1.4e-06 ***
## FirstAuthorFemale1  0.0977    0.0709   1.38 0.16853
## LastAuthorFemale1 -0.0788    0.0898  -0.88 0.38033
## Year1997         0.3522    0.2588   1.36 0.17387
## Year1998         0.5096    0.2283   2.23 0.02592 *
## Year1999         0.7928    0.2497   3.17 0.00156 **
## Year2000         0.4239    0.2940   1.44 0.14966
## Year2001         0.5628    0.1864   3.02 0.00261 **
## Year2002         0.8498    0.2099   4.05 5.7e-05 ***
## Year2003         0.8389    0.2081   4.03 6.1e-05 ***
```

```

## Year2004          0.6742      0.1739      3.88  0.00012 ***
## Year2005          0.5213      0.1722      3.03  0.00256 **
## Year2006          0.6736      0.1808      3.73  0.00021 ***
## Year2007          0.6213      0.1661      3.74  0.00020 ***
## Year2008          0.5641      0.1586      3.56  0.00040 ***
## Year2009          0.5692      0.1523      3.74  0.00020 ***
## Year2010          0.5202      0.1634      3.18  0.00152 **
## Year2011          0.4036      0.1612      2.50  0.01248 *
## Year2012          0.4054      0.1588      2.55  0.01087 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.663
## Multiple R-squared:  0.0712, Adjusted R-squared:  0.0491
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 2 observations c(32,34) are outliers with |weight| = 0 ( < 0.00013);
## 56 weights are ~= 1. The remaining 716 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0022 0.8630 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          1.29e-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 1.074 1          1.036
## Year              1.074 16          1.002

```

Residuals from first author



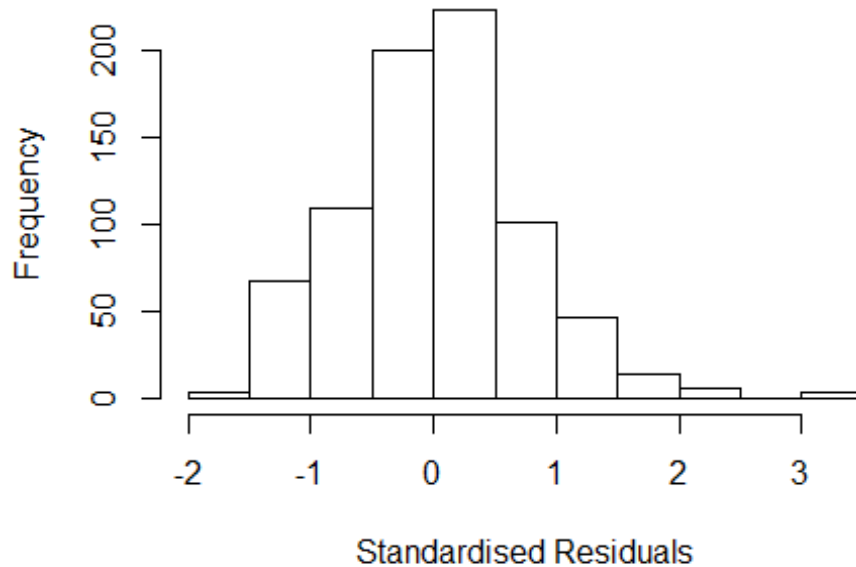
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 96  0030153193 3.713 1996      1909      1      3.030
## 105 0030209407 4.078 1996      1909      1      3.395
## 112 0030426687 3.967 1996      1909      1      3.284
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6144 -0.4626  0.0169  0.4332  3.3984
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6796    0.1399   4.86 1.5e-06 ***
## FirstAuthorFemale1 0.0863    0.0729   1.18 0.23709
## Year1997        0.3518    0.2616   1.34 0.17905
## Year1998        0.5104    0.2299   2.22 0.02671 *
## Year1999        0.7945    0.2501   3.18 0.00155 **
## Year2000        0.4266    0.2949   1.45 0.14844
## Year2001        0.5596    0.1855   3.02 0.00265 **
## Year2002        0.8485    0.2104   4.03 6.1e-05 ***
## Year2003        0.8416    0.2081   4.04 5.8e-05 ***
## Year2004        0.6640    0.1731   3.84 0.00013 ***
```

```

## Year2005          0.5199      0.1714      3.03  0.00251 **
## Year2006          0.6724      0.1813      3.71  0.00022 ***
## Year2007          0.6184      0.1654      3.74  0.00020 ***
## Year2008          0.5569      0.1580      3.53  0.00045 ***
## Year2009          0.5691      0.1519      3.75  0.00019 ***
## Year2010          0.5169      0.1625      3.18  0.00153 **
## Year2011          0.4039      0.1604      2.52  0.01202 *
## Year2012          0.3984      0.1583      2.52  0.01204 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.658
## Multiple R-squared:  0.0712, Adjusted R-squared:  0.0503
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(32,34) are outliers with |weight| = 0 ( < 0.00013);
## 54 weights are ~= 1. The remaining 718 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.001  0.860  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      1.29e-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 1.097 1      1.048
## Year      1.097 16      1.003

```

Residuals from last author



```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 96  0030153193 3.713 1996      1909      1      3.030
## 105 0030209407 4.078 1996      1909      1      3.395
## 112 0030426687 3.967 1996      1909      1      3.284
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.534 -0.477  0.039  0.431  3.394
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6844    0.1419   4.82 1.7e-06 ***
## LastAuthorFemale1 -0.0583    0.0903  -0.65 0.51844
## Year1997        0.3555    0.2618   1.36 0.17492
## Year1998        0.5142    0.2293   2.24 0.02520 *
## Year1999        0.7954    0.2509   3.17 0.00159 **
## Year2000        0.4296    0.2957   1.45 0.14668
## Year2001        0.5746    0.1876   3.06 0.00228 **
## Year2002        0.8575    0.2106   4.07 5.2e-05 ***
## Year2003        0.8498    0.2089   4.07 5.2e-05 ***
## Year2004        0.6796    0.1742   3.90 0.00010 ***
```

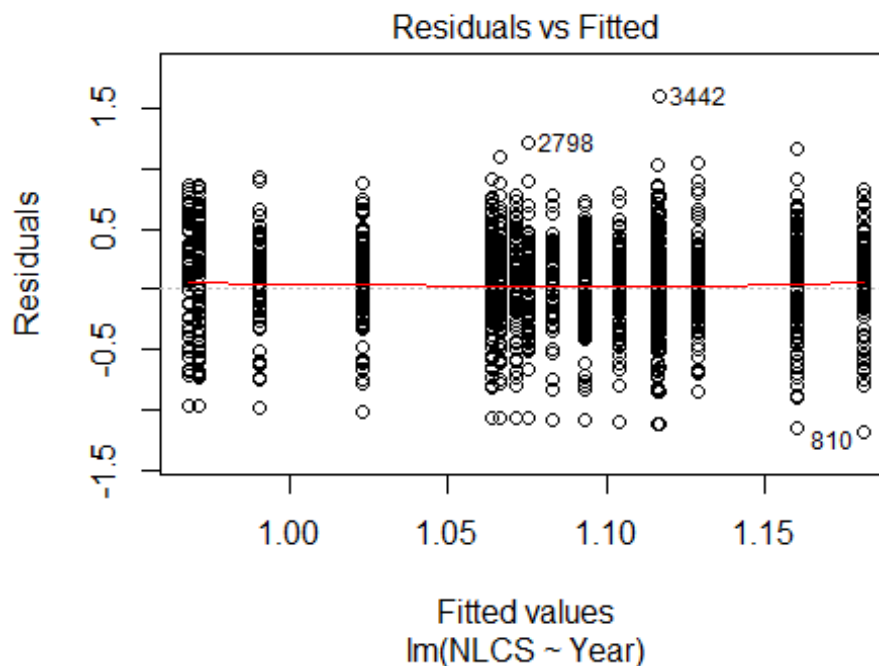
```

## Year2005          0.5294      0.1736      3.05  0.00237 **
## Year2006          0.6838      0.1826      3.75  0.00019 ***
## Year2007          0.6363      0.1668      3.81  0.00015 ***
## Year2008          0.5729      0.1599      3.58  0.00036 ***
## Year2009          0.5823      0.1528      3.81  0.00015 ***
## Year2010          0.5234      0.1649      3.17  0.00156 **
## Year2011          0.4229      0.1614      2.62  0.00895 **
## Year2012          0.4074      0.1603      2.54  0.01122 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.661
## Multiple R-squared:  0.0695, Adjusted R-squared:  0.0485
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 2 observations c(32,34) are outliers with |weight| = 0 ( < 0.00013);
## 58 weights are ~= 1. The remaining 714 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0018 0.8630 0.9500 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.29e-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: 774"
## [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
## 238 206 199 181 187 208 211 143 141 142 183 175 209 167 170
## 2011 2012
## 168 173
##

```

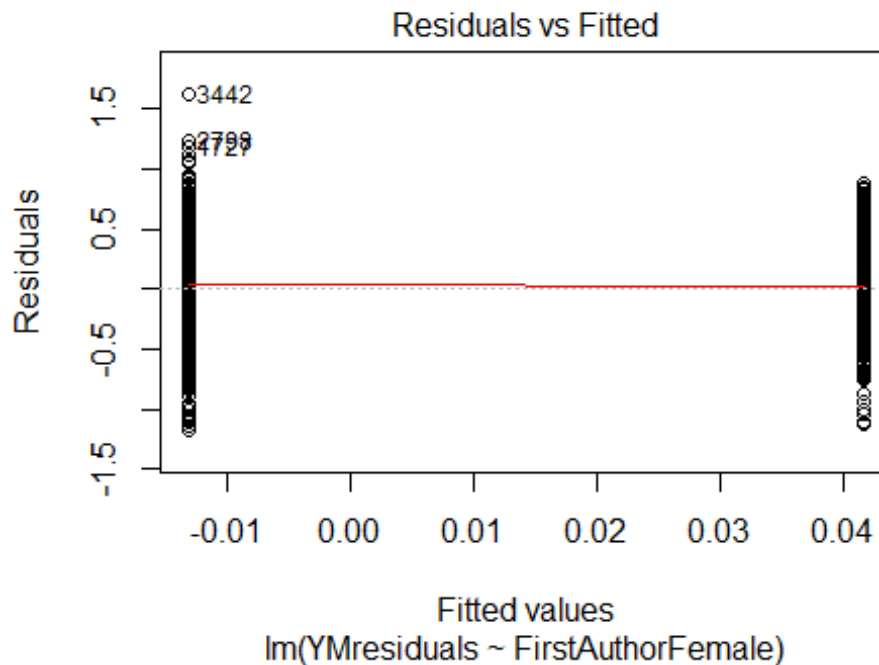


```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 118 120 101 93 62 80 120 85 86 89 119 109 121 114 97
## 2011 2012
## 104 125
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 109 106 92 87 52 72 102 68 74 71 99 93 92 97 84
## 2011 2012
## 93 110
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 62, df = 16, p-value = 3e-07
```



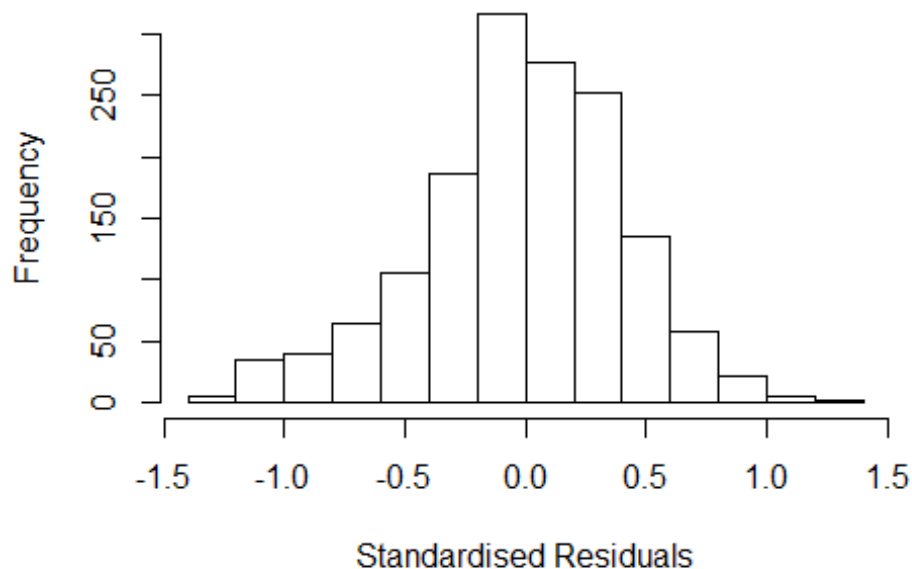
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 1e-04
## [1] "Female first author team size 2018 geometric mean: 3.31546814228346"
## [1] "Male first author team size 2018 geometric mean: 3.02427815551275"
```

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 870, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.41186772682494"
## [1] "Male last author team size 2018 geometric mean: 3.05130785760798"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 760, 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 1.144 1      1.070
## LastAuthorFemale  1.059 1      1.029
## UniqueAuthors    1.339 4      1.037
## Year              1.424 16     1.011
```

Residuals from first and last author and team size



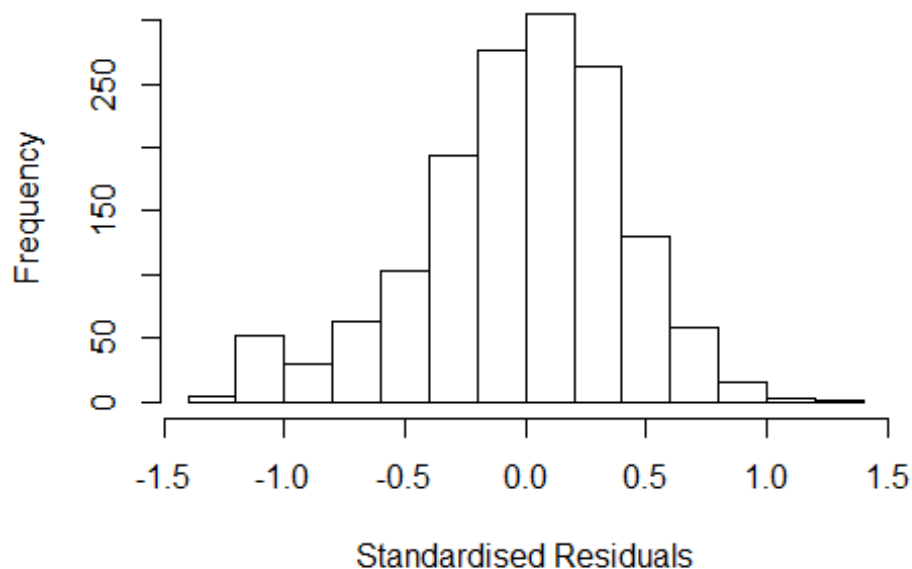
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.237463 -0.248739 -0.000281 0.261221 1.366935
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04331 0.05612 18.59 < 2e-16 ***
## FirstAuthorFemale1 0.00849 0.02420 0.35 0.72572
## LastAuthorFemale1 0.01485 0.03126 0.47 0.63487
## UniqueAuthors2 0.14054 0.03684 3.81 0.00014 ***
## UniqueAuthors3 0.16672 0.04000 4.17 3.2e-05 ***
## UniqueAuthors4 0.24700 0.04164 5.93 3.7e-09 ***
## UniqueAuthors5 0.26575 0.04443 5.98 2.8e-09 ***
## Year1997 -0.09572 0.06386 -1.50 0.13411
## Year1998 -0.18363 0.07958 -2.31 0.02117 *
## Year1999 0.02743 0.06502 0.42 0.67315
```

```

## Year2000      -0.06512      0.07125      -0.91      0.36090
## Year2001      -0.17654      0.09590      -1.84      0.06584 .
## Year2002       0.02298      0.07034       0.33      0.74396
## Year2003      -0.10669      0.06723      -1.59      0.11272
## Year2004      -0.15084      0.06687      -2.26      0.02423 *
## Year2005      -0.05342      0.06771      -0.79      0.43031
## Year2006      -0.09607      0.05928      -1.62      0.10532
## Year2007      -0.10824      0.05998      -1.80      0.07132 .
## Year2008      -0.06462      0.06011      -1.08      0.28255
## Year2009      -0.09804      0.05736      -1.71      0.08761 .
## Year2010      -0.10311      0.06389      -1.61      0.10678
## Year2011      -0.10671      0.06021      -1.77      0.07656 .
## Year2012      -0.11952      0.06237      -1.92      0.05552 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.0629, Adjusted R-squared:  0.049
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1365 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.190  0.863  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      6.66e-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 1.083 1      1.041
## LastAuthorFemale  1.048 1      1.024
## Year              1.107 16      1.003

```

Residuals from first and last author



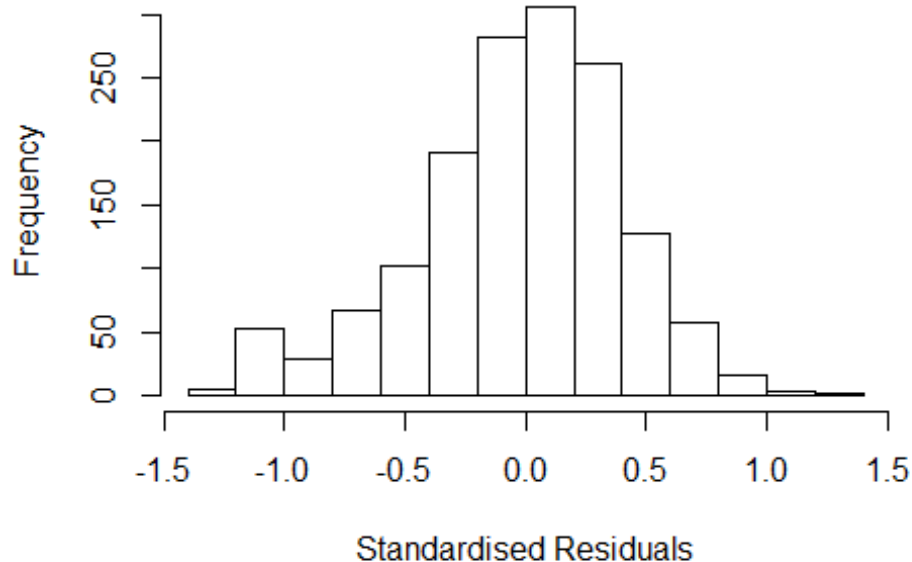
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2082 -0.2629  0.0121  0.2667  1.2250
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1686     0.0492   23.77  <2e-16 ***
## FirstAuthorFemale1  0.0423     0.0239    1.77   0.0774 .
## LastAuthorFemale1  0.0141     0.0317    0.44   0.6569
## Year1997          -0.1153     0.0660   -1.75   0.0807 .
## Year1998          -0.2214     0.0793   -2.79   0.0053 **
## Year1999           0.0396     0.0669    0.59   0.5543
## Year2000          -0.0725     0.0736   -0.99   0.3244
## Year2001          -0.1609     0.1036   -1.55   0.1208
## Year2002           0.0214     0.0707    0.30   0.7618
## Year2003          -0.0872     0.0673   -1.30   0.1952
## Year2004          -0.1455     0.0684   -2.13   0.0334 *
## Year2005          -0.0574     0.0680   -0.84   0.3984
```

```

## Year2006          -0.0716      0.0604   -1.19   0.2358
## Year2007          -0.0916      0.0613   -1.49   0.1356
## Year2008          -0.0385      0.0609   -0.63   0.5269
## Year2009          -0.0818      0.0586   -1.40   0.1626
## Year2010          -0.0650      0.0655   -0.99   0.3216
## Year2011          -0.0650      0.0623   -1.04   0.2969
## Year2012          -0.0978      0.0637   -1.54   0.1245
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.0249, Adjusted R-squared:  0.0131
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1380 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.307  0.862  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      6.66e-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 1.071 1      1.035
## Year              1.071 16      1.002

```

Residuals from first author



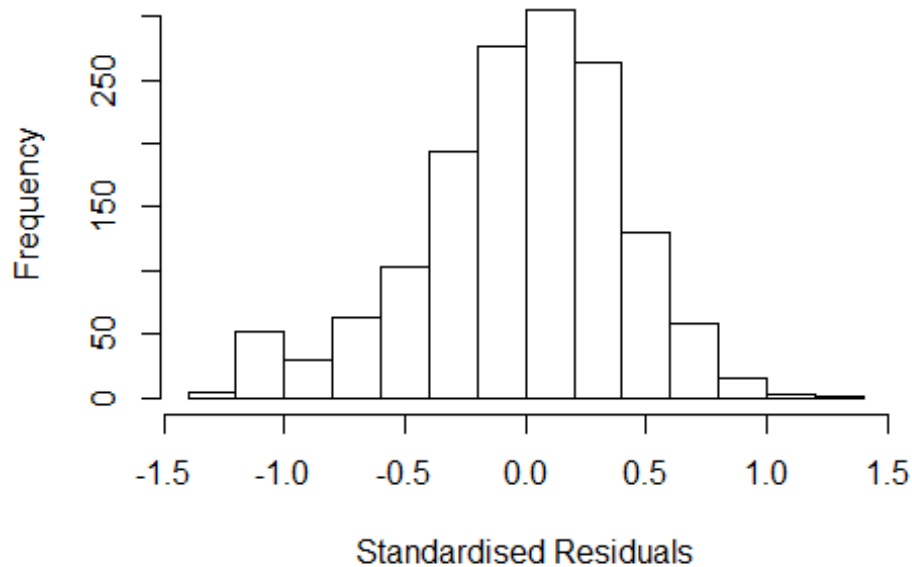
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2096 -0.2581 0.0132 0.2654 1.2242
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1696 0.0491 23.80 <2e-16 ***
## FirstAuthorFemale1 0.0445 0.0239 1.86 0.0633 .
## Year1997 -0.1154 0.0660 -1.75 0.0807 .
## Year1998 -0.2214 0.0792 -2.79 0.0053 **
## Year1999 0.0400 0.0668 0.60 0.5492
## Year2000 -0.0704 0.0734 -0.96 0.3372
## Year2001 -0.1588 0.1032 -1.54 0.1242
## Year2002 0.0211 0.0709 0.30 0.7658
## Year2003 -0.0866 0.0673 -1.29 0.1986
## Year2004 -0.1458 0.0683 -2.13 0.0330 *
## Year2005 -0.0571 0.0679 -0.84 0.4004
## Year2006 -0.0711 0.0603 -1.18 0.2390
```

```

## Year2007          -0.0917      0.0613   -1.50    0.1347
## Year2008          -0.0378      0.0607   -0.62    0.5339
## Year2009          -0.0807      0.0585   -1.38    0.1679
## Year2010          -0.0641      0.0655   -0.98    0.3277
## Year2011          -0.0642      0.0622   -1.03    0.3024
## Year2012          -0.0969      0.0635   -1.53    0.1273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.0248, Adjusted R-squared:  0.0136
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1377 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.307  0.861  0.949  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      6.66e-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 1.037 1          1.019
## Year            1.037 16          1.001

```


Residuals from last author



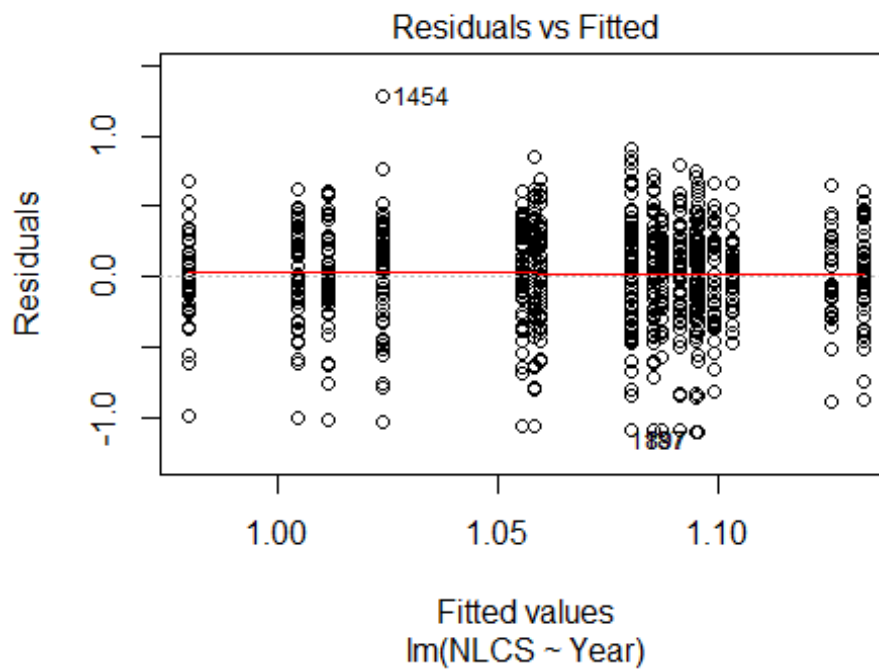
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.216 -0.255 0.015 0.266 1.217
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1720 0.0490 23.91 <2e-16 ***
## LastAuthorFemale1 0.0242 0.0317 0.76 0.4458
## Year1997 -0.1134 0.0661 -1.72 0.0862 .
## Year1998 -0.2203 0.0793 -2.78 0.0055 **
## Year1999 0.0438 0.0670 0.65 0.5139
## Year2000 -0.0710 0.0735 -0.97 0.3344
## Year2001 -0.1584 0.1032 -1.53 0.1252
## Year2002 0.0266 0.0707 0.38 0.7066
## Year2003 -0.0819 0.0674 -1.22 0.2240
## Year2004 -0.1383 0.0683 -2.02 0.0433 *
## Year2005 -0.0516 0.0678 -0.76 0.4467
## Year2006 -0.0635 0.0601 -1.06 0.2910
```

```

## Year2007          -0.0868      0.0611   -1.42   0.1557
## Year2008          -0.0309      0.0606   -0.51   0.6107
## Year2009          -0.0705      0.0581   -1.21   0.2247
## Year2010          -0.0541      0.0651   -0.83   0.4060
## Year2011          -0.0529      0.0618   -0.86   0.3923
## Year2012          -0.0878      0.0629   -1.39   0.1632
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.0231, Adjusted R-squared:  0.012
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 128 weights are ~ = 1. The remaining 1373 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.313  0.861  0.949  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      6.66e-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: 1501"
## [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
## 123 101 96 91 88 84 95 83 60 56 98 99 117 80 99
## 2011 2012
## 82 78
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 74 57 61 45 29 37 63 58 39 39 58 60 74 47 63
## 2011 2012

```

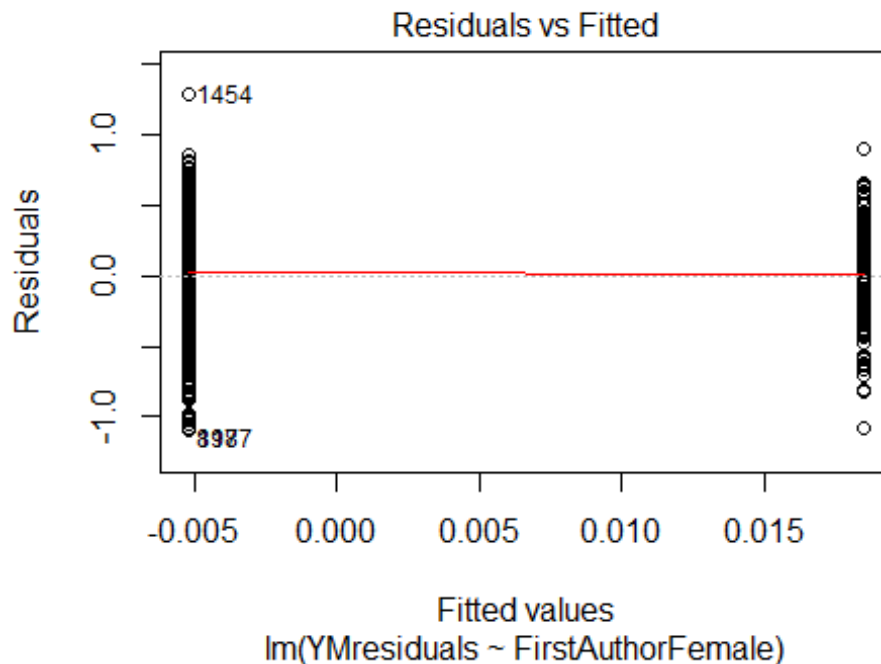
```
## 42 54
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 68 53 53 42 24 32 58 46 33 31 49 48 59 38 57
## 2011 2012
## 32 51
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 20, df = 16, p-value = 0.2
```



```
##
## 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: 2.38899080292393"
## [1] "Male first author team size 2018 geometric mean: 2.48818652562681"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.5949476326182"
## [1] "Male last author team size 2018 geometric mean: 2.40100330479099"

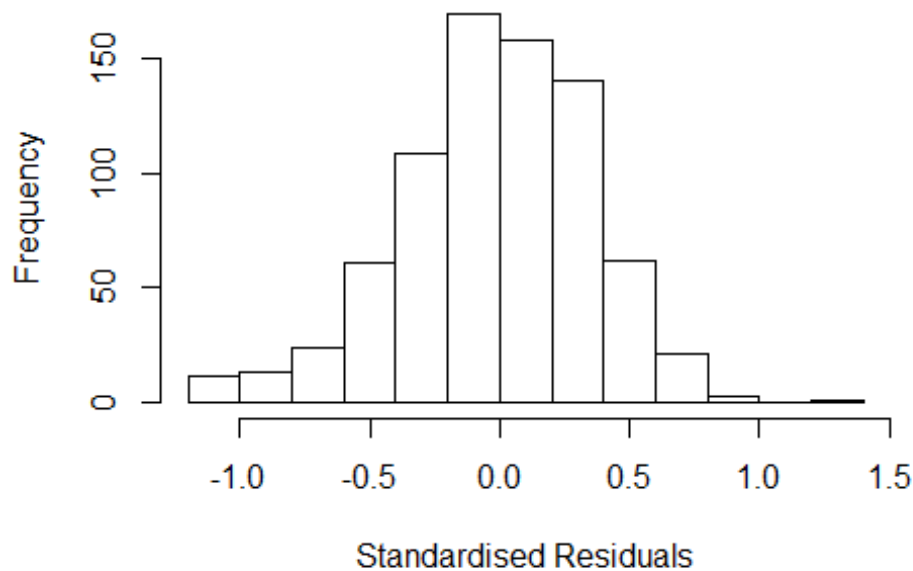
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 200, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	1.259	1	1.122
LastAuthorFemale	1.175	1	1.084
UniqueAuthors	1.915	4	1.085
Year	2.206	16	1.025

Residuals from first and last author and team size



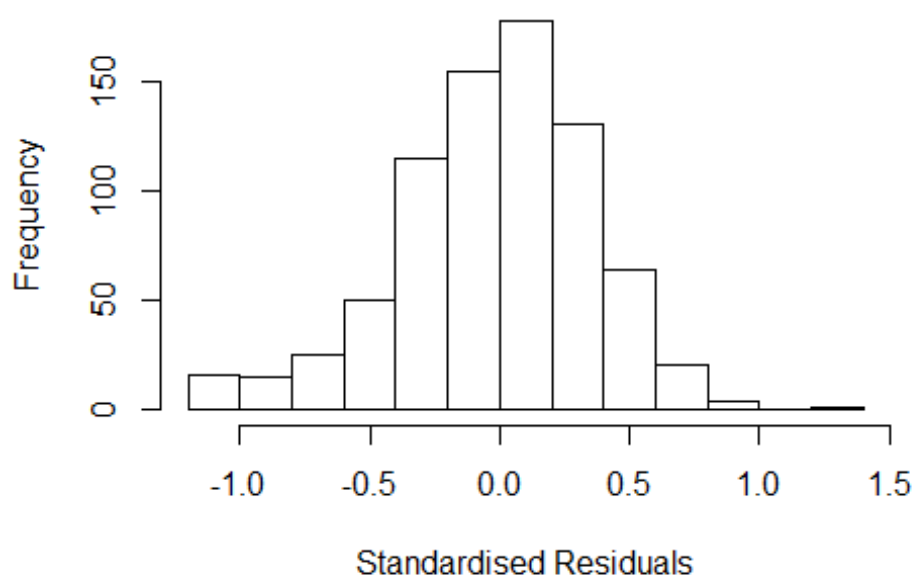
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.12116 -0.22699 -0.00344 0.24154 1.37636
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02e+00 5.45e-02 18.75 < 2e-16 ***
## FirstAuthorFemale1 -2.13e-03 3.28e-02 -0.07 0.9481
## LastAuthorFemale1 -6.64e-02 3.90e-02 -1.70 0.0888 .
## UniqueAuthors2 8.21e-02 3.47e-02 2.36 0.0183 *
## UniqueAuthors3 1.24e-01 4.13e-02 3.00 0.0028 **
## UniqueAuthors4 1.87e-01 4.66e-02 4.01 6.7e-05 ***
## UniqueAuthors5 2.61e-01 5.21e-02 5.00 7.0e-07 ***
## Year1997 -4.66e-03 7.34e-02 -0.06 0.9494
## Year1998 -4.60e-02 7.14e-02 -0.64 0.5195
## Year1999 4.46e-02 7.32e-02 0.61 0.5427
```

```

## Year2000          3.93e-02  7.98e-02   0.49  0.6230
## Year2001          1.40e-02  7.36e-02   0.19  0.8496
## Year2002          7.14e-03  6.75e-02   0.11  0.9158
## Year2003         -4.73e-02  6.82e-02  -0.69  0.4881
## Year2004         -1.33e-01  7.50e-02  -1.77  0.0768 .
## Year2005          6.40e-02  8.33e-02   0.77  0.4427
## Year2006         -2.48e-02  7.62e-02  -0.33  0.7447
## Year2007         -9.66e-02  8.61e-02  -1.12  0.2622
## Year2008         -8.91e-05  6.87e-02   0.00  0.9990
## Year2009         -2.48e-02  7.09e-02  -0.35  0.7266
## Year2010         -7.32e-02  7.69e-02  -0.95  0.3420
## Year2011         -2.10e-02  7.83e-02  -0.27  0.7884
## Year2012         -6.67e-02  7.28e-02  -0.92  0.3599
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.34
## Multiple R-squared:  0.058, Adjusted R-squared:  0.0304
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 705 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0636 0.8700 0.9480 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          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 1.213 1 1.101
## LastAuthorFemale 1.147 1 1.071
## Year 1.242 16 1.007

```

Residuals from first and last author



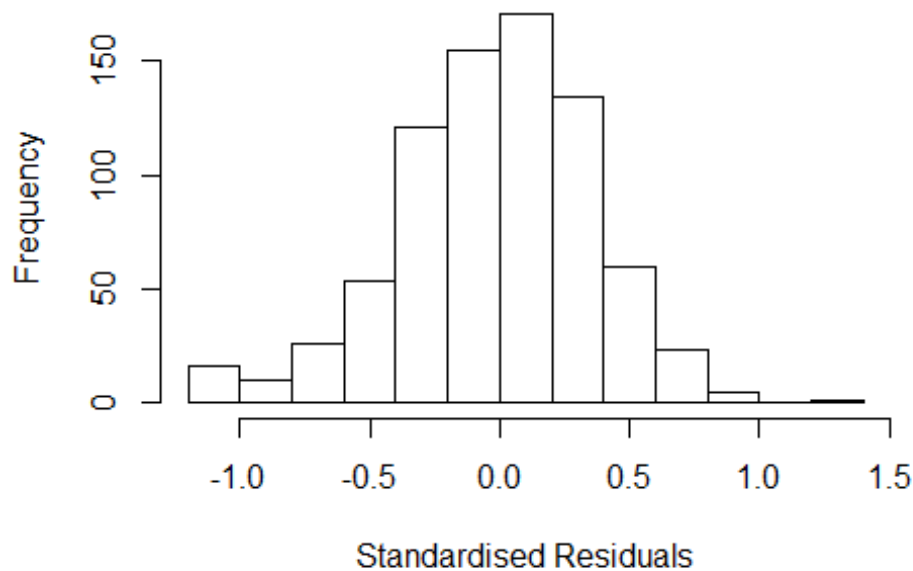
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1488 -0.2296 0.0107 0.2439 1.2815
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.080130 0.051910 20.81 <2e-16 ***
## FirstAuthorFemale1 0.031830 0.033096 0.96 0.336
## LastAuthorFemale1 -0.085161 0.038507 -2.21 0.027 *
## Year1997 0.000761 0.075754 0.01 0.992
## Year1998 -0.058308 0.070894 -0.82 0.411
## Year1999 0.055812 0.073997 0.75 0.451
## Year2000 0.028160 0.081843 0.34 0.731
## Year2001 0.024421 0.073676 0.33 0.740
## Year2002 0.028049 0.068918 0.41 0.684
## Year2003 -0.036422 0.071614 -0.51 0.611
## Year2004 -0.129870 0.073585 -1.76 0.078 .
## Year2005 0.068654 0.084375 0.81 0.416
```

```

## Year2006      -0.005046    0.077768   -0.06    0.948
## Year2007      -0.059669    0.084323   -0.71    0.479
## Year2008       0.037252    0.068531    0.54    0.587
## Year2009       0.019509    0.070572    0.28    0.782
## Year2010      -0.024590    0.076963   -0.32    0.749
## Year2011       0.041507    0.081016    0.51    0.609
## Year2012      -0.006190    0.072438   -0.09    0.932
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.349
## Multiple R-squared:  0.0206, Adjusted R-squared:  -0.00276
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 710 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.148  0.878  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      1.29e-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 1.152 1      1.073
## Year      1.152 16      1.004

```


Residuals from first author



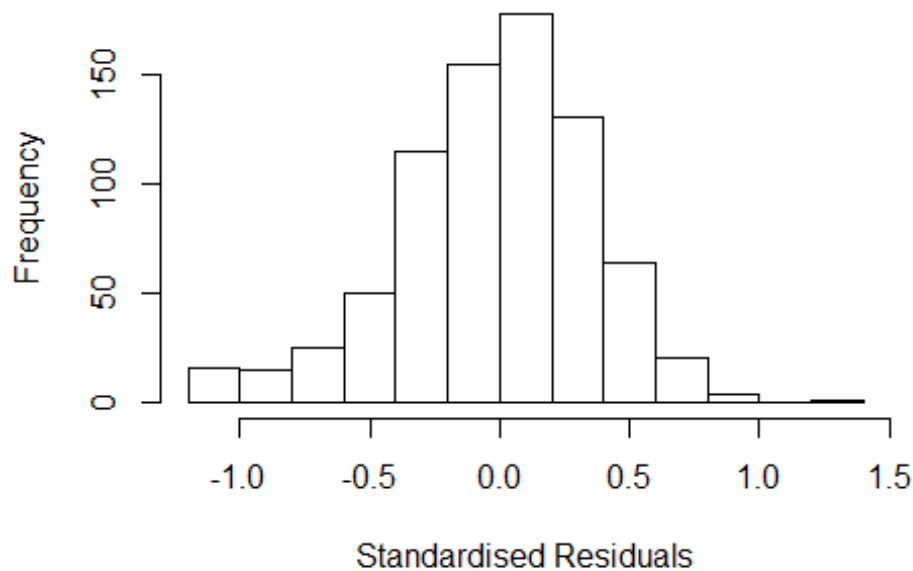
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.14354 -0.22866 0.00873 0.24286 1.28136
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.073250 0.052031 20.63 <2e-16 ***
## FirstAuthorFemale1 0.011599 0.032634 0.36 0.722
## Year1997 0.000411 0.075834 0.01 0.996
## Year1998 -0.053902 0.071274 -0.76 0.450
## Year1999 0.058033 0.075215 0.77 0.441
## Year2000 0.026489 0.081795 0.32 0.746
## Year2001 0.016949 0.074094 0.23 0.819
## Year2002 0.022874 0.069585 0.33 0.742
## Year2003 -0.034793 0.072568 -0.48 0.632
## Year2004 -0.130600 0.074090 -1.76 0.078 .
## Year2005 0.070288 0.085727 0.82 0.413
## Year2006 -0.006700 0.078177 -0.09 0.932
```

```

## Year2007      -0.052610    0.084781   -0.62    0.535
## Year2008      0.039847    0.068737    0.58    0.562
## Year2009      0.027936    0.070128    0.40    0.690
## Year2010     -0.024312    0.077895   -0.31    0.755
## Year2011      0.041451    0.080787    0.51    0.608
## Year2012     -0.011736    0.073387   -0.16    0.873
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.353
## Multiple R-squared:  0.0152, Adjusted R-squared:  -0.00695
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 711 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.160  0.880  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      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 1.091 1      1.044
## Year      1.091 16      1.003

```

Residuals from last author



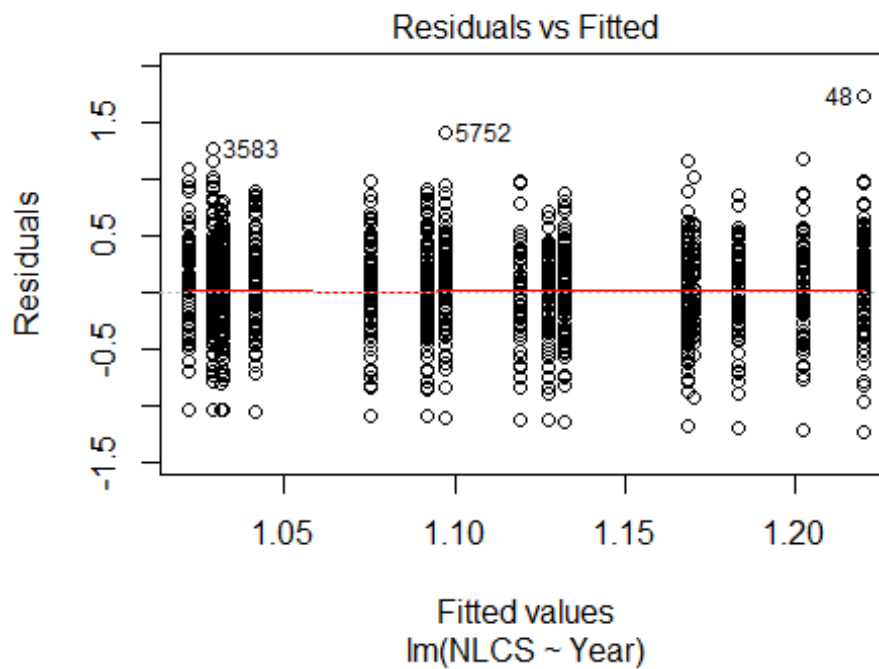
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1564 -0.2324 0.0148 0.2374 1.2771
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08e+00 5.16e-02 20.99 <2e-16 ***
## LastAuthorFemale1 -7.34e-02 3.77e-02 -1.95 0.052 .
## Year1997 9.79e-05 7.55e-02 0.00 0.999
## Year1998 -5.82e-02 7.05e-02 -0.83 0.409
## Year1999 5.94e-02 7.38e-02 0.81 0.421
## Year2000 3.06e-02 8.18e-02 0.37 0.708
## Year2001 2.53e-02 7.38e-02 0.34 0.732
## Year2002 2.82e-02 6.90e-02 0.41 0.682
## Year2003 -3.36e-02 7.10e-02 -0.47 0.636
## Year2004 -1.29e-01 7.33e-02 -1.75 0.080 .
## Year2005 7.34e-02 8.48e-02 0.87 0.387
## Year2006 -1.54e-03 7.78e-02 -0.02 0.984
```

```

## Year2007          -5.81e-02   8.39e-02   -0.69   0.489
## Year2008          4.22e-02   6.78e-02    0.62   0.533
## Year2009          2.61e-02   6.92e-02    0.38   0.706
## Year2010         -2.28e-02   7.69e-02   -0.30   0.767
## Year2011          5.08e-02   8.05e-02    0.63   0.528
## Year2012          6.83e-04   7.07e-02    0.01   0.992
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared:  0.0196, Adjusted R-squared:  -0.00243
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 707 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.150  0.875   0.950   0.897   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.29e-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: 774"
## [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
## 193 195 205 203 187 181 185 174 166 186 192 196 223 195 217
## 2011 2012
## 224 206
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 82 68 93 81 82 53 89 81 69 81 89 94 107 98 110
## 2011 2012

```

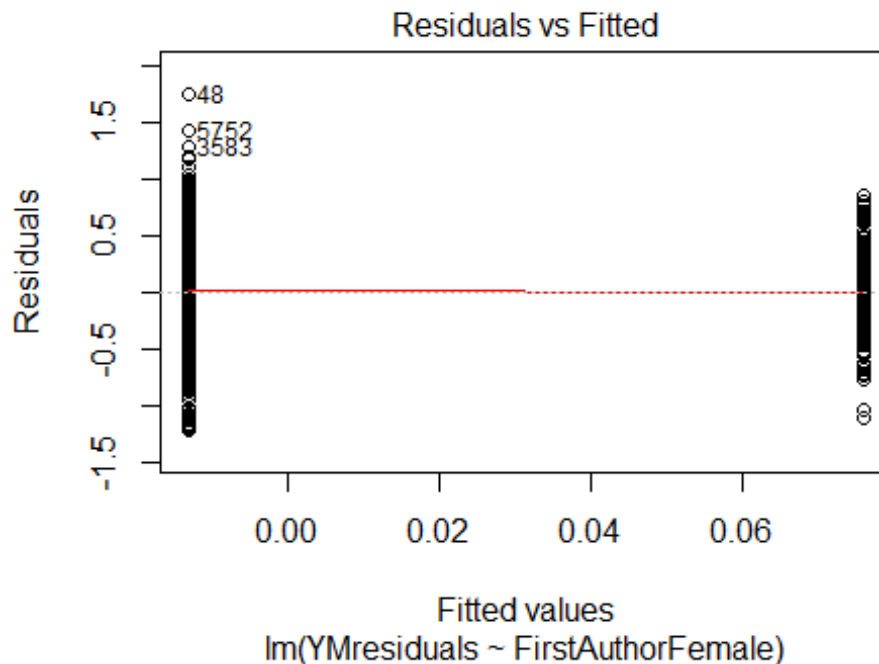
```
## 107 104
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 73 59 83 69 69 46 77 64 58 63 74 77 90 85 99
## 2011 2012
## 95 90
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.06
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 25, df = 1, p-value = 7e-07
## [1] "Female first author team size 2018 geometric mean: 2.02790655709495"
## [1] "Male first author team size 2018 geometric mean: 2.52946934213227"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

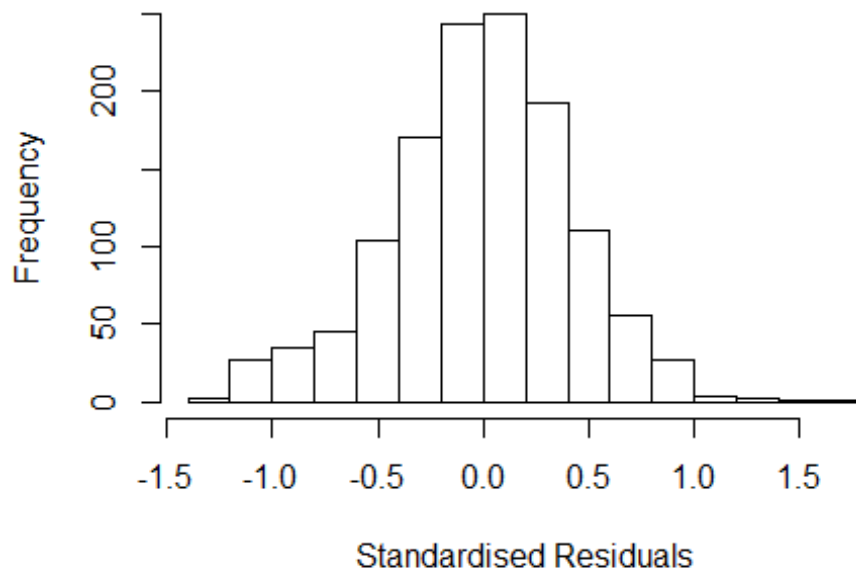
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 250, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.96934660191403"
## [1] "Male last author team size 2018 geometric mean: 2.4760239399401"

## Warning in wilcox.test.default(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] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.135  1      1.066
## LastAuthorFemale  1.131  1      1.064
## UniqueAuthors    1.320  4      1.035
## Year              1.484 16      1.012
```

Residuals from first and last author and team size



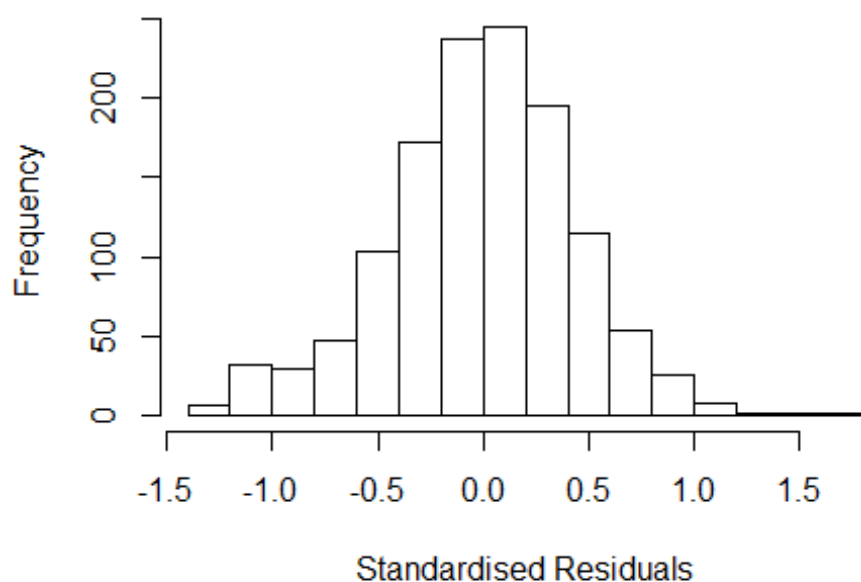
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27009 -0.27040  0.00382  0.28233  1.76999
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1840    0.0750   15.79 < 2e-16 ***
## FirstAuthorFemale1  0.0860    0.0291    2.96  0.0031 **
## LastAuthorFemale1 -0.0833    0.0396   -2.10  0.0356 *
## UniqueAuthors2     0.0602    0.0329    1.83  0.0671 .
## UniqueAuthors3     0.0861    0.0381    2.26  0.0242 *
## UniqueAuthors4     0.1331    0.0525    2.54  0.0113 *
## UniqueAuthors5     0.2396    0.0557    4.30 1.8e-05 ***
## Year1997          -0.0148    0.0905   -0.16  0.8701
## Year1998          -0.1800    0.0898   -2.00  0.0453 *
## Year1999          -0.1719    0.0869   -1.98  0.0482 *
```

```

## Year2000          -0.0286      0.0844   -0.34   0.7349
## Year2001          -0.0686      0.0897   -0.77   0.4441
## Year2002          -0.0690      0.0903   -0.76   0.4448
## Year2003          -0.1712      0.0885   -1.93   0.0533 .
## Year2004          -0.1554      0.0868   -1.79   0.0734 .
## Year2005          -0.0695      0.0886   -0.78   0.4333
## Year2006          -0.1025      0.0845   -1.21   0.2252
## Year2007          -0.2336      0.0866   -2.70   0.0070 **
## Year2008          -0.1675      0.0852   -1.97   0.0495 *
## Year2009          -0.2231      0.0845   -2.64   0.0084 **
## Year2010          -0.2348      0.0833   -2.82   0.0049 **
## Year2011          -0.2535      0.0838   -3.03   0.0025 **
## Year2012          -0.1964      0.0830   -2.37   0.0182 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0537, Adjusted R-squared:  0.037
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 1169 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0229 0.8640 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          7.87e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "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 1.129 1 1.062
## Year 1.207 16 1.006

```


Residuals from first and last author



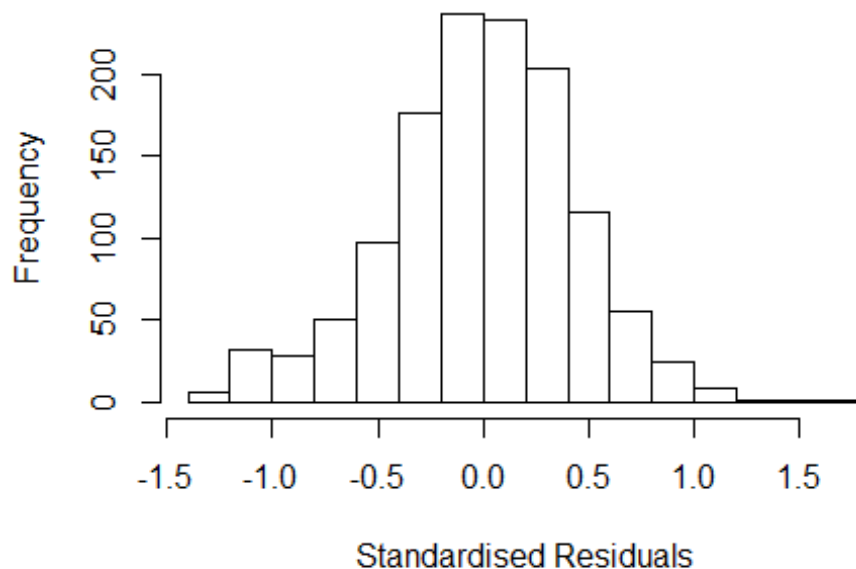
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23959 -0.26834  0.00764  0.27070  1.71441
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2396    0.0711   17.43 < 2e-16 ***
## FirstAuthorFemale1  0.1124    0.0290    3.87  0.00011 ***
## LastAuthorFemale1 -0.0727    0.0400   -1.82  0.06918 .
## Year1997          -0.0122    0.0900   -0.14  0.89189
## Year1998          -0.1870    0.0897   -2.08  0.03730 *
## Year1999          -0.1772    0.0880   -2.01  0.04433 *
## Year2000          -0.0352    0.0850   -0.41  0.67871
## Year2001          -0.0834    0.0908   -0.92  0.35853
## Year2002          -0.0816    0.0900   -0.91  0.36459
## Year2003          -0.1693    0.0890   -1.90  0.05741 .
## Year2004          -0.1411    0.0871   -1.62  0.10561
## Year2005          -0.0865    0.0888   -0.97  0.33037
```

```

## Year2006          -0.0894      0.0852   -1.05   0.29452
## Year2007          -0.2212      0.0874   -2.53   0.01148 *
## Year2008          -0.1634      0.0863   -1.89   0.05863 .
## Year2009          -0.2157      0.0844   -2.56   0.01072 *
## Year2010          -0.2245      0.0839   -2.68   0.00754 **
## Year2011          -0.2509      0.0839   -2.99   0.00285 **
## Year2012          -0.1740      0.0839   -2.07   0.03828 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.0369, Adjusted R-squared:  0.0231
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 100 weights are ~= 1. The remaining 1171 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0438 0.8650 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      7.87e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



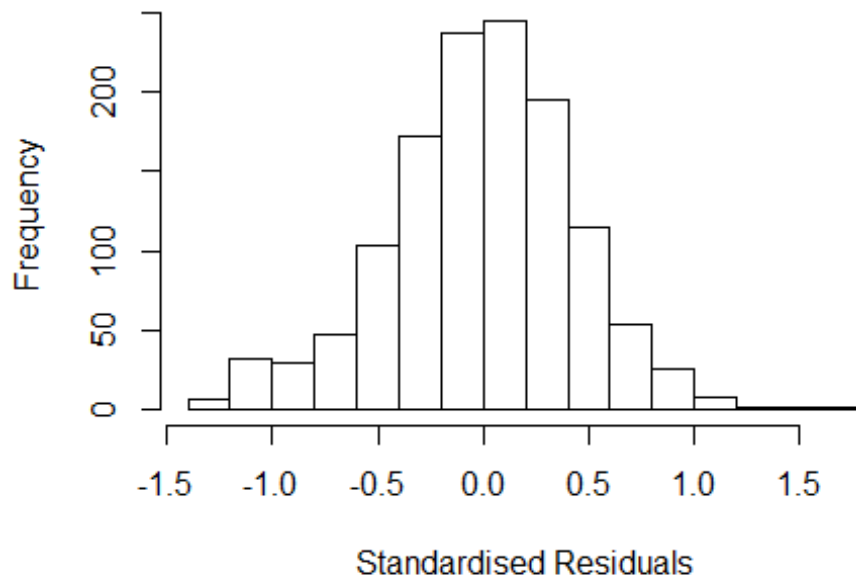
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23898 -0.26894 0.00697 0.27412 1.71502
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2390 0.0708 17.51 < 2e-16 ***
## FirstAuthorFemale1 0.1040 0.0291 3.57 0.00036 ***
## Year1997 -0.0177 0.0896 -0.20 0.84339
## Year1998 -0.1909 0.0894 -2.13 0.03304 *
## Year1999 -0.1901 0.0875 -2.17 0.02994 *
## Year2000 -0.0411 0.0849 -0.48 0.62843
## Year2001 -0.0970 0.0897 -1.08 0.27998
## Year2002 -0.0943 0.0892 -1.06 0.29067
## Year2003 -0.1759 0.0884 -1.99 0.04679 *
## Year2004 -0.1442 0.0867 -1.66 0.09646 .
## Year2005 -0.0990 0.0879 -1.13 0.26026
## Year2006 -0.0904 0.0852 -1.06 0.28885
```

```

## Year2007          -0.2249      0.0871   -2.58  0.00989 **
## Year2008          -0.1649      0.0860   -1.92  0.05544 .
## Year2009          -0.2174      0.0841   -2.58  0.00989 **
## Year2010          -0.2295      0.0834   -2.75  0.00605 **
## Year2011          -0.2564      0.0834   -3.07  0.00216 **
## Year2012          -0.1791      0.0838   -2.14  0.03275 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0346, Adjusted R-squared:  0.0215
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 100 weights are ~= 1. The remaining 1171 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0459 0.8650 0.9550 0.8980 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.87e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



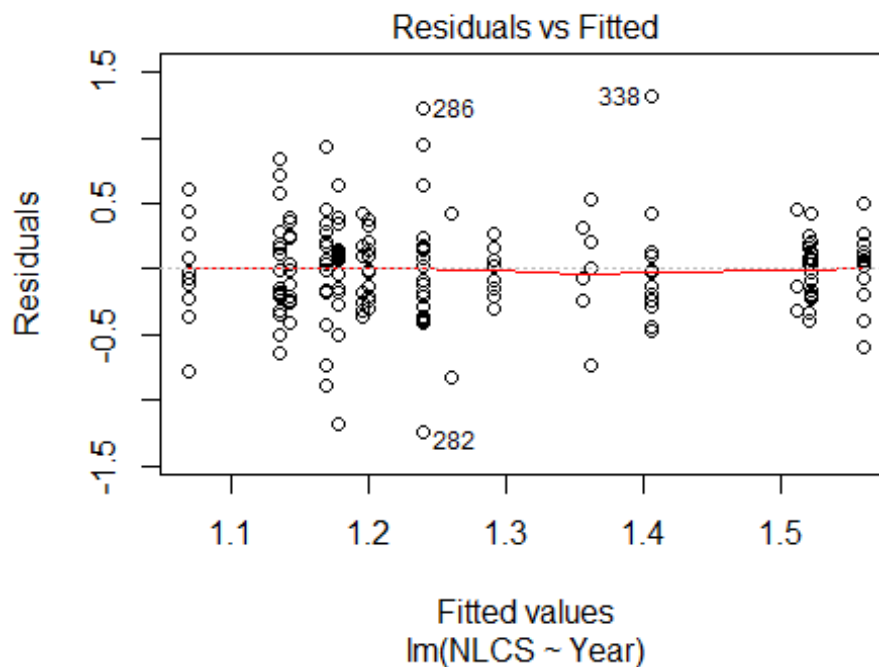
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2482 -0.2724 0.0049 0.2811 1.7058
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2482 0.0695 17.95 <2e-16 ***
## LastAuthorFemale1 -0.0531 0.0402 -1.32 0.1867
## Year1997 -0.0141 0.0885 -0.16 0.8733
## Year1998 -0.1811 0.0886 -2.04 0.0411 *
## Year1999 -0.1754 0.0871 -2.02 0.0441 *
## Year2000 -0.0359 0.0842 -0.43 0.6703
## Year2001 -0.0835 0.0889 -0.94 0.3478
## Year2002 -0.0734 0.0885 -0.83 0.4075
## Year2003 -0.1676 0.0870 -1.93 0.0544 .
## Year2004 -0.1351 0.0860 -1.57 0.1162
## Year2005 -0.0882 0.0880 -1.00 0.3164
## Year2006 -0.0783 0.0846 -0.93 0.3547
```

```

## Year2007          -0.2088      0.0862   -2.42   0.0156 *
## Year2008          -0.1490      0.0848   -1.76   0.0789 .
## Year2009          -0.2062      0.0826   -2.49   0.0127 *
## Year2010          -0.2115      0.0829   -2.55   0.0108 *
## Year2011          -0.2360      0.0834   -2.83   0.0047 **
## Year2012          -0.1563      0.0816   -1.92   0.0557 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0283, Adjusted R-squared:  0.0151
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 95 weights are ~= 1. The remaining 1176 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0492 0.8690 0.9510 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      7.87e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1271"
## [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
##   27   12   15   15   13   24   23    9   15   16   27   24   15   16   25
## 2011 2012
##   20   29
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   13    7    3    3    3   13   18    4    8   10   16   20   11   13   17
## 2011 2012

```

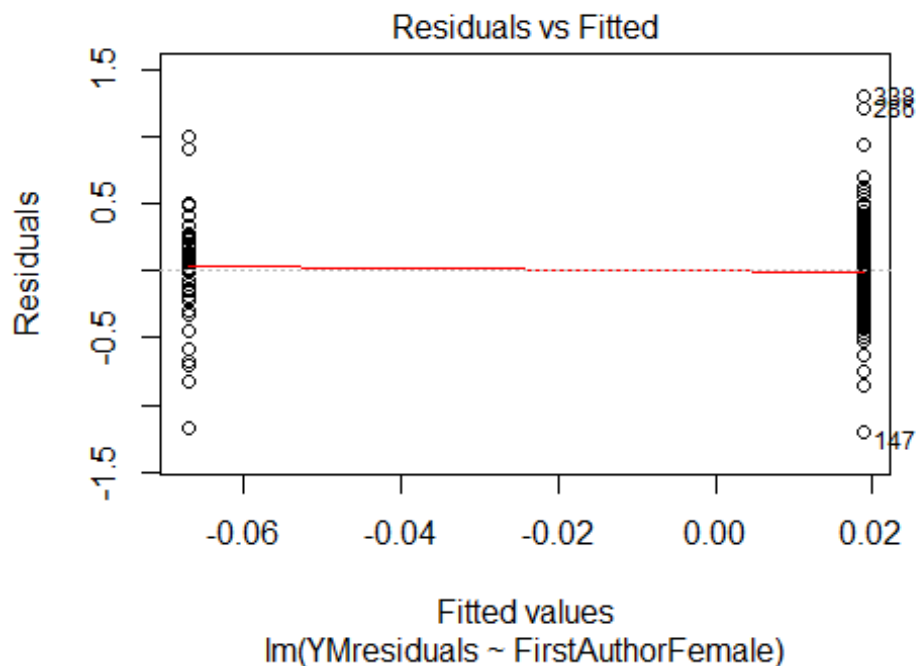
```
## 13 20
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 6 2 2 3 7 17 4 8 9 13 15 8 10 15
## 2011 2012
## 11 15
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 34, df = 16, p-value = 0.005
```



```
##
## 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: 4.37951913988789"
## [1] "Male first author team size 2018 geometric mean: 3.13671639180246"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

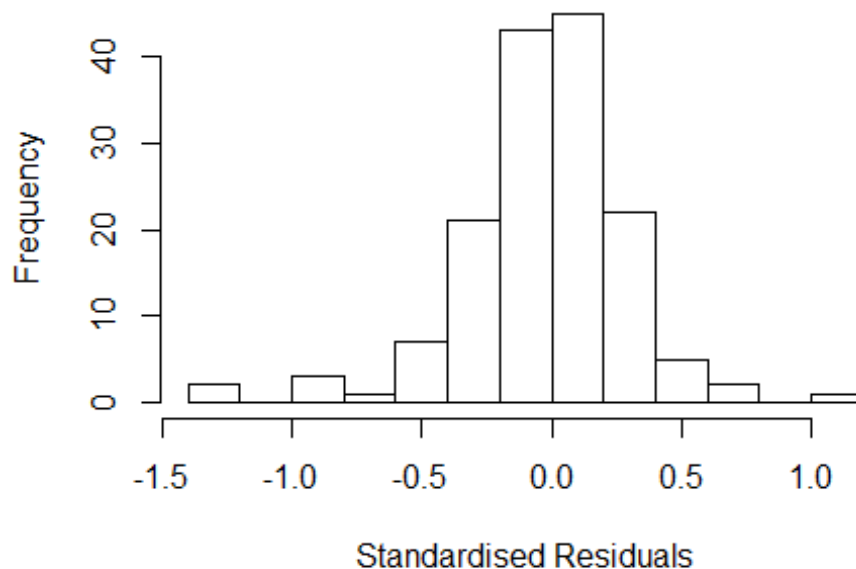
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.1924941018511"
## [1] "Male last author team size 2018 geometric mean: 3.00599225761114"

## Warning in wilcox.test.default(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 = 0.1
## 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 2.519 1      1.587
## LastAuthorFemale  3.061 1      1.750
## UniqueAuthors    20.113 4      1.455
## Year              68.640 16     1.141
```


Residuals from first and last author and team size



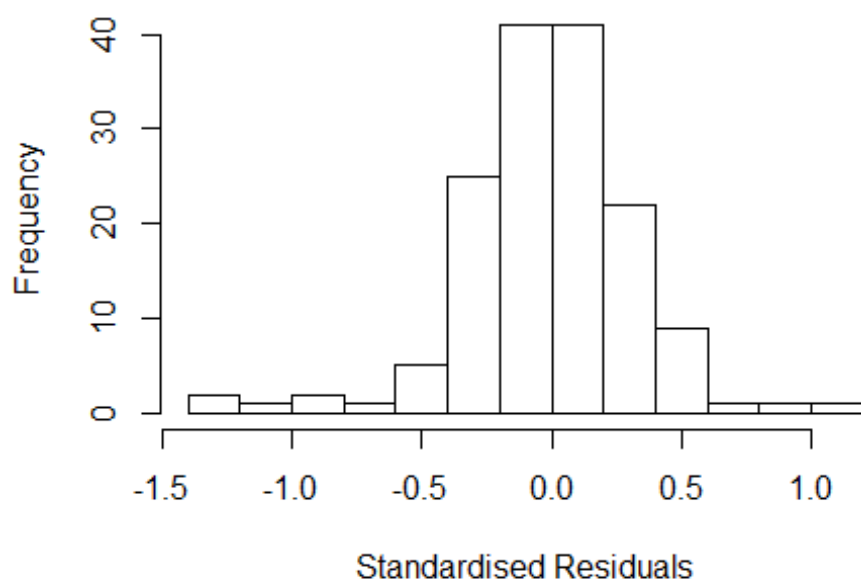
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23657 -0.16224 -0.00768  0.15971  1.04538
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.650218   0.101827   16.21 < 2e-16 ***
## FirstAuthorFemale1  0.026961   0.074839    0.36  0.71924
## LastAuthorFemale1 -0.165422   0.074223   -2.23  0.02756 *
## UniqueAuthors2     0.000613   0.062035    0.01  0.99214
## UniqueAuthors3     0.142296   0.089648    1.59  0.11490
## UniqueAuthors4     0.236423   0.103165    2.29  0.02354 *
## UniqueAuthors5     0.303301   0.117136    2.59  0.01072 *
## Year1997          -0.205632   0.160936   -1.28  0.20364
## Year1998           0.018282   0.290677    0.06  0.94995
## Year1999          -0.451524   0.120770   -3.74  0.00028 ***
```

```

## Year2000      -0.018522    0.131605    -0.14    0.88829
## Year2001      -0.173780    0.118031    -1.47    0.14337
## Year2002      -0.414261    0.128891    -3.21    0.00165 **
## Year2003      -0.128217    0.272488    -0.47    0.63876
## Year2004      -0.449858    0.175208    -2.57    0.01138 *
## Year2005      -0.379507    0.123548    -3.07    0.00260 **
## Year2006      -0.431015    0.123000    -3.50    0.00063 ***
## Year2007      -0.550053    0.140824    -3.91    0.00015 ***
## Year2008      -0.657637    0.153380    -4.29    3.5e-05 ***
## Year2009      -0.388856    0.128837    -3.02    0.00306 **
## Year2010      -0.602429    0.140919    -4.27    3.7e-05 ***
## Year2011      -0.487100    0.146488    -3.33    0.00115 **
## Year2012      -0.733821    0.146416    -5.01    1.7e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.274
## Multiple R-squared:  0.388, Adjusted R-squared:  0.283
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 142 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0053 0.8960 0.9610 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.58e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.164 1          1.471
## LastAuthorFemale 2.042 1          1.429
## Year      4.270 16          1.046

```

Residuals from first and last author



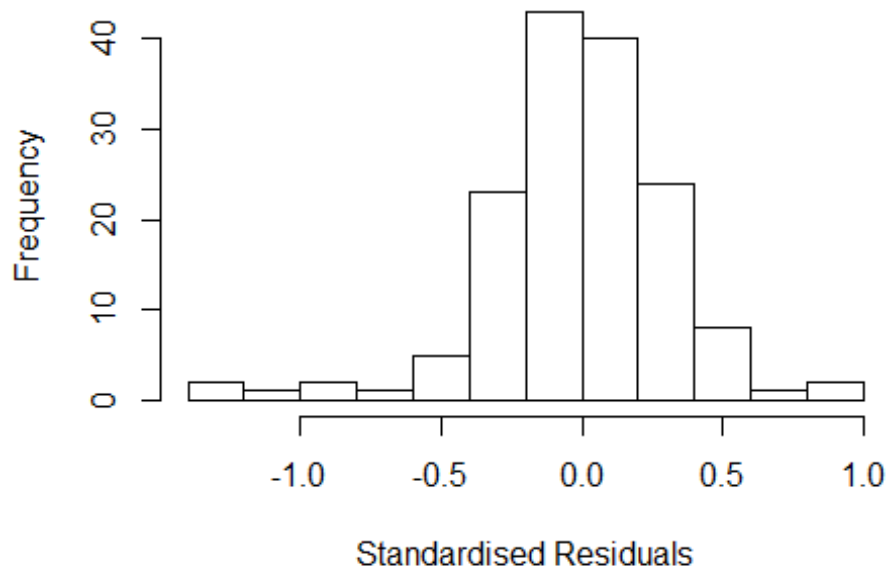
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.319 -0.167 -0.015  0.164  1.067
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.6501     0.1027   16.06 < 2e-16 ***
## FirstAuthorFemale1  0.0192     0.0796    0.24  0.81009
## LastAuthorFemale1 -0.1426     0.0697   -2.04  0.04285 *
## Year1997          -0.1472     0.1530   -0.96  0.33778
## Year1998           0.0184     0.2854    0.06  0.94863
## Year1999          -0.4511     0.1195   -3.77  0.00024 ***
## Year2000           0.0969     0.1205    0.80  0.42264
## Year2001          -0.1732     0.1175   -1.47  0.14266
## Year2002          -0.3765     0.1308   -2.88  0.00467 **
## Year2003          -0.1490     0.2835   -0.53  0.60017
## Year2004          -0.4033     0.1758   -2.29  0.02333 *
## Year2005          -0.3286     0.1177   -2.79  0.00603 **
```

```

## Year2006          -0.4188      0.1236   -3.39  0.00092 ***
## Year2007          -0.5126      0.1301   -3.94  0.00013 ***
## Year2008          -0.5557      0.1571   -3.54  0.00056 ***
## Year2009          -0.3253      0.1310   -2.48  0.01429 *
## Year2010          -0.4966      0.1482   -3.35  0.00105 **
## Year2011          -0.4620      0.1544   -2.99  0.00331 **
## Year2012          -0.6276      0.1357   -4.62  8.8e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.285
## Multiple R-squared:  0.322, Adjusted R-squared:  0.231
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 20 is an outlier with |weight| <= 0.00048 ( < 0.00066);
## 8 weights are ~= 1. The remaining 143 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0078 0.9000 0.9620 0.9000 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          6.58e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.708 1          1.646
## Year              2.708 16          1.032

```

Residuals from first author



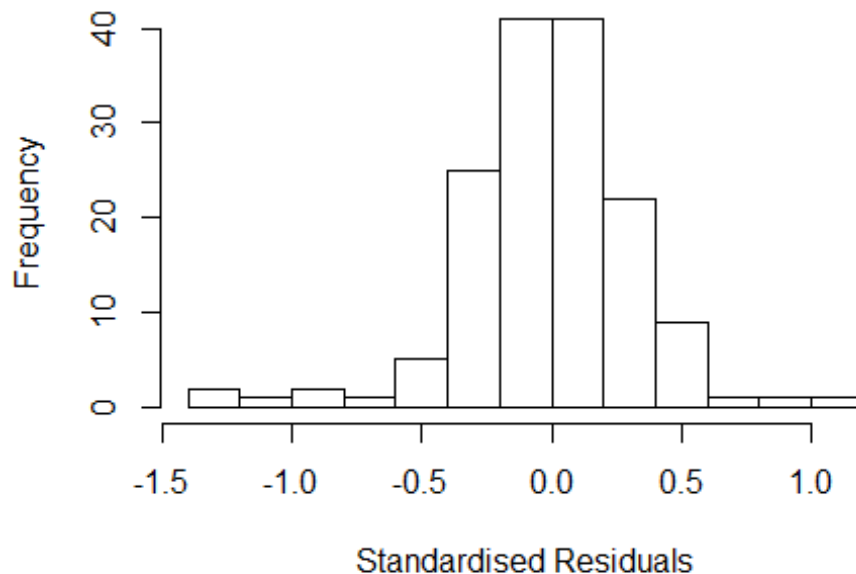
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2431 -0.1833 -0.0013 0.1616 0.9256
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.65078 0.10129 16.30 < 2e-16 ***
## FirstAuthorFemale1 0.00268 0.08167 0.03 0.97391
## Year1997 -0.14806 0.15192 -0.97 0.33154
## Year1998 0.01772 0.28349 0.06 0.95024
## Year1999 -0.45178 0.11828 -3.82 0.00020 ***
## Year2000 0.00923 0.11271 0.08 0.93487
## Year2001 -0.17390 0.11616 -1.50 0.13674
## Year2002 -0.40763 0.12583 -3.24 0.00151 **
## Year2003 -0.15946 0.28670 -0.56 0.57901
## Year2004 -0.46515 0.17272 -2.69 0.00799 **
## Year2005 -0.32576 0.11634 -2.80 0.00586 **
## Year2006 -0.45882 0.12038 -3.81 0.00021 ***
```

```

## Year2007          -0.51120      0.13113      -3.90      0.00015 ***
## Year2008          -0.62816      0.14433      -4.35      2.7e-05 ***
## Year2009          -0.35344      0.13259      -2.67      0.00863 **
## Year2010          -0.48208      0.14821      -3.25      0.00145 **
## Year2011          -0.51452      0.15102      -3.41      0.00087 ***
## Year2012          -0.65632      0.14136      -4.64      8.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.287
## Multiple R-squared:  0.293, Adjusted R-squared:  0.203
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 136 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0213 0.8800 0.9550 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      6.58e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 2.131 1      1.460
## Year              2.131 16      1.024

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3158 -0.1704 -0.0177 0.1781 1.0817
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.6508 0.1017 16.23 < 2e-16 ***
## LastAuthorFemale1 -0.1400 0.0725 -1.93 0.05541 .
## Year1997 -0.1483 0.1519 -0.98 0.33072
## Year1998 0.0177 0.2825 0.06 0.95021
## Year1999 -0.4518 0.1186 -3.81 0.00021 ***
## Year2000 0.0930 0.1235 0.75 0.45275
## Year2001 -0.1739 0.1165 -1.49 0.13800
## Year2002 -0.3742 0.1265 -2.96 0.00367 **
## Year2003 -0.1634 0.2778 -0.59 0.55731
## Year2004 -0.3953 0.1666 -2.37 0.01909 *
## Year2005 -0.3253 0.1138 -2.86 0.00496 **
## Year2006 -0.4167 0.1198 -3.48 0.00068 ***
```

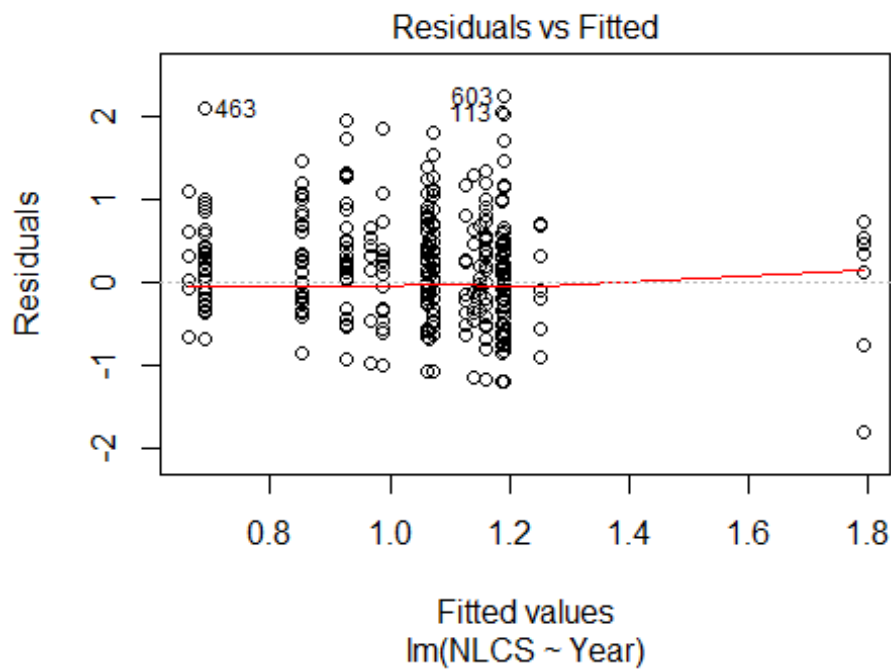
```

## Year2007          -0.5104      0.1275    -4.00   0.00010 ***
## Year2008          -0.5474      0.1427    -3.84   0.00019 ***
## Year2009          -0.3243      0.1295    -2.50   0.01347 *
## Year2010          -0.4955      0.1478    -3.35   0.00104 **
## Year2011          -0.4575      0.1490    -3.07   0.00259 **
## Year2012          -0.6200      0.1271    -4.88   3e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.289
## Multiple R-squared:  0.317, Adjusted R-squared:  0.23
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 144 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0033 0.9020 0.9630 0.8960 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      6.58e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 152"
## [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
##    9    7   11   11   12   13   11   40   21   18   54   42   61   68   62
## 2011 2012
##   68   71
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8    7    8   10    8   10    9   32   18    8   45   33   45   48   42
## 2011 2012

```



```
## 50 51
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 7 8 10 7 10 8 29 16 8 39 29 41 42 40
## 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 = 19, df = 16, p-value = 0.3
```



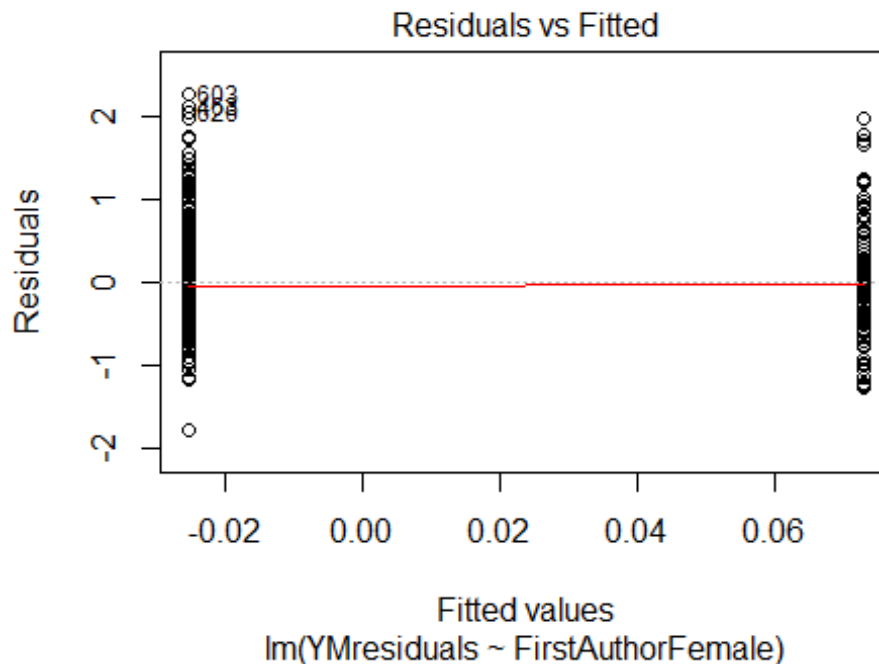
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.37, df = 1, p-value = 0.5

## [1] "Female first author team size 2018 geometric mean: 1.59713775274758"
## [1] "Male first author team size 2018 geometric mean: 1.41175892323576"

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

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.36426160182137"
## [1] "Male last author team size 2018 geometric mean: 1.47985984489079"

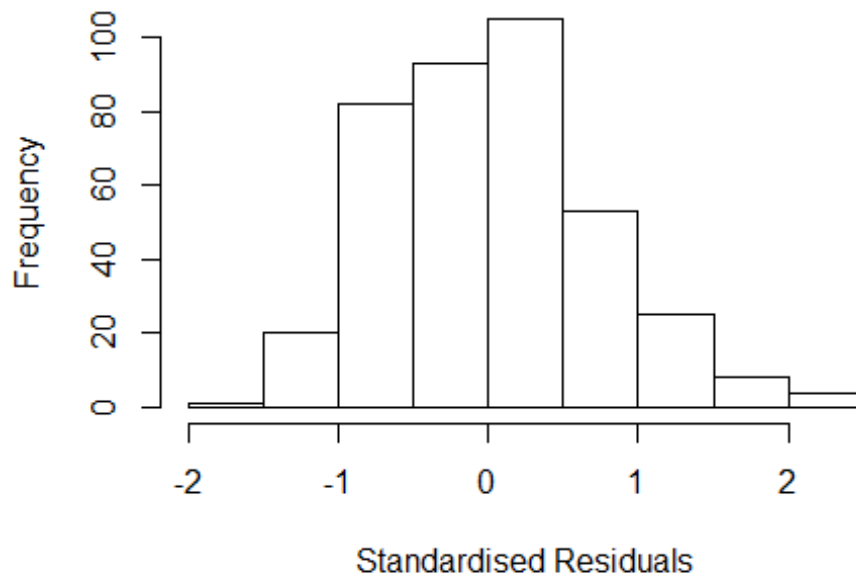
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120, p-value = 0.7
## 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	8.667	1	2.944
LastAuthorFemale	10.908	1	3.303
UniqueAuthors	48.661	4	1.625
Year	36.299	16	1.119

Residuals from first and last author and team size



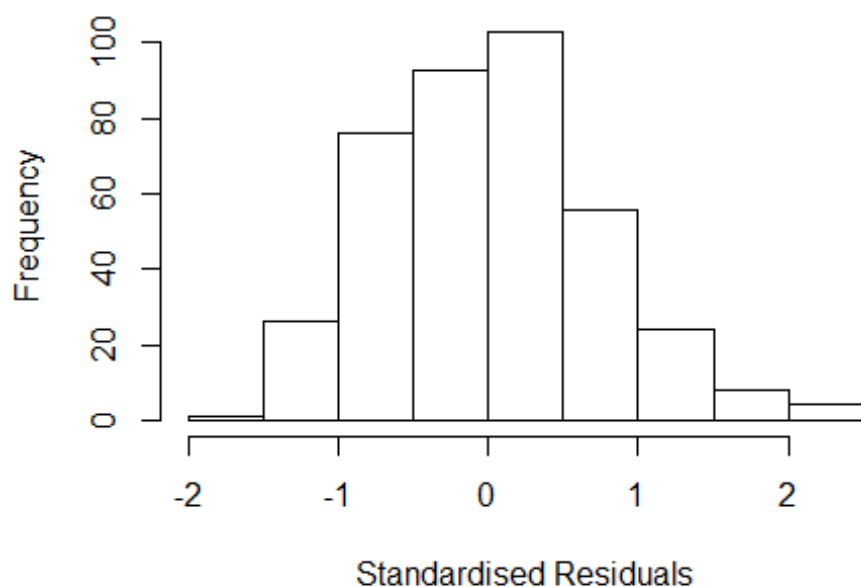
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.77e+00 -5.12e-01 -3.33e-16 4.38e-01 2.40e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5122 0.2474 2.07 0.0392 *
## FirstAuthorFemale1 0.1202 0.1205 1.00 0.3192
## LastAuthorFemale1 0.0243 0.1257 0.19 0.8467
## UniqueAuthors2 0.1821 0.1056 1.72 0.0855 .
## UniqueAuthors3 0.1990 0.1207 1.65 0.0999 .
## UniqueAuthors4 0.7059 0.1508 4.68 4e-06 ***
## UniqueAuthors5 0.0791 0.2004 0.39 0.6932
## Year1997 0.6782 0.3444 1.97 0.0496 *
## Year1998 0.5137 0.3924 1.31 0.1913
## Year1999 0.3306 0.3205 1.03 0.3029
```

```

## Year2000          1.2616      0.3939      3.20      0.0015 **
## Year2001          0.5095      0.2952      1.73      0.0852 .
## Year2002          0.3746      0.2749      1.36      0.1738
## Year2003          0.5417      0.2678      2.02      0.0438 *
## Year2004          0.2770      0.3085      0.90      0.3698
## Year2005          0.5237      0.2692      1.95      0.0525 .
## Year2006          0.4114      0.2764      1.49      0.1375
## Year2007          0.6852      0.2807      2.44      0.0151 *
## Year2008          0.4851      0.2746      1.77      0.0782 .
## Year2009          0.0848      0.2648      0.32      0.7489
## Year2010          0.3148      0.2794      1.13      0.2606
## Year2011          0.5234      0.2748      1.90      0.0577 .
## Year2012          0.2267      0.2766      0.82      0.4130
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.734
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0487
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 354 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.264  0.876  0.950  0.911  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.56e-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 1.970 1      1.403
## LastAuthorFemale  2.037 1      1.427
## Year              1.296 16      1.008

```

Residuals from first and last author



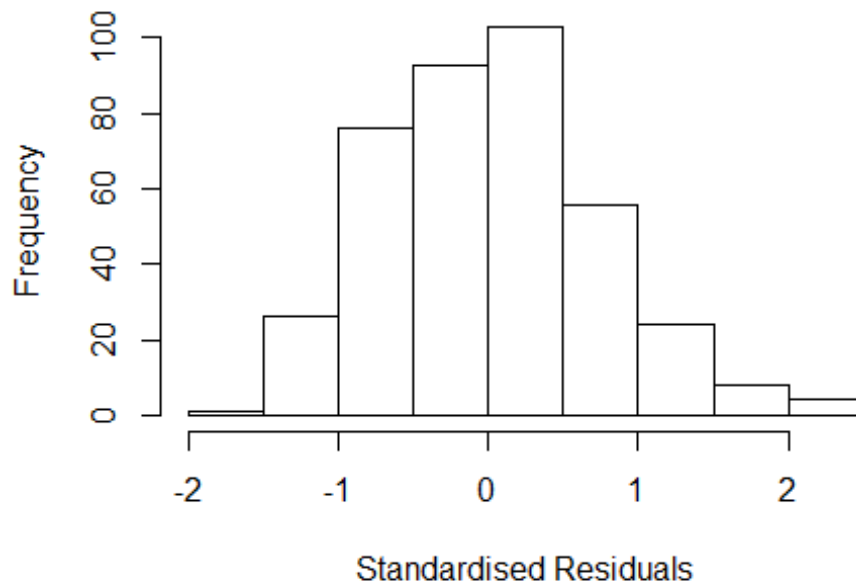
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.880079 -0.535878 -0.000454 0.466280 2.338551
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.56227 0.26199 2.15 0.0325 *
## FirstAuthorFemale1 0.14076 0.11194 1.26 0.2094
## LastAuthorFemale1 -0.00302 0.11728 -0.03 0.9795
## Year1997 0.65962 0.34877 1.89 0.0594 .
## Year1998 0.52779 0.38612 1.37 0.1725
## Year1999 0.38870 0.32395 1.20 0.2310
## Year2000 1.31781 0.40126 3.28 0.0011 **
## Year2001 0.49292 0.31694 1.56 0.1207
## Year2002 0.39025 0.28656 1.36 0.1741
## Year2003 0.56207 0.27868 2.02 0.0444 *
## Year2004 0.31484 0.32494 0.97 0.3332
## Year2005 0.56335 0.28369 1.99 0.0478 *
```

```

## Year2006          0.43933      0.29142      1.51      0.1325
## Year2007          0.71516      0.29402      2.43      0.0155 *
## Year2008          0.47544      0.28794      1.65      0.0995 .
## Year2009          0.07599      0.27837      0.27      0.7850
## Year2010          0.29518      0.29024      1.02      0.3098
## Year2011          0.53218      0.28860      1.84      0.0660 .
## Year2012          0.24983      0.29094      0.86      0.3911
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.743
## Multiple R-squared:  0.0887, Adjusted R-squared:  0.0446
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 29 weights are ~= 1. The remaining 362 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.301  0.882  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      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 1.133 1      1.064
## Year              1.133 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
## -1.88126 -0.53468 -0.00024 0.46608 2.33955
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5618 0.2618 2.15 0.0325 *
## FirstAuthorFemale1 0.1387 0.0852 1.63 0.1042
## Year1997 0.6599 0.3491 1.89 0.0595 .
## Year1998 0.5279 0.3865 1.37 0.1728
## Year1999 0.3889 0.3251 1.20 0.2323
## Year2000 1.3195 0.4009 3.29 0.0011 **
## Year2001 0.4928 0.3170 1.55 0.1209
## Year2002 0.3911 0.2852 1.37 0.1711
## Year2003 0.5620 0.2789 2.02 0.0446 *
## Year2004 0.3148 0.3248 0.97 0.3332
## Year2005 0.5640 0.2827 2.00 0.0467 *
## Year2006 0.4395 0.2911 1.51 0.1319
```

```

## Year2007          0.7158      0.2940      2.43      0.0154 *
## Year2008          0.4758      0.2880      1.65      0.0993 .
## Year2009          0.0757      0.2784      0.27      0.7858
## Year2010          0.2954      0.2902      1.02      0.3094
## Year2011          0.5316      0.2885      1.84      0.0661 .
## Year2012          0.2496      0.2911      0.86      0.3918
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.74
## Multiple R-squared:  0.0889, Adjusted R-squared:  0.0474
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 29 weights are ~= 1. The remaining 362 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.297  0.882  0.948  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      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 1.168 1      1.081
## Year      1.168 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
## -1.87956 -0.54137 -0.00145  0.48616  2.33218

```



```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5685    0.2620    2.17  0.0307 *
## LastAuthorFemale1 0.1031    0.0889    1.16  0.2473
## Year1997        0.6615    0.3485    1.90  0.0585 .
## Year1998        0.5289    0.3830    1.38  0.1681
## Year1999        0.3858    0.3266    1.18  0.2382
## Year2000        1.3110    0.4012    3.27  0.0012 **
## Year2001        0.4977    0.3184    1.56  0.1189
## Year2002        0.4236    0.2916    1.45  0.1471
## Year2003        0.5664    0.2801    2.02  0.0438 *
## Year2004        0.3105    0.3234    0.96  0.3376
## Year2005        0.5755    0.2834    2.03  0.0430 *
## Year2006        0.4592    0.2907    1.58  0.1150
## Year2007        0.7339    0.2932    2.50  0.0127 *
## Year2008        0.4778    0.2879    1.66  0.0978 .
## Year2009        0.0695    0.2787    0.25  0.8033
## Year2010        0.2981    0.2903    1.03  0.3052
## Year2011        0.5323    0.2890    1.84  0.0663 .
## Year2012        0.2647    0.2921    0.91  0.3655
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.744
## Multiple R-squared:  0.0859, Adjusted R-squared:  0.0442
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 360 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.305  0.880   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      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: 391"
## [1] ""

```

```

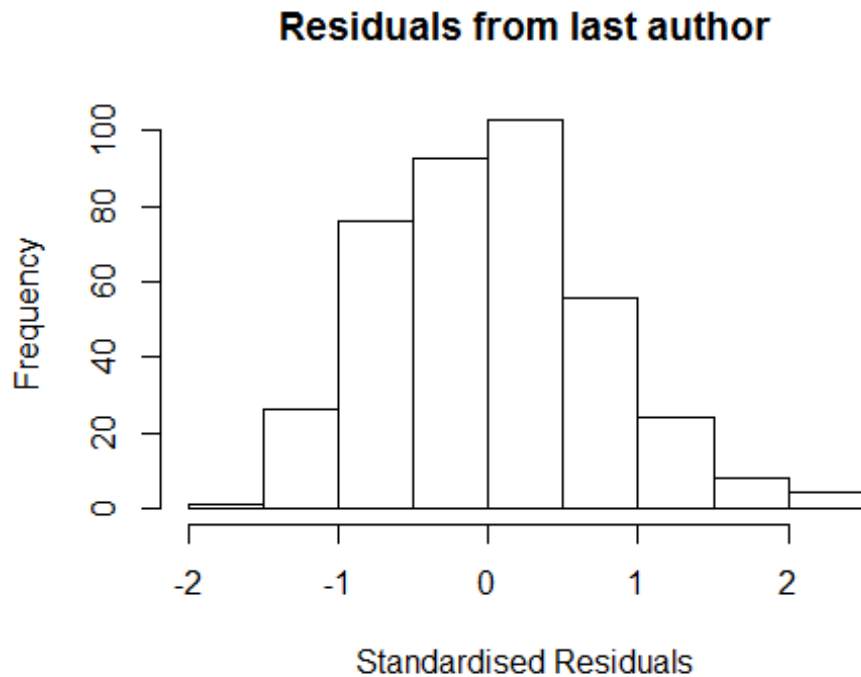
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2001"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4   12    6    9   10    6    7    6    7   15   19   18   16   33   39
## 2011 2012
##   40   28
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    7    6    9    9    1    5    3    5   13   18   16   13   32   31
## 2011 2012
##   32   21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    7    6    9    8    1    5    2    4   13   17   13   12   31   25
## 2011 2012
##   30   19
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.78160847012492"
## [1] "Male first author team size 2018 geometric mean: 1.49451163002088"

## Warning in wilcox.test.default(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 = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.1748021039364"
## [1] "Male last author team size 2018 geometric mean: 1.63577418436487"

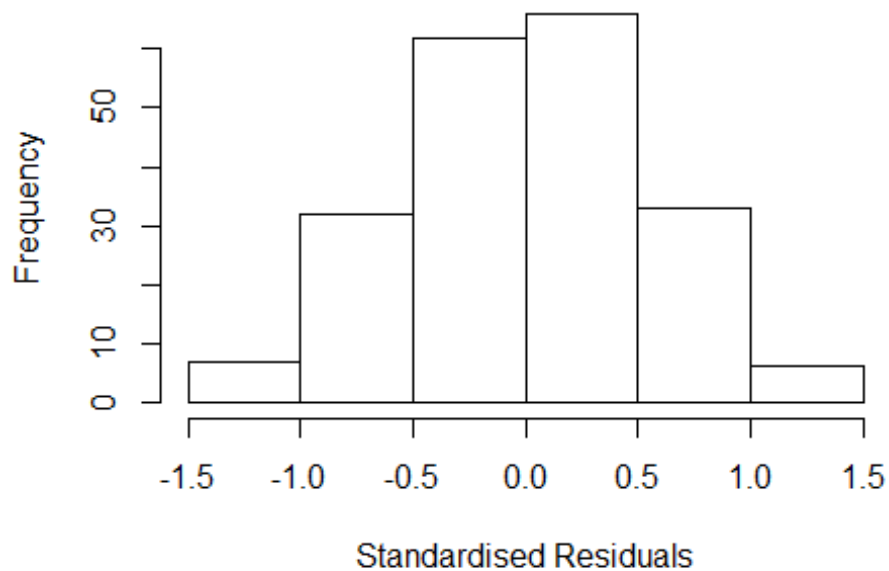
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 70, p-value = 0.1
## 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 2.117  1      1.455
## LastAuthorFemale  2.285  1      1.512
## UniqueAuthors    3.818  3      1.250
## Year              8.116 16      1.068
```

Residuals from first and last author and team size



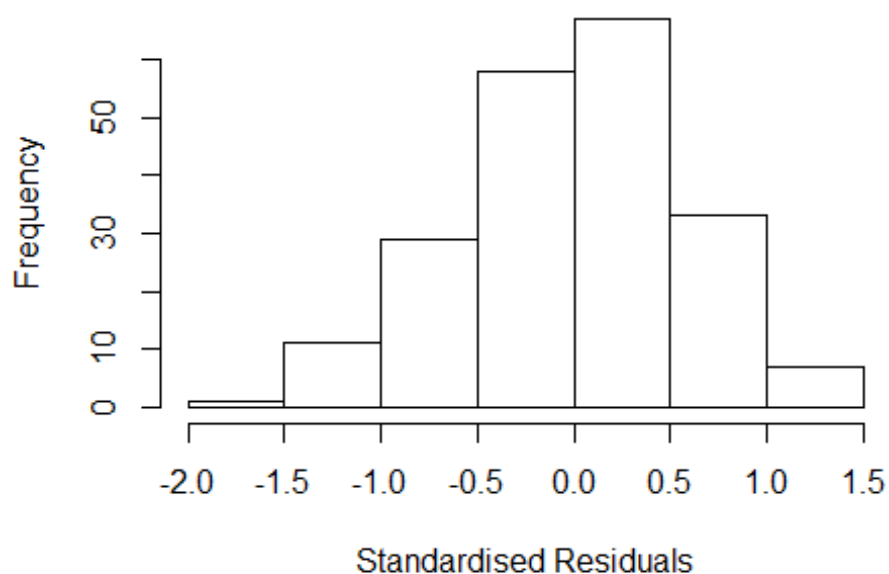
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3289 -0.3506 0.0149 0.3443 1.3093
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3446 0.1629 2.12 0.03572 *
## FirstAuthorFemale1 -0.2243 0.1083 -2.07 0.03977 *
## LastAuthorFemale1 0.0809 0.1148 0.70 0.48200
## UniqueAuthors2 0.2316 0.0967 2.39 0.01769 *
## UniqueAuthors3 0.4587 0.1194 3.84 0.00017 ***
## UniqueAuthors5 0.7066 0.1391 5.08 9.2e-07 ***
## Year1997 0.1672 0.2459 0.68 0.49746
## Year1998 0.3361 0.2672 1.26 0.20993
## Year1999 0.6452 0.2182 2.96 0.00352 **
## Year2000 0.7083 0.2408 2.94 0.00369 **
```

```

## Year2001          0.8194      0.1629      5.03  1.2e-06 ***
## Year2002          0.3201      0.2567      1.25  0.21397
## Year2003          0.9894      0.7557      1.31  0.19208
## Year2004          0.9120      0.3514      2.60  0.01021 *
## Year2005          0.9843      0.2035      4.84  2.8e-06 ***
## Year2006          0.8135      0.1883      4.32  2.6e-05 ***
## Year2007          0.7914      0.2334      3.39  0.00085 ***
## Year2008          0.3952      0.2466      1.60  0.11076
## Year2009          0.8023      0.1950      4.12  5.8e-05 ***
## Year2010          0.3198      0.1844      1.73  0.08453 .
## Year2011          0.6356      0.1921      3.31  0.00113 **
## Year2012          0.5434      0.1960      2.77  0.00612 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.589
## Multiple R-squared:  0.257, Adjusted R-squared:  0.173
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 189 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.590  0.879  0.965  0.921  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      4.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.130 1 1.459
## LastAuthorFemale 2.102 1 1.450
## Year 2.534 16 1.029

```

Residuals from first and last author



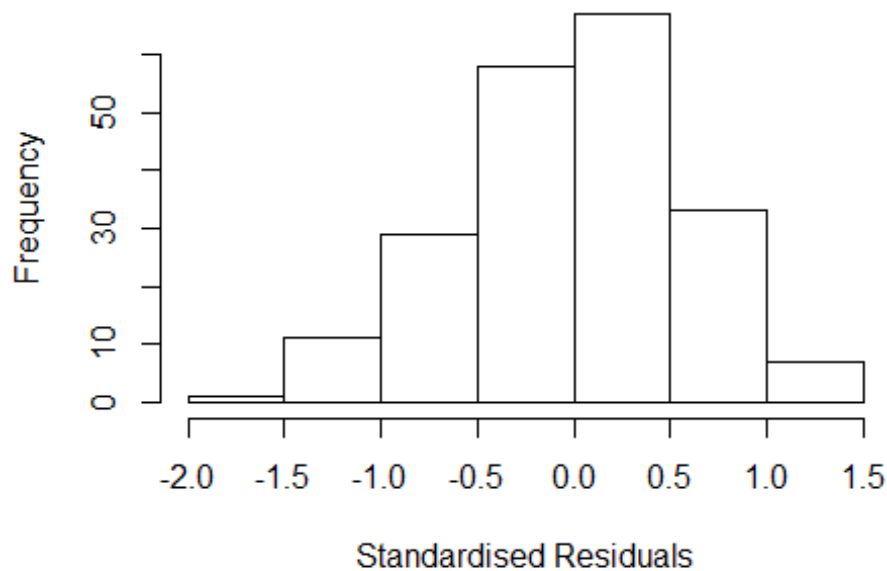
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.528 -0.373 0.014 0.391 1.218
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5239 0.2748 1.91 0.05810 .
## FirstAuthorFemale1 -0.1726 0.1137 -1.52 0.13067
## LastAuthorFemale1 0.0243 0.1251 0.19 0.84645
## Year1997 -0.0099 0.3315 -0.03 0.97621
## Year1998 0.2484 0.3353 0.74 0.45970
## Year1999 0.5855 0.3273 1.79 0.07525 .
## Year2000 0.5551 0.3190 1.74 0.08346 .
## Year2001 0.6401 0.2748 2.33 0.02089 *
## Year2002 0.1915 0.3574 0.54 0.59269
## Year2003 0.8101 0.7933 1.02 0.30845
## Year2004 0.7847 0.3816 2.06 0.04117 *
## Year2005 1.0037 0.2995 3.35 0.00098 ***
```

```

## Year2006          0.8042      0.2908      2.77  0.00625 **
## Year2007          0.7244      0.3327      2.18  0.03071 *
## Year2008          0.2661      0.3270      0.81  0.41685
## Year2009          0.7705      0.3114      2.47  0.01425 *
## Year2010          0.3097      0.2894      1.07  0.28596
## Year2011          0.6646      0.3038      2.19  0.02996 *
## Year2012          0.6184      0.3085      2.00  0.04641 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.584
## Multiple R-squared:  0.181, Adjusted R-squared:  0.102
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 183 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.474  0.863   0.949   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      4.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.841 1      1.357
## Year              1.841 16      1.019

```

Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.528 -0.371 0.019 0.390 1.211
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.53211 0.27684 1.92 0.0561 .
## FirstAuthorFemale1 -0.15989 0.10823 -1.48 0.1413
## Year1997 -0.00909 0.33293 -0.03 0.9783
## Year1998 0.24722 0.33944 0.73 0.4673
## Year1999 0.58120 0.33214 1.75 0.0818 .
## Year2000 0.54699 0.32053 1.71 0.0896 .
## Year2001 0.63189 0.27684 2.28 0.0236 *
## Year2002 0.18592 0.36024 0.52 0.6064
## Year2003 0.80189 0.78451 1.02 0.3080
## Year2004 0.77667 0.38233 2.03 0.0436 *
## Year2005 0.99547 0.30279 3.29 0.0012 **
## Year2006 0.79900 0.29286 2.73 0.0070 **
```

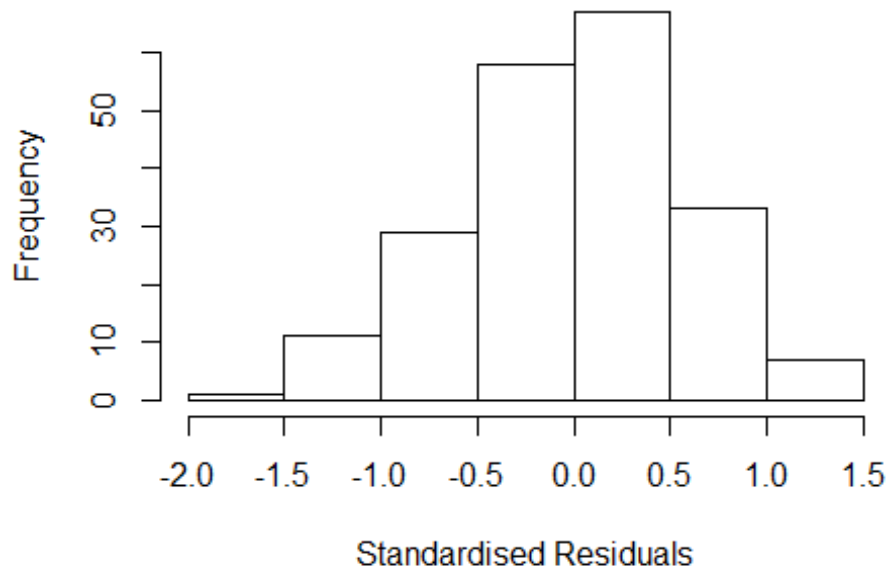


```

## Year2007          0.71613      0.33562      2.13      0.0342 *
## Year2008          0.25700      0.32547      0.79      0.4307
## Year2009          0.76456      0.31534      2.42      0.0163 *
## Year2010          0.30357      0.29180      1.04      0.2995
## Year2011          0.65824      0.30367      2.17      0.0314 *
## Year2012          0.61585      0.31131      1.98      0.0494 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.593
## Multiple R-squared:  0.179, Adjusted R-squared:  0.105
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 183 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.487  0.869  0.952  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      4.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.752 1      1.324
## Year              1.752 16      1.018

```

Residuals from last author



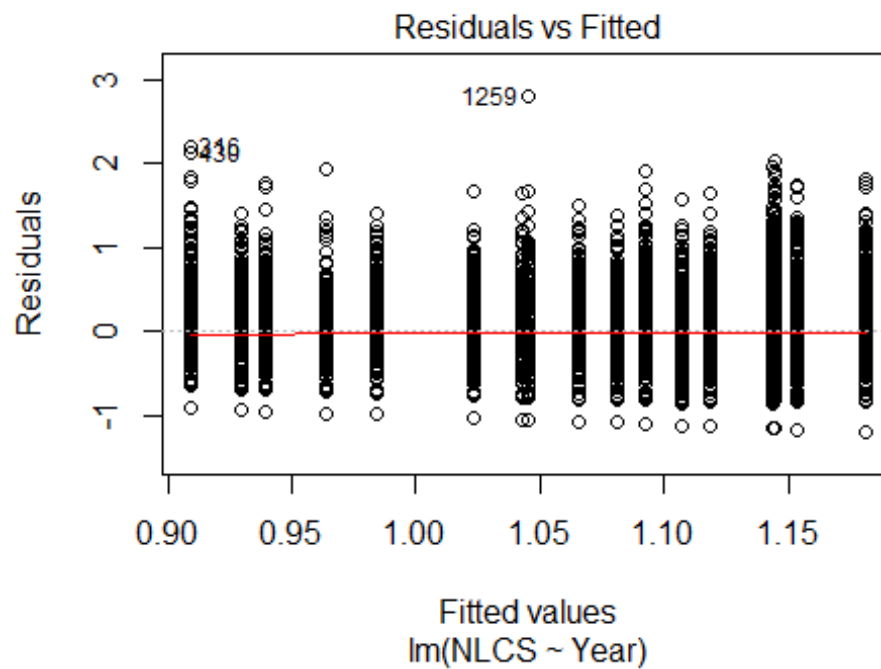
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5181 -0.3997 0.0275 0.4067 1.2527
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5193 0.3074 1.69 0.0928 .
## LastAuthorFemale1 -0.0687 0.1190 -0.58 0.5644
## Year1997 -0.0582 0.3658 -0.16 0.8738
## Year1998 0.2180 0.3601 0.61 0.5457
## Year1999 0.5486 0.3596 1.53 0.1288
## Year2000 0.5596 0.3475 1.61 0.1090
## Year2001 0.6447 0.3074 2.10 0.0373 *
## Year2002 0.1770 0.3869 0.46 0.6479
## Year2003 0.8147 0.8121 1.00 0.3171
## Year2004 0.7891 0.4059 1.94 0.0534 .
## Year2005 0.9988 0.3299 3.03 0.0028 **
## Year2006 0.8026 0.3218 2.49 0.0135 *
```

```

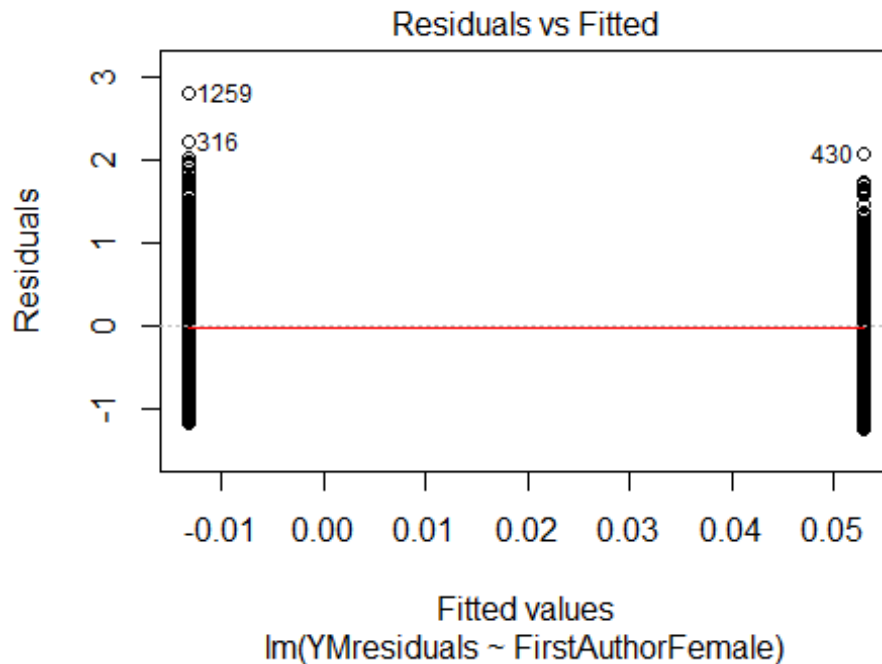
## Year2007          0.7261      0.3627      2.00      0.0467 *
## Year2008          0.2238      0.3548      0.63      0.5289
## Year2009          0.7544      0.3448      2.19      0.0299 *
## Year2010          0.2927      0.3208      0.91      0.3628
## Year2011          0.6396      0.3350      1.91      0.0577 .
## Year2012          0.6036      0.3469      1.74      0.0835 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.578
## Multiple R-squared:  0.174, Adjusted R-squared:  0.0995
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 184 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.471  0.864  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      4.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 206"
## [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
##  362  306  299  247  304  216  292  277  307  298  366  398  478  457  486
## 2011 2012
##  439  467
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  285  226  220  182  223  154  228  217  241  210  295  309  374  360  353
## 2011 2012

```

```
## 325 339
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 265 204 207 164 208 142 207 203 226 194 269 289 346 324 306
## 2011 2012
## 279 300
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 91, df = 16, p-value = 2e-12
```

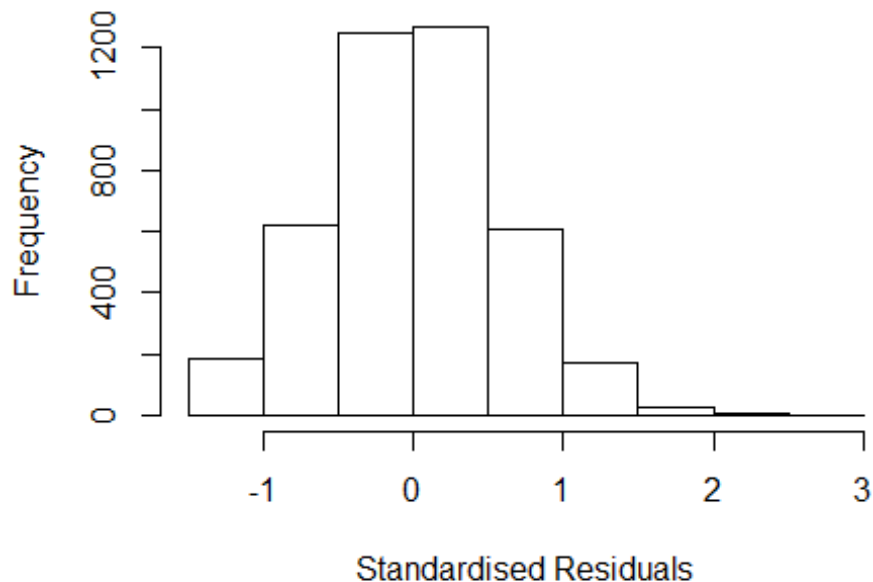


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.24, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 1.80589921424843"
## [1] "Male first author team size 2018 geometric mean: 1.55685916493143"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10000, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.75928927915549"
## [1] "Male last author team size 2018 geometric mean: 1.57353526680879"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9100, 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 1.636 1      1.279
## LastAuthorFemale  1.594 1      1.262
## UniqueAuthors    1.083 4      1.010
## Year              1.130 16     1.004
```

Residuals from first and last author and team size



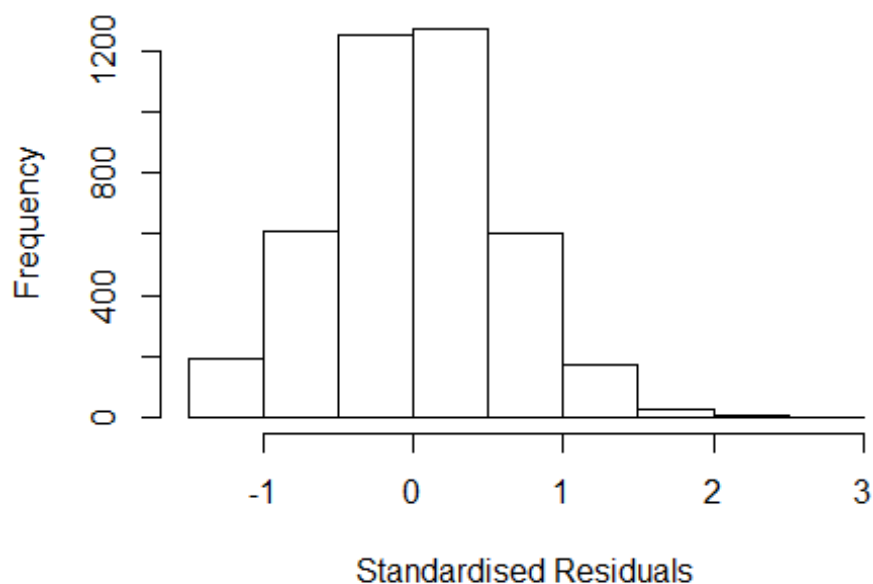
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 1259 0032702341 3.84 1999      2002      2      2.906
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26860 -0.38633  0.00506  0.39560  2.90552
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7803     0.0388   20.11 < 2e-16 ***
## FirstAuthorFemale1 0.0633     0.0298    2.12  0.03382 *
## LastAuthorFemale1 0.0124     0.0296    0.42  0.67642
## UniqueAuthors2    0.1469     0.0210    7.01  2.8e-12 ***
## UniqueAuthors3    0.1807     0.0330    5.47  4.7e-08 ***
## UniqueAuthors4    0.1506     0.0591    2.55  0.01091 *
## UniqueAuthors5    0.0899     0.0875    1.03  0.30385
## Year1997          0.0612     0.0549    1.11  0.26533
## Year1998          0.1453     0.0563    2.58  0.00986 **
## Year1999          0.1542     0.0590    2.61  0.00905 **
```

```

## Year2000          0.0784      0.0554      1.42  0.15696
## Year2001          0.0821      0.0615      1.34  0.18169
## Year2002          0.1993      0.0529      3.76  0.00017 ***
## Year2003          0.1126      0.0541      2.08  0.03753 *
## Year2004          0.2352      0.0528      4.46  8.5e-06 ***
## Year2005          0.1606      0.0545      2.95  0.00325 **
## Year2006          0.2091      0.0497      4.21  2.6e-05 ***
## Year2007          0.2469      0.0501      4.93  8.7e-07 ***
## Year2008          0.2416      0.0501      4.83  1.4e-06 ***
## Year2009          0.2102      0.0515      4.08  4.6e-05 ***
## Year2010          0.2657      0.0530      5.01  5.6e-07 ***
## Year2011          0.2325      0.0557      4.17  3.1e-05 ***
## Year2012          0.2395      0.0599      4.00  6.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.576
## Multiple R-squared:  0.0402, Adjusted R-squared:  0.035
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 701 is an outlier with |weight| = 0 ( < 2.4e-05);
## 364 weights are ~= 1. The remaining 3768 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0641 0.8700 0.9490 0.9060 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.42e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev mts compute.rd
##      0      1000      0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.632 1      1.278
## LastAuthorFemale 1.593 1      1.262
## Year 1.055 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
## 1259 0032702341 3.84 1999      2002      2      2.857
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.20305 -0.39287  0.00467  0.39348  2.85652
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8314    0.0381   21.81 < 2e-16 ***
## FirstAuthorFemale1 0.0713    0.0300    2.38  0.01752 *
## LastAuthorFemale1 0.0124    0.0299    0.41  0.67853
## Year1997        0.0682    0.0548    1.24  0.21383
## Year1998        0.1578    0.0561    2.81  0.00490 **
## Year1999        0.1521    0.0593    2.57  0.01034 *
## Year2000        0.0900    0.0558    1.61  0.10673
## Year2001        0.0892    0.0621    1.44  0.15123
## Year2002        0.2106    0.0528    3.99  6.8e-05 ***
## Year2003        0.1216    0.0539    2.26  0.02402 *
## Year2004        0.2505    0.0532    4.71  2.5e-06 ***
## Year2005        0.1845    0.0543    3.40  0.00069 ***
```

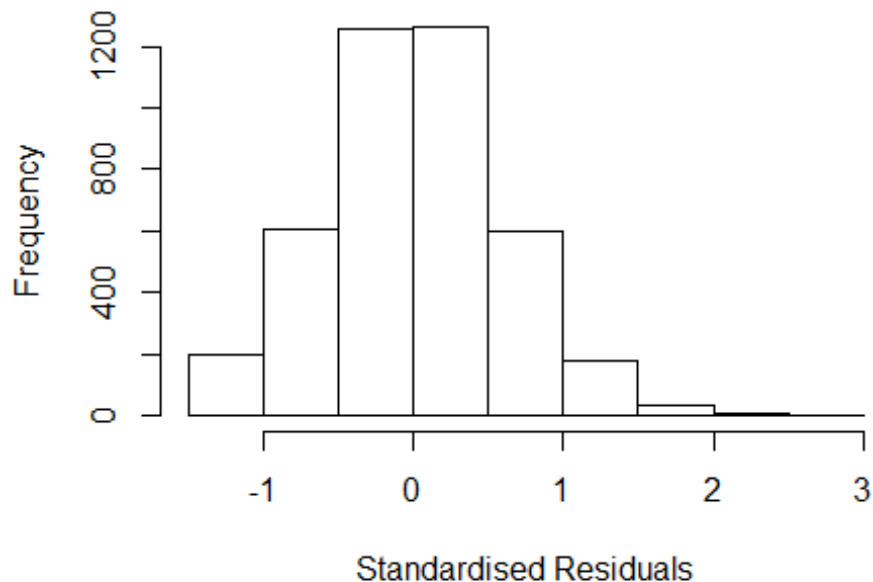


```

## Year2006          0.2229      0.0498      4.47  7.9e-06 ***
## Year2007          0.2582      0.0497      5.19  2.2e-07 ***
## Year2008          0.2537      0.0503      5.04  4.7e-07 ***
## Year2009          0.2301      0.0514      4.48  7.7e-06 ***
## Year2010          0.2880      0.0528      5.45  5.2e-08 ***
## Year2011          0.2524      0.0560      4.50  6.8e-06 ***
## Year2012          0.2611      0.0606      4.30  1.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.579
## Multiple R-squared:  0.0233, Adjusted R-squared:  0.019
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 701 is an outlier with |weight| = 0 ( < 2.4e-05);
## 401 weights are ~ = 1. The remaining 3731 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0862 0.8630 0.9470 0.9050 0.9850 0.9990
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          2.42e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev    mts    compute.rd
##           0         1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1          1.021
## Year              1.042 16          1.001

```

Residuals from first author



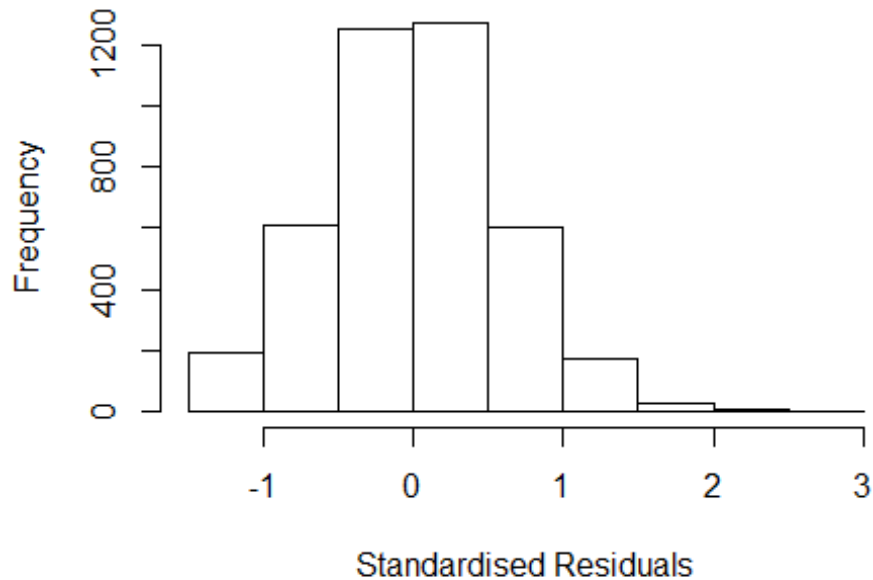
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 1259 0032702341 3.84 1999      2002      2      2.857
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19883 -0.39065  0.00398  0.39231  2.85501
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8323    0.0381   21.85 < 2e-16 ***
## FirstAuthorFemale1 0.0788    0.0240    3.29  0.00102 **
## Year1997        0.0682    0.0548    1.24  0.21402
## Year1998        0.1576    0.0560    2.81  0.00495 **
## Year1999        0.1527    0.0592    2.58  0.00998 **
## Year2000        0.0897    0.0558    1.61  0.10769
## Year2001        0.0892    0.0622    1.43  0.15150
## Year2002        0.2106    0.0528    3.99  6.8e-05 ***
## Year2003        0.1217    0.0539    2.26  0.02401 *
## Year2004        0.2502    0.0532    4.71  2.6e-06 ***
## Year2005        0.1846    0.0544    3.40  0.00069 ***
## Year2006        0.2224    0.0498    4.46  8.3e-06 ***
```

```

## Year2007          0.2584      0.0497      5.20  2.1e-07 ***
## Year2008          0.2536      0.0503      5.04  4.8e-07 ***
## Year2009          0.2304      0.0514      4.49  7.4e-06 ***
## Year2010          0.2877      0.0528      5.45  5.4e-08 ***
## Year2011          0.2523      0.0560      4.50  6.8e-06 ***
## Year2012          0.2609      0.0607      4.30  1.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.579
## Multiple R-squared:  0.0232, Adjusted R-squared:  0.0192
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 701 is an outlier with |weight| = 0 ( < 2.4e-05);
## 400 weights are ~= 1. The remaining 3732 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0864 0.8620 0.9480 0.9050 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.016 1      1.008
## Year      1.016 16      1.000

```

Residuals from last author



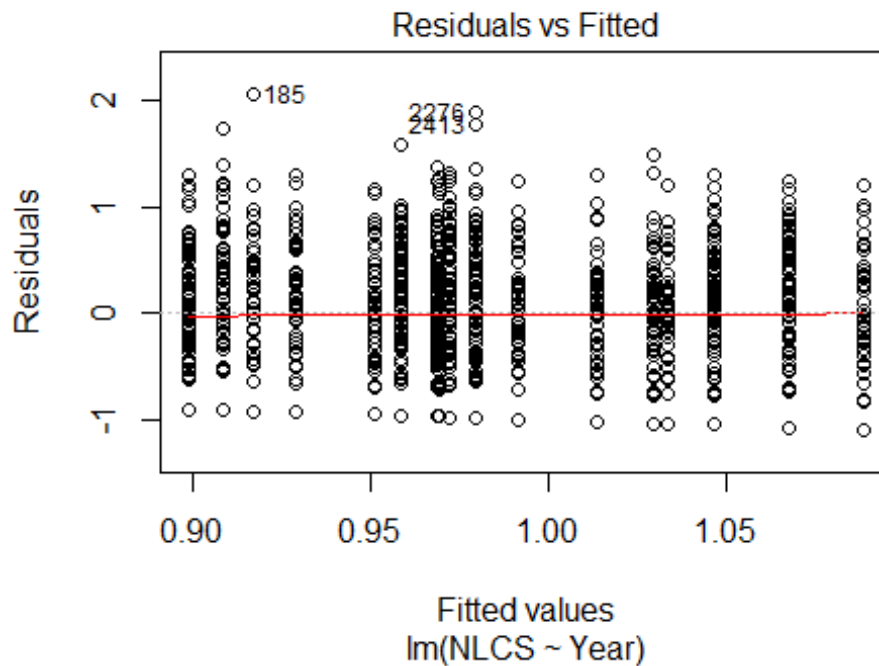
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 1259 0032702341 3.84 1999      2002      2      2.857
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.18473 -0.39191  0.00346  0.39412  2.85288
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8342    0.0381   21.90 < 2e-16 ***
## LastAuthorFemale1 0.0560    0.0238    2.36  0.01856 *
## Year1997        0.0693    0.0548    1.26  0.20608
## Year1998        0.1593    0.0561    2.84  0.00454 **
## Year1999        0.1530    0.0595    2.57  0.01013 *
## Year2000        0.0897    0.0558    1.61  0.10817
## Year2001        0.0902    0.0619    1.46  0.14521
## Year2002        0.2095    0.0528    3.96  7.5e-05 ***
## Year2003        0.1247    0.0538    2.32  0.02047 *
## Year2004        0.2531    0.0532    4.75  2.1e-06 ***
## Year2005        0.1863    0.0541    3.44  0.00058 ***
## Year2006        0.2298    0.0497    4.62  4.0e-06 ***
```

```

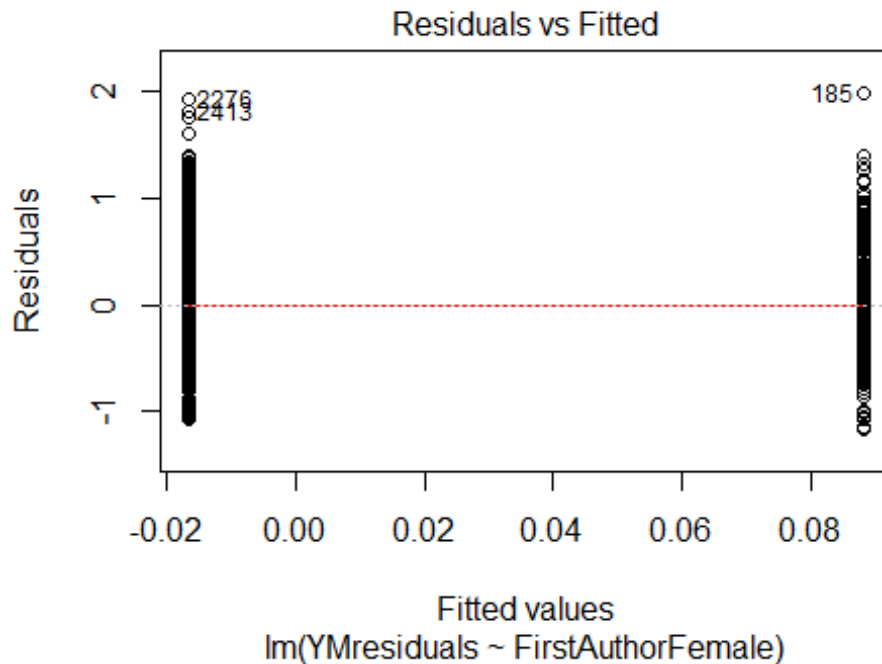
## Year2007          0.2594      0.0497      5.22  1.9e-07 ***
## Year2008          0.2611      0.0501      5.21  2.0e-07 ***
## Year2009          0.2307      0.0514      4.49  7.2e-06 ***
## Year2010          0.2945      0.0528      5.57  2.6e-08 ***
## Year2011          0.2595      0.0561      4.63  3.8e-06 ***
## Year2012          0.2674      0.0608      4.40  1.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.58
## Multiple R-squared:  0.0218, Adjusted R-squared:  0.0178
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 701 is an outlier with |weight| = 0 ( < 2.4e-05);
## 392 weights are ~= 1. The remaining 3740 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.088  0.863  0.948  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4133"
## [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
##   82   74   78   73   66   76   69   72   97   93  124  142  149  153  146
## 2011 2012
##  151  148
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   67   59   57   50   45   52   55   59   81   63   98  103  117  117   94

```

```
## 2011 2012
## 117 99
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 62 55 53 46 43 48 50 55 72 56 88 97 108 101 75
## 2011 2012
## 99 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 = 12, df = 16, p-value = 0.8
```

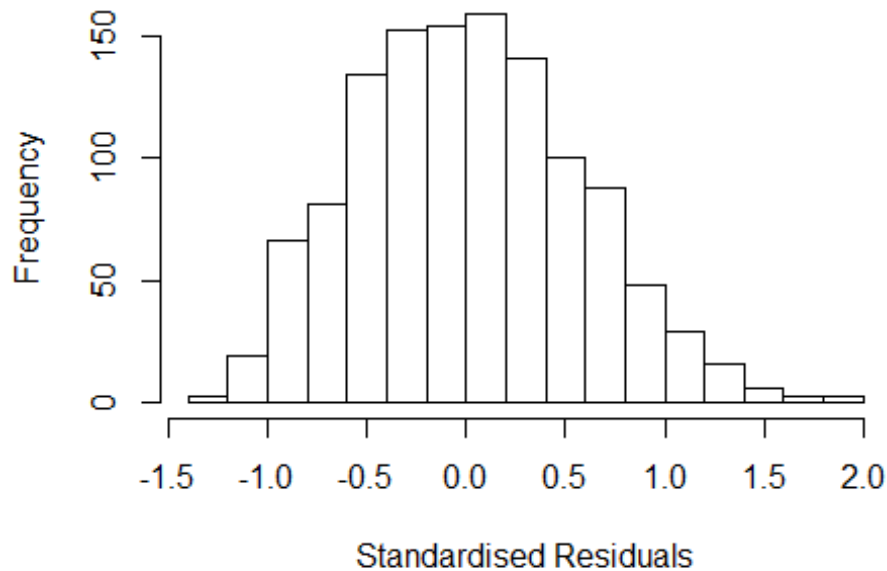


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.044, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 1.8881750225898"
## [1] "Male first author team size 2018 geometric mean: 1.51474728076334"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.82626029960213"
## [1] "Male last author team size 2018 geometric mean: 1.53889494803019"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.458 1          1.208
## LastAuthorFemale  1.422 1          1.192
## UniqueAuthors     1.247 4          1.028
## 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
## -1.24305 -0.40108 -0.00613 0.39034 1.85876
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8123 0.0805 10.09 < 2e-16 ***
## FirstAuthorFemale1 0.1690 0.0529 3.19 0.0014 **
## LastAuthorFemale1 -0.1308 0.0541 -2.42 0.0158 *
## UniqueAuthors2 0.1922 0.0377 5.09 4.1e-07 ***
## UniqueAuthors3 0.2301 0.0576 3.99 7.0e-05 ***
## UniqueAuthors4 0.0729 0.1215 0.60 0.5487
## UniqueAuthors5 0.2208 0.1885 1.17 0.2418
## Year1997 -0.0359 0.1174 -0.31 0.7601
## Year1998 0.0899 0.1098 0.82 0.4131
## Year1999 0.1286 0.1145 1.12 0.2618
```

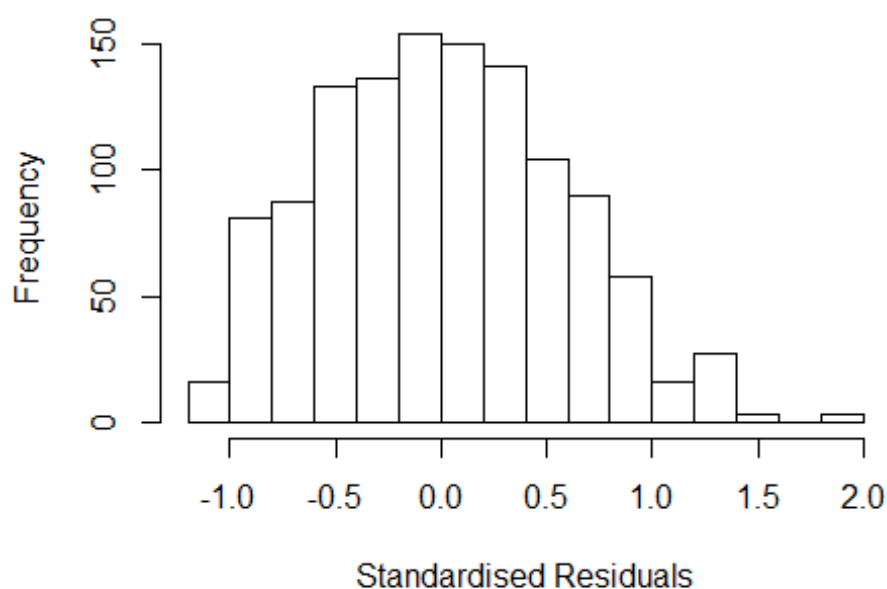


```

## Year2000          0.1406      0.1093      1.29      0.1986
## Year2001          0.0759      0.1111      0.68      0.4943
## Year2002          0.0204      0.1084      0.19      0.8505
## Year2003          0.0580      0.1092      0.53      0.5953
## Year2004          0.1190      0.1061      1.12      0.2622
## Year2005          0.0664      0.1015      0.65      0.5130
## Year2006          0.0574      0.0988      0.58      0.5617
## Year2007         -0.0402      0.0964     -0.42      0.6764
## Year2008          0.0501      0.0979      0.51      0.6088
## Year2009          0.0695      0.0982      0.71      0.4790
## Year2010          0.0925      0.1107      0.84      0.4036
## Year2011          0.0117      0.1019      0.11      0.9086
## Year2012         -0.0481      0.1016     -0.47      0.6356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.581
## Multiple R-squared:  0.0474, Adjusted R-squared:  0.0295
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 98 weights are ~= 1. The remaining 1101 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.284  0.873   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      8.34e-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 1.467 1      1.211
## LastAuthorFemale  1.427 1      1.195
## Year              1.097 16      1.003

```

Residuals from first and last author



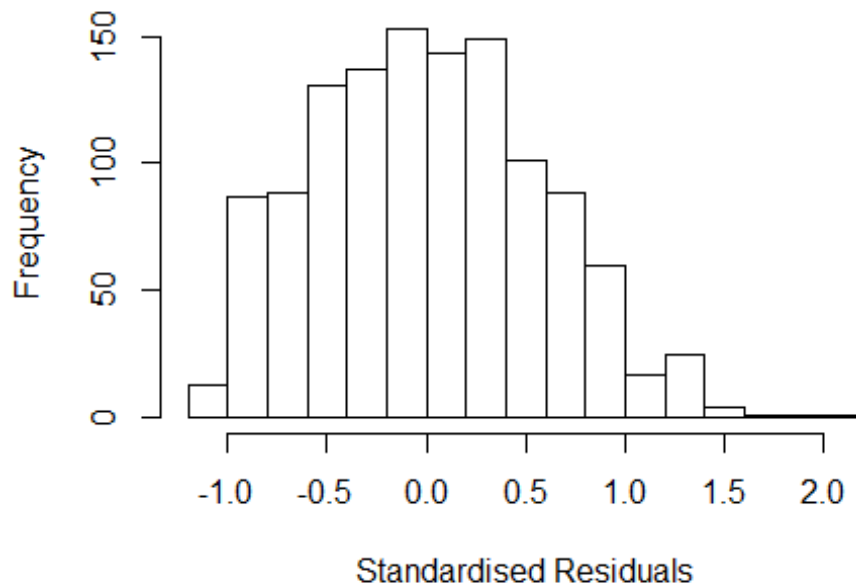
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17038 -0.42545 -0.00912 0.40657 1.94327
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8985 0.0784 11.45 < 2e-16 ***
## FirstAuthorFemale1 0.1785 0.0539 3.31 0.00096 ***
## LastAuthorFemale1 -0.1253 0.0551 -2.27 0.02316 *
## Year1997 -0.0235 0.1175 -0.20 0.84122
## Year1998 0.0737 0.1103 0.67 0.50400
## Year1999 0.1247 0.1177 1.06 0.28994
## Year2000 0.1489 0.1103 1.35 0.17702
## Year2001 0.0970 0.1126 0.86 0.38915
## Year2002 0.0295 0.1105 0.27 0.78956
## Year2003 0.0369 0.1090 0.34 0.73473
## Year2004 0.1053 0.1098 0.96 0.33755
## Year2005 0.0963 0.1032 0.93 0.35073
```

```

## Year2006          0.0614      0.0989      0.62  0.53495
## Year2007         -0.0291      0.0964     -0.30  0.76296
## Year2008          0.0387      0.0981      0.39  0.69344
## Year2009          0.0934      0.0974      0.96  0.33737
## Year2010          0.1178      0.1108      1.06  0.28809
## Year2011          0.0333      0.1015      0.33  0.74325
## Year2012         -0.0407      0.1024     -0.40  0.69098
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.018, Adjusted R-squared:  0.00303
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 100 weights are ~= 1. The remaining 1099 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.283  0.885   0.951   0.918   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.34e-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 1.065 1      1.032
## Year              1.065 16      1.002

```

Residuals from first author



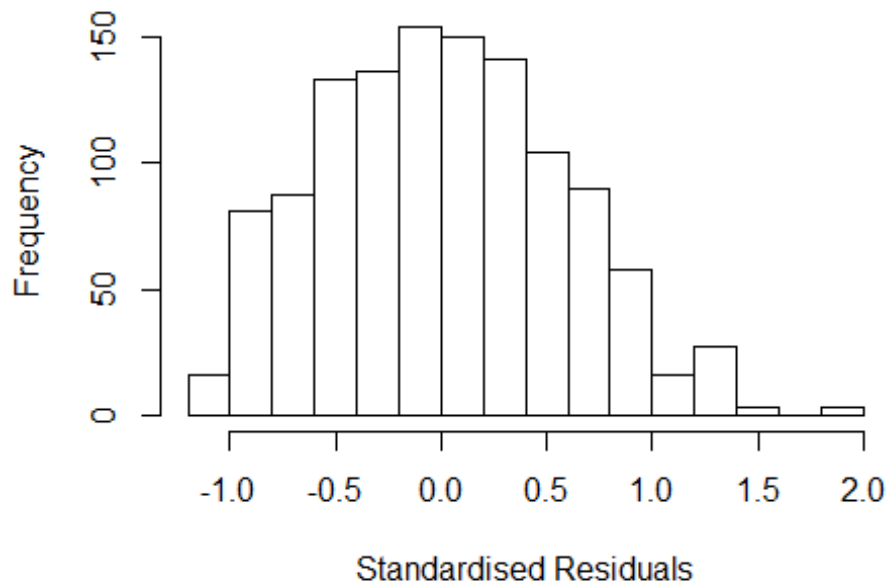
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1216 -0.4189 -0.0119  0.3959  2.0013
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8864    0.0779   11.38  <2e-16 ***
## FirstAuthorFemale1 0.1124    0.0472    2.38   0.018 *
## Year1997      -0.0221    0.1169   -0.19   0.850
## Year1998       0.0784    0.1096    0.72   0.474
## Year1999       0.1321    0.1178    1.12   0.263
## Year2000       0.1484    0.1102    1.35   0.178
## Year2001       0.0999    0.1130    0.88   0.377
## Year2002       0.0415    0.1103    0.38   0.707
## Year2003       0.0485    0.1091    0.44   0.657
## Year2004       0.1094    0.1098    1.00   0.319
## Year2005       0.1048    0.1027    1.02   0.307
## Year2006       0.0623    0.0989    0.63   0.529
```

```

## Year2007          -0.0337      0.0959   -0.35    0.725
## Year2008           0.0395      0.0979    0.40    0.687
## Year2009           0.0911      0.0972    0.94    0.349
## Year2010           0.1227      0.1103    1.11    0.266
## Year2011           0.0386      0.1018    0.38    0.704
## Year2012          -0.0348      0.1015   -0.34    0.732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.61
## Multiple R-squared:  0.0136, Adjusted R-squared:  -0.000593
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 90 weights are ~= 1. The remaining 1109 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.259  0.886  0.952  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      8.34e-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 1.037 1      1.018
## Year             1.037 16      1.001

```

Residuals from last author



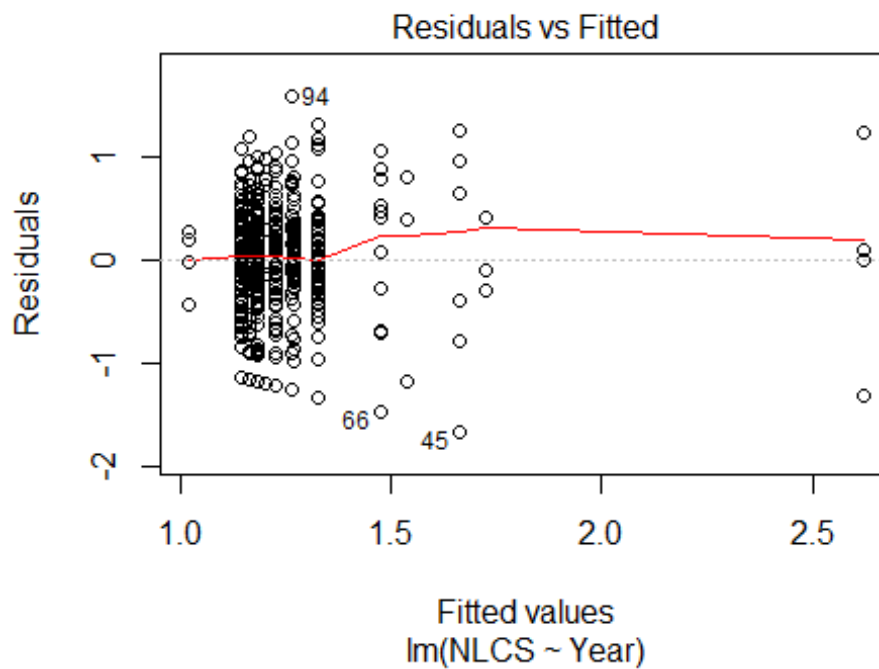
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.04461 -0.42534 -0.00145 0.41137 2.09786
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9036 0.0785 11.51 <2e-16 ***
## LastAuthorFemale1 -0.0290 0.0462 -0.63 0.53
## Year1997 -0.0234 0.1170 -0.20 0.84
## Year1998 0.0728 0.1102 0.66 0.51
## Year1999 0.1410 0.1179 1.20 0.23
## Year2000 0.1405 0.1097 1.28 0.20
## Year2001 0.1131 0.1126 1.00 0.32
## Year2002 0.0381 0.1119 0.34 0.73
## Year2003 0.0494 0.1098 0.45 0.65
## Year2004 0.1130 0.1092 1.03 0.30
## Year2005 0.1119 0.1032 1.08 0.28
## Year2006 0.0792 0.0989 0.80 0.42
```

```

## Year2007          -0.0240      0.0959    -0.25      0.80
## Year2008           0.0460      0.0977      0.47      0.64
## Year2009           0.0940      0.0985      0.95      0.34
## Year2010           0.1322      0.1118      1.18      0.24
## Year2011           0.0450      0.1027      0.44      0.66
## Year2012          -0.0238      0.1027     -0.23      0.82
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.611
## Multiple R-squared:  0.00913,    Adjusted R-squared:  -0.00513
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 1095 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.214  0.885  0.950  0.918  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.34e-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: 1199"
## [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
##    4    4   13    9    9   12   17   10   29   17   18   96  101  118  130
## 2011 2012
##   132  134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    0    2    3    4    3    6   13    4   20    9   13   54   67   75   82
## 2011 2012

```

```
##      87      74
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      0      2      2      3      3      4     13      4     20      6     11     46     50     61     71
## 2011 2012
##      73     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 = 68, df = 15, p-value = 1e-08
```



```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8, df = 1, p-value = 0.005
## [1] "Female first author team size 2018 geometric mean: 3.78349943554901"
## [1] "Male first author team size 2018 geometric mean: 2.49229070211985"
##
## Wilcoxon rank sum test with continuity correction
##
## data:  FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 600, p-value = 0.02
```



```
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.67652670306138"
## [1] "Male last author team size 2018 geometric mean: 2.55313681238118"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 470, 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"

## 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"

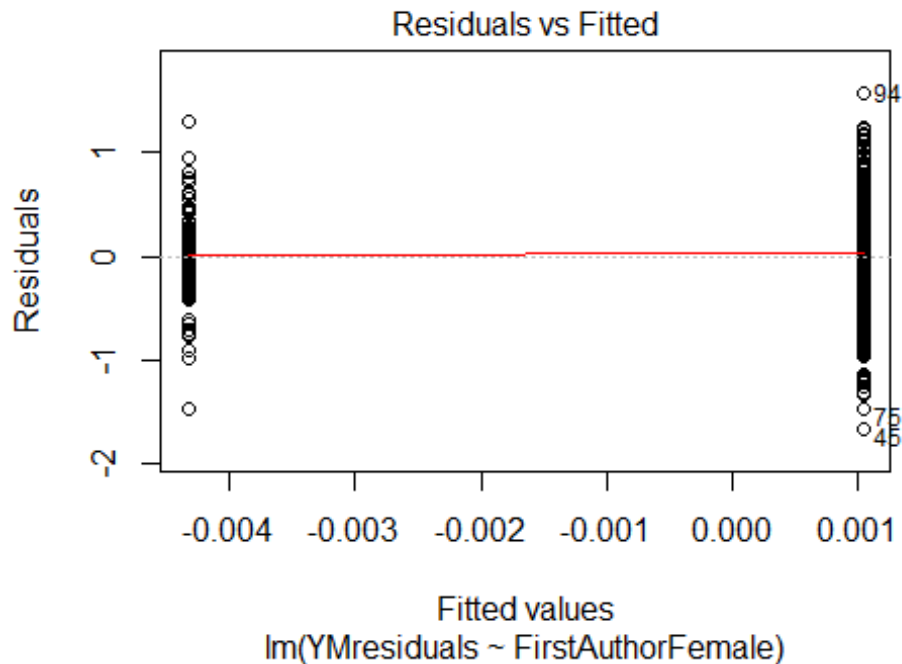
## Warning in lmrob.fit(x, y, control, init = init, mf = mf): M-step did NOT
## converge. Returning unconverged SM-estimate

## [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 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: 429"
## [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 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    5    4    5    1    5    3    8    4   13   16   16   22    5    4    5
## 2012
##    5
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    3    3    0    1    1    2    4    2    7    3    8    9    4    2    3
## 2012
##    4
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    3    3    0    1    1    2    4    2    6    2    5    9    4    2    3
## 2012
##    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: 1.88734887819666"
```

```
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale -2.098e+14  1      NaN
## LastAuthorFemale  -1.415e+14  1      NaN
## UniqueAuthors    -8.812e+15  4      NaN
## Year              -5.253e+16 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
## -1.7636 -0.2391  0.0471  0.2065  1.3208
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.2200     0.2569   -0.86  0.39866
## FirstAuthorFemale1  0.4667     0.2859    1.63  0.11311
## LastAuthorFemale1 -0.4246     0.2292   -1.85  0.07380 .
## UniqueAuthors2     0.7577     0.2355    3.22  0.00309 **
## UniqueAuthors3     0.8517     0.3508    2.43  0.02140 *
## UniqueAuthors4     0.7187     0.1866    3.85  0.00057 ***
## UniqueAuthors5     0.9584     0.1521    6.30  6.0e-07 ***
## Year1998           0.1112     0.2745    0.41  0.68829
## Year2000            0.2200     0.2569    0.86  0.39866
## Year2001            2.3960     0.2569    9.33  2.3e-10 ***
## Year2002           -0.4286     0.3046   -1.41  0.16977
## Year2003            0.1595     0.4183    0.38  0.70567
## Year2004            1.0288     0.2950    3.49  0.00153 **
## Year2005            1.0772     0.3288    3.28  0.00266 **
## Year2006            0.4738     0.3751    1.26  0.21632
## Year2007            1.4515     0.3084    4.71  5.3e-05 ***
## Year2008            0.7043     0.2945    2.39  0.02324 *
## Year2009            0.4769     0.2754    1.73  0.09365 .
## Year2010            0.0673     0.9994    0.07  0.94679
## Year2011            0.8562     0.3142    2.73  0.01062 *
## Year2012            0.5585     0.2807    1.99  0.05585 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.342
```

```

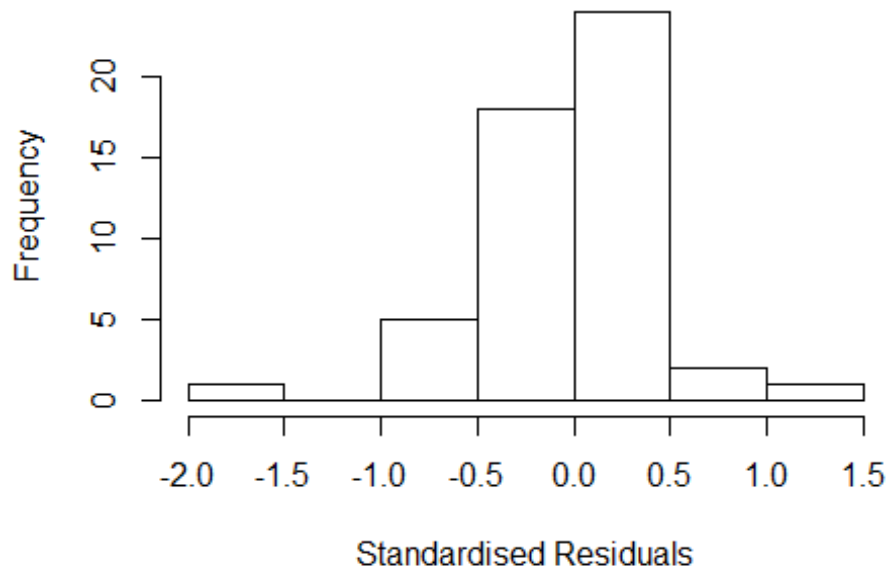
## Multiple R-squared:  0.836, Adjusted R-squared:  0.726
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## observation 25 is an outlier with |weight| = 0 ( < 0.002);
## 6 weights are ~= 1. The remaining 44 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.103  0.895   0.960   0.902   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           1.96e-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 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 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
## -1.29e+00 -2.74e-01 -6.48e-17  3.78e-01  2.00e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.48e-17  4.36e-09   0.00  1.00000
## FirstAuthorFemale1 6.60e-01  3.46e-01   1.91  0.06514 .
## LastAuthorFemale1  2.10e-01  1.55e-01   1.35  0.18452
## Year1998        -6.97e-02  7.13e-02  -0.98  0.33513
## Year2000        -1.41e-17  0.00e+00  -Inf < 2e-16 ***
## Year2001         2.18e+00  0.00e+00   Inf < 2e-16 ***
## Year2002         1.23e-02  1.73e-01   0.07  0.94373
```

```

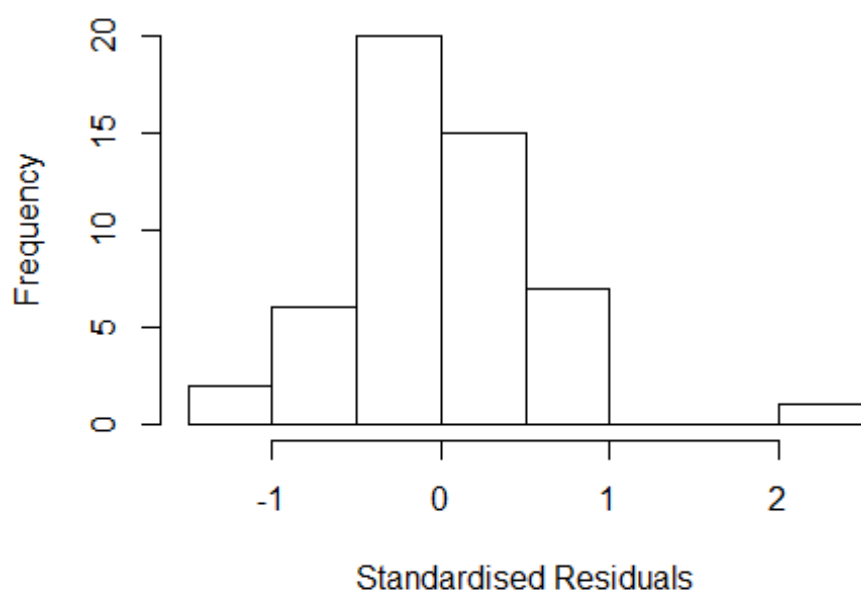
## Year2003          1.34e-02    2.99e-01    0.04  0.96440
## Year2004          1.09e+00    3.75e-01    2.91  0.00630 **
## Year2005          1.48e+00    3.41e-01    4.34  0.00012 ***
## Year2006          5.83e-01    5.78e-01    1.01  0.32065
## Year2007          1.52e+00    4.39e-01    3.46  0.00146 **
## Year2008          8.70e-01    2.01e-01    4.32  0.00013 ***
## Year2009          6.28e-01    2.88e-01    2.19  0.03586 *
## Year2010          5.86e-01    5.41e-01    1.08  0.28664
## Year2011          1.17e+00    2.15e-01    5.47  4.3e-06 ***
## Year2012          1.01e+00    2.25e-01    4.50  7.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.631, Adjusted R-squared:  0.457
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~ = 1. The remaining 39 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.123  0.892  0.941  0.895  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      1.96e-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 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
## -1.40e+00 -2.86e-01 -2.77e-17  3.54e-01  1.96e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.75e-16   0.00e+00  -Inf < 2e-16 ***
## FirstAuthorFemale1  6.76e-01   4.15e-01   1.63  0.11207
## Year1998        4.03e-16   0.00e+00    Inf < 2e-16 ***
## Year2000        4.15e-16   0.00e+00    Inf < 2e-16 ***
## Year2001        2.18e+00   0.00e+00    Inf < 2e-16 ***
## Year2002        4.31e-03   2.08e-01   0.02  0.98355
## Year2003        5.57e-02   2.41e-01   0.23  0.81860
## Year2004        1.08e+00   4.26e-01   2.54  0.01552 *
```

```

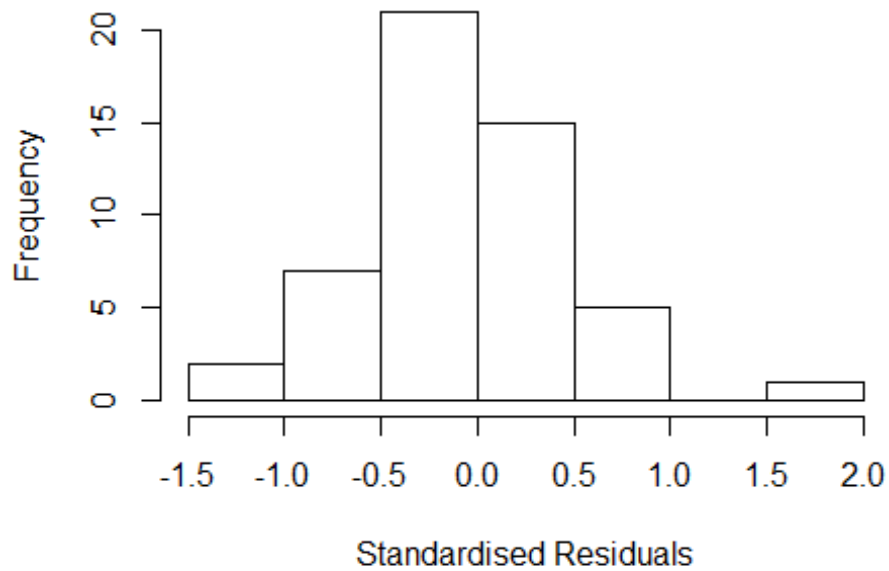
## Year2005          1.49e+00  3.53e-01  4.22  0.00016 ***
## Year2006          5.75e-01  6.11e-01  0.94  0.35298
## Year2007          1.62e+00  5.14e-01  3.15  0.00337 **
## Year2008          8.89e-01  2.22e-01  4.01  0.00031 ***
## Year2009          6.29e-01  2.95e-01  2.13  0.04011 *
## Year2010          5.85e-01  5.67e-01  1.03  0.30849
## Year2011          1.17e+00  2.17e-01  5.41  4.6e-06 ***
## Year2012          1.01e+00  2.28e-01  4.45  8.3e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.646, Adjusted R-squared:  0.494
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 39 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.085  0.878  0.935  0.882  0.968  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          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"

## 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

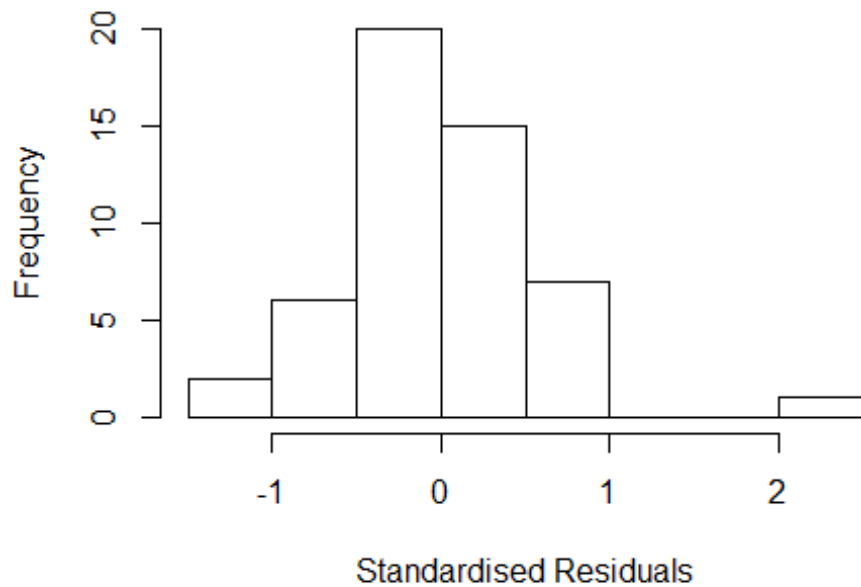
```


Residuals from first author



##	GVIF	Df	$GVIF^{1/(2 \cdot Df)}$
## LastAuthorFemale	NaN	1	NaN
## Year	NaN	14	NaN

Residuals from last author

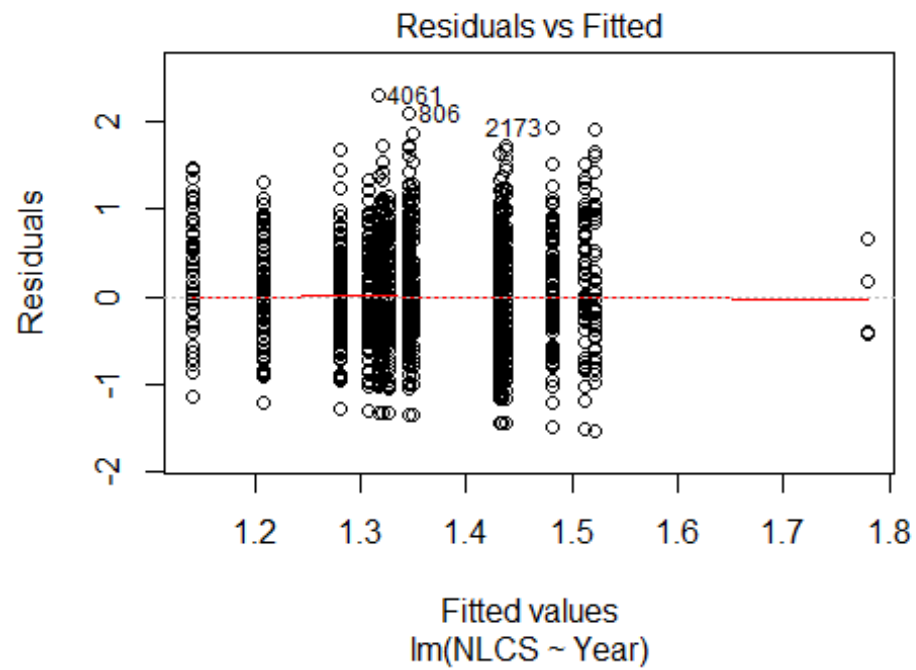


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24e+00 -3.33e-01 -4.22e-16 3.52e-01 1.82e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.22e-16 1.31e-08 0.00e+00 1.00000
## LastAuthorFemale1 9.25e-02 1.76e-01 5.20e-01 0.60340
## Year1998 -3.08e-02 6.27e-02 -4.90e-01 0.62625
## Year2000 1.40e-16 0.00e+00 Inf < 2e-16 ***
## Year2001 2.18e+00 7.96e-09 2.73e+08 < 2e-16 ***
## Year2002 3.42e-01 2.57e-01 1.34e+00 0.19044
## Year2003 1.99e-01 4.53e-01 4.40e-01 0.66333
## Year2004 1.42e+00 7.60e-01 1.87e+00 0.06989 .
## Year2005 1.57e+00 3.72e-01 4.23e+00 0.00016 ***
## Year2006 9.13e-01 1.14e+00 8.00e-01 0.42895
## Year2007 1.58e+00 2.61e-01 6.06e+00 6.4e-07 ***
## Year2008 1.06e+00 1.55e-01 6.80e+00 6.8e-08 ***
## Year2009 6.27e-01 2.74e-01 2.29e+00 0.02824 *
## Year2010 5.85e-01 4.97e-01 1.18e+00 0.24718
## Year2011 1.17e+00 2.10e-01 5.57e+00 2.9e-06 ***
## Year2012 1.01e+00 2.20e-01 4.59e+00 5.5e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.623
## Multiple R-squared: 0.524, Adjusted R-squared: 0.32
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 41 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.374 0.892 0.952 0.918 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 1.96e-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
```

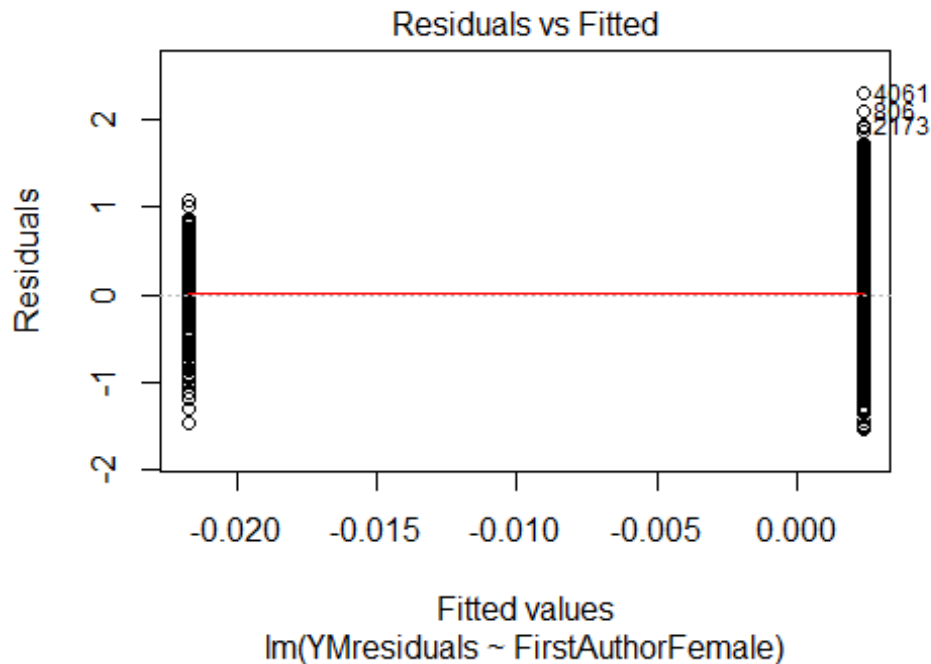
```

##           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
## 200 213 190 21 185 171 151 173 211 228 298 276 328 368 373
## 2011 2012
## 409 324
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 76 80 57 4 59 45 59 84 88 89 95 104 132 139 170
## 2011 2012
## 191 144
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 64 53 3 46 32 50 62 72 69 69 75 93 102 133
## 2011 2012
## 147 113
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 69, df = 16, p-value = 1e-08

```

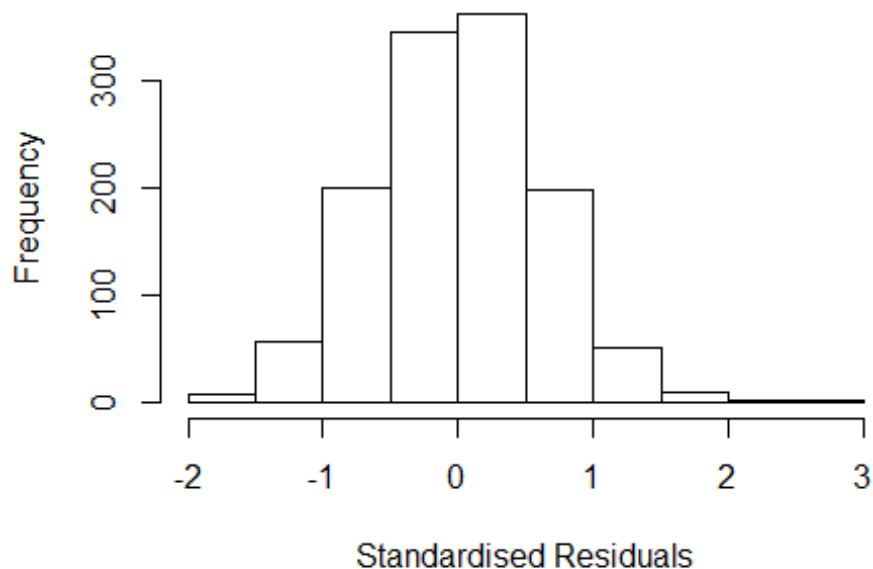


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 10, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 3.23874739820928"
## [1] "Male first author team size 2018 geometric mean: 2.74512738636787"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2500, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.32908790988426"
## [1] "Male last author team size 2018 geometric mean: 2.76977748261053"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, 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 1.171 1      1.082
## LastAuthorFemale  1.203 1      1.097
## UniqueAuthors    1.425 4      1.045
## Year              1.455 16     1.012
```

Residuals from first and last author and team size



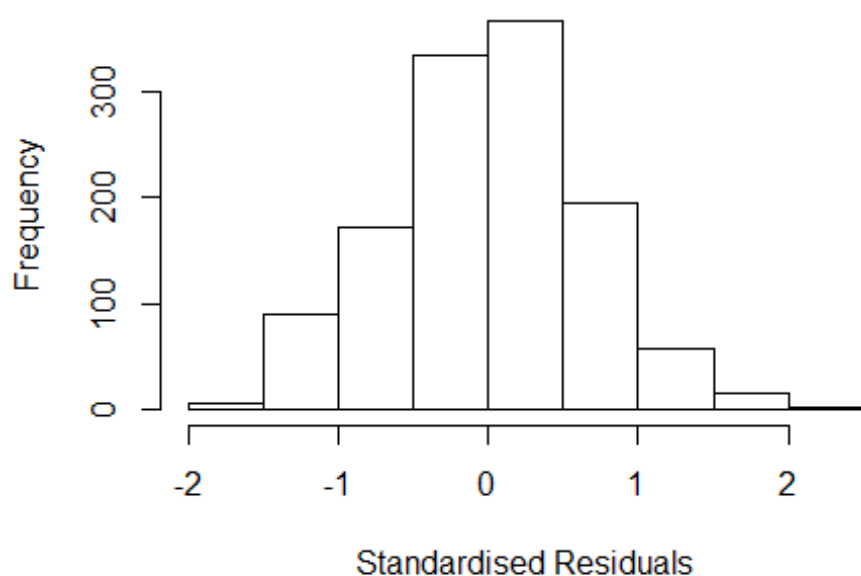
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 806    0033888092 3.438 2000    2102    4    2.509
## 4061 73649137161 3.611 2010    2102    2    2.771
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6808 -0.4140  0.0112  0.4264  2.7705
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9728    0.1029   9.45 < 2e-16 ***
## FirstAuthorFemale1 -0.0493    0.0559  -0.88    0.38
## LastAuthorFemale1 -0.0149    0.0781  -0.19    0.85
## UniqueAuthors2    0.4151    0.0633   6.56 8.0e-11 ***
## UniqueAuthors3    0.5538    0.0663   8.36 < 2e-16 ***
## UniqueAuthors4    0.4399    0.0716   6.14 1.1e-09 ***
## UniqueAuthors5    0.5206    0.0744   6.99 4.4e-12 ***
## Year1997         0.0239    0.1323   0.18    0.86
## Year1998        -0.1586    0.1637  -0.97    0.33
```

```

## Year1999      0.2749      0.2075      1.32      0.19
## Year2000     -0.0441      0.1875     -0.24      0.81
## Year2001      0.1932      0.1876      1.03      0.30
## Year2002      0.1542      0.1691      0.91      0.36
## Year2003     -0.1616      0.1285     -1.26      0.21
## Year2004      0.0293      0.1321      0.22      0.82
## Year2005      0.0641      0.1273      0.50      0.61
## Year2006      0.0328      0.1216      0.27      0.79
## Year2007     -0.0929      0.1177     -0.79      0.43
## Year2008      0.0096      0.1060      0.09      0.93
## Year2009     -0.0227      0.1099     -0.21      0.84
## Year2010     -0.1323      0.1012     -1.31      0.19
## Year2011     -0.1052      0.1013     -1.04      0.30
## Year2012     -0.1070      0.1027     -1.04      0.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.0971, Adjusted R-squared:  0.0807
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1128 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0026 0.8590 0.9460 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      8.08e-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 1.103 1      1.050
## LastAuthorFemale  1.110 1      1.054
## Year              1.096 16      1.003

```

Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5130 -0.4275  0.0252  0.4050  2.3306
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.32212    0.09199   14.37  <2e-16 ***
## FirstAuthorFemale1 -0.00140    0.05589   -0.02    0.98
## LastAuthorFemale1 -0.00168    0.07622   -0.02    0.98
## Year1997        -0.03610    0.13214   -0.27    0.78
## Year1998        -0.22233    0.16368   -1.36    0.17
## Year1999         0.23338    0.18633    1.25    0.21
## Year2000        -0.10789    0.18299   -0.59    0.56
## Year2001         0.18747    0.18031    1.04    0.30
## Year2002         0.19086    0.17109    1.12    0.26
## Year2003        -0.13610    0.12805   -1.06    0.29
## Year2004         0.06427    0.13586    0.47    0.64
## Year2005         0.08109    0.12769    0.64    0.53
```

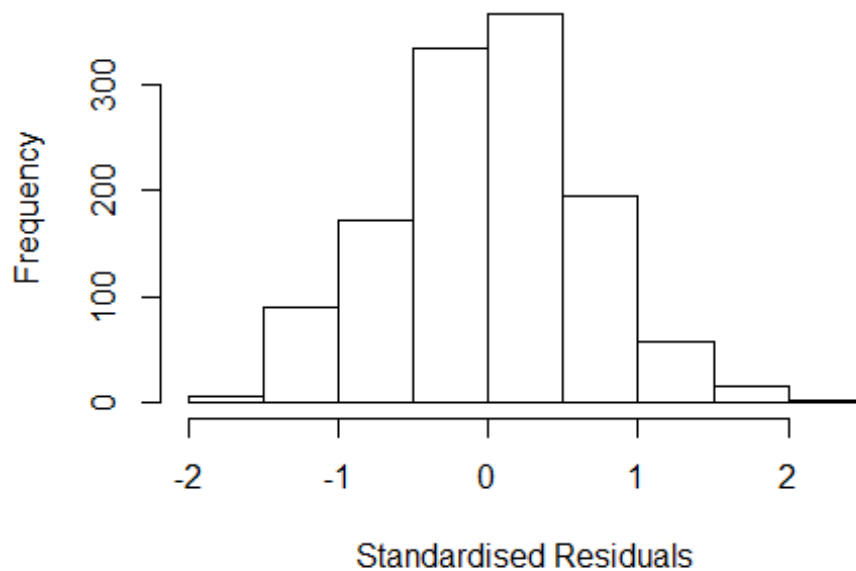


```

## Year2006          0.11838    0.12309    0.96    0.34
## Year2007          -0.06924    0.12226   -0.57    0.57
## Year2008          0.05178    0.11052    0.47    0.64
## Year2009          0.03271    0.11447    0.29    0.78
## Year2010          -0.04175    0.10139   -0.41    0.68
## Year2011          -0.01267    0.10230   -0.12    0.90
## Year2012          -0.00903    0.10249   -0.09    0.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.616
## Multiple R-squared:  0.0187, Adjusted R-squared:  0.00421
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1138 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.120  0.848  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      8.08e-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 1.04 1          1.020
## Year              1.04 16          1.001

```

Residuals from first author



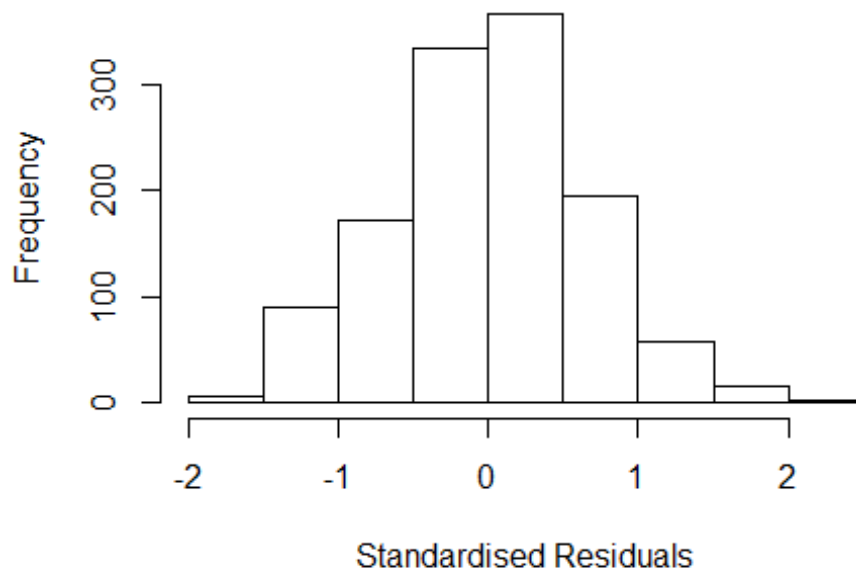
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5129 -0.4282 0.0253 0.4051 2.3308
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.32202 0.09176 14.41 <2e-16 ***
## FirstAuthorFemale1 -0.00165 0.05432 -0.03 0.98
## Year1997 -0.03610 0.13216 -0.27 0.78
## Year1998 -0.22231 0.16366 -1.36 0.17
## Year1999 0.23355 0.18613 1.25 0.21
## Year2000 -0.10778 0.18287 -0.59 0.56
## Year2001 0.18742 0.18022 1.04 0.30
## Year2002 0.19087 0.17109 1.12 0.26
## Year2003 -0.13608 0.12801 -1.06 0.29
## Year2004 0.06426 0.13586 0.47 0.64
## Year2005 0.08113 0.12763 0.64 0.53
## Year2006 0.11839 0.12312 0.96 0.34
```

```

## Year2007      -0.06913    0.12188   -0.57    0.57
## Year2008      0.05183    0.11050    0.47    0.64
## Year2009      0.03270    0.11449    0.29    0.78
## Year2010     -0.04180    0.10142   -0.41    0.68
## Year2011     -0.01274    0.10240   -0.12    0.90
## Year2012     -0.00901    0.10242   -0.09    0.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.616
## Multiple R-squared:  0.0187, Adjusted R-squared:  0.00503
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1138 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.120  0.848  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      8.08e-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 1.048 1          1.024
## Year            1.048 16          1.001

```

Residuals from last author



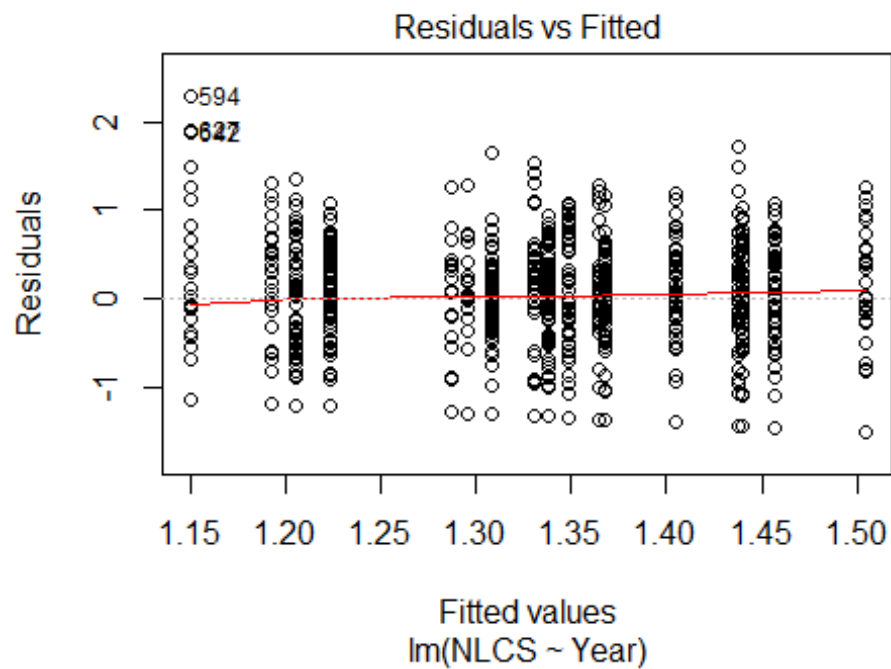
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5129 -0.4271 0.0253 0.4051 2.3307
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.32197 0.09165 14.42 <2e-16 ***
## LastAuthorFemale1 -0.00198 0.07424 -0.03 0.98
## Year1997 -0.03603 0.13204 -0.27 0.79
## Year1998 -0.22243 0.16355 -1.36 0.17
## Year1999 0.23309 0.18617 1.25 0.21
## Year2000 -0.10789 0.18300 -0.59 0.56
## Year2001 0.18745 0.18020 1.04 0.30
## Year2002 0.19093 0.17108 1.12 0.26
## Year2003 -0.13606 0.12804 -1.06 0.29
## Year2004 0.06430 0.13584 0.47 0.64
## Year2005 0.08116 0.12773 0.64 0.53
## Year2006 0.11840 0.12312 0.96 0.34
```

```

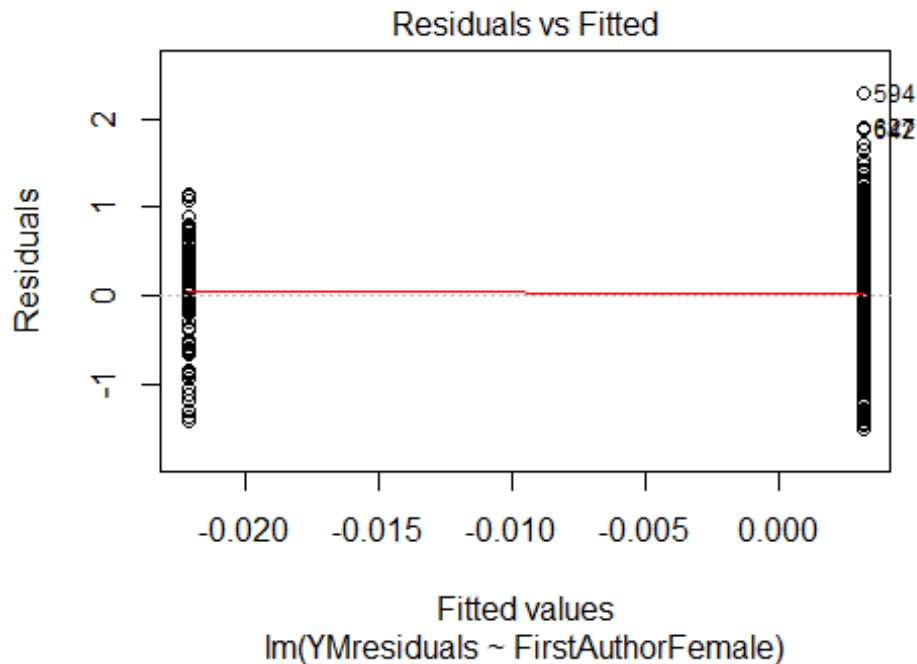
## Year2007          -0.06923      0.12222      -0.57      0.57
## Year2008           0.05172      0.11050       0.47      0.64
## Year2009           0.03276      0.11446       0.29      0.77
## Year2010          -0.04172      0.10138      -0.41      0.68
## Year2011          -0.01263      0.10230      -0.12      0.90
## Year2012          -0.00904      0.10248      -0.09      0.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.0187, Adjusted R-squared:  0.00503
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1138 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.120  0.848  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      8.08e-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: 1237"
## [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
## 120 111 110 122 95 121 93 123 122 140 190 153 199 207 215
## 2011 2012
## 225 193
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 42 39 34 20 35 23 31 60 47 51 57 57 71 59 84
## 2011 2012

```

```
## 105 89
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 26 32 31 19 29 14 25 49 39 42 48 45 53 47 70
## 2011 2012
## 82 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 = 71, df = 16, p-value = 7e-09
```

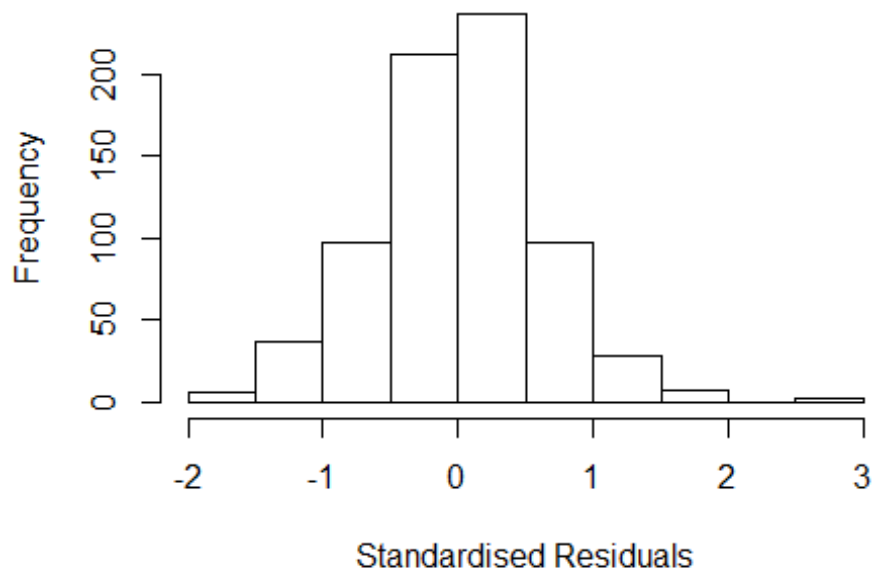


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.5, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.04889184291939"
## [1] "Male first author team size 2018 geometric mean: 2.85929986330132"
##
## 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] "Female last author team size 2018 geometric mean: 2.5544929375461"
## [1] "Male last author team size 2018 geometric mean: 2.95673731437817"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 740, 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 1.136 1      1.066
## LastAuthorFemale  1.132 1      1.064
## UniqueAuthors    1.798 4      1.076
## Year              1.928 16     1.021
```

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
## 594 0033888092 3.438 2000      2102      4      2.967
## 642 0033894333 3.035 2000      1500      3      2.564
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6355 -0.3718  0.0163  0.3896  2.9666
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9836    0.1433   6.87 1.5e-11 ***
## FirstAuthorFemale1 -0.1334    0.0640  -2.09 0.03742 *
## LastAuthorFemale1  0.0758    0.0947   0.80 0.42398
## UniqueAuthors2    0.3886    0.1047   3.71 0.00022 ***
## UniqueAuthors3    0.6267    0.1051   5.96 3.9e-09 ***
## UniqueAuthors4    0.4528    0.1084   4.18 3.4e-05 ***
## UniqueAuthors5    0.4105    0.1165   3.52 0.00046 ***
## Year1997        -0.0506    0.1877  -0.27 0.78745
## Year1998        -0.0185    0.1846  -0.10 0.92040
```

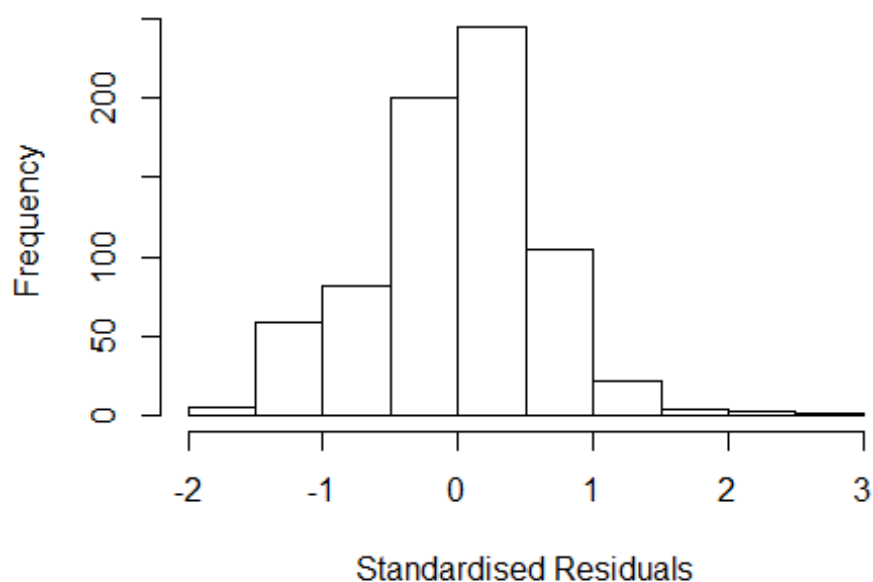


```

## Year1999          -0.0559      0.1956    -0.29  0.77533
## Year2000          -0.5122      0.1882    -2.72  0.00667 **
## Year2001          -0.1909      0.2000    -0.95  0.33997
## Year2002           0.1717      0.2326     0.74  0.46081
## Year2003          -0.1467      0.1478    -0.99  0.32120
## Year2004           0.1037      0.1768     0.59  0.55751
## Year2005           0.0372      0.1526     0.24  0.80723
## Year2006           0.0802      0.1414     0.57  0.57082
## Year2007           0.0761      0.1434     0.53  0.59572
## Year2008          -0.0186      0.1253    -0.15  0.88192
## Year2009           0.1095      0.1328     0.82  0.40977
## Year2010          -0.1376      0.1239    -1.11  0.26696
## Year2011          -0.1314      0.1216    -1.08  0.28030
## Year2012          -0.0948      0.1179    -0.80  0.42147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.561
## Multiple R-squared:  0.142, Adjusted R-squared:  0.115
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## observation 117 is an outlier with |weight| = 0 ( < 0.00014);
## 58 weights are ~= 1. The remaining 664 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0025 0.8560 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          1.38e-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 1.130 1          1.063
## LastAuthorFemale  1.099 1          1.048
## Year              1.119 16          1.004

```

Residuals from first and last author



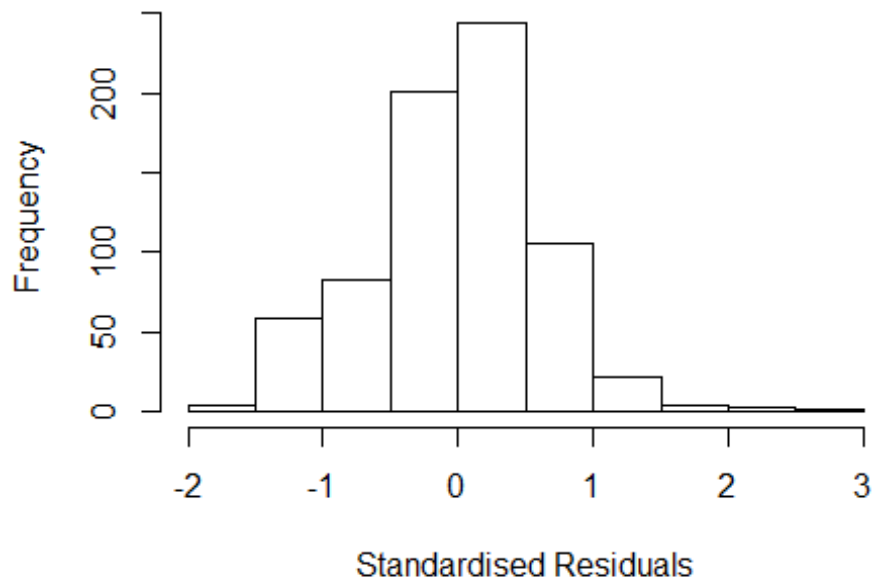
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 594 0033888092 3.438 2000      2102      4      2.636
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5664 -0.3643  0.0281  0.3882  2.6364
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3876     0.1155   12.02  <2e-16 ***
## FirstAuthorFemale1 -0.0725     0.0655   -1.11   0.2690
## LastAuthorFemale1  0.0935     0.0924    1.01   0.3114
## Year1997          -0.1050     0.1971   -0.53   0.5943
## Year1998          -0.1488     0.1964   -0.76   0.4489
## Year1999          -0.1022     0.2078   -0.49   0.6230
## Year2000          -0.5860     0.1857   -3.16   0.0017 **
## Year2001          -0.1454     0.1846   -0.79   0.4310
## Year2002           0.1788     0.2202    0.81   0.4171
## Year2003          -0.2234     0.1512   -1.48   0.1400
## Year2004           0.0620     0.1833    0.34   0.7354
## Year2005          -0.0369     0.1639   -0.23   0.8217
```

```

## Year2006          0.1015      0.1469      0.69      0.4899
## Year2007          0.0585      0.1600      0.37      0.7148
## Year2008         -0.0291      0.1337     -0.22      0.8278
## Year2009          0.1030      0.1431      0.72      0.4720
## Year2010         -0.0534      0.1291     -0.41      0.6793
## Year2011         -0.0646      0.1306     -0.49      0.6209
## Year2012         -0.0580      0.1246     -0.47      0.6419
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.0533, Adjusted R-squared:  0.0291
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 117 is an outlier with |weight| = 0 ( < 0.00014);
## 66 weights are ~= 1. The remaining 656 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0706 0.8480 0.9480 0.8850 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.38e-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 1.072 1          1.035
## Year              1.072 16          1.002

```

Residuals from first author



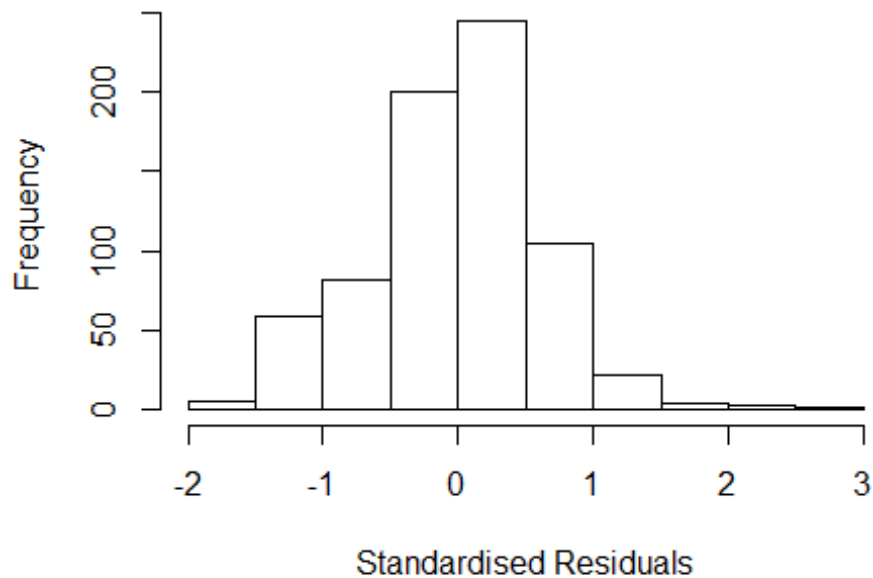
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 594 0033888092 3.438 2000      2102      4      2.636
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5701 -0.3686  0.0366  0.3875  2.6314
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3912     0.1156   12.04  <2e-16 ***
## FirstAuthorFemale1 -0.0592     0.0645   -0.92   0.3590
## Year1997         -0.1065     0.1960   -0.54   0.5872
## Year1998         -0.1498     0.1964   -0.76   0.4459
## Year1999         -0.1023     0.2094   -0.49   0.6255
## Year2000         -0.5846     0.1878   -3.11   0.0019 **
## Year2001         -0.1424     0.1858   -0.77   0.4438
## Year2002          0.1789     0.2202    0.81   0.4169
## Year2003         -0.2204     0.1514   -1.46   0.1459
## Year2004          0.0634     0.1837    0.35   0.7300
## Year2005         -0.0403     0.1643   -0.25   0.8064
## Year2006          0.1056     0.1467    0.72   0.4717
```

```

## Year2007          0.0562      0.1598      0.35      0.7253
## Year2008         -0.0290      0.1337     -0.22      0.8281
## Year2009          0.1038      0.1444      0.72      0.4724
## Year2010         -0.0532      0.1290     -0.41      0.6804
## Year2011         -0.0647      0.1311     -0.49      0.6220
## Year2012         -0.0552      0.1250     -0.44      0.6587
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.0514, Adjusted R-squared:  0.0285
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 663 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8500 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      1.38e-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 1.045 1          1.022
## Year            1.045 16          1.001

```

Residuals from last author



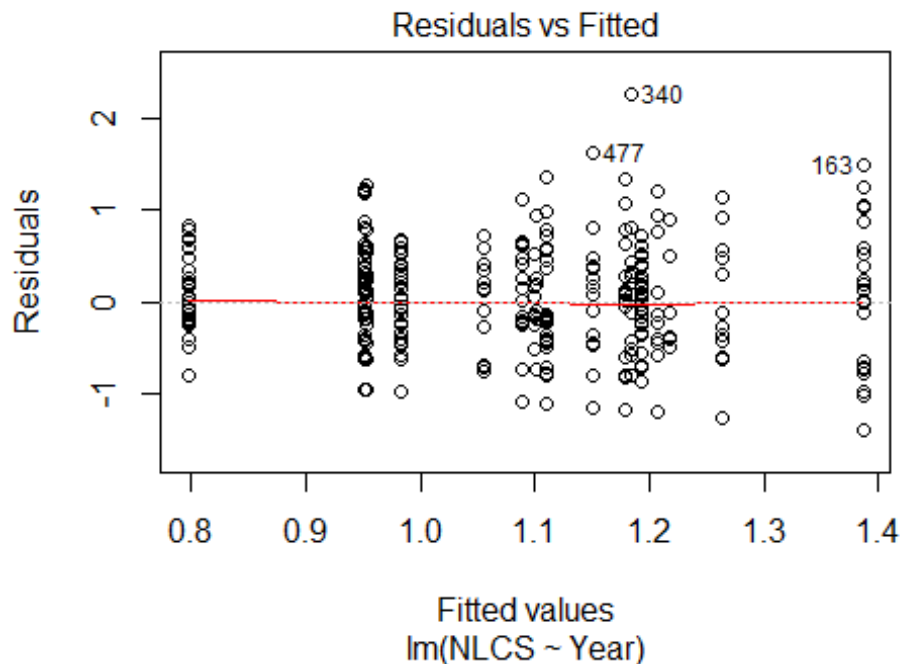
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 594 0033888092 3.438 2000      2102      4      2.636
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5590 -0.3659  0.0303  0.3663  2.6453
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3811     0.1156   11.95  <2e-16 ***
## LastAuthorFemale1  0.0739     0.0895    0.83  0.4092
## Year1997        -0.1053     0.1974   -0.53  0.5939
## Year1998        -0.1525     0.1973   -0.77  0.4400
## Year1999        -0.0997     0.2073   -0.48  0.6305
## Year2000        -0.5884     0.1868   -3.15  0.0017 **
## Year2001        -0.1495     0.1865   -0.80  0.4229
## Year2002         0.1779     0.2201    0.81  0.4193
## Year2003        -0.2263     0.1521   -1.49  0.1371
## Year2004         0.0641     0.1848    0.35  0.7289
## Year2005        -0.0368     0.1638   -0.22  0.8222
## Year2006         0.1010     0.1472    0.69  0.4928
```

```

## Year2007          0.0551      0.1612      0.34      0.7326
## Year2008          -0.0350      0.1340     -0.26      0.7940
## Year2009           0.0947      0.1431      0.66      0.5082
## Year2010          -0.0583      0.1296     -0.45      0.6527
## Year2011          -0.0689      0.1314     -0.52      0.6006
## Year2012          -0.0579      0.1251     -0.46      0.6436
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.0516, Adjusted R-squared:  0.0287
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 117 is an outlier with |weight| = 0 ( < 0.00014);
## 64 weights are ~= 1. The remaining 658 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0687 0.8500 0.9500 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          1.38e-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: 723"
## [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
##   78   85   57   65   52   42   50   46   38   64   60   47   96  103   62
## 2011 2012
##   91   88
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   22   22    6    9   14    8   17   12   10   25   22   14   33   39   20

```

```
## 2011 2012
## 41 36
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 16 3 8 11 6 13 11 9 19 16 12 25 31 16
## 2011 2012
## 32 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 = 31, df = 16, p-value = 0.01
```

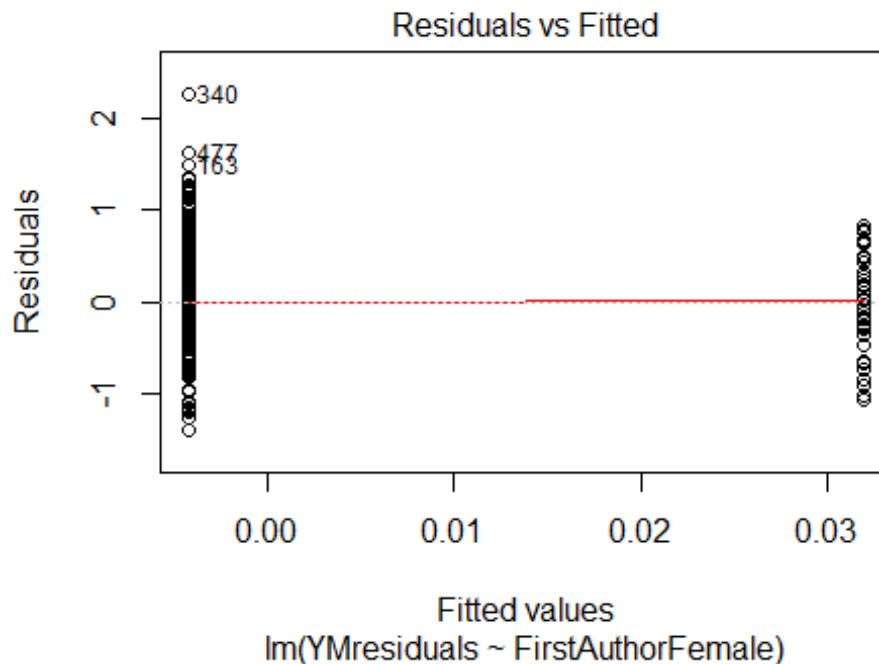


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.9, df = 1, p-value = 0.2
## [1] "Female first author team size 2018 geometric mean: 3.30192724889463"
## [1] "Male first author team size 2018 geometric mean: 2.44733216879806"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



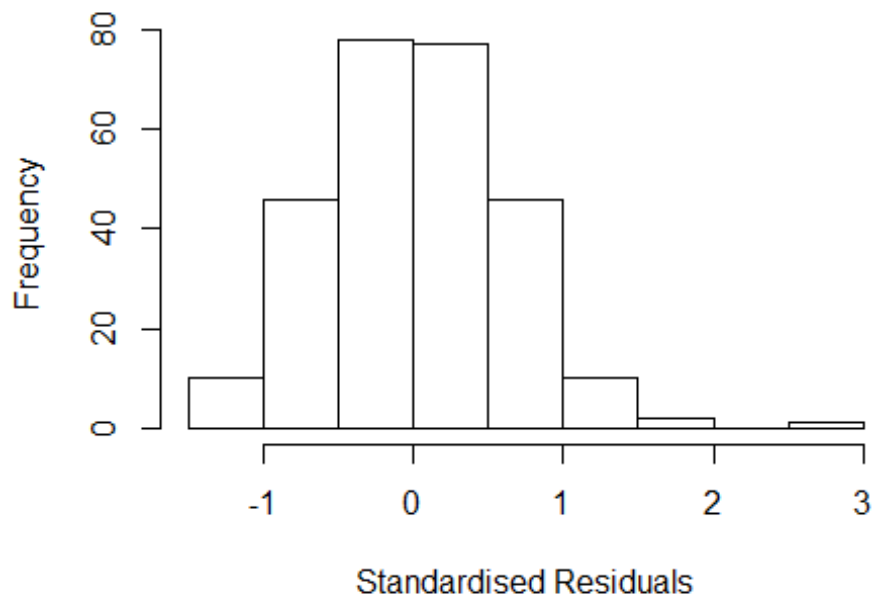
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 230, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.1192680453341"
## [1] "Male last author team size 2018 geometric mean: 2.69109692648093"

## Warning in wilcox.test.default(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 = 0.2
## 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 1.250  1      1.118
## LastAuthorFemale  1.379  1      1.174
## UniqueAuthors    3.597  4      1.174
## Year              4.387 16      1.047
```

Residuals from first and last author and team size



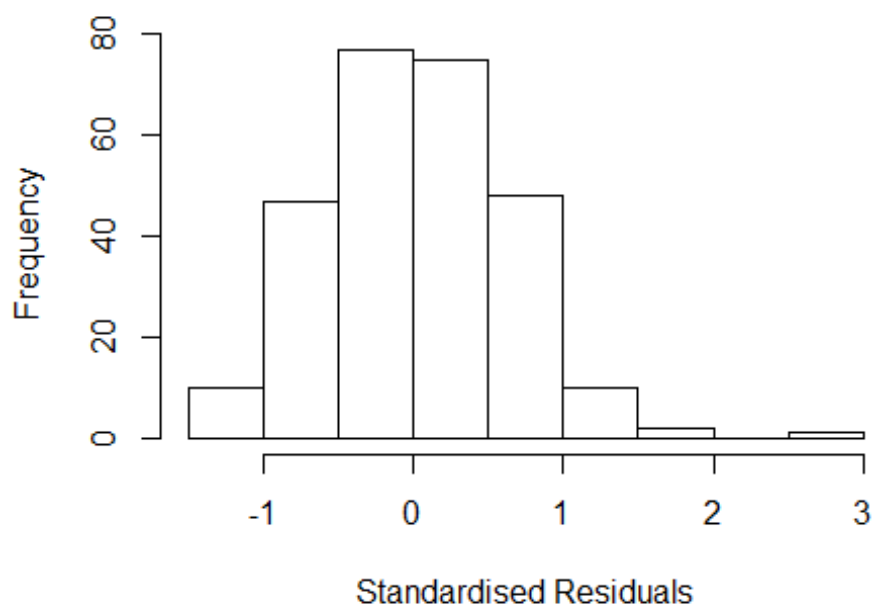
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 340 0033888092 3.438 2000      2102      4      2.58
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41005 -0.40211  0.00278  0.44532  2.58014
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03879    0.26765   3.88 0.00013 ***
## FirstAuthorFemale1  0.07803    0.11209   0.70 0.48697
## LastAuthorFemale1  0.03012    0.23055   0.13 0.89615
## UniqueAuthors2     0.00631    0.12644   0.05 0.96025
## UniqueAuthors3    -0.05937    0.15535  -0.38 0.70266
## UniqueAuthors4    -0.05135    0.15769  -0.33 0.74500
## UniqueAuthors5     0.14056    0.16060   0.88 0.38229
## Year1997          0.18293    0.40451   0.45 0.65150
## Year1998          0.11247    0.36779   0.31 0.76001
## Year1999          0.04303    0.29019   0.15 0.88223
```

```

## Year2000      -0.18093    0.31208   -0.58  0.56261
## Year2001      0.05928    0.35890    0.17  0.86895
## Year2002      0.00774    0.35543    0.02  0.98264
## Year2003      0.06640    0.31218    0.21  0.83174
## Year2004      0.14749    0.41517    0.36  0.72271
## Year2005     -0.05915    0.30933   -0.19  0.84850
## Year2006     -0.10939    0.32024   -0.34  0.73294
## Year2007      0.23070    0.36516    0.63  0.52812
## Year2008      0.18225    0.28815    0.63  0.52765
## Year2009     -0.23871    0.30080   -0.79  0.42820
## Year2010      0.06994    0.33115    0.21  0.83291
## Year2011     -0.03466    0.28520   -0.12  0.90338
## Year2012     -0.21075    0.29286   -0.72  0.47244
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.602
## Multiple R-squared:  0.0627, Adjusted R-squared:  -0.0208
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 251 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0269 0.8760 0.9490 0.9080 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.70e-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 1.155 1      1.075
## LastAuthorFemale  1.259 1      1.122
## Year              1.360 16      1.010

```

Residuals from first and last author

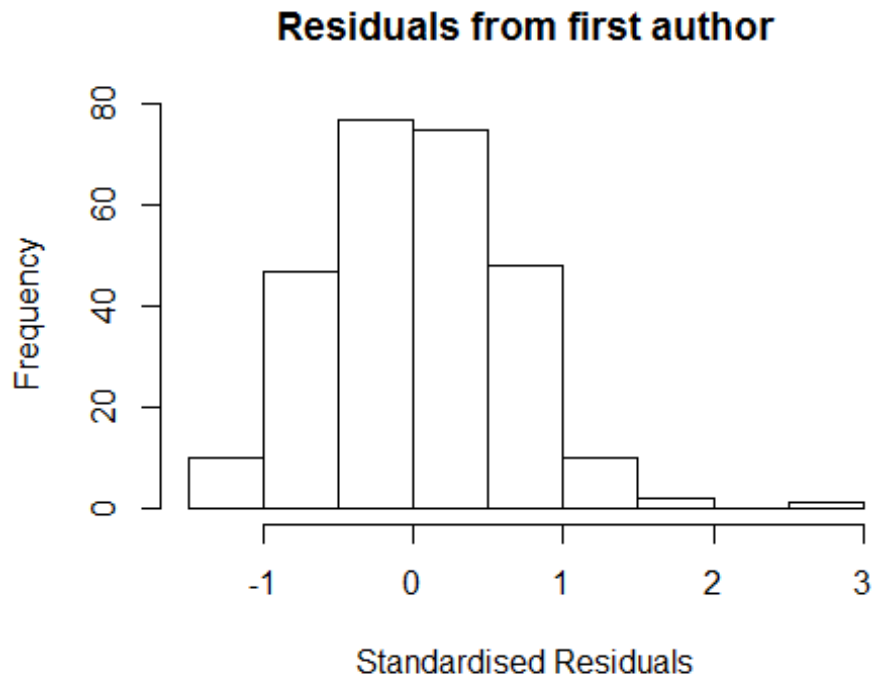


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 340 0033888092 3.438 2000      2102      4      2.577
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26175 -0.40549  0.00801  0.43181  2.57707
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0331     0.2653   3.89 0.00013 ***
## FirstAuthorFemale1  0.0881     0.1104   0.80 0.42542
## LastAuthorFemale1  0.0744     0.2286   0.33 0.74518
## Year1997          0.1772     0.3886   0.46 0.64873
## Year1998          0.1719     0.3510   0.49 0.62470
## Year1999          0.0441     0.2858   0.15 0.87748
## Year2000         -0.1722     0.3001  -0.57 0.56666
## Year2001          0.0483     0.3390   0.14 0.88694
## Year2002          0.0292     0.3328   0.09 0.93018
## Year2003          0.0773     0.3004   0.26 0.79705
## Year2004          0.1157     0.3899   0.30 0.76698
## Year2005         -0.0457     0.3014  -0.15 0.87968
```

```

## Year2006          -0.0960      0.3116   -0.31  0.75820
## Year2007           0.2286      0.3571    0.64  0.52260
## Year2008           0.1906      0.2829    0.67  0.50112
## Year2009          -0.2010      0.2854   -0.70  0.48203
## Year2010           0.0687      0.3116    0.22  0.82566
## Year2011          -0.0407      0.2783   -0.15  0.88381
## Year2012          -0.1977      0.2798   -0.71  0.48057
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.0534, Adjusted R-squared:  -0.0145
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 255 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.040  0.885   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      3.70e-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 1.11 1         1.053
## Year              1.11 16         1.003

```

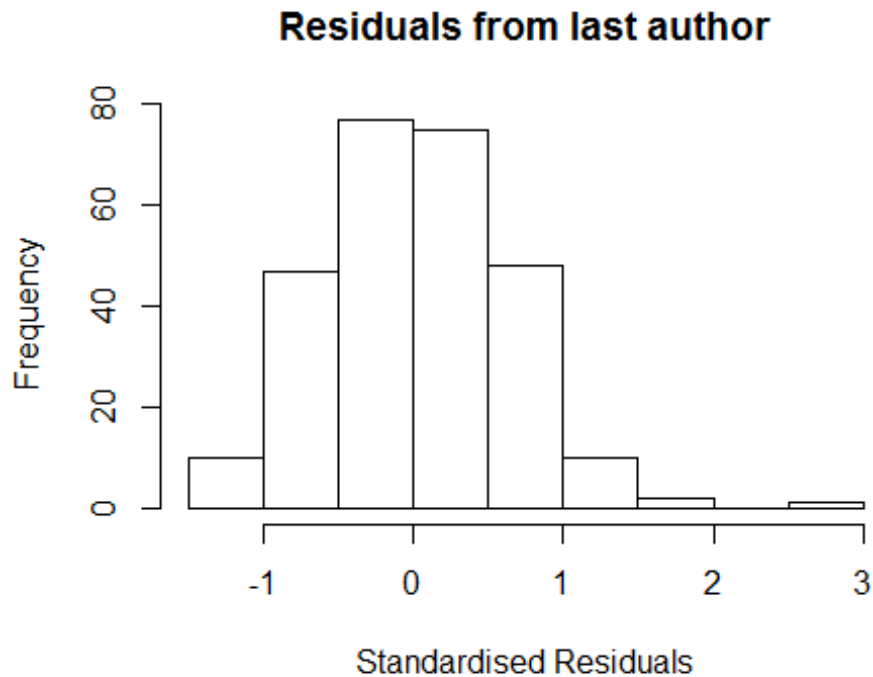


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 340 0033888092 3.438 2000      2102      4      2.577
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2702 -0.4046  0.0068  0.4269  2.5703
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0426     0.2579   4.04    7e-05 ***
## FirstAuthorFemale1  0.0961     0.1078   0.89     0.37
## Year1997          0.1783     0.3834   0.47     0.64
## Year1998          0.1628     0.3454   0.47     0.64
## Year1999          0.0347     0.2790   0.12     0.90
## Year2000         -0.1749     0.3005  -0.58     0.56
## Year2001          0.0399     0.3341   0.12     0.91
## Year2002          0.0214     0.3279   0.07     0.95
## Year2003          0.0688     0.2906   0.24     0.81
## Year2004          0.1169     0.3927   0.30     0.77
## Year2005         -0.0464     0.2975  -0.16     0.88
## Year2006         -0.1022     0.3048  -0.34     0.74
```

```

## Year2007          0.2277      0.3551      0.64      0.52
## Year2008          0.1818      0.2754      0.66      0.51
## Year2009         -0.2066      0.2811     -0.73      0.46
## Year2010          0.0578      0.3039      0.19      0.85
## Year2011         -0.0519      0.2700     -0.19      0.85
## Year2012         -0.2066      0.2731     -0.76      0.45
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.632
## Multiple R-squared:  0.0525, Adjusted R-squared:  -0.0115
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 250 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0608 0.8880 0.9500 0.9150 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.70e-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 1.201 1          1.096
## Year              1.201 16          1.006

```



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 340 0033888092 3.438 2000      2102      4      2.577
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26896 -0.40043  0.00549  0.44801  2.57847
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0407     0.2652   3.92 0.00011 ***
## LastAuthorFemale1  0.0987     0.2231   0.44 0.65869
## Year1997          0.1711     0.3858   0.44 0.65779
## Year1998          0.1643     0.3511   0.47 0.64023
## Year1999          0.0365     0.2858   0.13 0.89833
## Year2000         -0.1811     0.2998  -0.60 0.54630
## Year2001          0.0404     0.3390   0.12 0.90514
## Year2002          0.0213     0.3327   0.06 0.94911
## Year2003          0.0895     0.3030   0.30 0.76797
## Year2004          0.1047     0.3881   0.27 0.78747
## Year2005         -0.0508     0.3032  -0.17 0.86714
## Year2006         -0.0993     0.3109  -0.32 0.74978
```

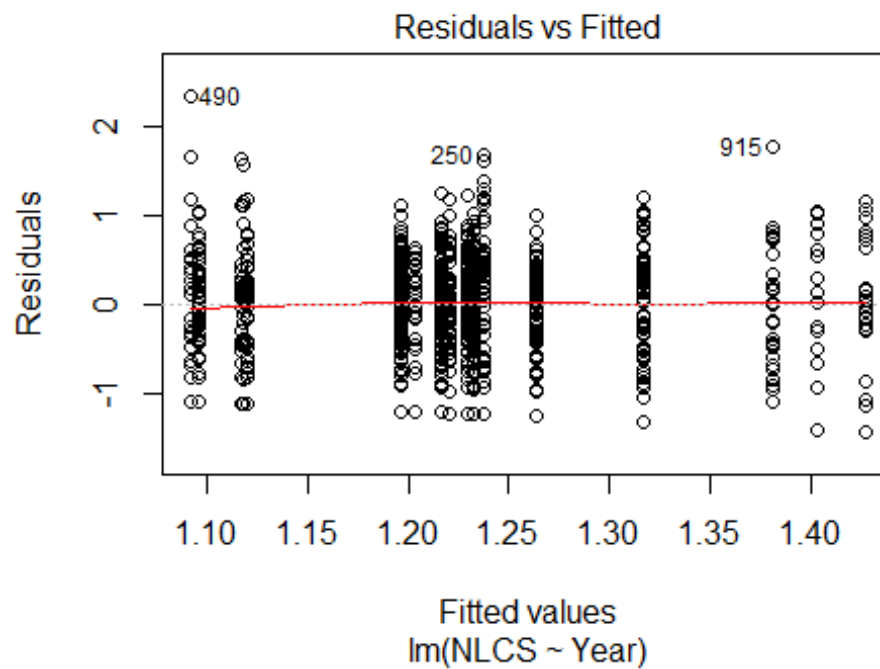


```

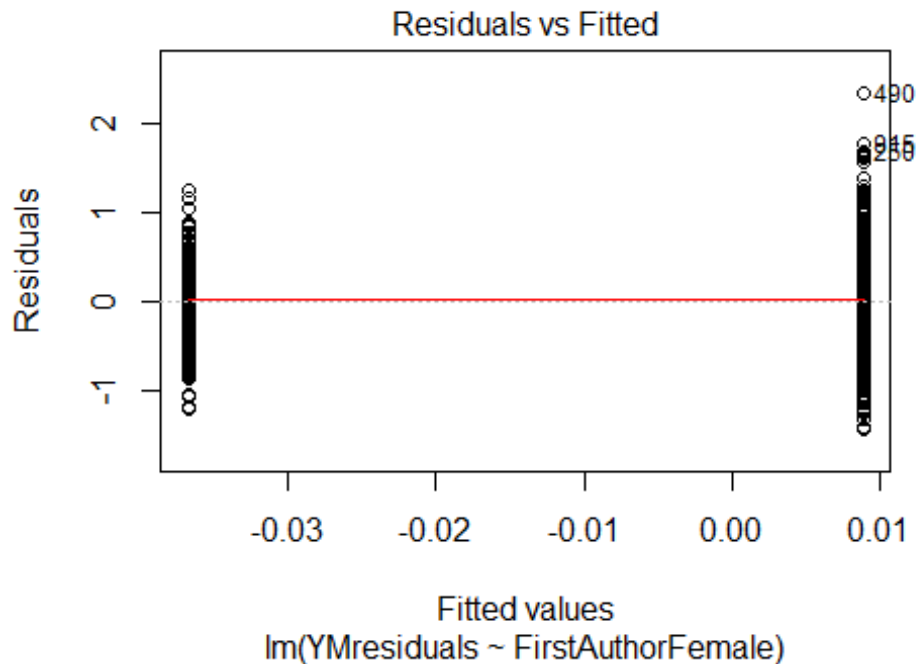
## Year2007          0.2283      0.3599      0.63  0.52650
## Year2008          0.1955      0.2840      0.69  0.49175
## Year2009         -0.1977      0.2875     -0.69  0.49241
## Year2010          0.0686      0.3128      0.22  0.82650
## Year2011         -0.0338      0.2799     -0.12  0.90412
## Year2012         -0.1925      0.2810     -0.68  0.49398
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.611
## Multiple R-squared:  0.0517, Adjusted R-squared:  -0.0123
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 251 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0358 0.8800 0.9460 0.9100 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.70e-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: 270"
## [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
##   90  141   98   88   75   81   73   93   87  118  158  157  212  245  287
## 2011 2012
##  329  331
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   28   55   28   18   30   28   25   41   31   49   66   61   82  117  143
## 2011 2012

```

```
## 162 170
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 17 42 21 12 27 22 22 32 28 39 51 46 64 97 116
## 2011 2012
## 133 142
## [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-15
```

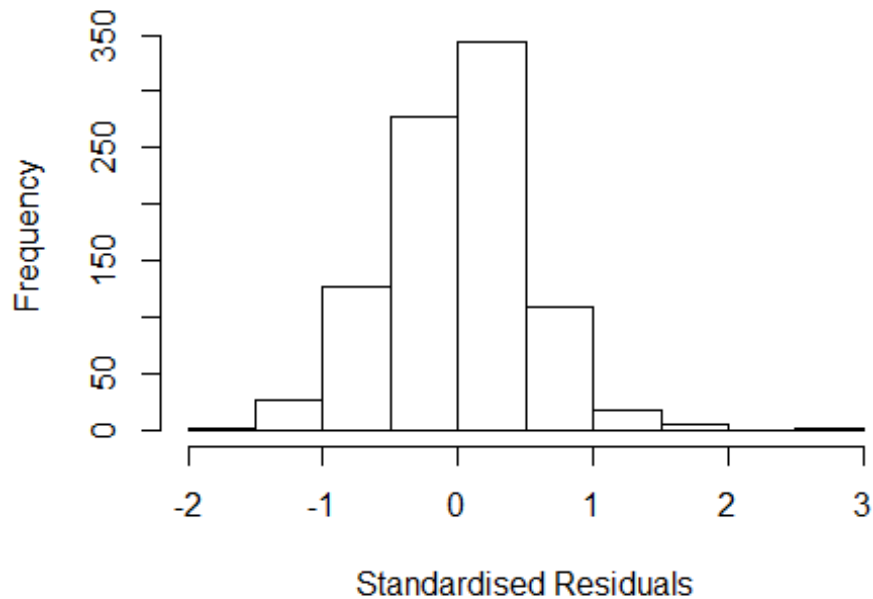


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.5, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 3.04671098698968"
## [1] "Male first author team size 2018 geometric mean: 2.80249937988468"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5200, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.62824932806061"
## [1] "Male last author team size 2018 geometric mean: 2.89847054610875"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2900, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.212 1          1.101
## LastAuthorFemale  1.226 1          1.107
## UniqueAuthors    1.623 4          1.062
## Year              1.705 16         1.017
```

Residuals from first and last author and team size



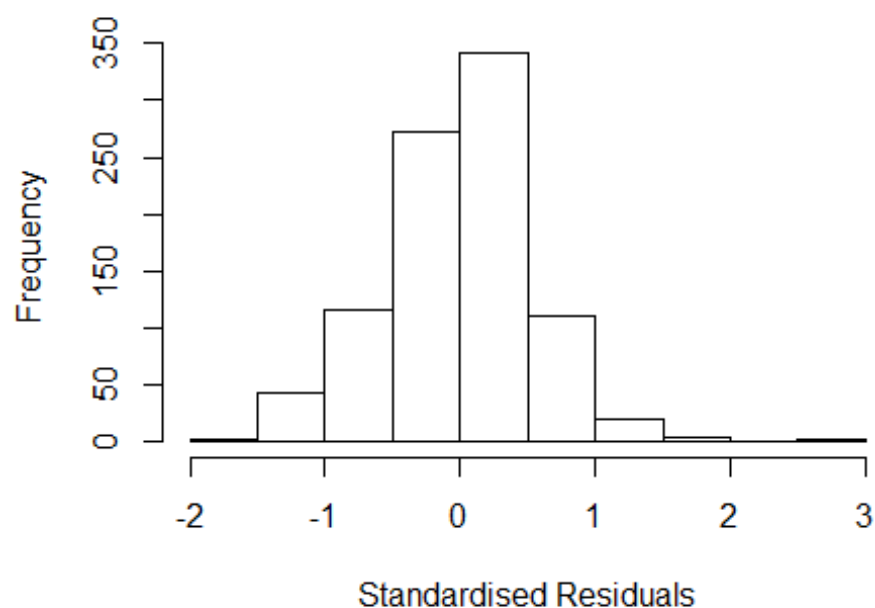
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 490 0033888092 3.438 2000      2102      4      2.791
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6028 -0.3339  0.0308  0.3243  2.7915
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2854    0.2215   5.80 9.1e-09 ***
## FirstAuthorFemale1 -0.0950    0.0433  -2.19  0.0286 *
## LastAuthorFemale1  0.0980    0.0538   1.82  0.0687 .
## UniqueAuthors2     0.2194    0.0669   3.28  0.0011 **
## UniqueAuthors3     0.3436    0.0682   5.04 5.7e-07 ***
## UniqueAuthors4     0.3461    0.0715   4.84 1.5e-06 ***
## UniqueAuthors5     0.3012    0.0742   4.06 5.3e-05 ***
## Year1997          -0.5019    0.2550  -1.97  0.0493 *
## Year1998          -0.2352    0.2365  -0.99  0.3202
## Year1999          -0.2359    0.2875  -0.82  0.4121
```

```

## Year2000          -0.6389      0.2719    -2.35    0.0190 *
## Year2001          -0.4407      0.2547    -1.73    0.0840 .
## Year2002          -0.4457      0.2524    -1.77    0.0778 .
## Year2003          -0.3862      0.2326    -1.66    0.0972 .
## Year2004          -0.1705      0.2639    -0.65    0.5186
## Year2005          -0.4562      0.2234    -2.04    0.0414 *
## Year2006          -0.2244      0.2255    -0.99    0.3201
## Year2007          -0.3261      0.2178    -1.50    0.1346
## Year2008          -0.1837      0.2127    -0.86    0.3882
## Year2009          -0.2772      0.2124    -1.31    0.1921
## Year2010          -0.3749      0.2103    -1.78    0.0750 .
## Year2011          -0.3106      0.2078    -1.49    0.1353
## Year2012          -0.3362      0.2096    -1.60    0.1091
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.0953, Adjusted R-squared:  0.0729
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 97 is an outlier with |weight| = 0 ( < 0.00011);
## 76 weights are ~= 1. The remaining 834 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0865 0.8630 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          1.10e-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 1.152 1          1.073
## LastAuthorFemale 1.117 1          1.057
## Year 1.145 16          1.004

```

Residuals from first and last author

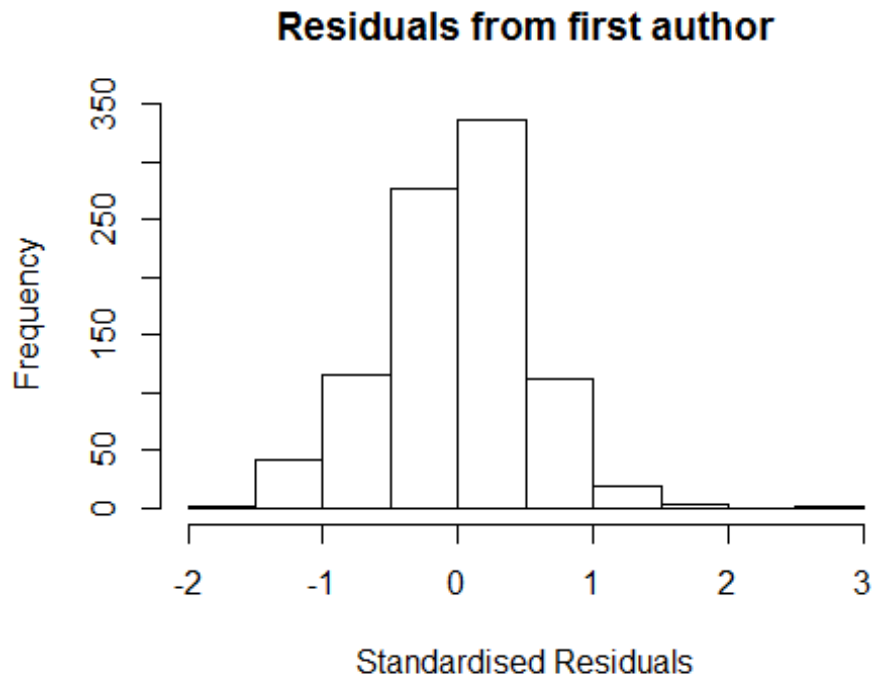


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 490 0033888092 3.438 2000      2102      4      2.582
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5908 -0.3285  0.0347  0.3227  2.5824
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5120     0.2050   7.38 3.7e-13 ***
## FirstAuthorFemale1 -0.0848     0.0434  -1.95  0.051 .
## LastAuthorFemale1  0.0788     0.0523   1.51  0.132
## Year1997          -0.5060     0.2615  -1.94  0.053 .
## Year1998          -0.2693     0.2337  -1.15  0.250
## Year1999          -0.2100     0.2910  -0.72  0.471
## Year2000          -0.6564     0.2586  -2.54  0.011 *
## Year2001          -0.4547     0.2565  -1.77  0.077 .
## Year2002          -0.4501     0.2525  -1.78  0.075 .
## Year2003          -0.4107     0.2313  -1.78  0.076 .
## Year2004          -0.1689     0.2655  -0.64  0.525
## Year2005          -0.4716     0.2198  -2.15  0.032 *
```

```

## Year2006          -0.2343      0.2247   -1.04    0.297
## Year2007          -0.3103      0.2174   -1.43    0.154
## Year2008          -0.1770      0.2109   -0.84    0.402
## Year2009          -0.2567      0.2104   -1.22    0.223
## Year2010          -0.3374      0.2081   -1.62    0.105
## Year2011          -0.2620      0.2065   -1.27    0.205
## Year2012          -0.2838      0.2077   -1.37    0.172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.0468, Adjusted R-squared:  0.0275
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 97 is an outlier with |weight| = 0 ( < 0.00011);
## 67 weights are ~= 1. The remaining 843 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0993 0.8590 0.9520 0.8950 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.087 1          1.043
## Year              1.087 16          1.003

```



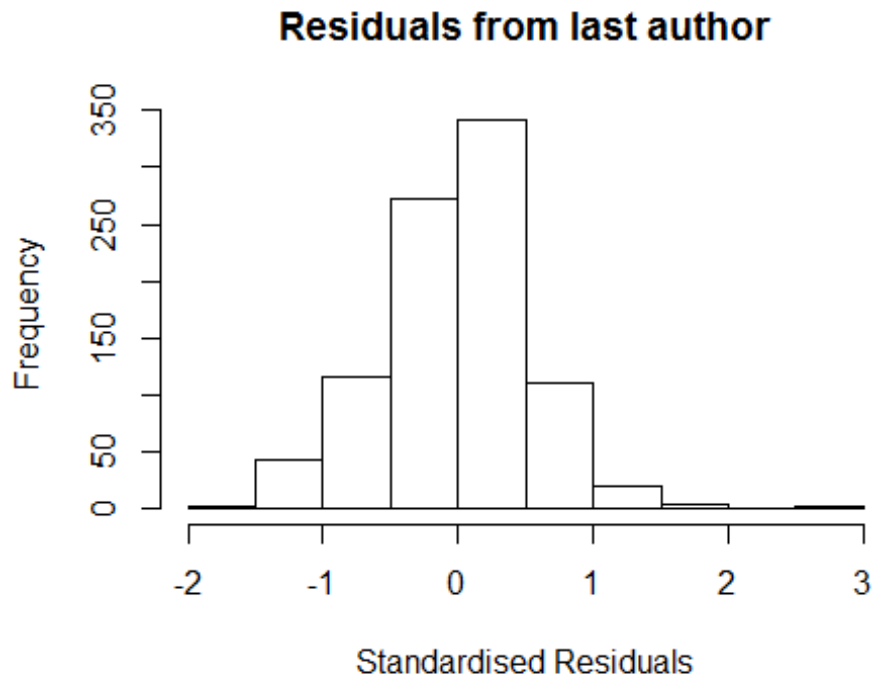
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 490 0033888092 3.438 2000      2102      4      2.582
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5074 -0.3369  0.0404  0.3287  2.5798
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5074     0.2052   7.34 4.7e-13 ***
## FirstAuthorFemale1 -0.0690     0.0426  -1.62  0.106
## Year1997         -0.4851     0.2587  -1.87  0.061 .
## Year1998         -0.2629     0.2334  -1.13  0.260
## Year1999         -0.1910     0.2873  -0.66  0.506
## Year2000         -0.6492     0.2576  -2.52  0.012 *
## Year2001         -0.4364     0.2550  -1.71  0.087 .
## Year2002         -0.4366     0.2511  -1.74  0.082 .
## Year2003         -0.3962     0.2314  -1.71  0.087 .
## Year2004         -0.1515     0.2619  -0.58  0.563
## Year2005         -0.4598     0.2203  -2.09  0.037 *
## Year2006         -0.2238     0.2245  -1.00  0.319
```



```

## Year2007          -0.2997      0.2173   -1.38    0.168
## Year2008          -0.1708      0.2110   -0.81    0.419
## Year2009          -0.2411      0.2104   -1.15    0.252
## Year2010          -0.3261      0.2082   -1.57    0.118
## Year2011          -0.2493      0.2064   -1.21    0.228
## Year2012          -0.2722      0.2079   -1.31    0.191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.043, Adjusted R-squared:  0.0247
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 97 is an outlier with |weight| = 0 ( < 0.00011);
## 61 weights are ~= 1. The remaining 849 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.114  0.866  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      1.10e-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 1.052 1          1.026
## Year            1.052 16          1.002

```



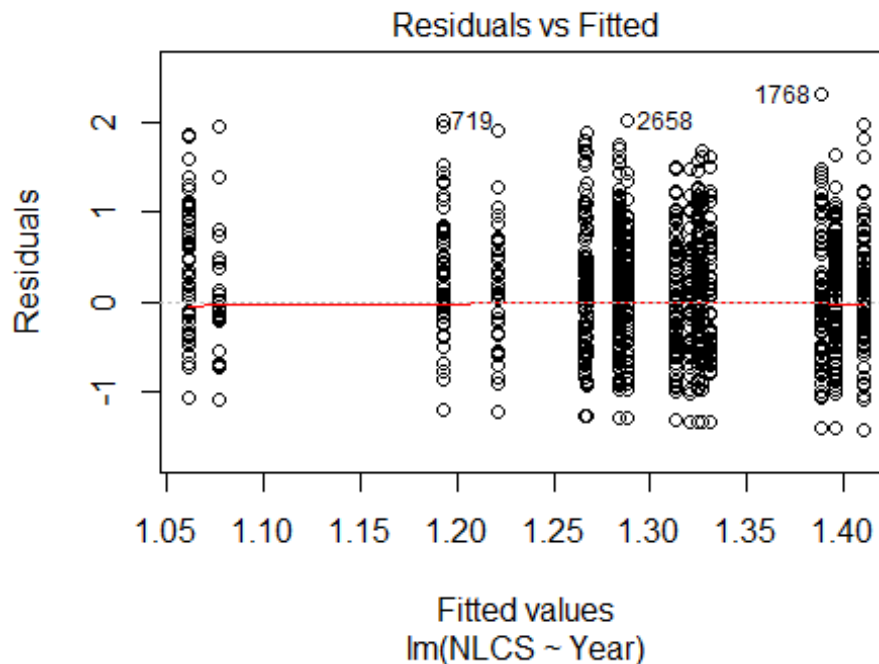
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 490 0033888092 3.438 2000      2102      4      2.582
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5398 -0.3480  0.0394  0.3175  2.5933
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4834     0.2069   7.17 1.6e-12 ***
## LastAuthorFemale1  0.0564     0.0505   1.12  0.264
## Year1997        -0.4857     0.2626  -1.85  0.065 .
## Year1998        -0.2464     0.2363  -1.04  0.297
## Year1999        -0.2123     0.2927  -0.73  0.468
## Year2000        -0.6387     0.2622  -2.44  0.015 *
## Year2001        -0.4515     0.2607  -1.73  0.084 .
## Year2002        -0.4331     0.2575  -1.68  0.093 .
## Year2003        -0.3925     0.2341  -1.68  0.094 .
## Year2004        -0.1466     0.2616  -0.56  0.575
## Year2005        -0.4511     0.2227  -2.03  0.043 *
## Year2006        -0.2244     0.2280  -0.98  0.325
```

```

## Year2007          -0.2824      0.2193    -1.29      0.198
## Year2008          -0.1739      0.2146    -0.81      0.418
## Year2009          -0.2446      0.2139    -1.14      0.253
## Year2010          -0.3216      0.2113    -1.52      0.128
## Year2011          -0.2537      0.2102    -1.21      0.228
## Year2012          -0.2658      0.2107    -1.26      0.207
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.0419, Adjusted R-squared:  0.0237
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 97 is an outlier with |weight| = 0 ( < 0.00011);
## 59 weights are ~= 1. The remaining 851 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.106  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      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 911"
## [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
## 309 198 156 183 144 100 101 98 93 152 170 140 160 178 193
## 2011 2012
## 185 186
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 150 78 75 87 71 26 53 45 43 73 81 73 87 89 112

```

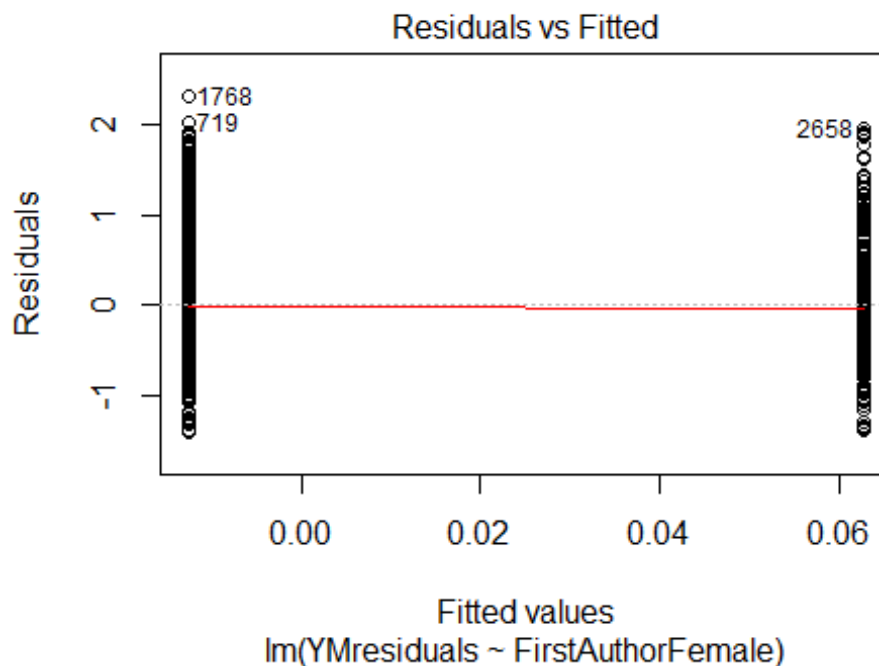
```
## 2011 2012
## 102 114
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 134 61 64 76 62 24 49 37 35 64 67 55 66 73 93
## 2011 2012
## 77 95
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 17, df = 16, p-value = 0.4
```



```
##
## 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: 2.68296613833544"
## [1] "Male first author team size 2018 geometric mean: 2.70862962797663"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

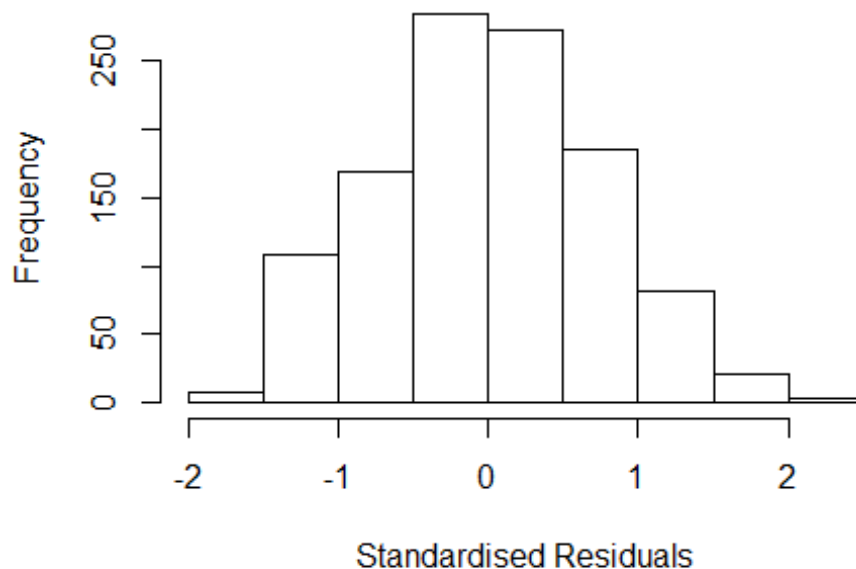
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 270, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.0356543087298"
## [1] "Male last author team size 2018 geometric mean: 2.57305422260652"

## Warning in wilcox.test.default(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 = 0.06
## 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 1.225  1      1.107
## LastAuthorFemale  1.262  1      1.124
## UniqueAuthors    1.248  4      1.028
## Year              1.364 16      1.010
```

Residuals from first and last author and team size



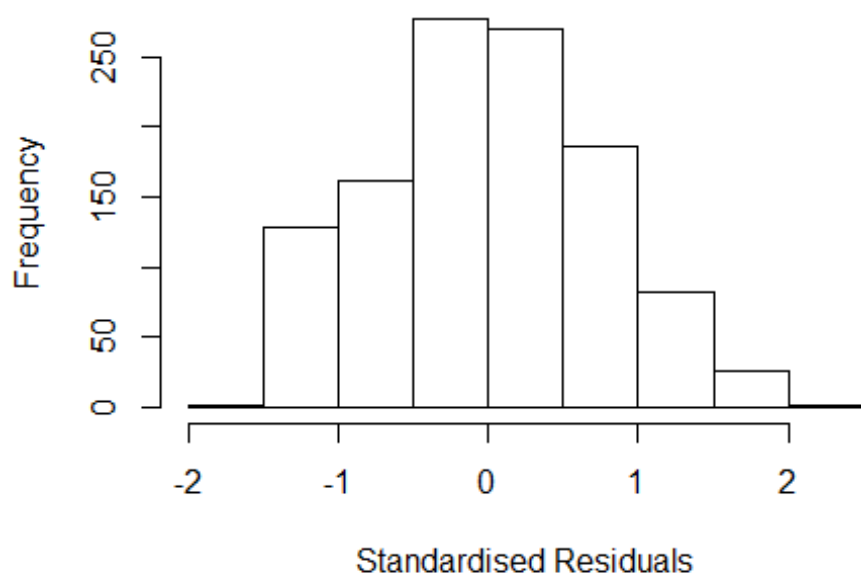
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.57346 -0.50404 -0.00219  0.51617  2.30566
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.085569   0.075635   14.35 < 2e-16 ***
## FirstAuthorFemale1  0.052905   0.068115    0.78  0.43750
## LastAuthorFemale1  0.134596   0.072066    1.87  0.06207 .
## UniqueAuthors2     0.201543   0.058228    3.46  0.00056 ***
## UniqueAuthors3     0.358831   0.068083    5.27  1.6e-07 ***
## UniqueAuthors4     0.201218   0.084093    2.39  0.01689 *
## UniqueAuthors5     0.370612   0.119520    3.10  0.00198 **
## Year1997          -0.053245   0.131832   -0.40  0.68637
## Year1998          -0.108982   0.131522   -0.83  0.40750
## Year1999          -0.206895   0.121838   -1.70  0.08977 .
```

```

## Year2000      0.091558    0.113376    0.81  0.41952
## Year2001     -0.308318    0.142429   -2.16  0.03062 *
## Year2002      0.058314    0.113962    0.51  0.60896
## Year2003     -0.076223    0.157077   -0.49  0.62759
## Year2004     -0.097994    0.145579   -0.67  0.50100
## Year2005      0.104232    0.121726    0.86  0.39202
## Year2006      0.111127    0.114715    0.97  0.33289
## Year2007      0.082774    0.127350    0.65  0.51585
## Year2008      0.000185    0.111999    0.00  0.99868
## Year2009     -0.038209    0.098695   -0.39  0.69873
## Year2010     -0.072049    0.110708   -0.65  0.51531
## Year2011      0.047924    0.107642    0.45  0.65625
## Year2012     -0.070221    0.099673   -0.70  0.48126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.773
## Multiple R-squared:  0.0543, Adjusted R-squared:  0.0355
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 90 weights are ~= 1. The remaining 1042 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.354  0.873  0.951  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          8.83e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.202 1          1.097
## LastAuthorFemale  1.248 1          1.117
## Year              1.121 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
## -1.52643 -0.51429 -0.00643 0.52676 2.34557
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2612 0.0682 18.48 <2e-16 ***
## FirstAuthorFemale1 0.0530 0.0674 0.79 0.432
## LastAuthorFemale1 0.1355 0.0725 1.87 0.062 .
## Year1997 -0.0675 0.1365 -0.49 0.621
## Year1998 -0.1268 0.1364 -0.93 0.353
## Year1999 -0.2286 0.1243 -1.84 0.066 .
## Year2000 0.0791 0.1116 0.71 0.479
## Year2001 -0.3378 0.1542 -2.19 0.029 *
## Year2002 0.0534 0.1159 0.46 0.645
## Year2003 -0.0808 0.1560 -0.52 0.604
## Year2004 -0.1075 0.1372 -0.78 0.433
## Year2005 0.0902 0.1222 0.74 0.460
```

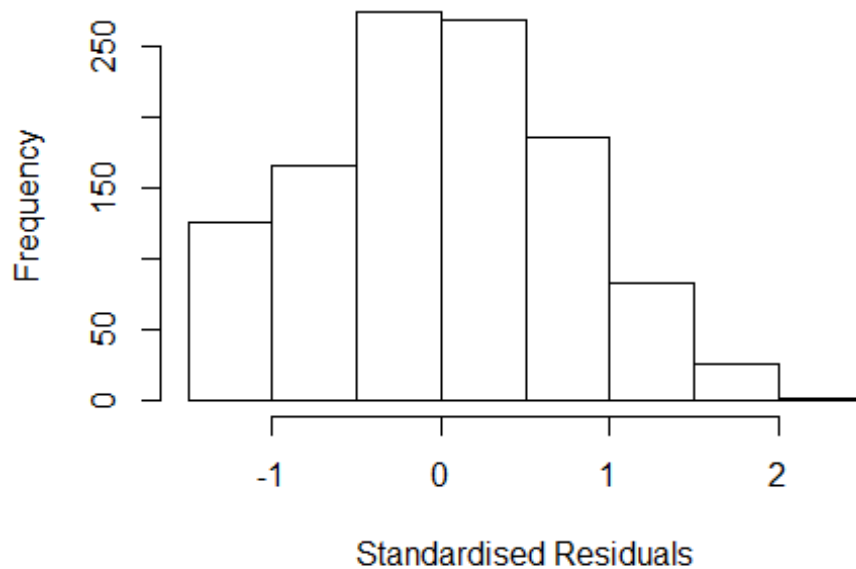


```

## Year2006          0.1297      0.1147      1.13      0.258
## Year2007          0.0818      0.1288      0.63      0.526
## Year2008          0.0176      0.1147      0.15      0.878
## Year2009         -0.0144      0.1008     -0.14      0.887
## Year2010         -0.0180      0.1114     -0.16      0.871
## Year2011          0.0903      0.1119      0.81      0.420
## Year2012         -0.0193      0.0992     -0.19      0.846
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.784
## Multiple R-squared:  0.0232, Adjusted R-squared:  0.0074
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1035 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.350  0.872  0.950  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      8.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



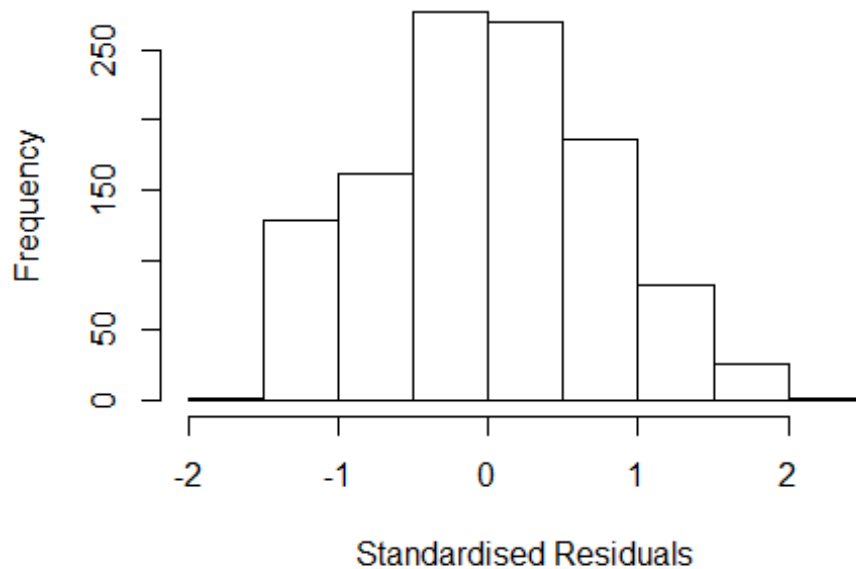
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.458 -0.515 -0.007 0.543 2.336
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26502 0.06844 18.48 <2e-16 ***
## FirstAuthorFemale1 0.10037 0.06268 1.60 0.110
## Year1997 -0.04482 0.13645 -0.33 0.743
## Year1998 -0.11396 0.13814 -0.82 0.410
## Year1999 -0.22982 0.12492 -1.84 0.066 .
## Year2000 0.09227 0.11133 0.83 0.407
## Year2001 -0.30300 0.15873 -1.91 0.057 .
## Year2002 0.06352 0.11553 0.55 0.583
## Year2003 -0.08272 0.15740 -0.53 0.599
## Year2004 -0.09329 0.14031 -0.66 0.506
## Year2005 0.09582 0.12127 0.79 0.430
## Year2006 0.15170 0.11366 1.33 0.182
```

```

## Year2007          0.08394    0.13032    0.64    0.520
## Year2008          0.01916    0.11557    0.17    0.868
## Year2009         -0.00923    0.10113   -0.09    0.927
## Year2010         -0.00642    0.11010   -0.06    0.954
## Year2011          0.10032    0.11155    0.90    0.369
## Year2012         -0.01293    0.09868   -0.13    0.896
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.787
## Multiple R-squared:  0.0199, Adjusted R-squared:  0.00499
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 1030 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.358  0.870  0.950  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      8.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.081 1          1.040
## Year              1.081 16          1.002

```

Residuals from last author



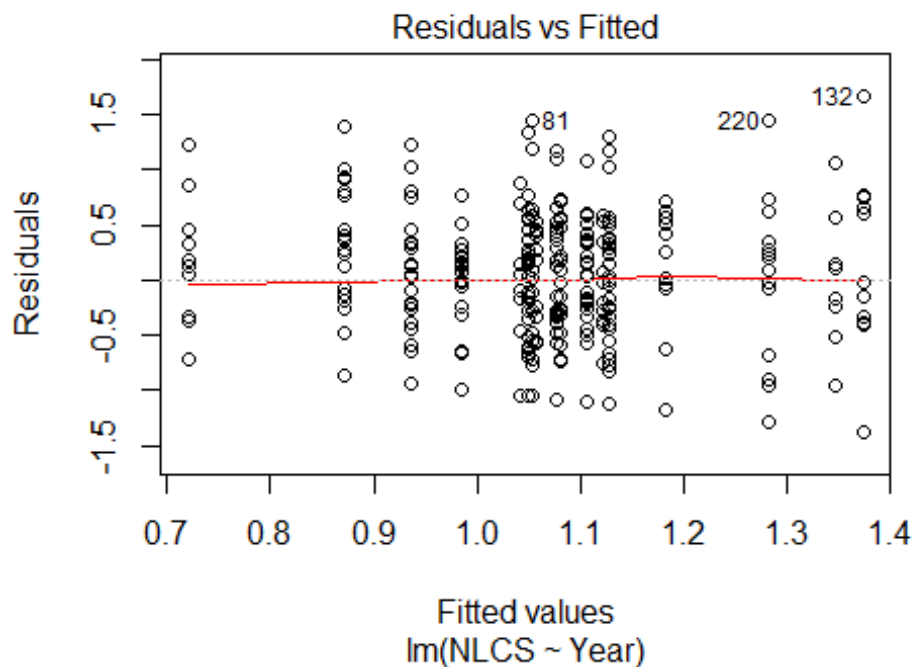
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.55241 -0.51724 -0.00383 0.52629 2.34133
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2669 0.0677 18.71 <2e-16 ***
## LastAuthorFemale1 0.1561 0.0673 2.32 0.021 *
## Year1997 -0.0703 0.1360 -0.52 0.605
## Year1998 -0.1268 0.1367 -0.93 0.354
## Year1999 -0.2282 0.1244 -1.83 0.067 .
## Year2000 0.0770 0.1114 0.69 0.490
## Year2001 -0.3441 0.1531 -2.25 0.025 *
## Year2002 0.0508 0.1157 0.44 0.661
## Year2003 -0.0782 0.1547 -0.51 0.613
## Year2004 -0.1115 0.1365 -0.82 0.414
## Year2005 0.0887 0.1225 0.72 0.469
## Year2006 0.1294 0.1144 1.13 0.258
```

```

## Year2007          0.0829      0.1292      0.64      0.521
## Year2008          0.0193      0.1144      0.17      0.866
## Year2009         -0.0140      0.1007     -0.14      0.889
## Year2010         -0.0191      0.1112     -0.17      0.864
## Year2011          0.0950      0.1121      0.85      0.397
## Year2012         -0.0164      0.0990     -0.17      0.869
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.783
## Multiple R-squared:  0.0227, Adjusted R-squared:  0.0078
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1033 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.351  0.873  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      8.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1132"
## [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
##   26   17   16   23   18   27   21   15   24   33   25   48   45   39   37
## 2011 2012
##   44   45
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   13   11    8   13   11   13   12    8   14   17   19   27   30   29   28
## 2011 2012

```

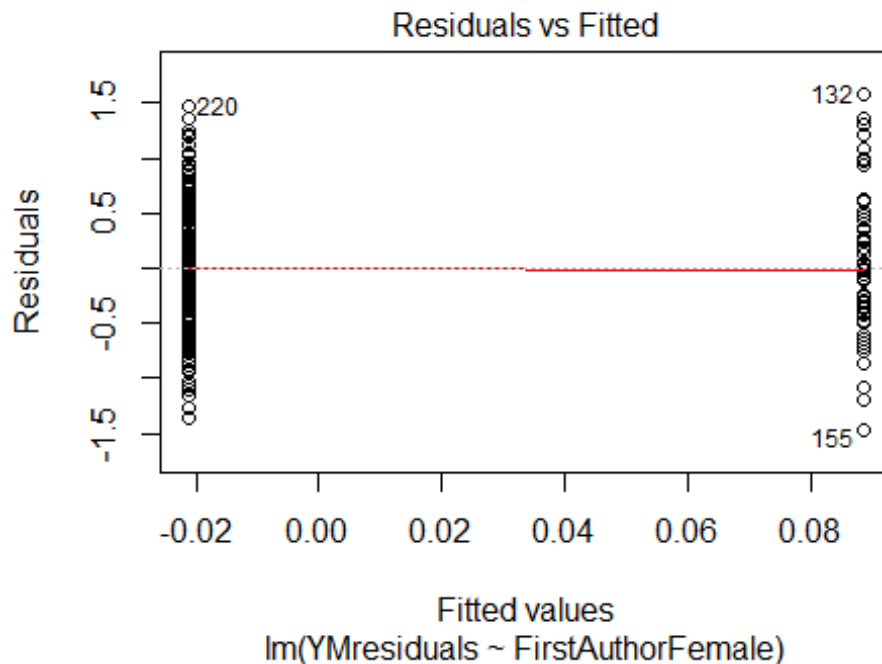
```
## 27 29
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 8 6 11 8 13 12 7 13 15 16 23 25 24 26
## 2011 2012
## 25 25
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 18, df = 16, p-value = 0.3
```



```
##
## 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: 2.02369588183562"
## [1] "Male first author team size 2018 geometric mean: 2.89235989936565"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

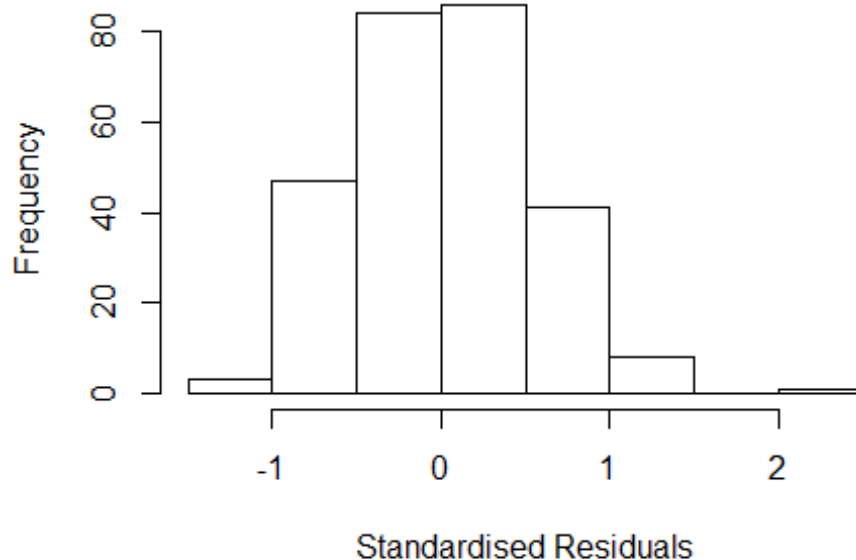
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 110, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.53746267757577"
## [1] "Male last author team size 2018 geometric mean: 3.09667884697756"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 54, p-value = 0.005
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.816 1      1.348
## LastAuthorFemale  1.715 1      1.310
## UniqueAuthors    4.111 4      1.193
## Year              6.005 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
## -1.35517 -0.38303 0.00426 0.39310 2.26729
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4745 0.1693 2.80 0.0055 **
## FirstAuthorFemale1 0.1691 0.1047 1.61 0.1077
## LastAuthorFemale1 -0.1648 0.1175 -1.40 0.1621
## UniqueAuthors2 0.4223 0.0988 4.28 2.7e-05 ***
## UniqueAuthors3 0.5843 0.1196 4.89 1.9e-06 ***
## UniqueAuthors4 0.6401 0.1217 5.26 3.1e-07 ***
## UniqueAuthors5 0.7176 0.1402 5.12 6.2e-07 ***
## Year1997 0.2035 0.1929 1.05 0.2925
## Year1998 0.2679 0.2555 1.05 0.2954
## Year1999 0.3111 0.2224 1.40 0.1631
```

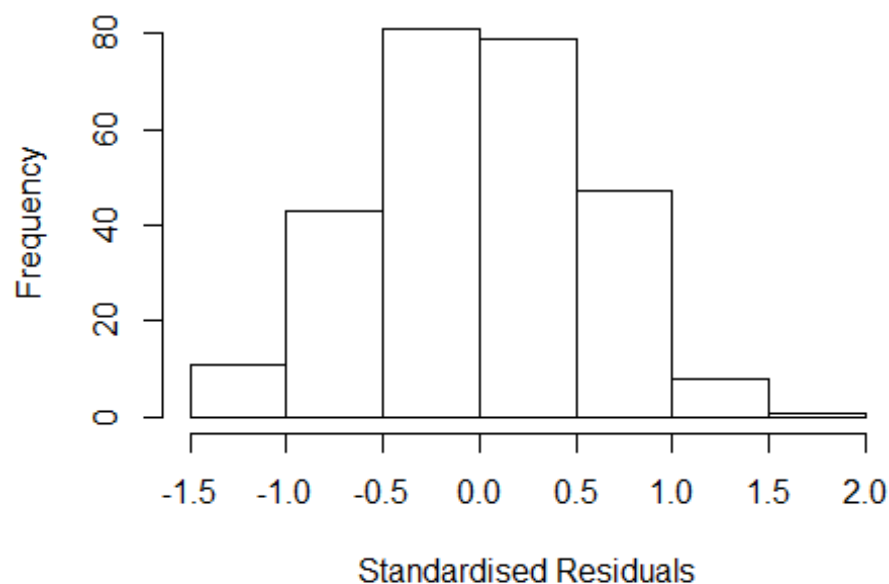


```

## Year2000          0.1773      0.1888      0.94      0.3488
## Year2001          0.2938      0.2486      1.18      0.2383
## Year2002          0.2963      0.2225      1.33      0.1842
## Year2003          0.6897      0.2733      2.52      0.0123 *
## Year2004          0.2543      0.2337      1.09      0.2776
## Year2005          0.2449      0.2203      1.11      0.2674
## Year2006          0.1272      0.1934      0.66      0.5112
## Year2007          0.0844      0.1927      0.44      0.6619
## Year2008          0.2540      0.2171      1.17      0.2432
## Year2009          0.1222      0.2026      0.60      0.5471
## Year2010          0.1301      0.2115      0.62      0.5391
## Year2011          0.2492      0.1819      1.37      0.1720
## Year2012         -0.0125      0.1867     -0.07      0.9465
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.51
## Multiple R-squared:  0.269, Adjusted R-squared:  0.204
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 249 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0097 0.8760 0.9380 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          3.70e-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 1.591 1          1.262
## LastAuthorFemale  1.601 1          1.265
## Year              1.821 16          1.019

```

Residuals from first and last author



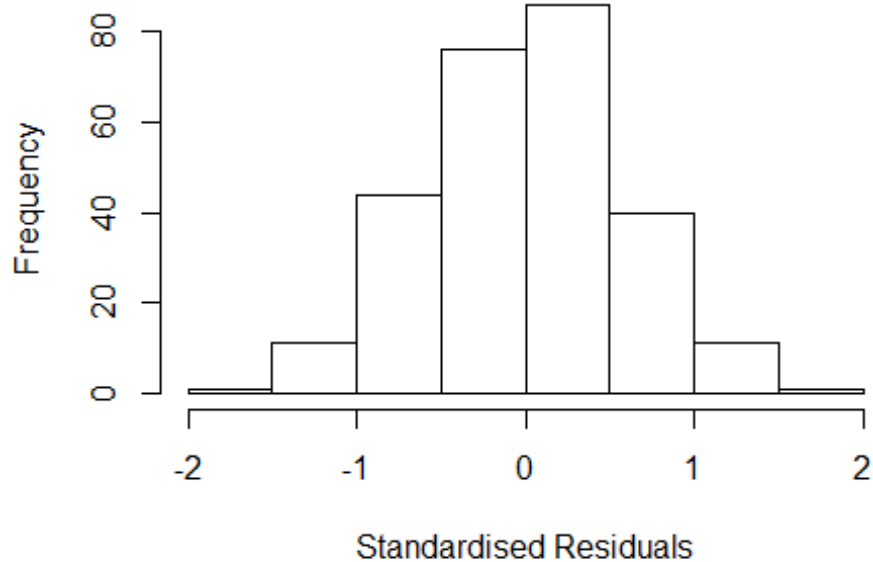
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35811 -0.38375 0.00602 0.41669 1.75839
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.692 0.175 3.95 0.0001 ***
## FirstAuthorFemale1 0.249 0.113 2.20 0.0289 *
## LastAuthorFemale1 -0.326 0.132 -2.47 0.0143 *
## Year1997 0.332 0.235 1.41 0.1601
## Year1998 0.501 0.279 1.80 0.0736 .
## Year1999 0.268 0.270 0.99 0.3226
## Year2000 0.231 0.266 0.87 0.3862
## Year2001 0.666 0.315 2.11 0.0355 *
## Year2002 0.526 0.261 2.02 0.0447 *
## Year2003 0.604 0.293 2.06 0.0401 *
## Year2004 0.525 0.260 2.02 0.0447 *
## Year2005 0.412 0.227 1.82 0.0705 .
```

```

## Year2006          0.402      0.196      2.05      0.0414 *
## Year2007          0.212      0.222      0.95      0.3413
## Year2008          0.436      0.223      1.95      0.0523 .
## Year2009          0.338      0.206      1.64      0.1024
## Year2010          0.213      0.234      0.91      0.3623
## Year2011          0.385      0.215      1.79      0.0745 .
## Year2012          0.313      0.208      1.50      0.1339
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.625
## Multiple R-squared:  0.0856, Adjusted R-squared:  0.02
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 254 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.408  0.885   0.958   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      3.70e-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 1.328 1      1.152
## Year              1.328 16      1.009

```

Residuals from first author



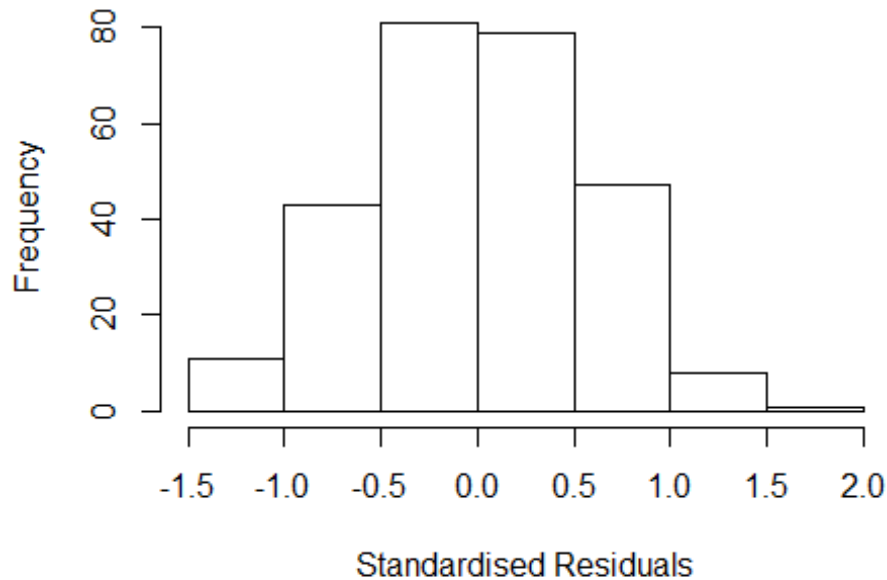
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5078 -0.3916 0.0382 0.4100 1.5322
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.693 0.175 3.97 9.5e-05 ***
## FirstAuthorFemale1 0.137 0.112 1.22 0.223
## Year1997 0.361 0.239 1.51 0.132
## Year1998 0.466 0.262 1.78 0.076 .
## Year1999 0.267 0.272 0.98 0.327
## Year2000 0.184 0.235 0.78 0.434
## Year2001 0.678 0.316 2.15 0.033 *
## Year2002 0.534 0.261 2.04 0.042 *
## Year2003 0.612 0.300 2.04 0.043 *
## Year2004 0.501 0.257 1.95 0.053 .
## Year2005 0.370 0.226 1.64 0.103
## Year2006 0.336 0.205 1.64 0.102
```

```

## Year2007          0.185      0.219      0.85      0.398
## Year2008          0.394      0.219      1.80      0.074 .
## Year2009          0.351      0.211      1.66      0.097 .
## Year2010          0.186      0.233      0.80      0.425
## Year2011          0.369      0.212      1.74      0.083 .
## Year2012          0.269      0.208      1.29      0.197
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.635
## Multiple R-squared:  0.0649, Adjusted R-squared:  0.00178
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 251 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.540  0.894  0.956  0.921  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.70e-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 1.38 1      1.175
## Year              1.38 16      1.010

```

Residuals from last author



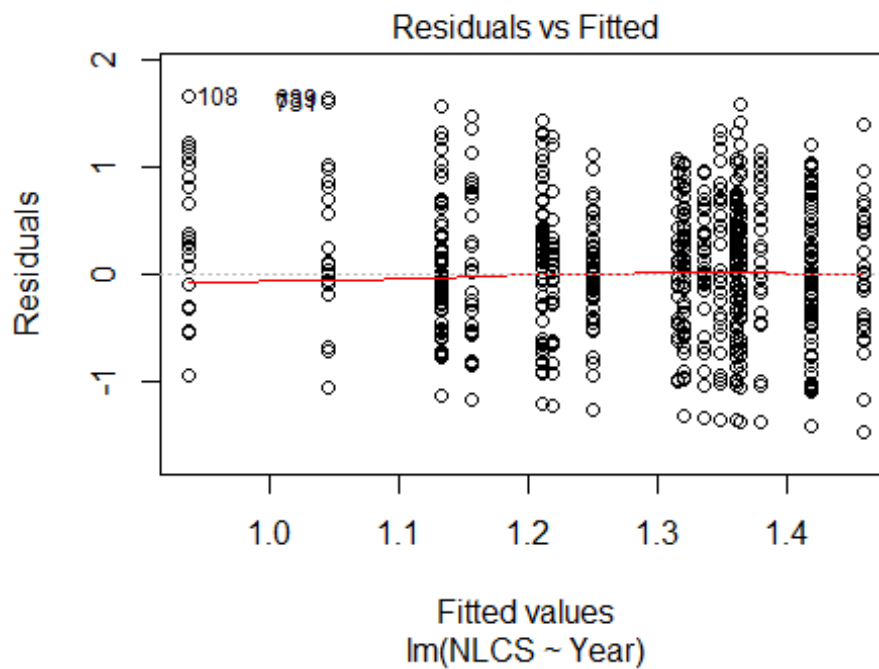
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3498 -0.3799 0.0276 0.4282 1.8963
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.691 0.175 3.95 0.0001 ***
## LastAuthorFemale1 -0.206 0.116 -1.78 0.0763 .
## Year1997 0.398 0.244 1.63 0.1038
## Year1998 0.519 0.291 1.79 0.0754 .
## Year1999 0.263 0.274 0.96 0.3389
## Year2000 0.266 0.257 1.03 0.3022
## Year2001 0.658 0.310 2.12 0.0349 *
## Year2002 0.551 0.265 2.08 0.0383 *
## Year2003 0.622 0.313 1.99 0.0478 *
## Year2004 0.538 0.261 2.06 0.0406 *
## Year2005 0.441 0.227 1.94 0.0533 .
## Year2006 0.425 0.194 2.19 0.0296 *
```

```

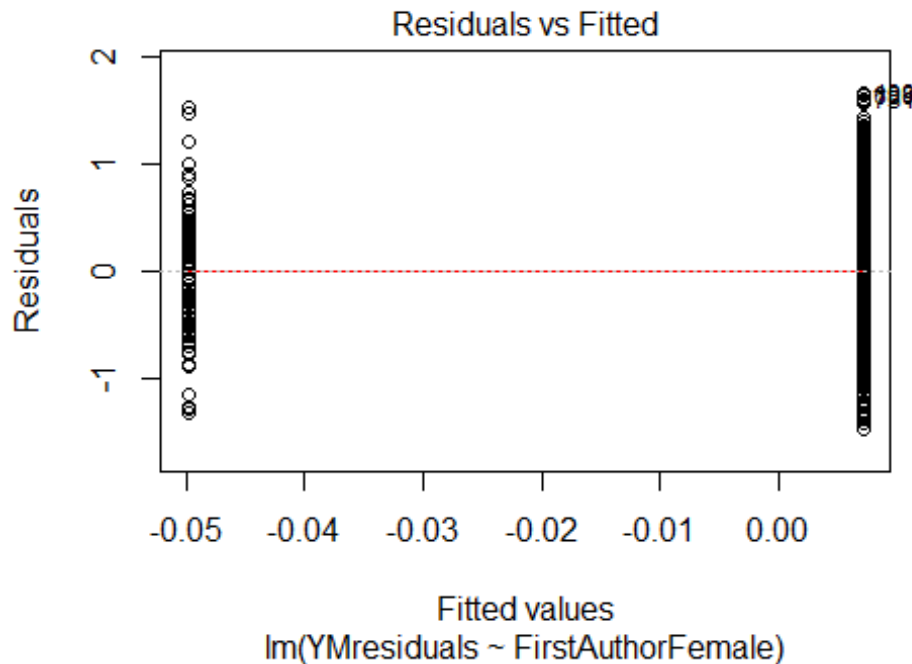
## Year2007          0.247      0.220      1.12      0.2630
## Year2008          0.467      0.221      2.11      0.0358 *
## Year2009          0.428      0.205      2.09      0.0377 *
## Year2010          0.214      0.237      0.91      0.3663
## Year2011          0.431      0.215      2.01      0.0460 *
## Year2012          0.368      0.206      1.79      0.0745 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.622
## Multiple R-squared:  0.0664, Adjusted R-squared:  0.00347
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 249 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.332  0.891  0.953  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      3.70e-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: 270"
## [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
##  111  101  112  125   82  129   79   88   77   75  103  131  133  183  169
## 2011 2012
##  145  161
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   30   32   29   32   27   27   35   33   27   36   54   61   84   89
## 2011 2012

```

```
## 59 77
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 27 24 27 23 30 21 23 29 23 19 27 43 49 62 65
## 2011 2012
## 48 55
## [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
```

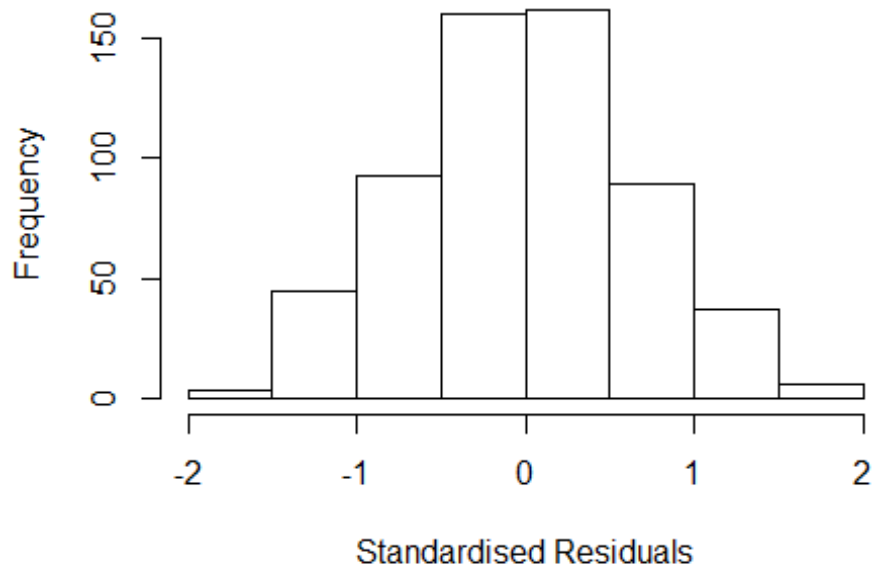


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.2, df = 1, p-value = 0.07
```

```
## [1] "Female first author team size 2018 geometric mean: 2"
## [1] "Male first author team size 2018 geometric mean: 2.36544380331277"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 21, p-value = 0.6
## 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: 2.32778574075409"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.635 1      1.279
## LastAuthorFemale  1.727 1      1.314
## UniqueAuthors     1.851 4      1.080
## Year              1.809 16     1.019
```

Residuals from first and last author and team size



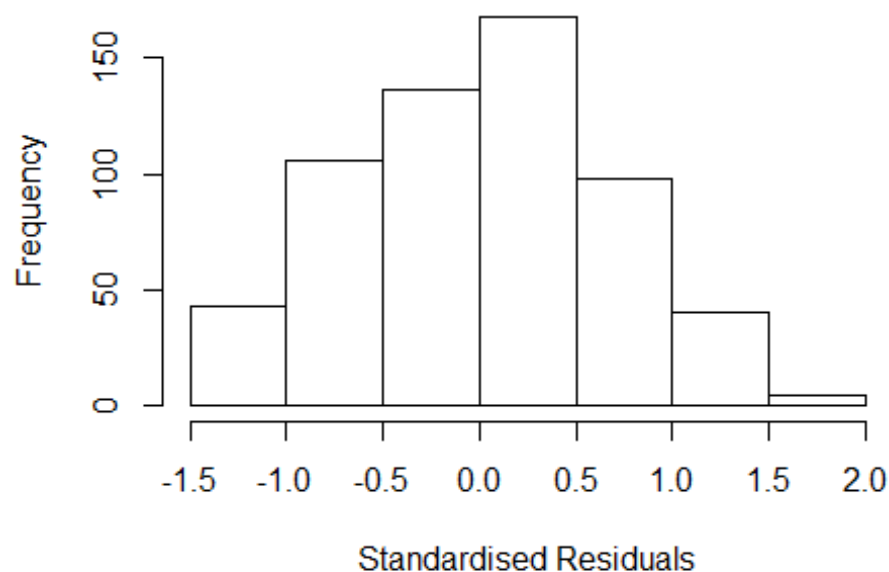
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5839 -0.4879 -0.0132 0.4486 1.7293
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5706 0.1920 2.97 0.0031 **
## FirstAuthorFemale1 -0.0446 0.0869 -0.51 0.6083
## LastAuthorFemale1 0.1432 0.1104 1.30 0.1949
## UniqueAuthors2 0.3568 0.0795 4.49 8.7e-06 ***
## UniqueAuthors3 0.4373 0.0889 4.92 1.1e-06 ***
## UniqueAuthors4 0.3150 0.1403 2.24 0.0252 *
## UniqueAuthors5 0.6413 0.1401 4.58 5.7e-06 ***
## Year1997 0.4620 0.2542 1.82 0.0697 .
## Year1998 0.5322 0.2374 2.24 0.0253 *
## Year1999 0.5678 0.2312 2.46 0.0143 *
```

```

## Year2000          0.4649      0.2390      1.95      0.0522 .
## Year2001          0.0802      0.2682      0.30      0.7650
## Year2002          0.2923      0.2219      1.32      0.1883
## Year2003          0.2997      0.2410      1.24      0.2142
## Year2004          0.0692      0.2326      0.30      0.7663
## Year2005          0.5737      0.2519      2.28      0.0232 *
## Year2006          0.4818      0.2210      2.18      0.0296 *
## Year2007          0.3301      0.2022      1.63      0.1031
## Year2008          0.4194      0.2077      2.02      0.0439 *
## Year2009          0.4297      0.2060      2.09      0.0374 *
## Year2010          0.5001      0.2037      2.45      0.0144 *
## Year2011          0.3721      0.2128      1.75      0.0809 .
## Year2012          0.1201      0.2041      0.59      0.5565
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.664
## Multiple R-squared:  0.121, Adjusted R-squared:  0.0873
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 537 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.478  0.867  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      1.68e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.384 1      1.176
## LastAuthorFemale  1.371 1      1.171
## Year              1.247 16      1.007

```

Residuals from first and last author



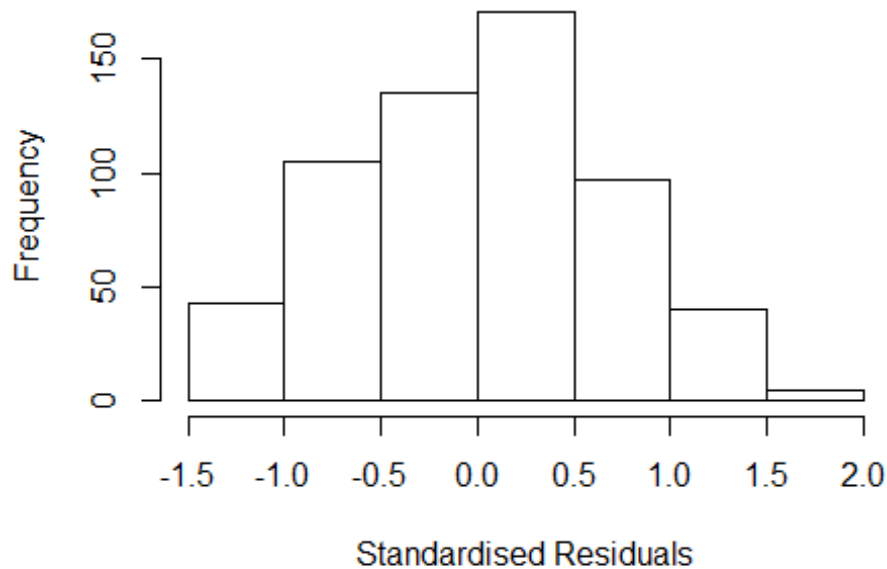
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4242 -0.4959 0.0351 0.4807 1.8314
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7646 0.1948 3.92 9.7e-05 ***
## FirstAuthorFemale1 -0.0239 0.0853 -0.28 0.7798
## LastAuthorFemale1 0.0156 0.1093 0.14 0.8866
## Year1997 0.5209 0.2648 1.97 0.0497 *
## Year1998 0.5850 0.2626 2.23 0.0263 *
## Year1999 0.5984 0.2412 2.48 0.0134 *
## Year2000 0.5711 0.2466 2.32 0.0209 *
## Year2001 0.1158 0.2885 0.40 0.6884
## Year2002 0.4055 0.2329 1.74 0.0822 .
## Year2003 0.3833 0.2437 1.57 0.1163
## Year2004 0.1473 0.2571 0.57 0.5670
## Year2005 0.6187 0.2580 2.40 0.0168 *
```

```

## Year2006          0.5898      0.2320      2.54      0.0113 *
## Year2007          0.4596      0.2108      2.18      0.0297 *
## Year2008          0.5609      0.2195      2.56      0.0109 *
## Year2009          0.5800      0.2170      2.67      0.0077 **
## Year2010          0.6596      0.2157      3.06      0.0023 **
## Year2011          0.5470      0.2223      2.46      0.0142 *
## Year2012          0.2849      0.2095      1.36      0.1745
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.685
## Multiple R-squared:  0.0612, Adjusted R-squared:  0.0318
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 540 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.455  0.862  0.942  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          1.68e-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 1.116 1      1.056
## Year              1.116 16      1.003

```

Residuals from first author



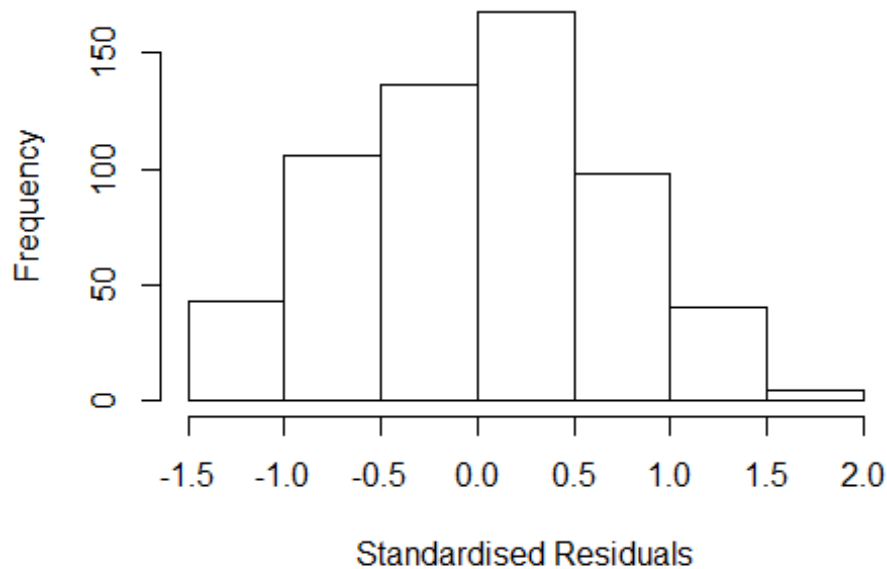
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4242 -0.4934 0.0337 0.4778 1.8296
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7664 0.1942 3.95 8.9e-05 ***
## FirstAuthorFemale1 -0.0187 0.0764 -0.25 0.8065
## Year1997 0.5191 0.2640 1.97 0.0498 *
## Year1998 0.5831 0.2614 2.23 0.0261 *
## Year1999 0.5969 0.2405 2.48 0.0133 *
## Year2000 0.5703 0.2458 2.32 0.0207 *
## Year2001 0.1160 0.2871 0.40 0.6862
## Year2002 0.4046 0.2327 1.74 0.0827 .
## Year2003 0.3829 0.2421 1.58 0.1143
## Year2004 0.1457 0.2562 0.57 0.5698
## Year2005 0.6167 0.2574 2.40 0.0169 *
## Year2006 0.5879 0.2310 2.55 0.0112 *
```

```

## Year2007          0.4584      0.2099      2.18      0.0294 *
## Year2008          0.5596      0.2185      2.56      0.0107 *
## Year2009          0.5787      0.2160      2.68      0.0076 **
## Year2010          0.6577      0.2148      3.06      0.0023 **
## Year2011          0.5462      0.2211      2.47      0.0138 *
## Year2012          0.2832      0.2090      1.36      0.1759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.691
## Multiple R-squared:  0.0607, Adjusted R-squared:  0.033
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 542 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.464  0.864  0.944  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.68e-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 1.109 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
## -1.4206 -0.4958 0.0365 0.4724 1.8315
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7645 0.1948 3.92 9.8e-05 ***
## LastAuthorFemale1 0.0042 0.0986 0.04 0.9660
## Year1997 0.5210 0.2647 1.97 0.0495 *
## Year1998 0.5853 0.2622 2.23 0.0260 *
## Year1999 0.5936 0.2402 2.47 0.0138 *
## Year2000 0.5678 0.2459 2.31 0.0213 *
## Year2001 0.1163 0.2878 0.40 0.6862
## Year2002 0.4035 0.2330 1.73 0.0838 .
## Year2003 0.3820 0.2435 1.57 0.1173
## Year2004 0.1448 0.2544 0.57 0.5695
## Year2005 0.6174 0.2577 2.40 0.0169 *
## Year2006 0.5892 0.2318 2.54 0.0113 *
```

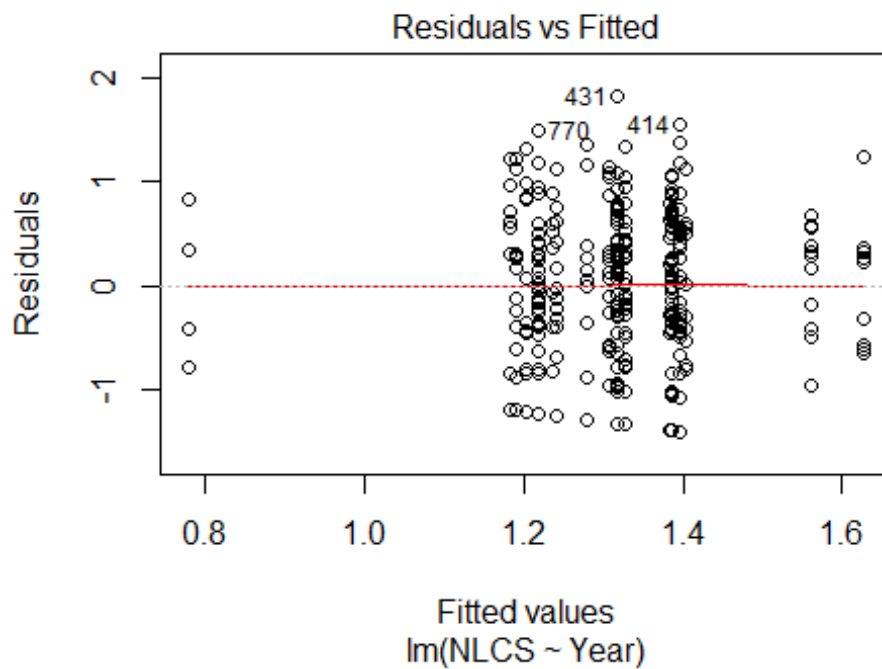


```

## Year2007          0.4586      0.2109      2.17      0.0300 *
## Year2008          0.5572      0.2200      2.53      0.0116 *
## Year2009          0.5791      0.2169      2.67      0.0078 **
## Year2010          0.6561      0.2151      3.05      0.0024 **
## Year2011          0.5430      0.2223      2.44      0.0149 *
## Year2012          0.2808      0.2087      1.35      0.1790
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.688
## Multiple R-squared:  0.0609, Adjusted R-squared:  0.0332
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 542 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.458  0.861  0.943  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      1.68e-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: 595"
## [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
##   23   19   23   36   18   15   19   19   20   30   28   37   64   95   87
## 2011 2012
##   67   74
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   10   11   11   18   10    4   12   10    7   11   12   13   36   50   46
## 2011 2012

```

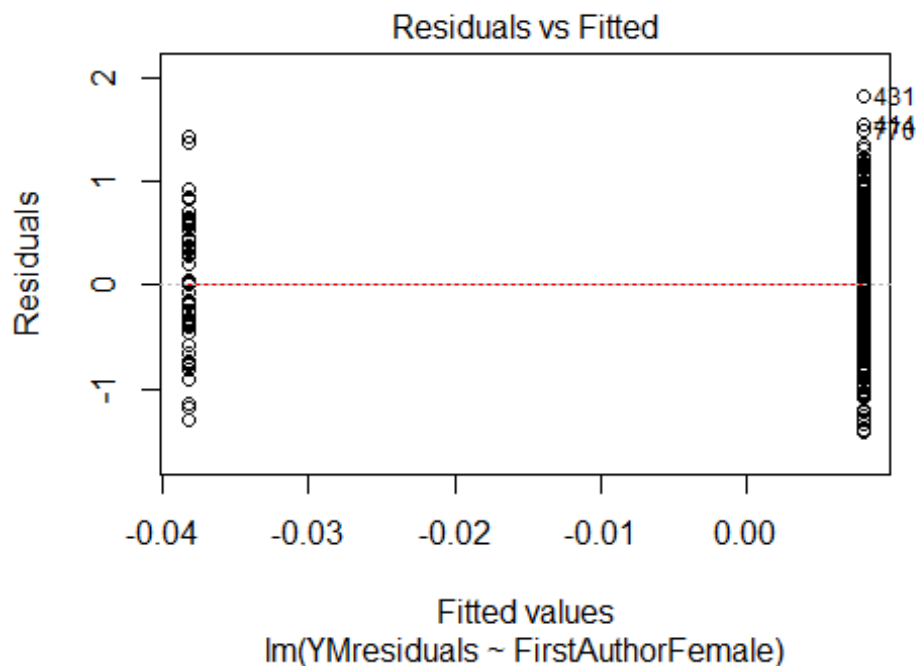
```
## 36 35
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 11 11 14 9 4 8 8 5 10 7 9 28 36 29
## 2011 2012
## 31 26
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 7.2, df = 16, p-value = 1
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1, df = 1, p-value = 0.3
## [1] "Female first author team size 2018 geometric mean: 2.20817902734762"
## [1] "Male first author team size 2018 geometric mean: 2.69736673839921"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

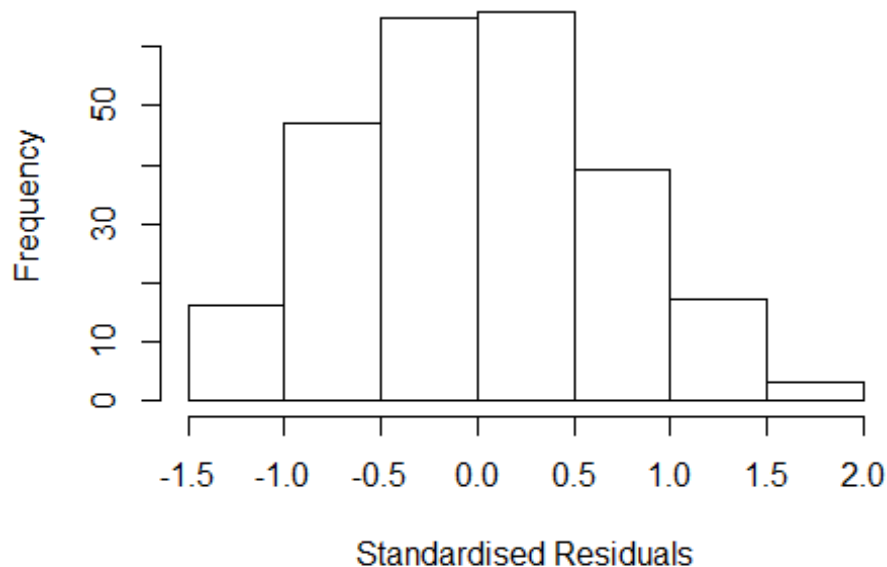
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 86, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.63214802590498"
## [1] "Male last author team size 2018 geometric mean: 2.61092541793714"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 81, 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 1.200  1      1.095
## LastAuthorFemale  1.484  1      1.218
## UniqueAuthors    2.555  4      1.124
## Year              3.480 16      1.040
```

Residuals from first and last author and team size



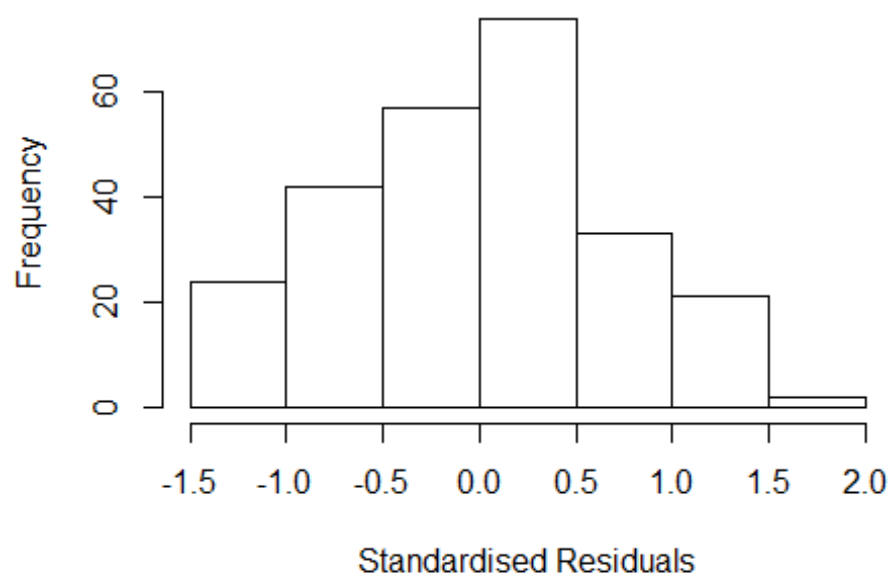
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4804 -0.4902 -0.0133 0.4470 1.8972
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.37138 0.41241 0.90 0.3688
## FirstAuthorFemale1 -0.03380 0.10837 -0.31 0.7554
## LastAuthorFemale1 -0.00173 0.16105 -0.01 0.9914
## UniqueAuthors2 0.67374 0.16004 4.21 3.7e-05 ***
## UniqueAuthors3 0.70253 0.15806 4.44 1.4e-05 ***
## UniqueAuthors4 0.68307 0.19001 3.59 0.0004 ***
## UniqueAuthors5 0.51725 0.20487 2.52 0.0123 *
## Year1997 0.58111 0.49759 1.17 0.2441
## Year1998 0.21432 0.45951 0.47 0.6414
## Year1999 0.57753 0.43232 1.34 0.1829
```

```

## Year2000      0.75585      0.48523      1.56      0.1207
## Year2001     -0.09610      0.47349     -0.20      0.8393
## Year2002      0.18277      0.40270      0.45      0.6503
## Year2003      0.62561      0.39748      1.57      0.1169
## Year2004      0.47432      0.46489      1.02      0.3087
## Year2005      0.61353      0.41395      1.48      0.1397
## Year2006      0.85119      0.40955      2.08      0.0388 *
## Year2007      0.20077      0.44265      0.45      0.6506
## Year2008      0.43524      0.39430      1.10      0.2708
## Year2009      0.19772      0.38212      0.52      0.6054
## Year2010      0.39443      0.39226      1.01      0.3157
## Year2011      0.37535      0.38522      0.97      0.3309
## Year2012      0.22367      0.38403      0.58      0.5608
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.704
## Multiple R-squared:  0.156, Adjusted R-squared:  0.0752
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 238 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.448  0.884  0.953   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      3.95e-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 1.186 1      1.089
## LastAuthorFemale  1.233 1      1.110
## Year              1.458 16      1.012

```

Residuals from first and last author

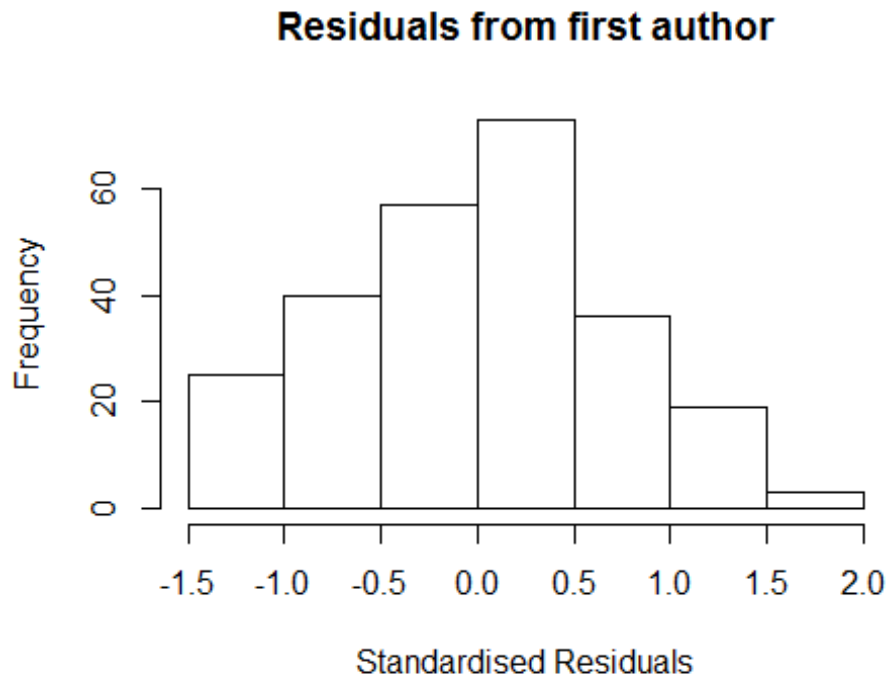


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4267 -0.5183 0.0112 0.4379 1.9829
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0062 0.2735 3.68 0.00029 ***
## FirstAuthorFemale1 0.0147 0.1151 0.13 0.89860
## LastAuthorFemale1 -0.0821 0.1649 -0.50 0.61907
## Year1997 0.2610 0.3754 0.70 0.48762
## Year1998 0.1896 0.4073 0.47 0.64203
## Year1999 0.2724 0.3376 0.81 0.42056
## Year2000 0.3834 0.4603 0.83 0.40577
## Year2001 -0.1916 0.4510 -0.42 0.67142
## Year2002 0.0496 0.3628 0.14 0.89141
## Year2003 0.5819 0.3475 1.67 0.09538 .
## Year2004 0.4205 0.4588 0.92 0.36030
## Year2005 0.5735 0.3316 1.73 0.08503 .
```

```

## Year2006          0.8607      0.3285      2.62  0.00937 **
## Year2007          0.1636      0.3566      0.46  0.64678
## Year2008          0.3952      0.3190      1.24  0.21667
## Year2009          0.1509      0.3011      0.50  0.61676
## Year2010          0.3060      0.3263      0.94  0.34936
## Year2011          0.3491      0.3060      1.14  0.25506
## Year2012          0.2033      0.2967      0.69  0.49380
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.757
## Multiple R-squared:  0.0538, Adjusted R-squared:  -0.019
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 232 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.472  0.876  0.957  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      3.95e-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 1.187 1      1.089
## Year              1.187 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 ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4127 -0.5099 0.0193 0.4331 1.9872
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00627 0.27343 3.68 0.00029 ***
## FirstAuthorFemale1 0.00481 0.11618 0.04 0.96703
## Year1997 0.26094 0.37531 0.70 0.48758
## Year1998 0.17040 0.40284 0.42 0.67268
## Year1999 0.26779 0.33641 0.80 0.42682
## Year2000 0.38615 0.46034 0.84 0.40241
## Year2001 -0.23111 0.44470 -0.52 0.60376
## Year2002 0.05251 0.36286 0.14 0.88507
## Year2003 0.56471 0.35347 1.60 0.11147
## Year2004 0.40643 0.45032 0.90 0.36769
## Year2005 0.56267 0.33148 1.70 0.09094 .
## Year2006 0.86212 0.32840 2.63 0.00923 **
```

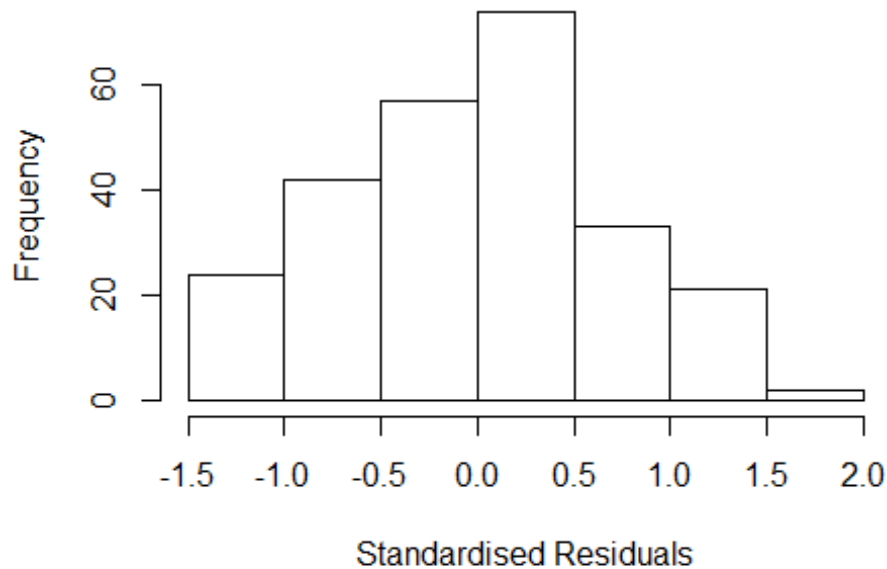


```

## Year2007          0.15521    0.35356    0.44  0.66106
## Year2008          0.38425    0.31781    1.21  0.22784
## Year2009          0.14652    0.30036    0.49  0.62613
## Year2010          0.30406    0.32613    0.93  0.35213
## Year2011          0.34060    0.30567    1.11  0.26631
## Year2012          0.20066    0.29745    0.67  0.50059
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.758
## Multiple R-squared:  0.0531, Adjusted R-squared:  -0.0154
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 235 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.471  0.875   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      3.95e-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 1.225  1          1.107
## Year              1.225 16          1.006

```

Residuals from last author



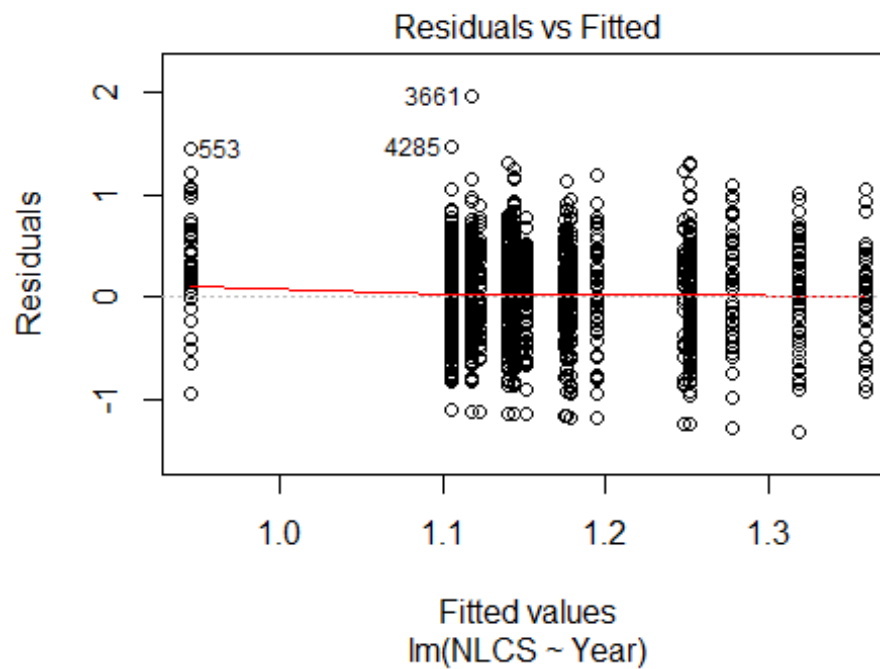
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4296 -0.5228 0.0192 0.4365 1.9815
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0059 0.2738 3.67 0.0003 ***
## LastAuthorFemale1 -0.0792 0.1653 -0.48 0.6324
## Year1997 0.2612 0.3758 0.70 0.4877
## Year1998 0.1903 0.4074 0.47 0.6410
## Year1999 0.2759 0.3357 0.82 0.4120
## Year2000 0.3893 0.4553 0.86 0.3934
## Year2001 -0.1894 0.4497 -0.42 0.6740
## Year2002 0.0540 0.3591 0.15 0.8805
## Year2003 0.5830 0.3475 1.68 0.0948 .
## Year2004 0.4238 0.4604 0.92 0.3583
## Year2005 0.5748 0.3310 1.74 0.0838 .
## Year2006 0.8638 0.3275 2.64 0.0089 **
```

```

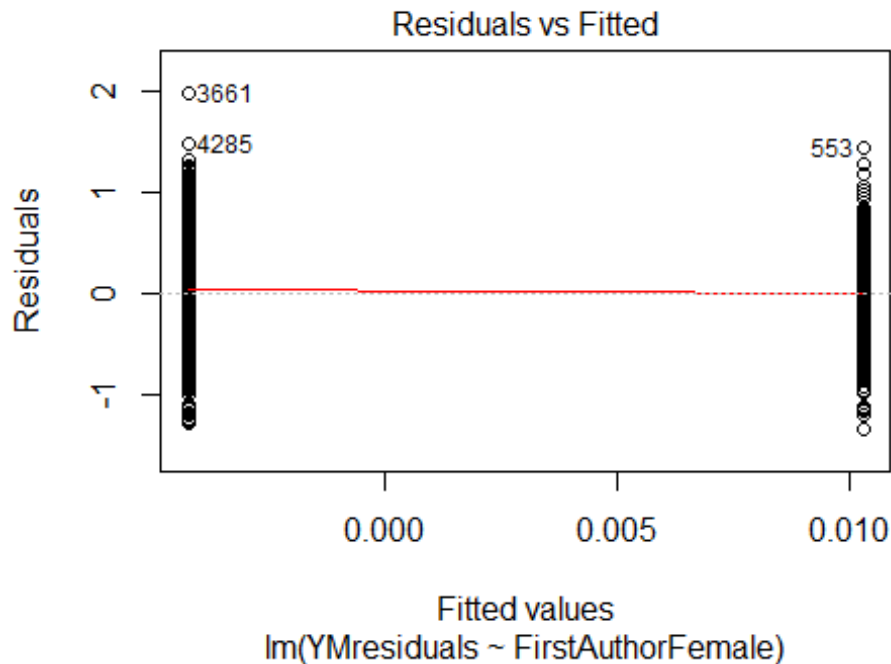
## Year2007          0.1672      0.3544      0.47      0.6375
## Year2008          0.3984      0.3182      1.25      0.2118
## Year2009          0.1527      0.3008      0.51      0.6123
## Year2010          0.3094      0.3248      0.95      0.3417
## Year2011          0.3539      0.3033      1.17      0.2444
## Year2012          0.2059      0.2966      0.69      0.4882
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.753
## Multiple R-squared:  0.0539, Adjusted R-squared:  -0.0145
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 233 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.469  0.876  0.957   0.916  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.95e-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: 253"
## [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
## 104 106 103 94 159 105 132 98 125 156 186 232 245 336 356
## 2011 2012
## 412 433
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 39 48 50 47 68 38 63 60 77 88 105 148 139 213 220
## 2011 2012

```

```
## 262 291
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 29 41 40 36 55 36 56 46 58 67 83 116 116 173 182
## 2011 2012
## 205 234
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 61, df = 16, p-value = 4e-07
```

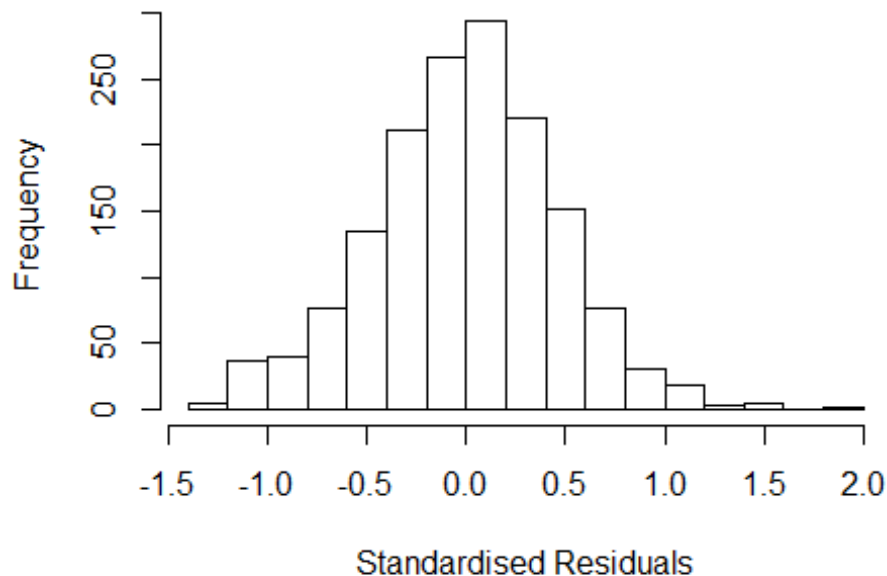


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 12, df = 1, p-value = 5e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.03130302680607"
## [1] "Male first author team size 2018 geometric mean: 4.00633017689752"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5800, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.69288503761937"
## [1] "Male last author team size 2018 geometric mean: 4.10408325722017"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4000, 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 1.109 1      1.053
## LastAuthorFemale  1.066 1      1.032
## UniqueAuthors    1.335 4      1.037
## Year              1.469 16     1.012
```

Residuals from first and last author and team size



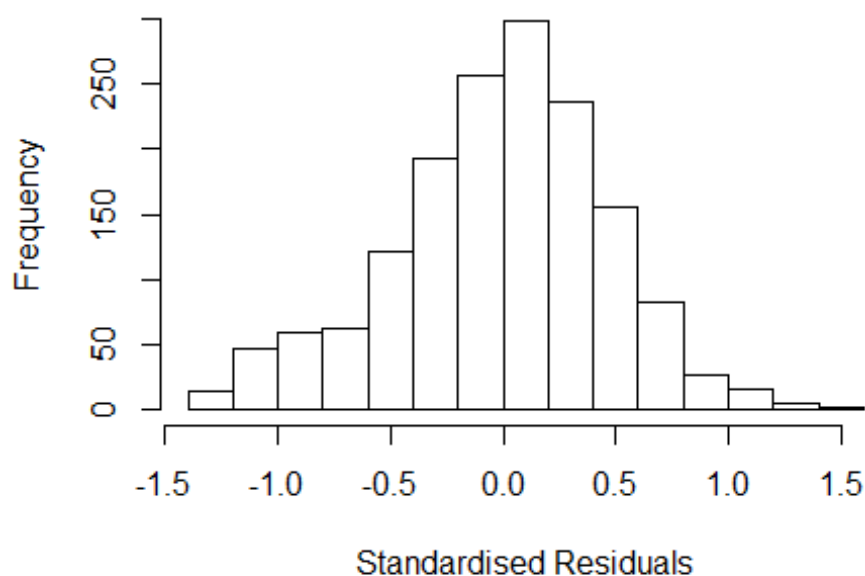
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3576 -0.2921 0.0134 0.2893 1.8233
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8514 0.1364 6.24 5.5e-10 ***
## FirstAuthorFemale1 -0.0208 0.0251 -0.83 0.4070
## LastAuthorFemale1 0.0219 0.0278 0.79 0.4314
## UniqueAuthors2 0.5376 0.0931 5.78 9.2e-09 ***
## UniqueAuthors3 0.5010 0.0935 5.36 9.5e-08 ***
## UniqueAuthors4 0.5752 0.0926 6.21 6.8e-10 ***
## UniqueAuthors5 0.5538 0.0913 6.07 1.6e-09 ***
## Year1997 -0.0879 0.1368 -0.64 0.5207
## Year1998 -0.0271 0.1269 -0.21 0.8311
## Year1999 -0.1736 0.1150 -1.51 0.1314
```

```

## Year2000          -0.3683      0.1594    -2.31    0.0210 *
## Year2001          -0.0555      0.1215    -0.46    0.6477
## Year2002          -0.1769      0.1174    -1.51    0.1322
## Year2003          -0.2661      0.1124    -2.37    0.0180 *
## Year2004          -0.1327      0.1054    -1.26    0.2086
## Year2005          -0.1516      0.1050    -1.44    0.1488
## Year2006          -0.1663      0.1025    -1.62    0.1049
## Year2007          -0.0689      0.1037    -0.66    0.5063
## Year2008          -0.2129      0.1020    -2.09    0.0370 *
## Year2009          -0.2002      0.0988    -2.03    0.0430 *
## Year2010          -0.2277      0.0969    -2.35    0.0189 *
## Year2011          -0.2642      0.0968    -2.73    0.0064 **
## Year2012          -0.2596      0.0971    -2.67    0.0076 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.0993, Adjusted R-squared:  0.0865
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1444 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0312 0.8570 0.9490 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.36e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.090 1 1.044
## LastAuthorFemale 1.048 1 1.024
## 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
## -1.301 -0.296 0.026 0.290 1.436
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3538 0.0877 15.44 <2e-16 ***
## FirstAuthorFemale1 -0.0137 0.0257 -0.53 0.5936
## LastAuthorFemale1 0.0130 0.0281 0.46 0.6446
## Year1997 -0.0794 0.1330 -0.60 0.5506
## Year1998 -0.0540 0.1192 -0.45 0.6505
## Year1999 -0.1622 0.1057 -1.53 0.1250
## Year2000 -0.3837 0.1821 -2.11 0.0353 *
## Year2001 -0.0528 0.1277 -0.41 0.6793
## Year2002 -0.1514 0.1227 -1.23 0.2177
## Year2003 -0.2389 0.1089 -2.19 0.0284 *
## Year2004 -0.1206 0.1010 -1.19 0.2325
## Year2005 -0.1416 0.1015 -1.40 0.1631
```

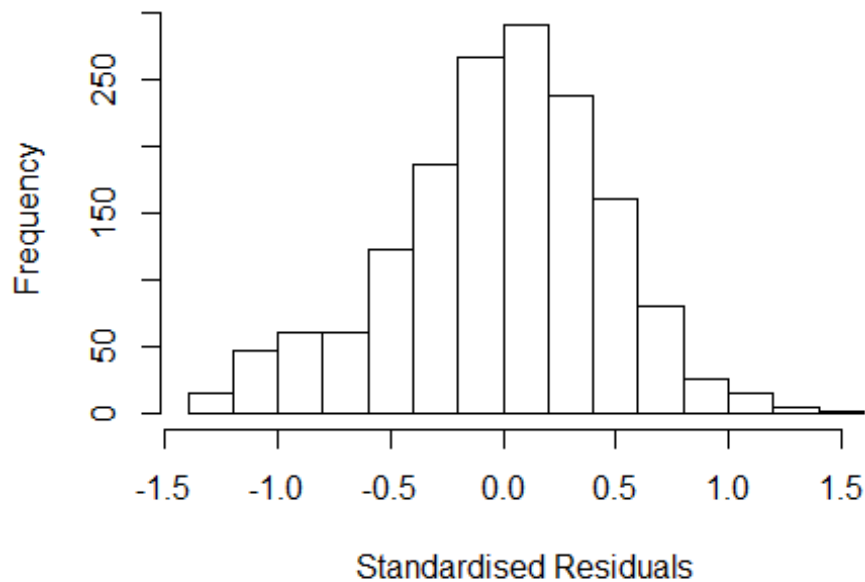


```

## Year2006          -0.1436      0.0992    -1.45    0.1478
## Year2007          -0.0647      0.0995    -0.65    0.5157
## Year2008          -0.1931      0.0988    -1.95    0.0508 .
## Year2009          -0.1879      0.0947    -1.98    0.0473 *
## Year2010          -0.1968      0.0927    -2.12    0.0339 *
## Year2011          -0.2393      0.0930    -2.57    0.0102 *
## Year2012          -0.2435      0.0929    -2.62    0.0089 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.0244, Adjusted R-squared:  0.0131
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1452 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.262  0.860  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      6.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.073 1      1.036
## Year              1.073 16      1.002

```

Residuals from first author



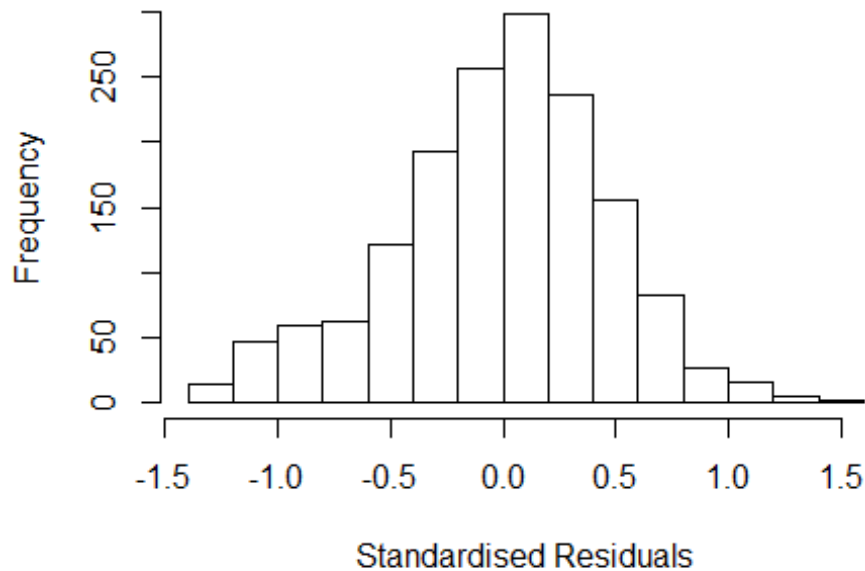
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3038 -0.2965 0.0239 0.2905 1.4313
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3565 0.0877 15.47 <2e-16 ***
## FirstAuthorFemale1 -0.0117 0.0255 -0.46 0.6456
## Year1997 -0.0804 0.1330 -0.60 0.5454
## Year1998 -0.0546 0.1194 -0.46 0.6478
## Year1999 -0.1638 0.1058 -1.55 0.1219
## Year2000 -0.3841 0.1819 -2.11 0.0349 *
## Year2001 -0.0528 0.1278 -0.41 0.6797
## Year2002 -0.1520 0.1230 -1.24 0.2166
## Year2003 -0.2391 0.1090 -2.19 0.0284 *
## Year2004 -0.1205 0.1011 -1.19 0.2334
## Year2005 -0.1420 0.1016 -1.40 0.1625
## Year2006 -0.1444 0.0993 -1.45 0.1462
```

```

## Year2007          -0.0662      0.0996   -0.67   0.5060
## Year2008          -0.1945      0.0989   -1.97   0.0495 *
## Year2009          -0.1876      0.0948   -1.98   0.0479 *
## Year2010          -0.1974      0.0928   -2.13   0.0336 *
## Year2011          -0.2393      0.0931   -2.57   0.0103 *
## Year2012          -0.2439      0.0930   -2.62   0.0088 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0243, Adjusted R-squared:  0.0136
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1452 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.264  0.861  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      6.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.033 1          1.016
## Year            1.033 16          1.001

```

Residuals from last author



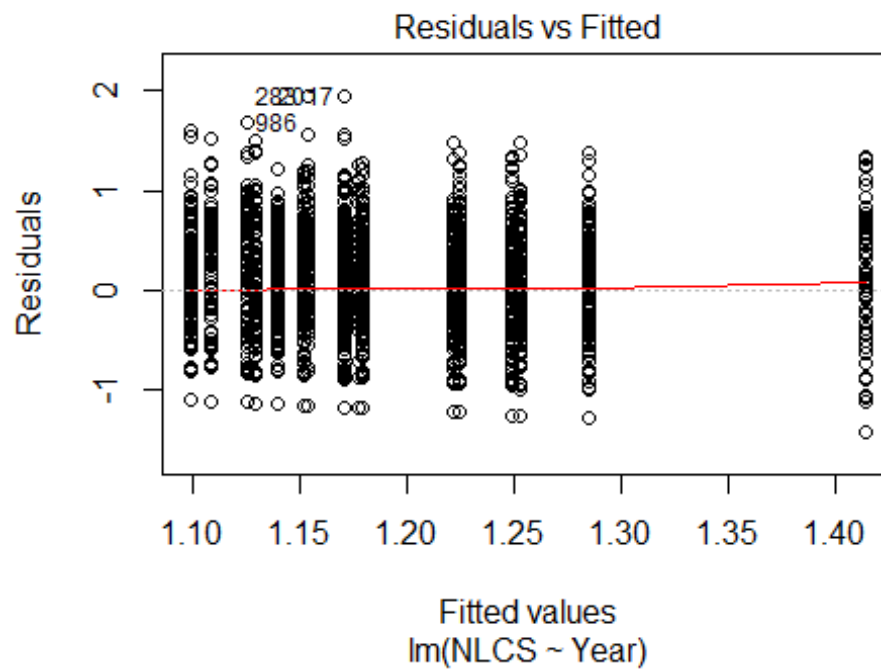
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3065 -0.2964  0.0218  0.2913  1.4252
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3520     0.0877   15.42  <2e-16 ***
## LastAuthorFemale1  0.0103     0.0280    0.37   0.7137
## Year1997         -0.0814     0.1329   -0.61   0.5401
## Year1998         -0.0558     0.1195   -0.47   0.6406
## Year1999         -0.1616     0.1059   -1.53   0.1273
## Year2000         -0.3852     0.1810   -2.13   0.0335 *
## Year2001         -0.0555     0.1276   -0.43   0.6636
## Year2002         -0.1530     0.1229   -1.25   0.2133
## Year2003         -0.2400     0.1092   -2.20   0.0281 *
## Year2004         -0.1243     0.1009   -1.23   0.2179
## Year2005         -0.1427     0.1016   -1.41   0.1602
## Year2006         -0.1478     0.0989   -1.49   0.1353
```

```

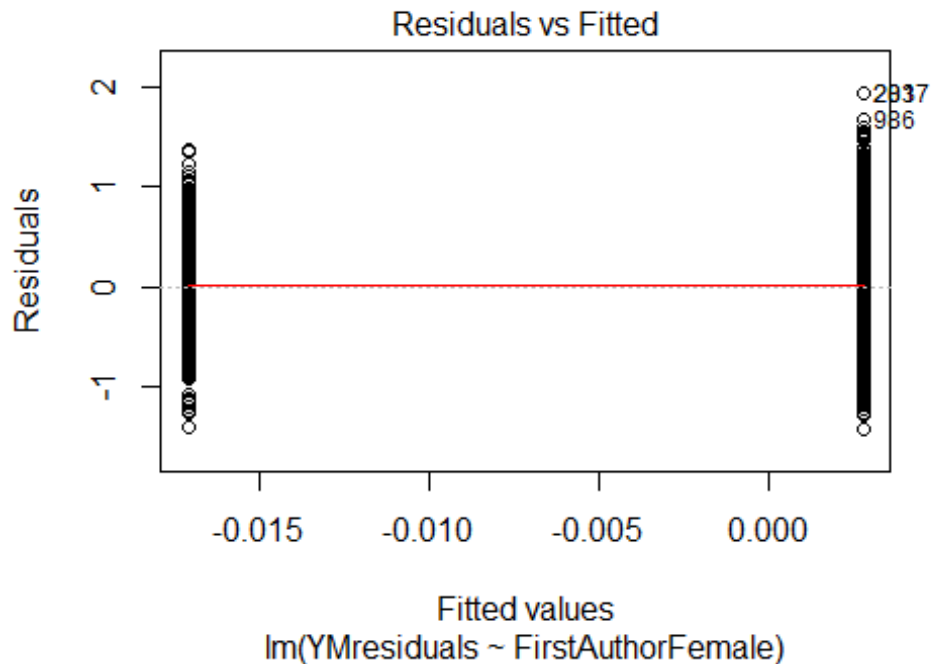
## Year2007          -0.0681      0.0993   -0.69    0.4931
## Year2008          -0.1956      0.0990   -1.98    0.0484 *
## Year2009          -0.1899      0.0948   -2.00    0.0454 *
## Year2010          -0.1991      0.0928   -2.15    0.0320 *
## Year2011          -0.2421      0.0930   -2.60    0.0093 **
## Year2012          -0.2453      0.0931   -2.64    0.0085 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.441
## Multiple R-squared:  0.0242, Adjusted R-squared:  0.0136
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1451 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.274  0.863  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      6.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1573"
## [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
##  302  216  276  279  269  345  228  248  219  247  311  311  312  355  392
## 2011 2012
##  368  420
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  138  103  123  120   81   69  105  133  110  126  138  153  159  182  193
## 2011 2012

```

```
## 216 234
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 112 85 105 98 68 54 85 110 90 99 111 117 129 151 161
## 2011 2012
## 186 204
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 69, df = 16, p-value = 1e-08
```

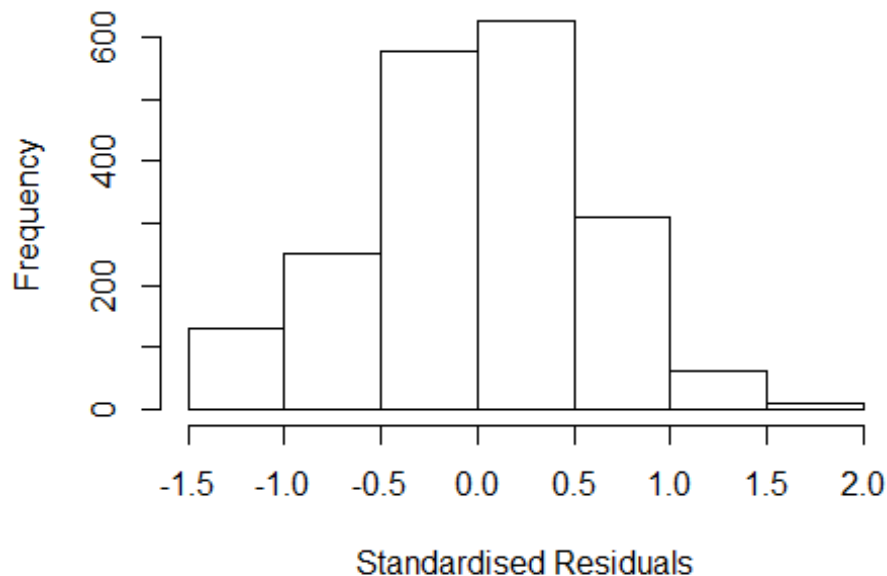


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.21, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 2.76961772165842"
## [1] "Male first author team size 2018 geometric mean: 2.55672341276643"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7500, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.52798661360767"
## [1] "Male last author team size 2018 geometric mean: 2.60343355337333"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4200, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1          1.021
## LastAuthorFemale  1.061 1          1.030
## UniqueAuthors    1.200 4          1.023
## Year              1.236 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
## -1.4952 -0.3810 0.0126 0.4070 1.8949
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08530 0.07030 15.44 <2e-16 ***
## FirstAuthorFemale1 0.00507 0.04021 0.13 0.8997
## LastAuthorFemale1 -0.03966 0.04883 -0.81 0.4168
## UniqueAuthors2 0.11380 0.04821 2.36 0.0183 *
## UniqueAuthors3 0.12950 0.04958 2.61 0.0091 **
## UniqueAuthors4 0.05825 0.06374 0.91 0.3609
## UniqueAuthors5 0.12528 0.06841 1.83 0.0672 .
## Year1997 -0.07859 0.09922 -0.79 0.4284
## Year1998 0.06905 0.09024 0.77 0.4443
## Year1999 -0.05050 0.10752 -0.47 0.6387
```

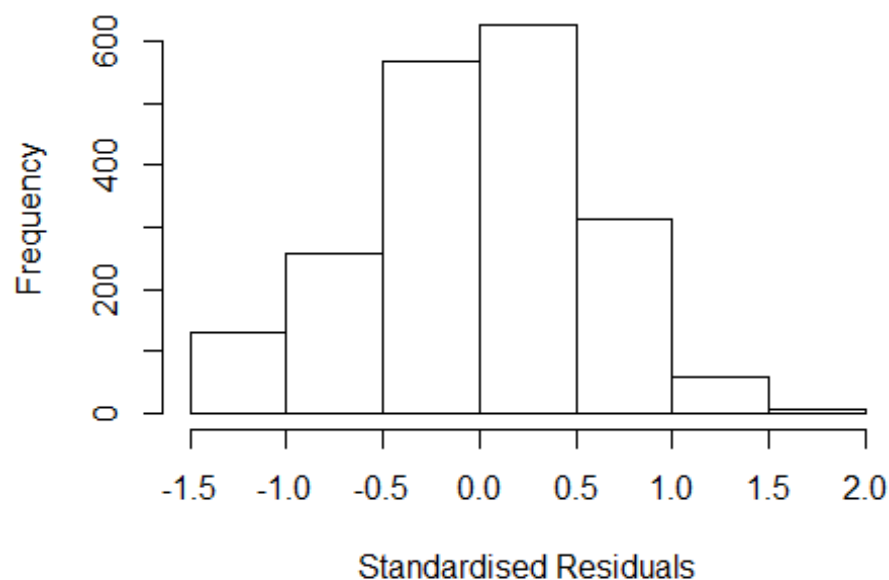


```

## Year2000      0.04314      0.11029      0.39      0.6958
## Year2001      0.27537      0.12334      2.23      0.0257 *
## Year2002     -0.07889      0.08885     -0.89      0.3747
## Year2003      0.11027      0.08400      1.31      0.1894
## Year2004     -0.02621      0.09301     -0.28      0.7781
## Year2005      0.00489      0.08168      0.06      0.9522
## Year2006      0.06414      0.08225      0.78      0.4356
## Year2007      0.06417      0.07804      0.82      0.4111
## Year2008     -0.04245      0.07951     -0.53      0.5934
## Year2009     -0.07984      0.07340     -1.09      0.2769
## Year2010      0.05137      0.07447      0.69      0.4904
## Year2011     -0.03403      0.06927     -0.49      0.6233
## Year2012     -0.02115      0.06971     -0.30      0.7617
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.577
## Multiple R-squared:  0.0209, Adjusted R-squared:  0.00985
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 162 weights are ~= 1. The remaining 1803 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.259  0.864  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      5.09e-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 1.023 1      1.011
## LastAuthorFemale  1.047 1      1.023
## Year              1.062 16      1.002

```

Residuals from first and last author



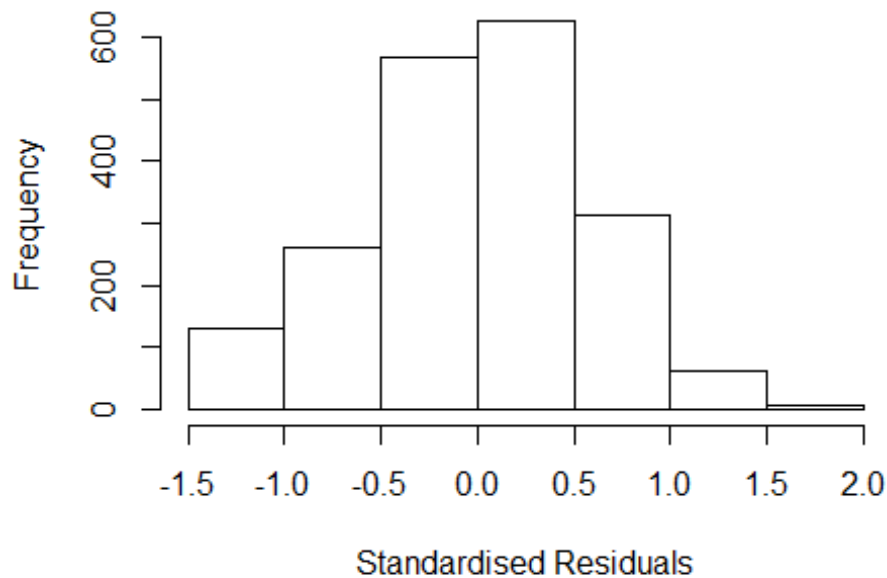
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4602 -0.3838 0.0199 0.3999 1.9206
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17343 0.05964 19.68 <2e-16 ***
## FirstAuthorFemale1 0.00905 0.03999 0.23 0.821
## LastAuthorFemale1 -0.04214 0.04839 -0.87 0.384
## Year1997 -0.08761 0.09970 -0.88 0.380
## Year1998 0.07008 0.09009 0.78 0.437
## Year1999 -0.05263 0.10695 -0.49 0.623
## Year2000 0.05686 0.11001 0.52 0.605
## Year2001 0.27768 0.12374 2.24 0.025 *
## Year2002 -0.07596 0.08900 -0.85 0.393
## Year2003 0.11236 0.08460 1.33 0.184
## Year2004 -0.01324 0.09340 -0.14 0.887
## Year2005 0.01349 0.08173 0.17 0.869
```

```

## Year2006          0.07305    0.08308    0.88    0.379
## Year2007          0.07487    0.07772    0.96    0.335
## Year2008         -0.03426    0.07916   -0.43    0.665
## Year2009         -0.07115    0.07375   -0.96    0.335
## Year2010          0.06413    0.07484    0.86    0.392
## Year2011         -0.01800    0.06924   -0.26    0.795
## Year2012         -0.00341    0.06959   -0.05    0.961
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.576
## Multiple R-squared:  0.015, Adjusted R-squared:  0.00593
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 156 weights are ~= 1. The remaining 1809 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.244  0.863  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      5.09e-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 1.018 1      1.009
## Year              1.018 16      1.001

```

Residuals from first author



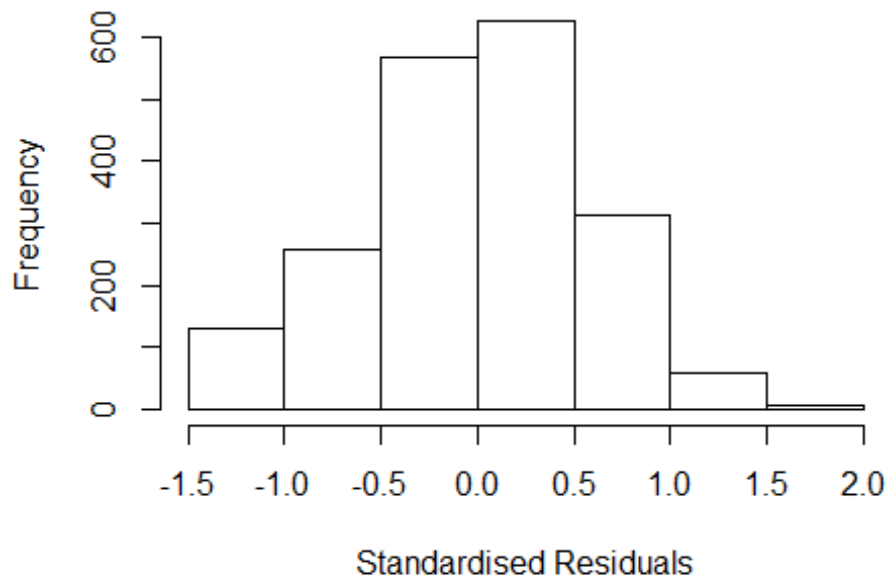
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4553 -0.3838 0.0245 0.4002 1.9229
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17115 0.05961 19.65 <2e-16 ***
## FirstAuthorFemale1 0.00483 0.03995 0.12 0.904
## Year1997 -0.08841 0.09964 -0.89 0.375
## Year1998 0.06938 0.09005 0.77 0.441
## Year1999 -0.05611 0.10731 -0.52 0.601
## Year2000 0.05624 0.10974 0.51 0.608
## Year2001 0.27933 0.12371 2.26 0.024 *
## Year2002 -0.07544 0.08900 -0.85 0.397
## Year2003 0.10985 0.08447 1.30 0.194
## Year2004 -0.01100 0.09320 -0.12 0.906
## Year2005 0.01252 0.08159 0.15 0.878
## Year2006 0.07119 0.08337 0.85 0.393
```

```

## Year2007          0.07427    0.07778    0.95    0.340
## Year2008          -0.03661    0.07918   -0.46    0.644
## Year2009          -0.07233    0.07370   -0.98    0.326
## Year2010           0.06125    0.07476    0.82    0.413
## Year2011          -0.01936    0.06913   -0.28    0.779
## Year2012          -0.00538    0.06950   -0.08    0.938
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.576
## Multiple R-squared:  0.0146, Adjusted R-squared:  0.00604
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 161 weights are ~= 1. The remaining 1804 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.242  0.861  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      5.09e-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 1.042 1          1.021
## Year            1.042 16          1.001

```

Residuals from last author



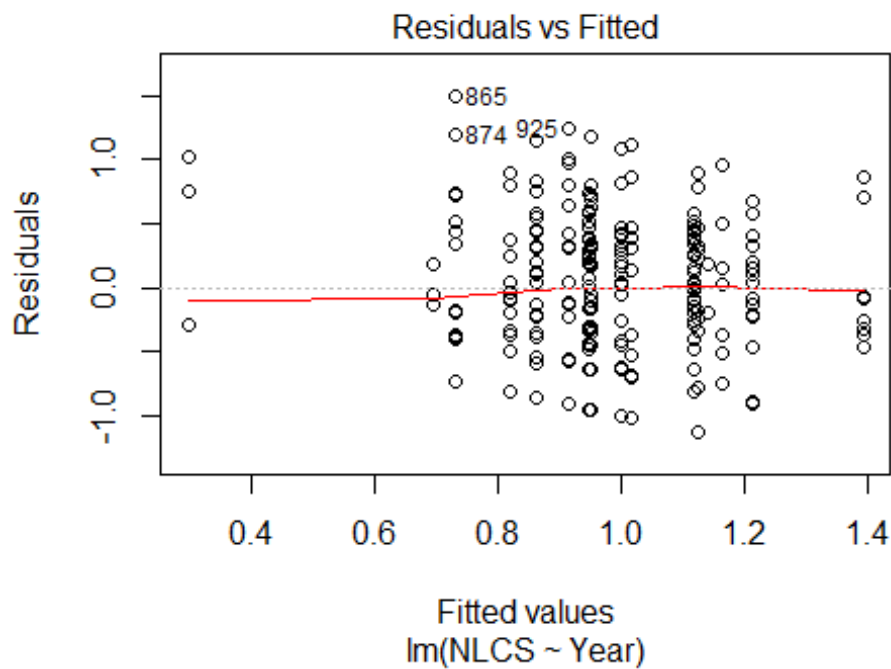
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4524 -0.3848  0.0215  0.3986  1.9194
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1746     0.0594   19.78  <2e-16 ***
## LastAuthorFemale1 -0.0409     0.0484   -0.85    0.397
## Year1997         -0.0878     0.0997   -0.88    0.379
## Year1998          0.0702     0.0902    0.78    0.436
## Year1999         -0.0530     0.1070   -0.50    0.621
## Year2000          0.0565     0.1101    0.51    0.608
## Year2001          0.2778     0.1235    2.25    0.025 *
## Year2002         -0.0763     0.0890   -0.86    0.391
## Year2003          0.1118     0.0845    1.32    0.186
## Year2004         -0.0133     0.0934   -0.14    0.887
## Year2005          0.0137     0.0817    0.17    0.867
## Year2006          0.0727     0.0830    0.88    0.381
```

```

## Year2007          0.0750      0.0777      0.97      0.335
## Year2008          -0.0342     0.0792     -0.43     0.666
## Year2009          -0.0711     0.0737     -0.96     0.335
## Year2010           0.0643     0.0748      0.86     0.390
## Year2011          -0.0179     0.0692     -0.26     0.796
## Year2012          -0.0034     0.0696     -0.05     0.961
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.576
## Multiple R-squared:  0.015, Adjusted R-squared:  0.00641
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 158 weights are ~= 1. The remaining 1807 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.244  0.863  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      5.09e-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: 1965"
## [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
##   35   31   24   36   35   44   40   57   52   61   80   65   86   33   41
## 2011 2012
##   48   50
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8    3    2    7    7    8   11   15   10   22   25   31   29   12   21
## 2011 2012

```

```
## 20 26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 3 1 4 6 4 8 11 7 21 22 22 23 10 15
## 2011 2012
## 15 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 = 17, df = 16, p-value = 0.4
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.25, df = 1, p-value = 0.6

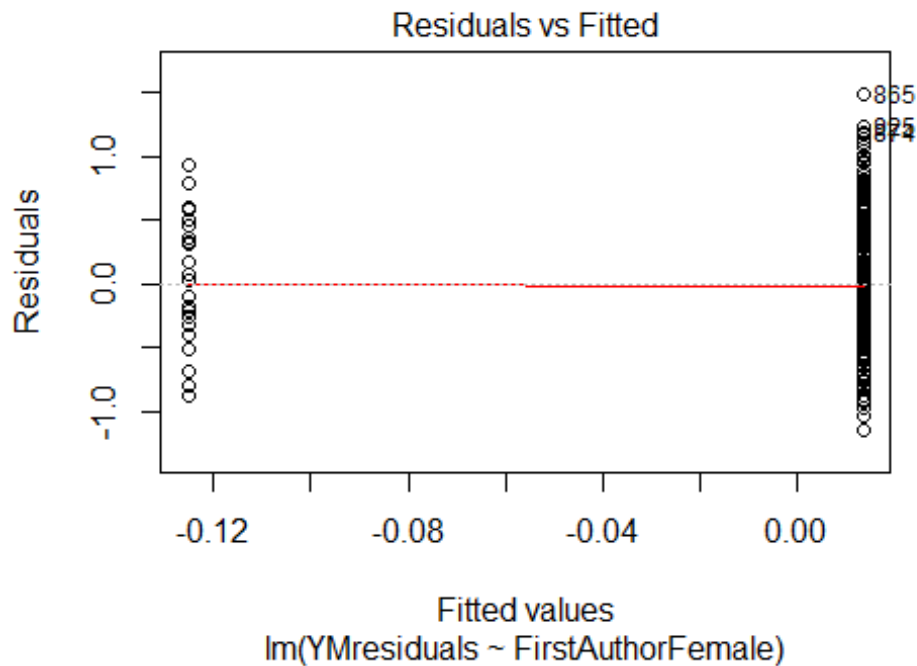
## [1] "Female first author team size 2018 geometric mean: 2"
## [1] "Male first author team size 2018 geometric mean: 3.00625306188766"

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



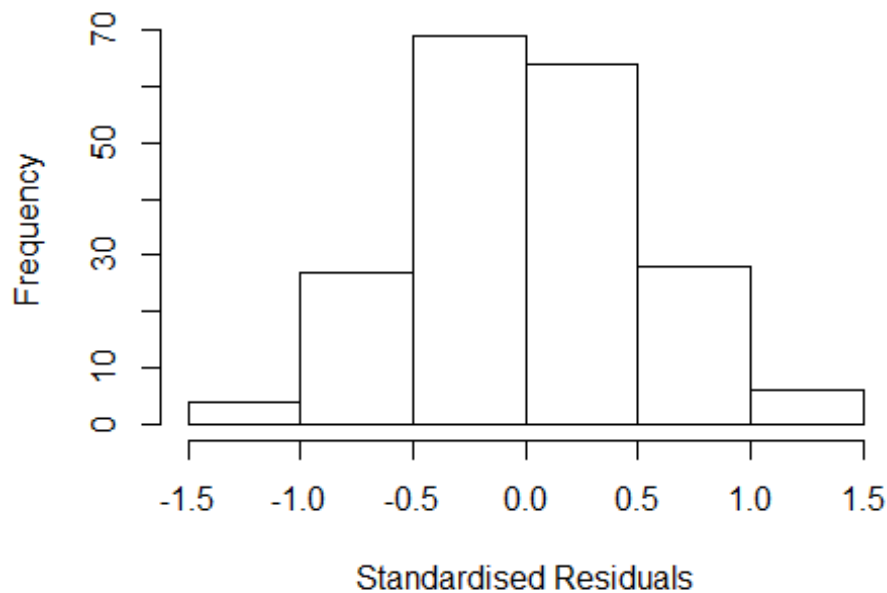
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3"
## [1] "Male last author team size 2018 geometric mean: 2.95372048071244"

## Warning in wilcox.test.default(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"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale  3.924  1      1.981
## LastAuthorFemale  1.900  1      1.378
## UniqueAuthors    22.110  4      1.473
## Year              86.228 16      1.149
```

Residuals from first and last author and team size



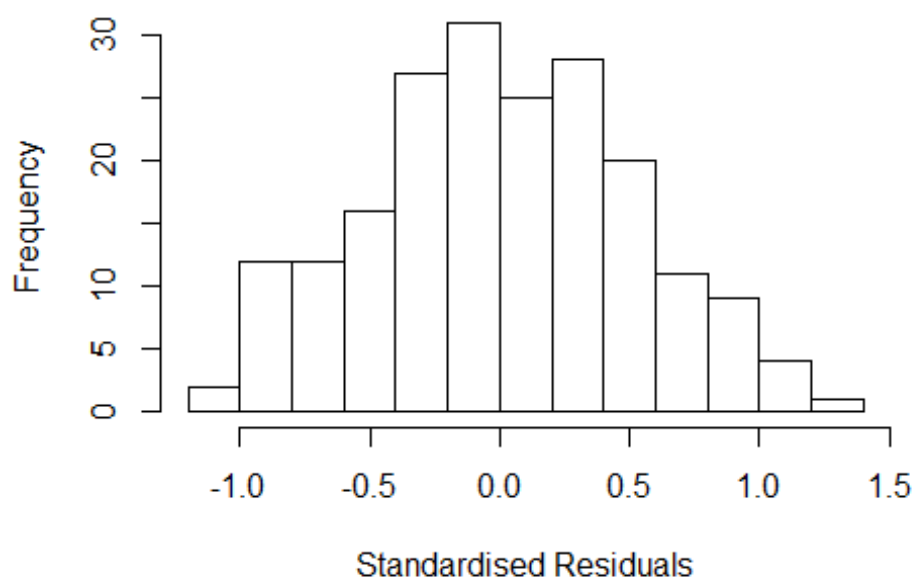
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.13602 -0.31055 -0.00336 0.33446 1.41231
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.1610 0.1878 0.86 0.39269
## FirstAuthorFemale1 -0.0645 0.1710 -0.38 0.70644
## LastAuthorFemale1 0.1635 0.1977 0.83 0.40937
## UniqueAuthors2 0.1498 0.1324 1.13 0.25944
## UniqueAuthors3 0.1941 0.1358 1.43 0.15466
## UniqueAuthors4 0.1471 0.1720 0.86 0.39367
## UniqueAuthors5 0.4380 0.2920 1.50 0.13537
## Year1997 0.4113 0.2496 1.65 0.10113
## Year1998 1.0222 0.2350 4.35 2.3e-05 ***
## Year1999 1.1371 0.4299 2.65 0.00891 **
```

```

## Year2000          1.0110      0.2821      3.58  0.00044 ***
## Year2001          0.9300      0.2272      4.09  6.5e-05 ***
## Year2002          0.4596      0.2879      1.60  0.11219
## Year2003          0.4756      0.2454      1.94  0.05428 .
## Year2004          0.9765      0.2972      3.29  0.00123 **
## Year2005          0.6618      0.2589      2.56  0.01144 *
## Year2006          0.6155      0.2276      2.70  0.00754 **
## Year2007          0.8088      0.2418      3.34  0.00101 **
## Year2008          0.5940      0.2538      2.34  0.02039 *
## Year2009          0.7390      0.3066      2.41  0.01695 *
## Year2010          0.2079      0.2541      0.82  0.41427
## Year2011          0.6433      0.3181      2.02  0.04469 *
## Year2012          0.5741      0.2415      2.38  0.01851 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.51
## Multiple R-squared:  0.207, Adjusted R-squared:  0.107
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 183 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.423  0.863  0.958  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      5.05e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.297 1      1.816
## LastAuthorFemale  1.827 1      1.352
## Year              3.814 16      1.043

```

Residuals from first and last author



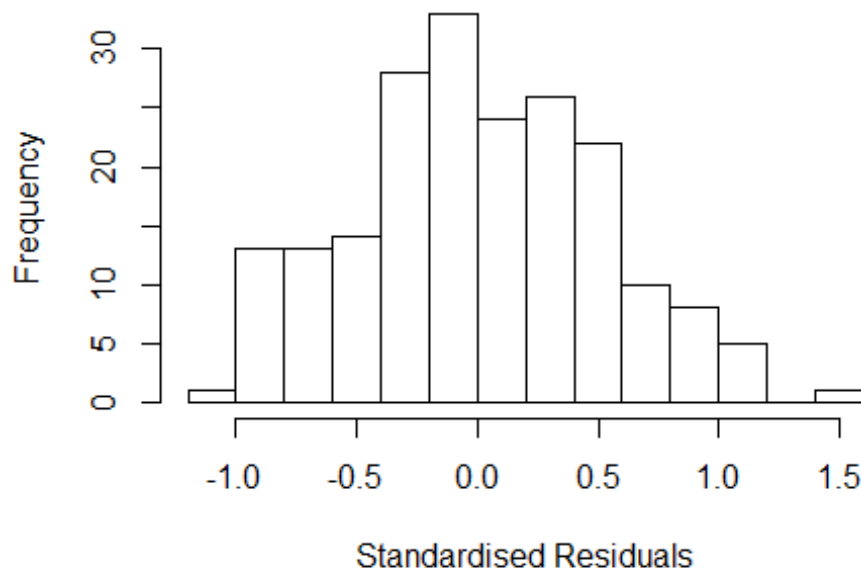
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1137 -0.3098 -0.0116 0.3445 1.3988
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.2065 0.2089 0.99 0.32418
## FirstAuthorFemale1 -0.0655 0.1676 -0.39 0.69637
## LastAuthorFemale1 0.1614 0.2101 0.77 0.44332
## Year1997 0.5309 0.2391 2.22 0.02769 *
## Year1998 1.1265 0.2089 5.39 2.2e-07 ***
## Year1999 1.1747 0.4330 2.71 0.00731 **
## Year2000 1.0672 0.2872 3.72 0.00027 ***
## Year2001 0.9675 0.2249 4.30 2.8e-05 ***
## Year2002 0.5344 0.3067 1.74 0.08316 .
## Year2003 0.5316 0.2432 2.19 0.03012 *
## Year2004 1.1273 0.2689 4.19 4.3e-05 ***
## Year2005 0.7458 0.2447 3.05 0.00265 **
```

```

## Year2006          0.7258      0.2231      3.25  0.00136 **
## Year2007          0.9199      0.2269      4.05  7.5e-05 ***
## Year2008          0.6904      0.2451      2.82  0.00540 **
## Year2009          0.8383      0.3009      2.79  0.00592 **
## Year2010          0.3257      0.2388      1.36  0.17422
## Year2011          0.7502      0.2962      2.53  0.01219 *
## Year2012          0.7030      0.2397      2.93  0.00380 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.546
## Multiple R-squared:  0.178, Adjusted R-squared:  0.0949
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 180 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.492  0.883  0.959  0.916  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      5.05e-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 2.877 1      1.696
## Year              2.877 16      1.034

```

Residuals from first author



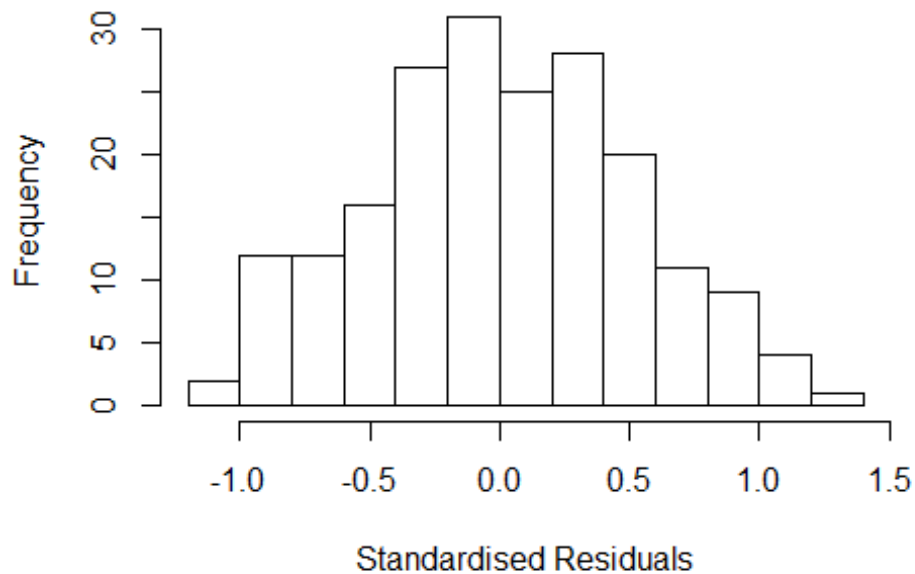
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0448 -0.3217 -0.0145 0.3460 1.4013
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.2450 0.1976 1.24 0.21651
## FirstAuthorFemale1 -0.0293 0.1521 -0.19 0.84766
## Year1997 0.4681 0.2013 2.33 0.02114 *
## Year1998 1.0880 0.1976 5.51 1.2e-07 ***
## Year1999 1.1359 0.4269 2.66 0.00850 **
## Year2000 1.0212 0.2714 3.76 0.00023 ***
## Year2001 0.9289 0.2144 4.33 2.4e-05 ***
## Year2002 0.4937 0.2988 1.65 0.10026
## Year2003 0.4932 0.2336 2.11 0.03611 *
## Year2004 1.1113 0.2598 4.28 3.1e-05 ***
## Year2005 0.7228 0.2388 3.03 0.00284 **
## Year2006 0.7057 0.2191 3.22 0.00152 **
```

```

## Year2007          0.8983      0.2177      4.13  5.6e-05 ***
## Year2008          0.6519      0.2354      2.77  0.00620 **
## Year2009          0.7998      0.2931      2.73  0.00699 **
## Year2010          0.2847      0.2267      1.26  0.21084
## Year2011          0.7187      0.2932      2.45  0.01520 *
## Year2012          0.6678      0.2278      2.93  0.00381 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.549
## Multiple R-squared:  0.173, Adjusted R-squared:  0.0946
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 181 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.494  0.878  0.959  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      5.05e-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 1.377 1      1.173
## Year              1.377 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
## -1.0914 -0.3170 -0.0115 0.3484 1.4049
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.191 0.208 0.92 0.35930
## LastAuthorFemale1 0.144 0.193 0.75 0.45517
## Year1997 0.503 0.223 2.26 0.02512 *
## Year1998 1.142 0.208 5.49 1.3e-07 ***
## Year1999 1.191 0.433 2.75 0.00661 **
## Year2000 1.070 0.287 3.73 0.00026 ***
## Year2001 0.983 0.224 4.39 1.9e-05 ***
## Year2002 0.539 0.313 1.72 0.08639 .
## Year2003 0.547 0.242 2.26 0.02524 *
## Year2004 1.138 0.263 4.33 2.5e-05 ***
## Year2005 0.756 0.247 3.07 0.00250 **
## Year2006 0.738 0.223 3.31 0.00114 **
```

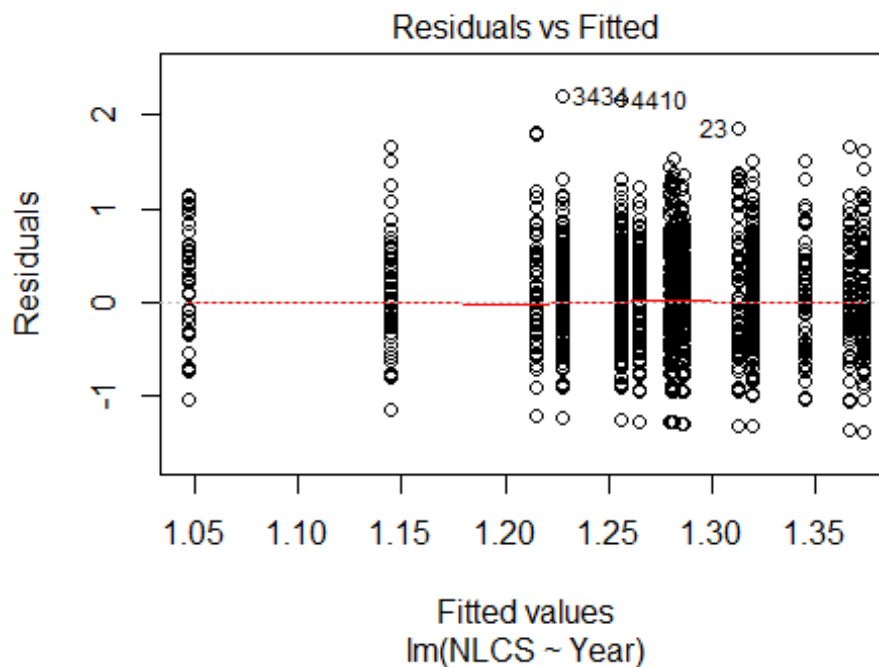


```

## Year2007          0.933      0.225      4.15  5.2e-05 ***
## Year2008          0.706      0.244      2.89  0.00435 **
## Year2009          0.854      0.301      2.84  0.00502 **
## Year2010          0.335      0.241      1.39  0.16545
## Year2011          0.766      0.297      2.58  0.01058 *
## Year2012          0.712      0.239      2.98  0.00325 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.543
## Multiple R-squared:  0.178, Adjusted R-squared:  0.1
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 180 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.483  0.882  0.958  0.915  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      5.05e-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: 198"
## [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
## 186 148 168 126 170 150 175 152 154 213 252 260 260 291 278
## 2011 2012
## 276 314
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 63 48 72 54 70 59 74 61 67 93 116 112 115 136 131
## 2011 2012

```

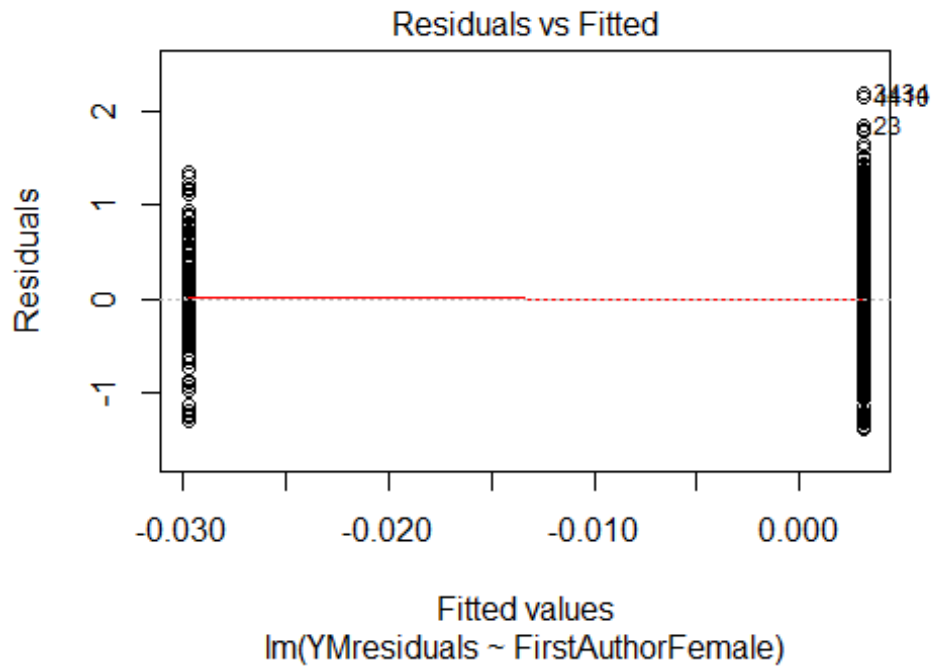
```
## 139 166
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 52 39 57 47 59 42 62 48 52 75 83 93 92 109 102
## 2011 2012
## 100 117
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 34, df = 16, p-value = 0.005
```



```
##
## 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: 2.82286471646455"
## [1] "Male first author team size 2018 geometric mean: 2.64470202664712"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 200, p-value = 1
```

```
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.53016547635958"
## [1] "Male last author team size 2018 geometric mean: 2.69836957008245"

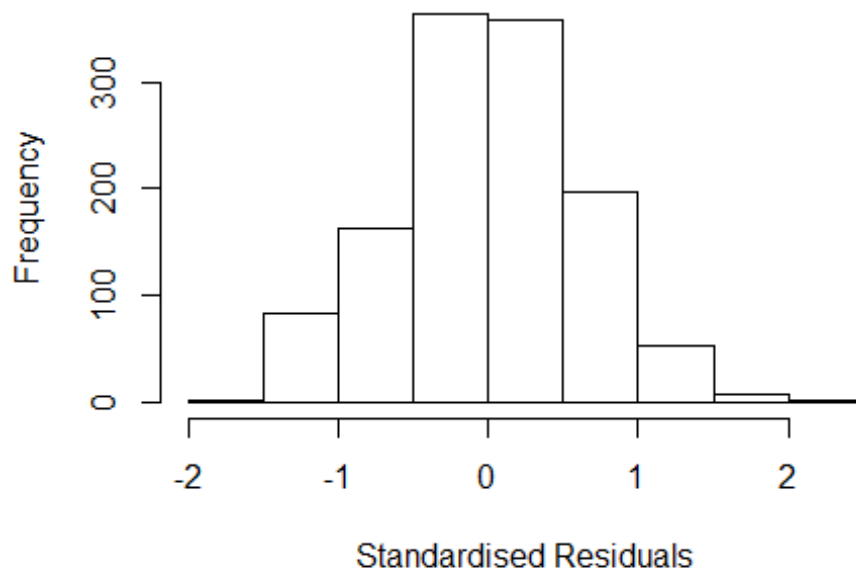
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 200, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "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	1.113	1	1.055
UniqueAuthors	1.403	4	1.043
Year	1.479	16	1.012

Residuals from first and last author and team size



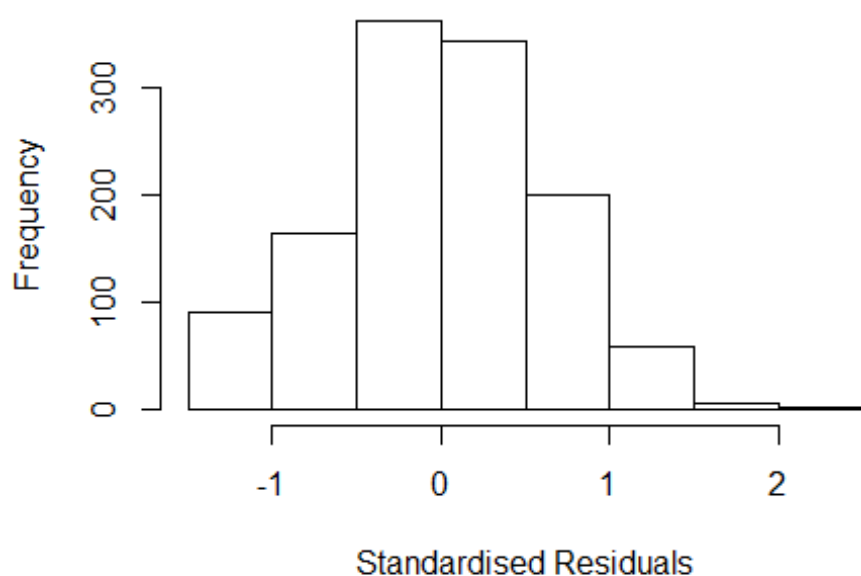
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5040 -0.3967 0.0078 0.4269 2.2808
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14812 0.11669 9.84 < 2e-16 ***
## FirstAuthorFemale1 -0.03800 0.06209 -0.61 0.54065
## LastAuthorFemale1 0.04640 0.08363 0.55 0.57912
## UniqueAuthors2 0.12015 0.05029 2.39 0.01705 *
## UniqueAuthors3 0.25388 0.05682 4.47 8.6e-06 ***
## UniqueAuthors4 0.26024 0.07449 3.49 0.00049 ***
## UniqueAuthors5 0.24992 0.08421 2.97 0.00306 **
## Year1997 0.07544 0.17979 0.42 0.67486
## Year1998 -0.00133 0.15196 -0.01 0.99304
## Year1999 -0.18612 0.16333 -1.14 0.25469
```

```

## Year2000      -0.08542    0.14583   -0.59  0.55816
## Year2001      -0.27252    0.15600   -1.75  0.08091 .
## Year2002      -0.02879    0.14708   -0.20  0.84485
## Year2003       0.02975    0.14874    0.20  0.84148
## Year2004       0.04068    0.15362    0.26  0.79119
## Year2005       0.01381    0.14104    0.10  0.92204
## Year2006      -0.03577    0.13525   -0.26  0.79145
## Year2007       0.05557    0.13091    0.42  0.67131
## Year2008      -0.02272    0.13505   -0.17  0.86643
## Year2009       0.02607    0.12937    0.20  0.84034
## Year2010      -0.10763    0.13111   -0.82  0.41184
## Year2011      -0.08627    0.12937   -0.67  0.50498
## Year2012      -0.14508    0.13006   -1.12  0.26484
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.0391, Adjusted R-squared:  0.0216
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 1122 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.139  0.863  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      8.14e-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 1.051 1      1.025
## LastAuthorFemale  1.081 1      1.040
## Year              1.097 16      1.003

```

Residuals from first and last author



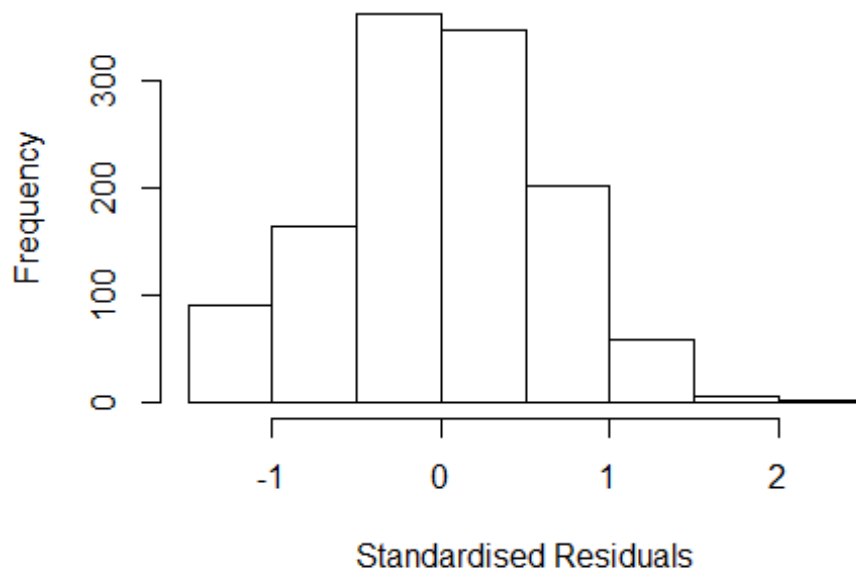
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.36915 -0.39231 -0.00341 0.43459 2.25098
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2251 0.1144 10.71 <2e-16 ***
## FirstAuthorFemale1 -0.0142 0.0618 -0.23 0.82
## LastAuthorFemale1 0.0307 0.0832 0.37 0.71
## Year1997 0.0951 0.1815 0.52 0.60
## Year1998 0.0265 0.1549 0.17 0.86
## Year1999 -0.1857 0.1667 -1.11 0.27
## Year2000 -0.0418 0.1444 -0.29 0.77
## Year2001 -0.2060 0.1555 -1.33 0.19
## Year2002 0.0131 0.1444 0.09 0.93
## Year2003 0.0713 0.1461 0.49 0.63
## Year2004 0.1008 0.1510 0.67 0.50
## Year2005 0.0874 0.1375 0.64 0.53
```

```

## Year2006          0.0265      0.1321      0.20      0.84
## Year2007          0.1133      0.1290      0.88      0.38
## Year2008          0.0453      0.1330      0.34      0.73
## Year2009          0.1029      0.1268      0.81      0.42
## Year2010         -0.0541      0.1302     -0.42      0.68
## Year2011         -0.0047      0.1272     -0.04      0.97
## Year2012         -0.0483      0.1269     -0.38      0.70
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.625
## Multiple R-squared:  0.0167, Adjusted R-squared:  0.00203
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 106 weights are ~= 1. The remaining 1123 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.167  0.866  0.950  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      8.14e-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 1.035 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
## -1.33939 -0.39296 -0.00196 0.43205 2.25047
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22454 0.11417 10.73 <2e-16 ***
## FirstAuthorFemale1 -0.00740 0.06201 -0.12 0.91
## Year1997 0.09763 0.18114 0.54 0.59
## Year1998 0.02811 0.15492 0.18 0.86
## Year1999 -0.18463 0.16670 -1.11 0.27
## Year2000 -0.03980 0.14420 -0.28 0.78
## Year2001 -0.20386 0.15499 -1.32 0.19
## Year2002 0.01442 0.14426 0.10 0.92
## Year2003 0.07327 0.14610 0.50 0.62
## Year2004 0.10242 0.15071 0.68 0.50
## Year2005 0.09109 0.13654 0.67 0.50
## Year2006 0.03031 0.13088 0.23 0.82
```

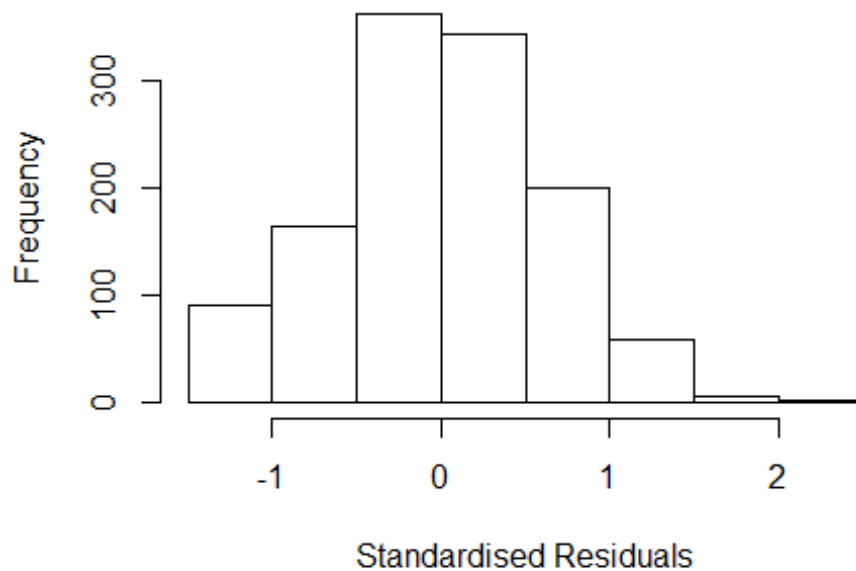


```

## Year2007          0.11485    0.12882    0.89    0.37
## Year2008          0.04838    0.13213    0.37    0.71
## Year2009          0.10484    0.12641    0.83    0.41
## Year2010         -0.05301    0.12988   -0.41    0.68
## Year2011         -0.00351    0.12692   -0.03    0.98
## Year2012         -0.04720    0.12662   -0.37    0.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.624
## Multiple R-squared:  0.0165, Adjusted R-squared:  0.00273
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 1121 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.166  0.866  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      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 1.06 1          1.030
## Year              1.06 16          1.002

```

Residuals from last author



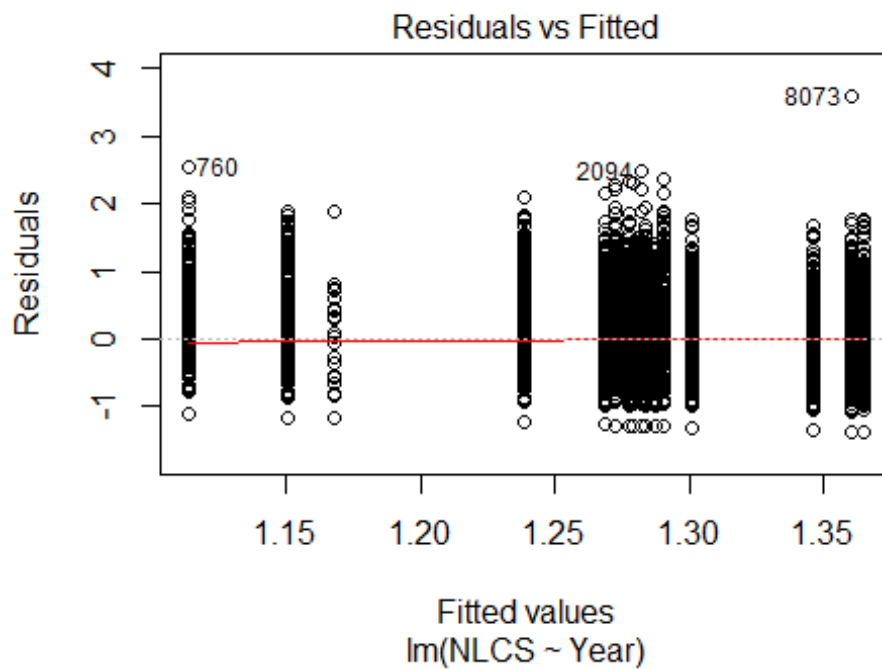
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.36292 -0.39118 -0.00191 0.42744 2.25244
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22483 0.11370 10.77 <2e-16 ***
## LastAuthorFemale1 0.02590 0.08293 0.31 0.75
## Year1997 0.09570 0.18067 0.53 0.60
## Year1998 0.02617 0.15454 0.17 0.87
## Year1999 -0.18684 0.16520 -1.13 0.26
## Year2000 -0.04183 0.14403 -0.29 0.77
## Year2001 -0.20699 0.15514 -1.33 0.18
## Year2002 0.01213 0.14406 0.08 0.93
## Year2003 0.07074 0.14586 0.48 0.63
## Year2004 0.09984 0.15071 0.66 0.51
## Year2005 0.08635 0.13727 0.63 0.53
## Year2006 0.02591 0.13171 0.20 0.84
```

```

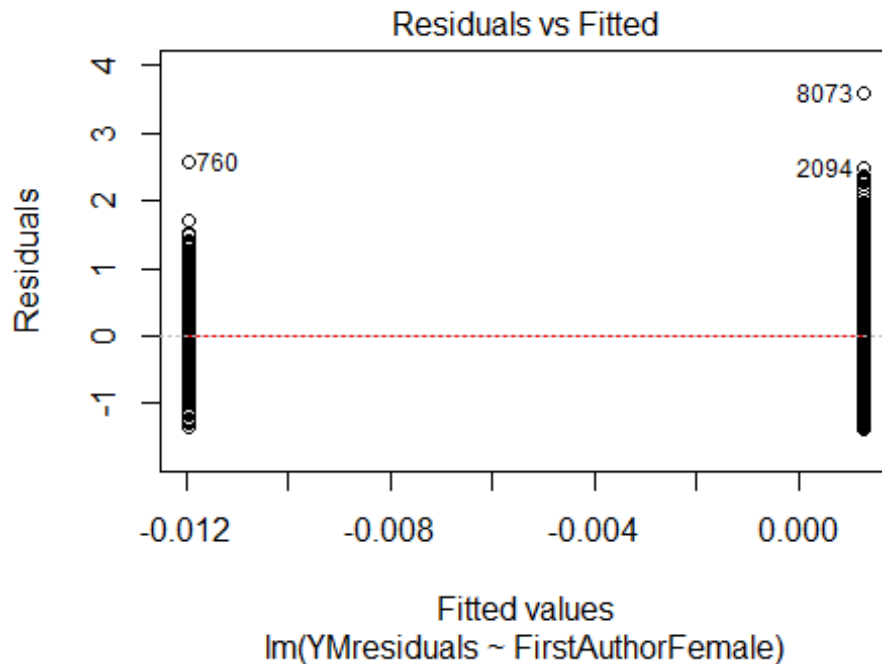
## Year2007      0.11219      0.12881      0.87      0.38
## Year2008      0.04450      0.13257      0.34      0.74
## Year2009      0.10244      0.12636      0.81      0.42
## Year2010     -0.05527      0.12993     -0.43      0.67
## Year2011     -0.00592      0.12690     -0.05      0.96
## Year2012     -0.04973      0.12674     -0.39      0.69
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.63
## Multiple R-squared:  0.0166, Adjusted R-squared:  0.00275
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 1122 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.174  0.868  0.951  0.907  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.14e-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 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
## 748 768 700 60 675 637 784 552 770 929 1121 1159 1134 1260 1232
## 2011 2012
## 1200 1215
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 265 275 285 21 293 246 383 291 371 439 524 552 578 645 618
## 2011 2012

```

```
## 619 636
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 204 208 221 19 227 189 271 212 279 319 379 386 423 459 430
## 2011 2012
## 429 450
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 210, df = 16, p-value <2e-16
```

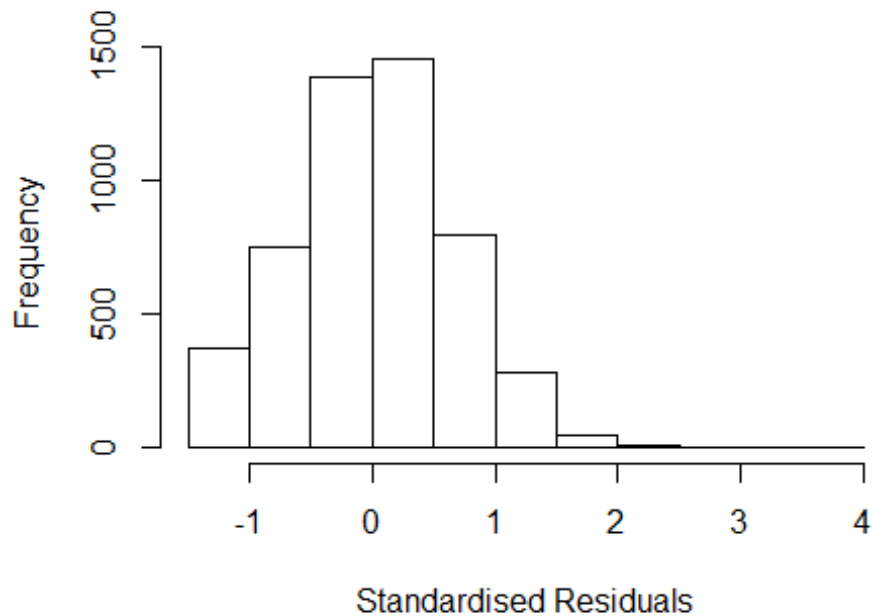


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.9, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 3.12639000651073"
## [1] "Male first author team size 2018 geometric mean: 2.82124334563565"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.7834411251135"
## [1] "Male last author team size 2018 geometric mean: 2.86280560700581"
##
## 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 1.024 1      1.012
## LastAuthorFemale  1.020 1      1.010
## UniqueAuthors    1.094 4      1.011
## Year             1.100 16      1.003
```

Residuals from first and last author and team size



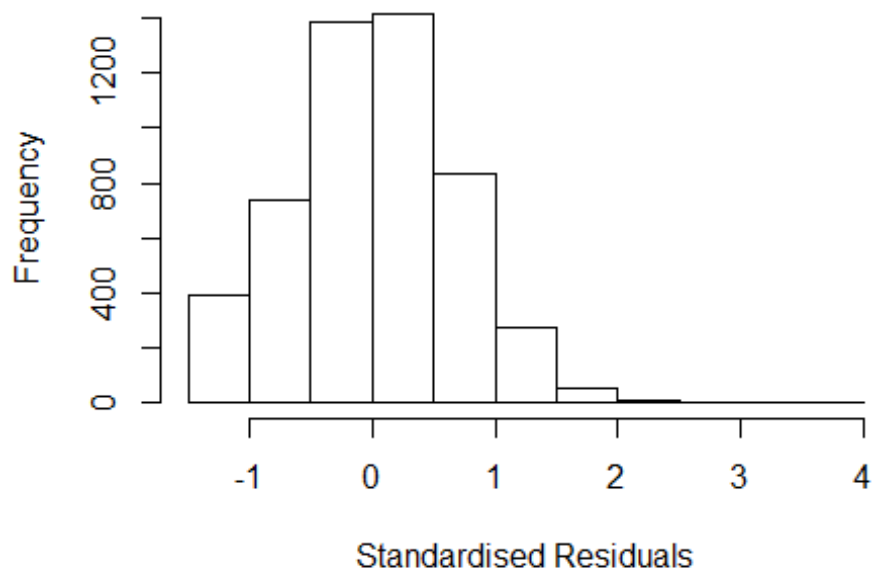
```
## [1] "List of 3 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 760      0030151578 3.673 1996    1711      3      2.537
## 8073    13844296408 4.951 2005    1705      2      3.781
## 15534   73649137161 3.611 2010    2102      2      2.545
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4120 -0.4377  0.0131  0.4407  3.7805
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.955085   0.068135   14.02  < 2e-16 ***
## FirstAuthorFemale1 0.001637   0.029364    0.06  0.9555
## LastAuthorFemale1 -0.000608   0.034745   -0.02  0.9860
## UniqueAuthors2    0.179160   0.030474    5.88 4.4e-09 ***
## UniqueAuthors3    0.222215   0.031666    7.02 2.6e-12 ***
## UniqueAuthors4    0.168135   0.039505    4.26 2.1e-05 ***
## UniqueAuthors5    0.241471   0.043592    5.54 3.2e-08 ***
## Year1997          0.166007   0.088962    1.87  0.0621 .
```

```

## Year1998          0.182983    0.085419    2.14    0.0322 *
## Year1999          -0.006644    0.204819   -0.03    0.9741
## Year2000          0.109033    0.088918    1.23    0.2202
## Year2001          0.124955    0.093996    1.33    0.1838
## Year2002          -0.045298    0.081049   -0.56    0.5763
## Year2003          0.226716    0.077850    2.91    0.0036 **
## Year2004          0.234010    0.077306    3.03    0.0025 **
## Year2005          0.215410    0.074030    2.91    0.0036 **
## Year2006          0.110608    0.072610    1.52    0.1277
## Year2007          0.120698    0.072750    1.66    0.0972 .
## Year2008          0.130645    0.070689    1.85    0.0646 .
## Year2009          0.159398    0.070431    2.26    0.0237 *
## Year2010          0.111380    0.070479    1.58    0.1141
## Year2011          0.159285    0.070582    2.26    0.0241 *
## Year2012          0.121359    0.070684    1.72    0.0861 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.658
## Multiple R-squared:  0.0231, Adjusted R-squared:  0.0189
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 2102 is an outlier with |weight| = 0 ( < 2e-05);
## 373 weights are ~= 1. The remaining 4731 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.102  0.873  0.953   0.909   0.986   0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.96e-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 1.013 1           1.006
## LastAuthorFemale  1.013 1           1.006
## Year              1.020 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
## 760   0030151578 3.673 1996    1711     3     2.582
## 2094  0031987382 3.783 1998    1706     4     2.514
## 8073 13844296408 4.951 2005    1705     2     3.612
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3508 -0.4369  0.0146  0.4498  3.6116
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.07983    0.06579   16.41 < 2e-16 ***
## FirstAuthorFemale1  0.01145    0.02952    0.39  0.69814
## LastAuthorFemale1 -0.00267    0.03522   -0.08  0.93957
## Year1997         0.17698    0.08948    1.98  0.04801 *
## Year1998         0.18925    0.08712    2.17  0.02989 *
## Year1999        -0.00192    0.19939   -0.01  0.99233
## Year2000         0.11620    0.08861    1.31  0.18978
## Year2001         0.15994    0.09462    1.69  0.09101 .
## Year2002        -0.00982    0.08191   -0.12  0.90459
## Year2003         0.25749    0.07900    3.26  0.00112 **
```

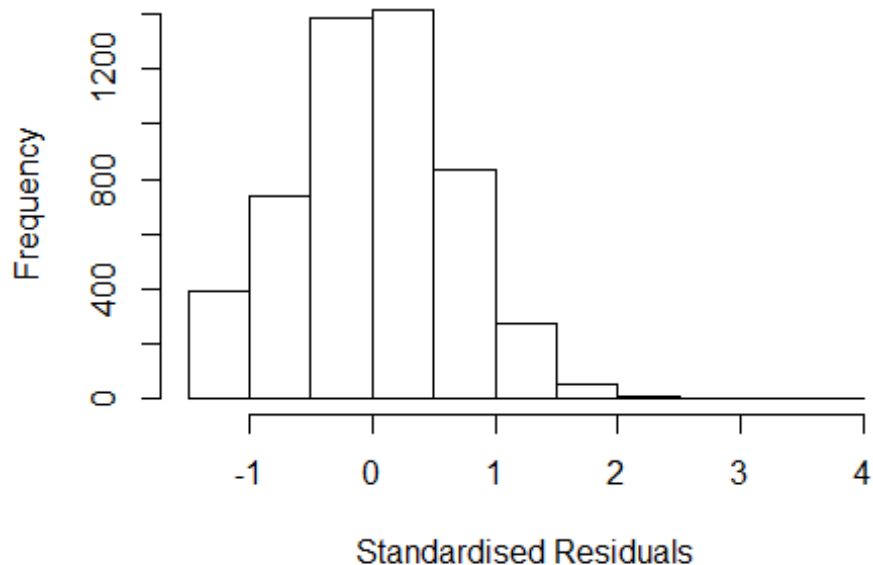


```

## Year2004          0.26821    0.07880    3.40  0.00067 ***
## Year2005          0.25954    0.07508    3.46  0.00055 ***
## Year2006          0.15461    0.07384    2.09  0.03633 *
## Year2007          0.16338    0.07381    2.21  0.02691 *
## Year2008          0.17713    0.07181    2.47  0.01367 *
## Year2009          0.21223    0.07164    2.96  0.00306 **
## Year2010          0.15835    0.07165    2.21  0.02715 *
## Year2011          0.20891    0.07168    2.91  0.00358 **
## Year2012          0.18074    0.07168    2.52  0.01171 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.662
## Multiple R-squared:  0.0106, Adjusted R-squared:  0.00709
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 2102 is an outlier with |weight| = 0 ( < 2e-05);
## 354 weights are ~= 1. The remaining 4750 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0942 0.8730 0.9530 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          1.96e-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 1.01 1          1.005
## Year              1.01 16          1.000

```

Residuals from first author



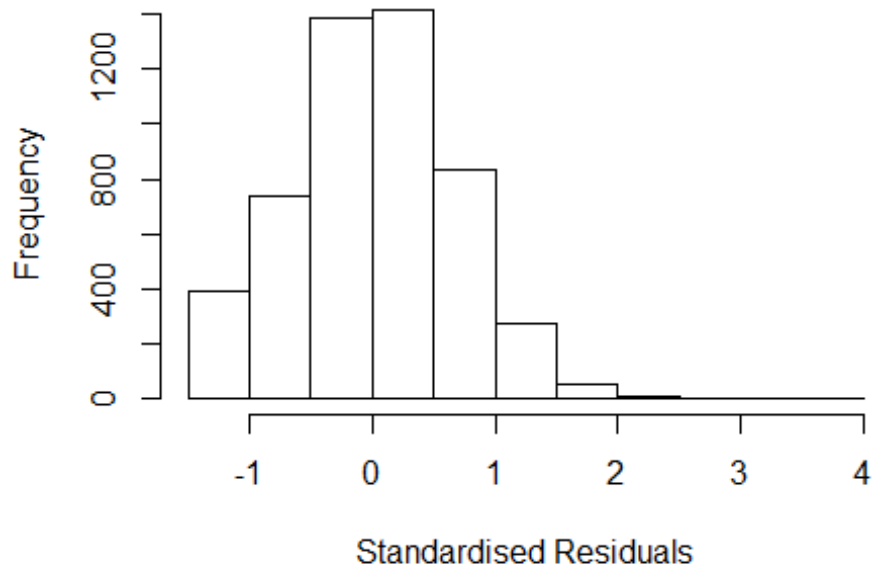
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 760   0030151578 3.673 1996    1711     3     2.582
## 2094  0031987382 3.783 1998    1706     4     2.514
## 8073 13844296408 4.951 2005    1705     2     3.612
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3503 -0.4373  0.0143  0.4499  3.6118
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.07967    0.06564   16.45 < 2e-16 ***
## FirstAuthorFemale1 0.01110    0.02964    0.37  0.70807
## Year1997        0.17698    0.08941    1.98  0.04783 *
## Year1998        0.18924    0.08703    2.17  0.02972 *
## Year1999       -0.00190    0.19946   -0.01  0.99238
## Year2000        0.11618    0.08856    1.31  0.18962
## Year2001        0.15997    0.09460    1.69  0.09091 .
## Year2002       -0.00977    0.08184   -0.12  0.90498
## Year2003        0.25746    0.07884    3.27  0.00110 **
## Year2004        0.26824    0.07865    3.41  0.00065 ***
```

```

## Year2005          0.25948    0.07495    3.46  0.00054 ***
## Year2006          0.15469    0.07369    2.10  0.03585 *
## Year2007          0.16343    0.07366    2.22  0.02654 *
## Year2008          0.17708    0.07166    2.47  0.01350 *
## Year2009          0.21219    0.07145    2.97  0.00300 **
## Year2010          0.15834    0.07148    2.22  0.02680 *
## Year2011          0.20882    0.07149    2.92  0.00350 **
## Year2012          0.18063    0.07147    2.53  0.01152 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.662
## Multiple R-squared:  0.0106, Adjusted R-squared:  0.00729
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2102 is an outlier with |weight| = 0 ( < 2e-05);
## 353 weights are ~= 1. The remaining 4751 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0938 0.8730 0.9530 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      1.96e-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 1.01 1          1.005
## Year            1.01 16          1.000

```

Residuals from last author



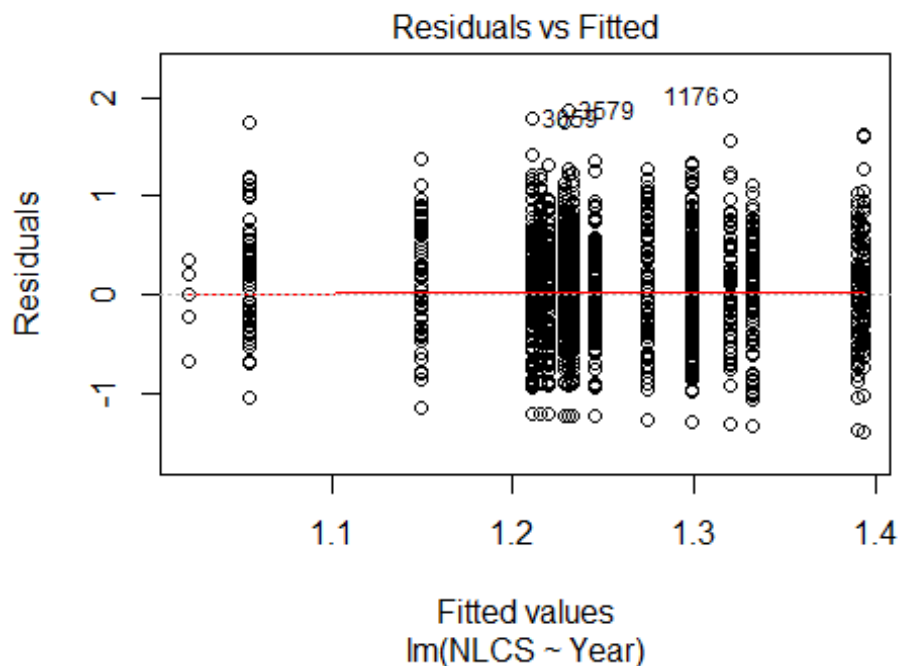
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 760   0030151578 3.673 1996    1711     3     2.582
## 2094  0031987382 3.783 1998    1706     4     2.514
## 8073 13844296408 4.951 2005    1705     2     3.612
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3489 -0.4380  0.0142  0.4492  3.6102
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.080771   0.065540   16.49 < 2e-16 ***
## LastAuthorFemale1 -0.000706   0.035299   -0.02  0.98403
## Year1997         0.176614   0.089373    1.98  0.04819 *
## Year1998         0.188855   0.086980    2.17  0.02996 *
## Year1999        -0.002388   0.199168   -0.01  0.99044
## Year2000         0.115739   0.088526    1.31  0.19114
## Year2001         0.159810   0.094562    1.69  0.09109 .
## Year2002        -0.009691   0.081836   -0.12  0.90574
## Year2003         0.257750   0.078833    3.27  0.00108 **
## Year2004         0.268149   0.078646    3.41  0.00066 ***
```

```

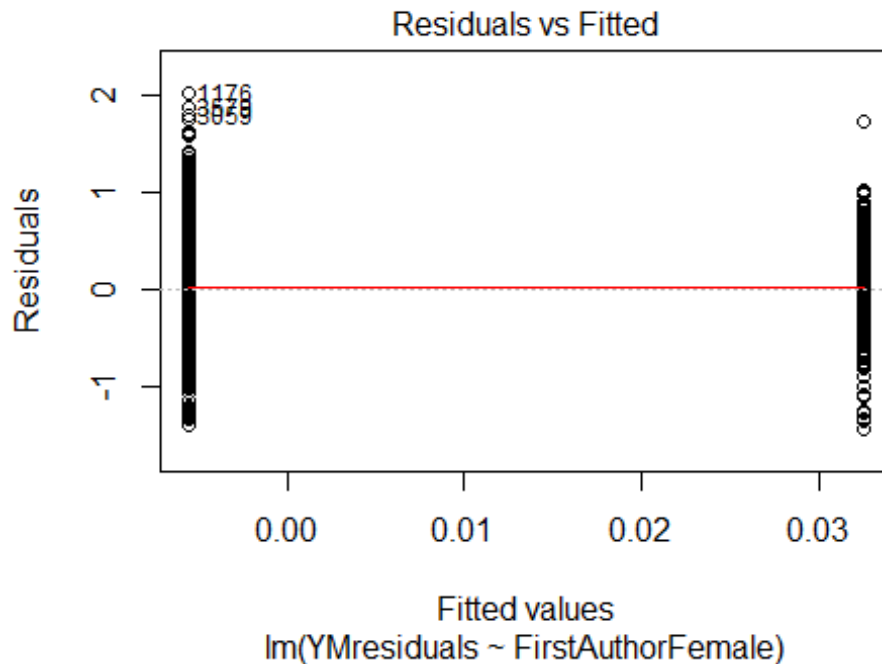
## Year2005      0.260000    0.074955    3.47  0.00053 ***
## Year2006      0.154753    0.073680    2.10  0.03575 *
## Year2007      0.163510    0.073630    2.22  0.02642 *
## Year2008      0.177439    0.071676    2.48  0.01333 *
## Year2009      0.212206    0.071461    2.97  0.00300 **
## Year2010      0.158808    0.071488    2.22  0.02636 *
## Year2011      0.209070    0.071518    2.92  0.00348 **
## Year2012      0.180925    0.071503    2.53  0.01143 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.662
## Multiple R-squared:  0.0106, Adjusted R-squared:  0.00726
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2102 is an outlier with |weight| = 0 ( < 2e-05);
## 354 weights are ~ = 1. The remaining 4750 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0905 0.8730 0.9520 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      1.96e-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: 5105"
## [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
##   16  252  210  222  212  248  204  217  231  269  308  317  327  394  384
## 2011 2012
##   322  411
##

```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      6   92   82   83   82   75   76   74   86  107  127  132  137  181  186
## 2011 2012
##  141  187
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      5   79   67   75   64   65   62   51   75   77   89  102  101  138  124
## 2011 2012
##  103  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 = 47, df = 16, p-value = 6e-05
```

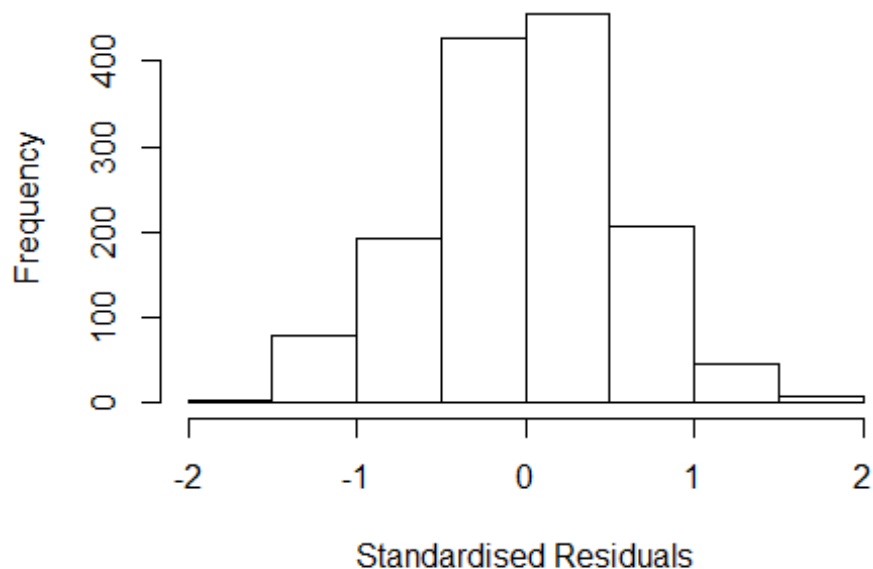


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 16, df = 1, p-value = 5e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 2.73070495331979"
## [1] "Male first author team size 2018 geometric mean: 2.86136957799797"
##
## 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: 2.70522263864908"
## [1] "Male last author team size 2018 geometric mean: 2.8598593018353"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.086 1      1.042
## LastAuthorFemale  1.105 1      1.051
## UniqueAuthors    1.227 4      1.026
## Year              1.380 16     1.010
```

Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5058 -0.3738 0.0104 0.3784 1.9630
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6332 0.1632 3.88 0.00011 ***
## FirstAuthorFemale1 0.0202 0.0373 0.54 0.58822
## LastAuthorFemale1 -0.0318 0.0525 -0.61 0.54524
## UniqueAuthors2 0.3025 0.0575 5.27 1.6e-07 ***
## UniqueAuthors3 0.3018 0.0556 5.42 6.8e-08 ***
## UniqueAuthors4 0.4009 0.0619 6.48 1.3e-10 ***
## UniqueAuthors5 0.3224 0.0664 4.85 1.4e-06 ***
## Year1997 0.4292 0.1728 2.48 0.01311 *
## Year1998 0.4502 0.1860 2.42 0.01564 *
## Year1999 0.2029 0.1808 1.12 0.26189
```

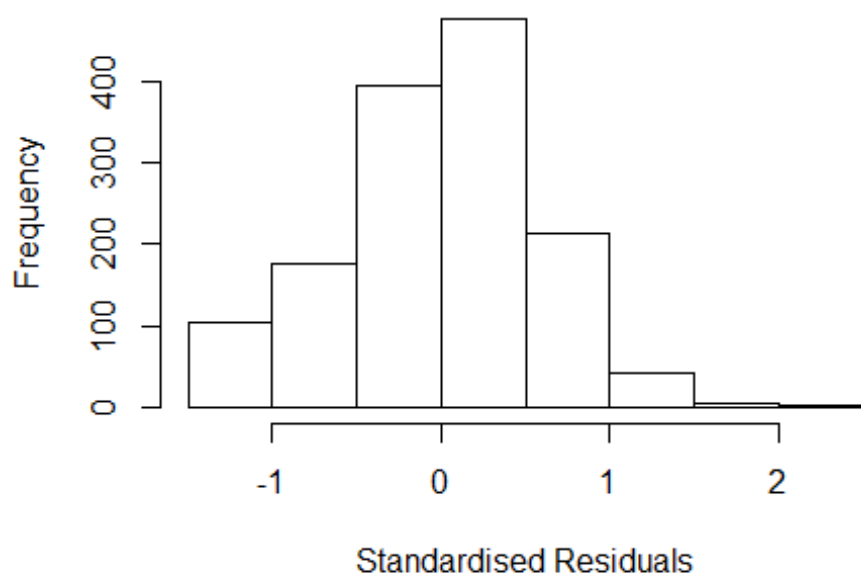


```

## Year2000          0.5701      0.1793      3.18  0.00151 **
## Year2001          0.4395      0.1861      2.36  0.01835 *
## Year2002          0.2300      0.1888      1.22  0.22321
## Year2003          0.5225      0.1739      3.00  0.00270 **
## Year2004          0.4505      0.1674      2.69  0.00721 **
## Year2005          0.4334      0.1662      2.61  0.00919 **
## Year2006          0.3035      0.1663      1.82  0.06826 .
## Year2007          0.2475      0.1634      1.51  0.13014
## Year2008          0.2937      0.1653      1.78  0.07575 .
## Year2009          0.3552      0.1622      2.19  0.02869 *
## Year2010          0.2344      0.1619      1.45  0.14782
## Year2011          0.3049      0.1619      1.88  0.05986 .
## Year2012          0.3168      0.1613      1.96  0.04970 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.0689, Adjusted R-squared:  0.0541
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 128 weights are ~= 1. The remaining 1284 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.199  0.860  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      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 1.059 1      1.029
## LastAuthorFemale  1.089 1      1.044
## Year              1.148 16      1.004

```

Residuals from first and last author



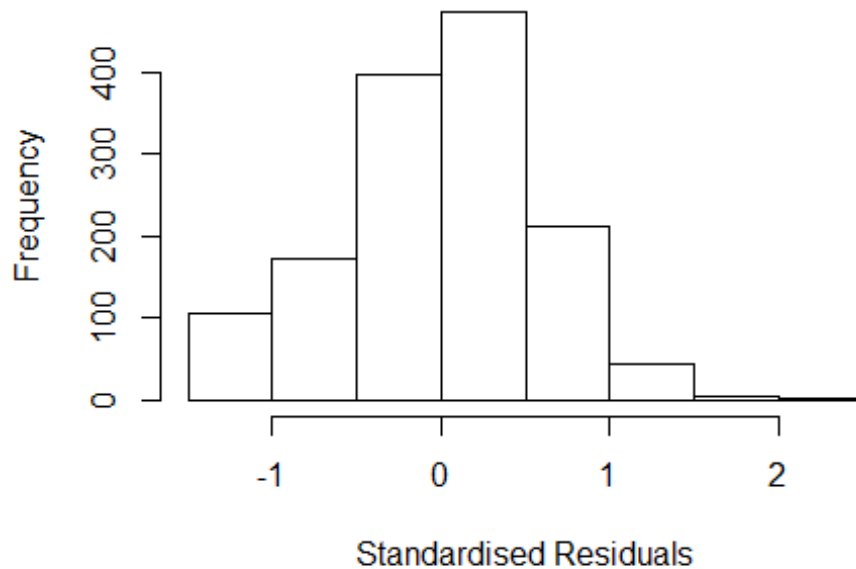
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4596 -0.3786 0.0288 0.3816 2.0320
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9479 0.1610 5.89 4.9e-09 ***
## FirstAuthorFemale1 0.0560 0.0370 1.51 0.1305
## LastAuthorFemale1 -0.0274 0.0527 -0.52 0.6030
## Year1997 0.3570 0.1815 1.97 0.0493 *
## Year1998 0.3292 0.1910 1.72 0.0851 .
## Year1999 0.0727 0.1834 0.40 0.6921
## Year2000 0.4507 0.1814 2.49 0.0131 *
## Year2001 0.3531 0.1902 1.86 0.0636 .
## Year2002 0.1386 0.1925 0.72 0.4717
## Year2003 0.4831 0.1814 2.66 0.0078 **
## Year2004 0.4025 0.1747 2.30 0.0214 *
## Year2005 0.3880 0.1715 2.26 0.0238 *
```

```

## Year2006          0.2523      0.1718      1.47      0.1422
## Year2007          0.1930      0.1695      1.14      0.2551
## Year2008          0.2412      0.1715      1.41      0.1597
## Year2009          0.3240      0.1682      1.93      0.0543 .
## Year2010          0.1944      0.1682      1.16      0.2479
## Year2011          0.2723      0.1680      1.62      0.1054
## Year2012          0.2947      0.1675      1.76      0.0787 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.568
## Multiple R-squared:  0.0281, Adjusted R-squared:  0.0155
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 132 weights are ~= 1. The remaining 1280 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.174  0.859   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      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 1.059 1          1.029
## Year              1.059 16          1.002

```

Residuals from first author



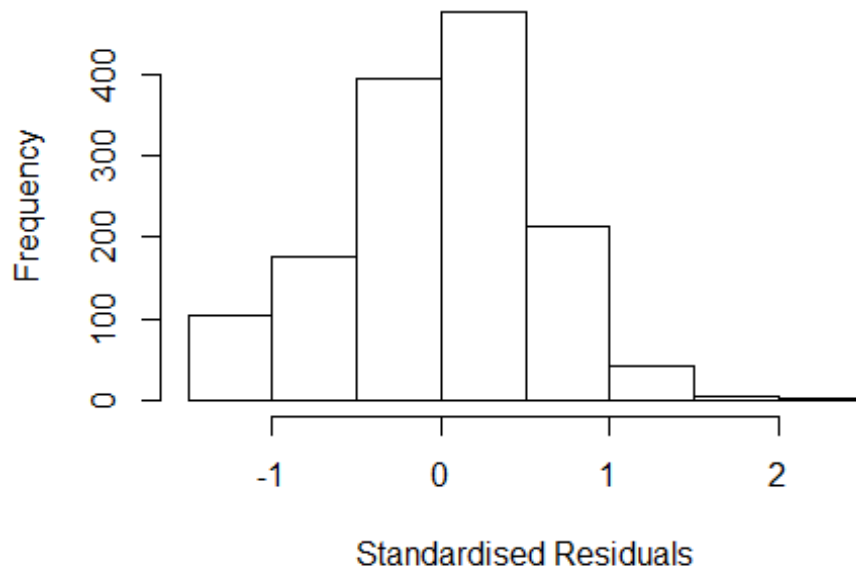
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4805 -0.3778 0.0281 0.3736 2.0349
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9486 0.1614 5.88 5.2e-09 ***
## FirstAuthorFemale1 0.0523 0.0374 1.40 0.1626
## Year1997 0.3552 0.1819 1.95 0.0510 .
## Year1998 0.3255 0.1912 1.70 0.0889 .
## Year1999 0.0712 0.1840 0.39 0.6989
## Year2000 0.4473 0.1814 2.47 0.0138 *
## Year2001 0.3494 0.1905 1.83 0.0668 .
## Year2002 0.1371 0.1927 0.71 0.4768
## Year2003 0.4796 0.1817 2.64 0.0084 **
## Year2004 0.3987 0.1751 2.28 0.0229 *
## Year2005 0.3868 0.1718 2.25 0.0245 *
## Year2006 0.2487 0.1720 1.45 0.1485
```

```

## Year2007          0.1923      0.1700      1.13      0.2579
## Year2008          0.2383      0.1717      1.39      0.1654
## Year2009          0.3200      0.1684      1.90      0.0576 .
## Year2010          0.1928      0.1686      1.14      0.2531
## Year2011          0.2685      0.1682      1.60      0.1106
## Year2012          0.2907      0.1676      1.73      0.0831 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.567
## Multiple R-squared:  0.0279, Adjusted R-squared:  0.0161
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 130 weights are ~= 1. The remaining 1282 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.171  0.860  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      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 1.088 1      1.043
## Year              1.088 16      1.003

```

Residuals from last author



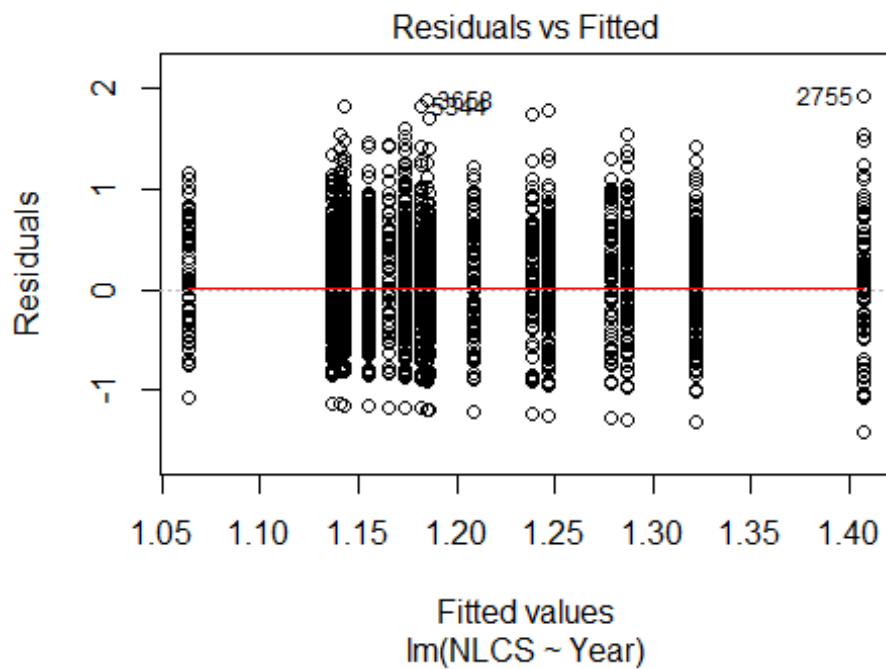
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4181 -0.3688 0.0288 0.3840 2.0215
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9589 0.1673 5.73 1.2e-08 ***
## LastAuthorFemale1 -0.0151 0.0526 -0.29 0.774
## Year1997 0.3479 0.1871 1.86 0.063 .
## Year1998 0.3234 0.1962 1.65 0.100 .
## Year1999 0.0688 0.1889 0.36 0.716
## Year2000 0.4457 0.1870 2.38 0.017 *
## Year2001 0.3527 0.1956 1.80 0.072 .
## Year2002 0.1387 0.1987 0.70 0.485
## Year2003 0.4743 0.1870 2.54 0.011 *
## Year2004 0.4009 0.1806 2.22 0.027 *
## Year2005 0.3874 0.1773 2.19 0.029 *
## Year2006 0.2474 0.1778 1.39 0.164
```

```

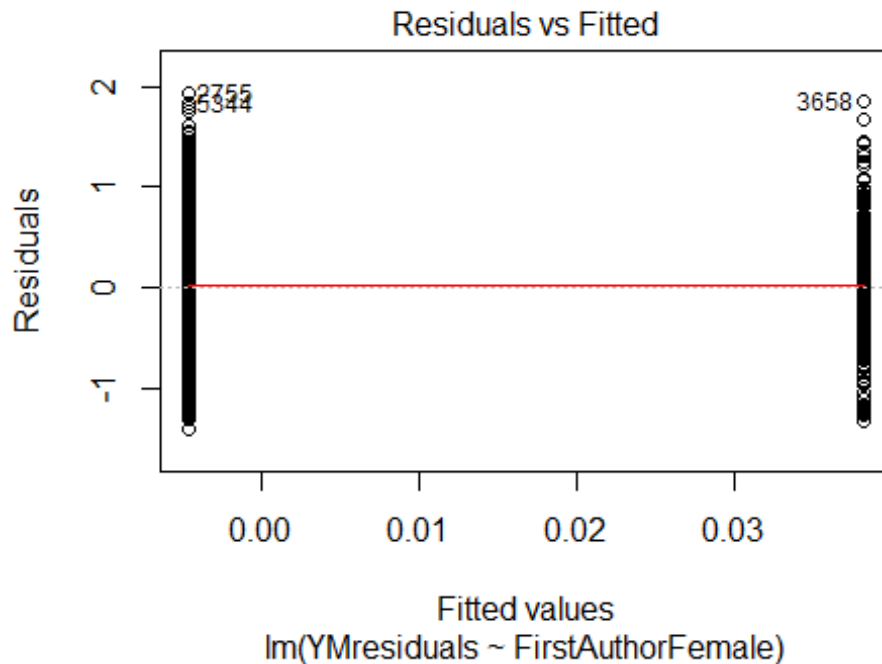
## Year2007          0.1927      0.1758      1.10      0.273
## Year2008          0.2442      0.1776      1.37      0.169
## Year2009          0.3231      0.1745      1.85      0.064 .
## Year2010          0.1911      0.1744      1.10      0.273
## Year2011          0.2720      0.1743      1.56      0.119
## Year2012          0.2935      0.1737      1.69      0.091 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.568
## Multiple R-squared:  0.027, Adjusted R-squared:  0.0151
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1283 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.178  0.859  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      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 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
##  513  494  511  319  355  449  428  439  455  577  585  620  761  770  747
## 2011 2012
##  796  808
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   99   85  104   68   79   83  101  126  120  168  163  193  260  285  282
## 2011 2012

```

```
## 345 380
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 84 68 89 58 64 63 85 98 95 122 120 145 196 231 221
## 2011 2012
## 280 295
## [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
```

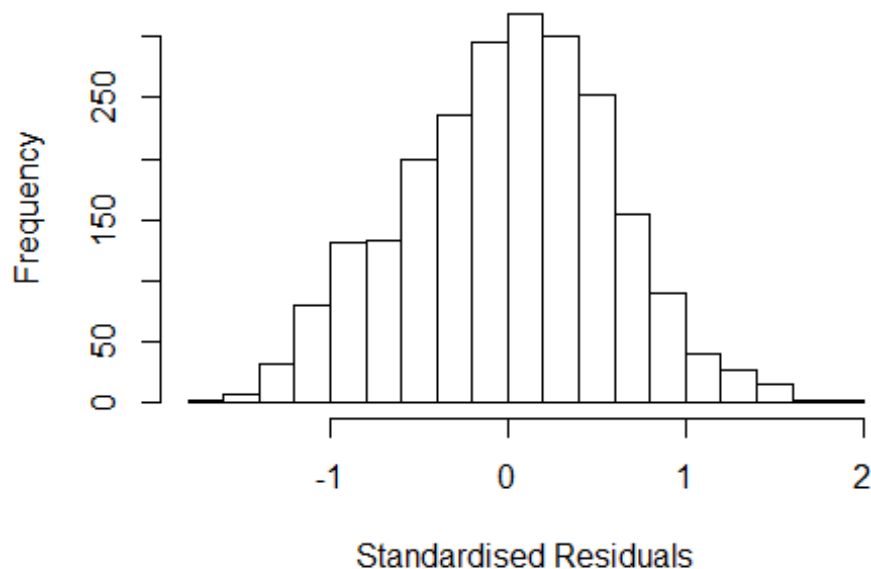


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.99, df = 1, p-value = 0.3
```

```
## [1] "Female first author team size 2018 geometric mean: 2.82414043385318"
## [1] "Male first author team size 2018 geometric mean: 2.70520082698148"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7600, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.92027960120015"
## [1] "Male last author team size 2018 geometric mean: 2.6970257748705"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.051 1 1.025
## LastAuthorFemale 1.037 1 1.018
## UniqueAuthors 1.282 4 1.032
## Year 1.320 16 1.009
```

Residuals from first and last author and team size



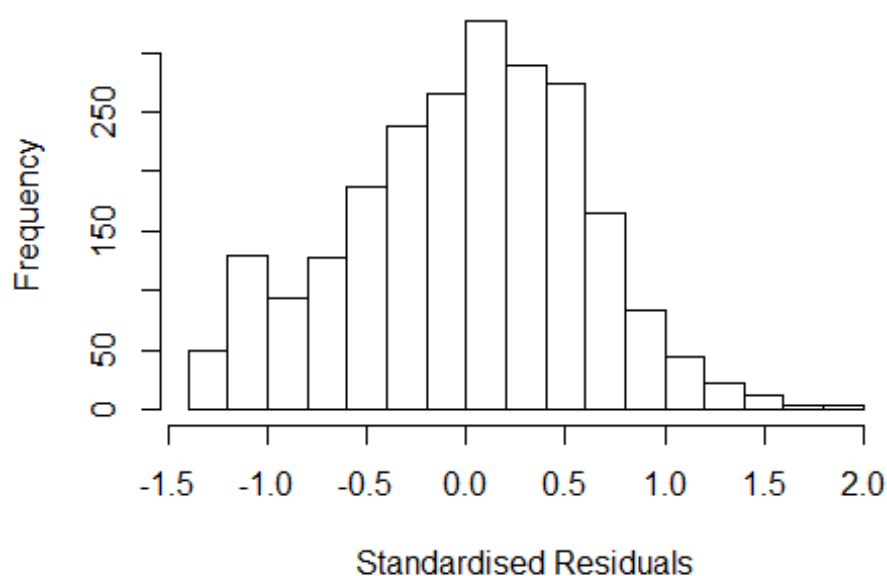
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6038 -0.4035 0.0261 0.4003 1.8581
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0118 0.0773 13.10 < 2e-16 ***
## FirstAuthorFemale1 0.0169 0.0428 0.39 0.694
## LastAuthorFemale1 -0.0635 0.0525 -1.21 0.227
## UniqueAuthors2 0.3243 0.0421 7.71 1.9e-14 ***
## UniqueAuthors3 0.3625 0.0452 8.02 1.7e-15 ***
## UniqueAuthors4 0.3512 0.0535 6.56 6.6e-11 ***
## UniqueAuthors5 0.5736 0.0599 9.57 < 2e-16 ***
## Year1997 0.0429 0.1210 0.35 0.723
## Year1998 -0.0686 0.1017 -0.67 0.500
## Year1999 -0.0372 0.1242 -0.30 0.765
```

```

## Year2000          -0.2219      0.1139    -1.95      0.052 .
## Year2001           0.1388      0.1212      1.15      0.252
## Year2002           0.0183      0.1135      0.16      0.872
## Year2003          -0.1041      0.0976    -1.07      0.286
## Year2004          -0.0619      0.1018    -0.61      0.543
## Year2005          -0.1483      0.0936    -1.59      0.113
## Year2006           0.0855      0.0897      0.95      0.341
## Year2007          -0.0396      0.0870    -0.45      0.649
## Year2008          -0.1644      0.0851    -1.93      0.054 .
## Year2009          -0.1226      0.0837    -1.47      0.143
## Year2010          -0.2126      0.0849    -2.50      0.012 *
## Year2011          -0.1929      0.0821    -2.35      0.019 *
## Year2012          -0.2128      0.0825    -2.58      0.010 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0754, Adjusted R-squared:  0.0666
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 200 weights are ~= 1. The remaining 2114 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.295  0.867  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          4.32e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.027 1          1.013
## LastAuthorFemale  1.041 1          1.020
## Year              1.067 16          1.002

```

Residuals from first and last author



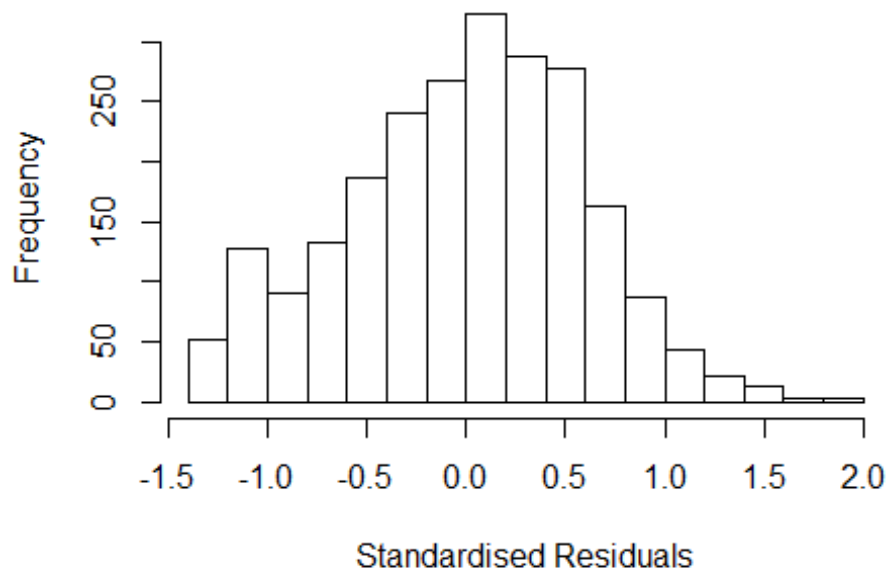
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3915 -0.4127 0.0405 0.4191 1.9415
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2072 0.0723 16.70 <2e-16 ***
## FirstAuthorFemale1 0.0659 0.0437 1.51 0.132
## LastAuthorFemale1 -0.0543 0.0548 -0.99 0.322
## Year1997 0.0862 0.1211 0.71 0.477
## Year1998 -0.0363 0.1054 -0.34 0.731
## Year1999 -0.0180 0.1269 -0.14 0.887
## Year2000 -0.1644 0.1147 -1.43 0.152
## Year2001 0.1843 0.1179 1.56 0.118
## Year2002 0.0972 0.1139 0.85 0.394
## Year2003 -0.0360 0.0980 -0.37 0.713
## Year2004 0.0501 0.1046 0.48 0.632
## Year2005 -0.0412 0.0931 -0.44 0.658
```

```

## Year2006          0.1625      0.0887      1.83      0.067 .
## Year2007          0.0658      0.0847      0.78      0.437
## Year2008         -0.0404      0.0840     -0.48      0.631
## Year2009         -0.0158      0.0831     -0.19      0.849
## Year2010         -0.0931      0.0830     -1.12      0.262
## Year2011         -0.0668      0.0799     -0.84      0.403
## Year2012         -0.0733      0.0794     -0.92      0.356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0173, Adjusted R-squared:  0.00955
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 200 weights are ~= 1. The remaining 2114 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.263  0.868  0.948  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      4.32e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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      1.013
## Year              1.025 16      1.001

```

Residuals from first author



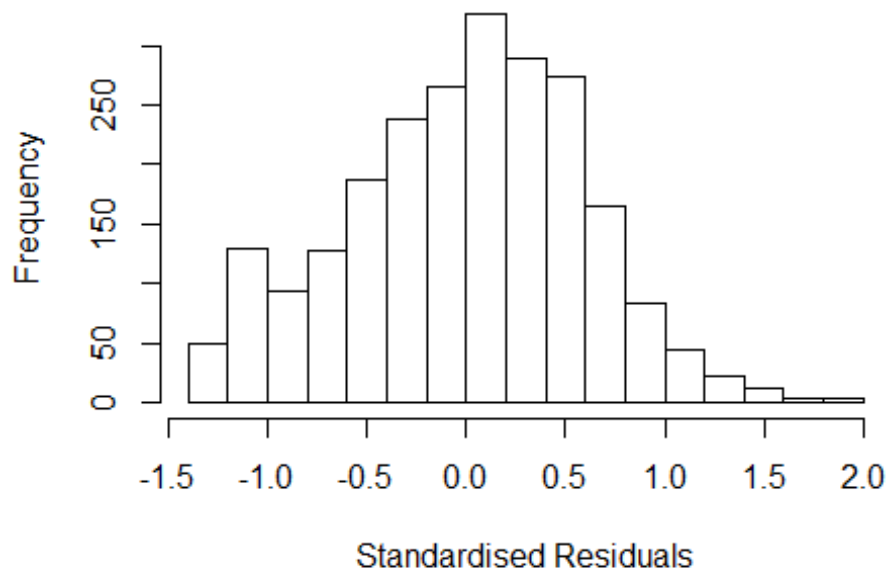
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3899 -0.4132  0.0415  0.4221  1.9431
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2055     0.0724   16.65  <2e-16 ***
## FirstAuthorFemale1  0.0608     0.0444    1.37   0.171
## Year1997          0.0857     0.1213    0.71   0.480
## Year1998         -0.0387     0.1054   -0.37   0.713
## Year1999         -0.0196     0.1271   -0.15   0.878
## Year2000         -0.1638     0.1149   -1.43   0.154
## Year2001          0.1845     0.1178    1.57   0.117
## Year2002          0.0985     0.1139    0.86   0.387
## Year2003         -0.0365     0.0981   -0.37   0.710
## Year2004          0.0492     0.1045    0.47   0.638
## Year2005         -0.0455     0.0930   -0.49   0.625
## Year2006          0.1581     0.0889    1.78   0.075 .
```

```

## Year2007          0.0664      0.0848      0.78      0.434
## Year2008          -0.0433      0.0840     -0.51      0.607
## Year2009          -0.0158      0.0832     -0.19      0.849
## Year2010          -0.0926      0.0831     -1.11      0.265
## Year2011          -0.0686      0.0799     -0.86      0.391
## Year2012          -0.0767      0.0792     -0.97      0.333
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0168, Adjusted R-squared:  0.00957
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 201 weights are ~= 1. The remaining 2113 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.262  0.868  0.948  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      4.32e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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
## -1.3958 -0.4105 0.0401 0.4209 1.9372
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2124 0.0725 16.73 <2e-16 ***
## LastAuthorFemale1 -0.0451 0.0550 -0.82 0.413
## Year1997 0.0828 0.1213 0.68 0.495
## Year1998 -0.0365 0.1057 -0.35 0.730
## Year1999 -0.0185 0.1264 -0.15 0.883
## Year2000 -0.1609 0.1151 -1.40 0.162
## Year2001 0.1834 0.1182 1.55 0.121
## Year2002 0.0939 0.1140 0.82 0.410
## Year2003 -0.0337 0.0976 -0.35 0.730
## Year2004 0.0489 0.1045 0.47 0.640
## Year2005 -0.0388 0.0933 -0.42 0.678
## Year2006 0.1606 0.0889 1.81 0.071 .
```

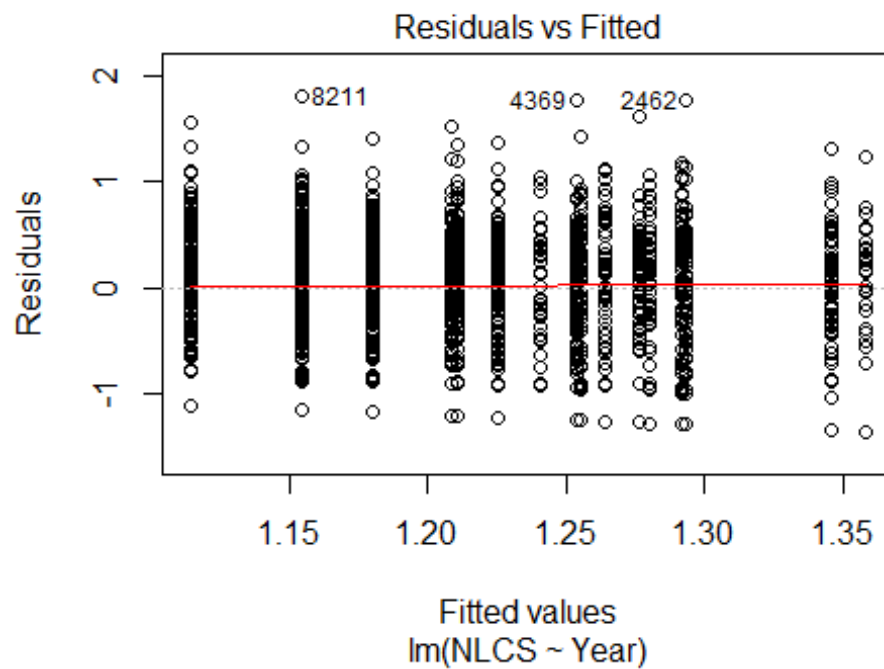


```

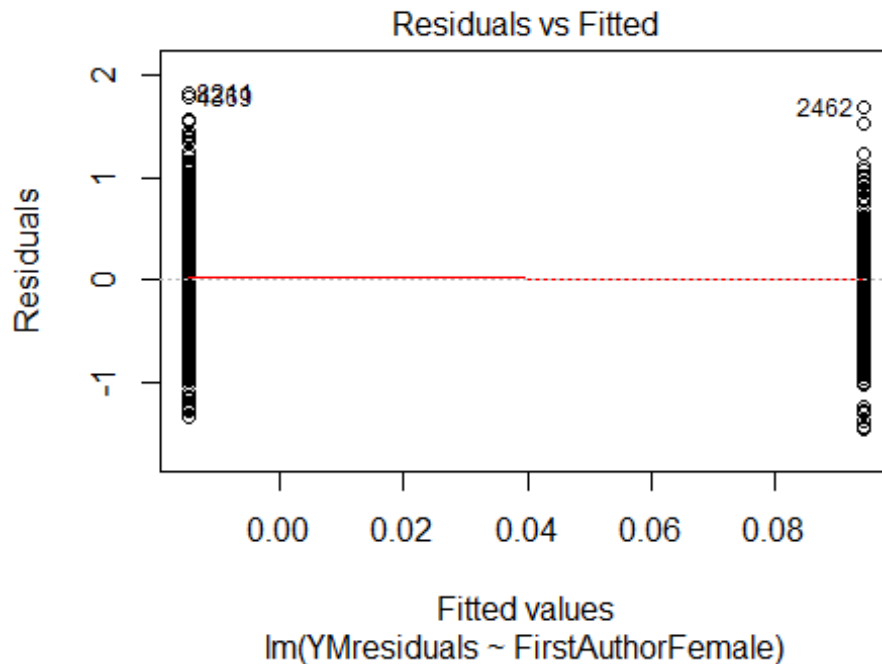
## Year2007          0.0660      0.0848      0.78      0.437
## Year2008          -0.0377      0.0842     -0.45      0.655
## Year2009          -0.0145      0.0832     -0.17      0.861
## Year2010          -0.0924      0.0833     -1.11      0.267
## Year2011          -0.0651      0.0800     -0.81      0.416
## Year2012          -0.0721      0.0795     -0.91      0.365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0163, Adjusted R-squared:  0.00898
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 201 weights are ~= 1. The remaining 2113 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.265  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      4.32e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2314"
## [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
## 339 322 359 167 201 323 268 303 328 385 391 403 513 512 568
## 2011 2012
## 601 553
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 66 57 78 32 44 55 67 70 83 116 119 141 175 178 227
## 2011 2012

```

```
## 254 250
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 53 46 62 24 35 41 52 59 66 85 93 113 138 147 184
## 2011 2012
## 206 194
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 43, df = 16, p-value = 3e-04
```

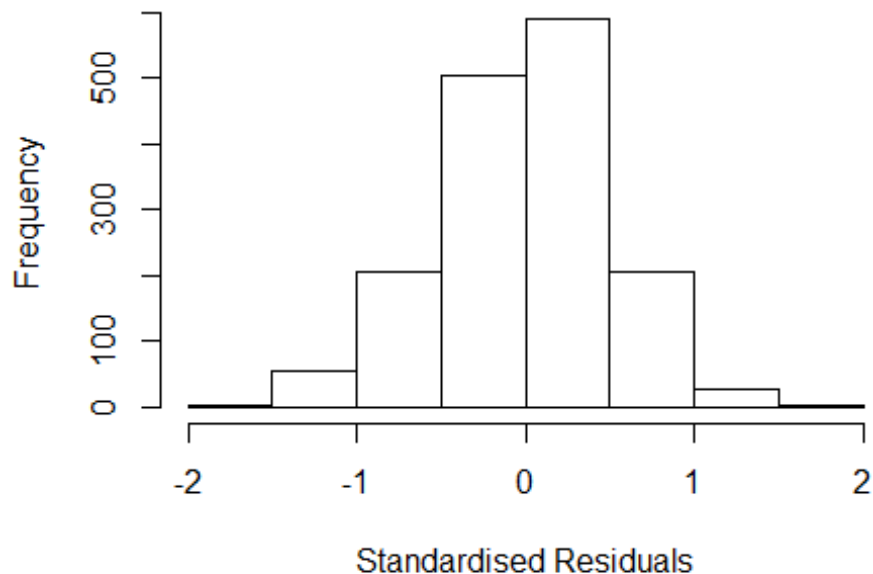


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.00059, df = 1, p-value = 1
```



```
## [1] "Female first author team size 2018 geometric mean: 3.17056577241056"
## [1] "Male first author team size 2018 geometric mean: 2.75049432572308"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3800, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.93432291602534"
## [1] "Male last author team size 2018 geometric mean: 2.78640240303011"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3200, 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 1.069 1          1.034
## LastAuthorFemale  1.087 1          1.043
## UniqueAuthors     1.377 4          1.041
## Year              1.366 16         1.010
```

Residuals from first and last author and team size



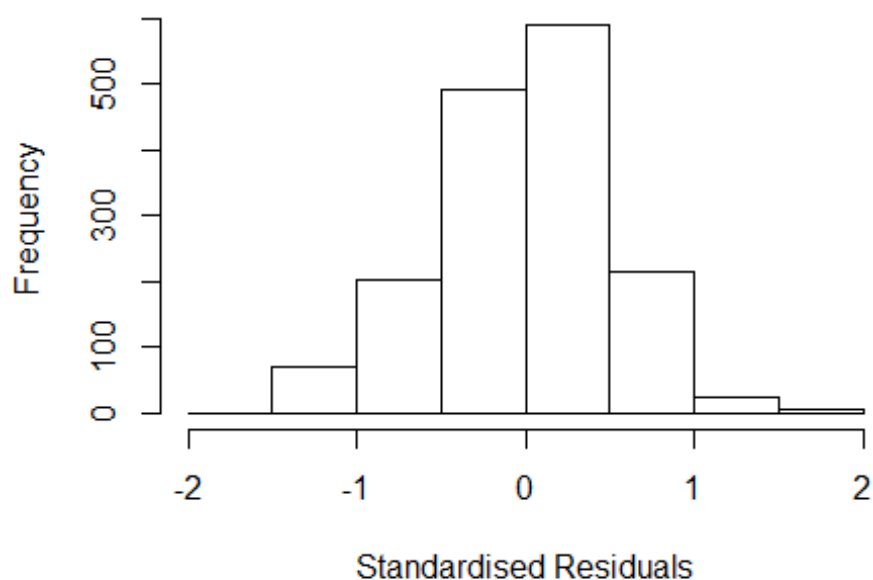
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6381 -0.3391 0.0206 0.3288 1.7367
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09621 0.07871 13.93 < 2e-16 ***
## FirstAuthorFemale1 0.06478 0.03943 1.64 0.1006
## LastAuthorFemale1 0.07184 0.04078 1.76 0.0783 .
## UniqueAuthors2 0.23813 0.05114 4.66 3.5e-06 ***
## UniqueAuthors3 0.25200 0.05370 4.69 2.9e-06 ***
## UniqueAuthors4 0.26891 0.06119 4.39 1.2e-05 ***
## UniqueAuthors5 0.42264 0.06078 6.95 5.2e-12 ***
## Year1997 0.00844 0.11850 0.07 0.9432
## Year1998 0.00226 0.11259 0.02 0.9840
## Year1999 0.22513 0.12902 1.74 0.0812 .
```

```

## Year2000      -0.04652    0.11909   -0.39    0.6961
## Year2001      0.02861    0.10965    0.26    0.7942
## Year2002     -0.06569    0.12475   -0.53    0.5986
## Year2003     -0.04756    0.11158   -0.43    0.6700
## Year2004     -0.01104    0.09914   -0.11    0.9113
## Year2005     -0.10184    0.08948   -1.14    0.2553
## Year2006     -0.08896    0.08892   -1.00    0.3172
## Year2007     -0.05694    0.08526   -0.67    0.5043
## Year2008     -0.16953    0.08800   -1.93    0.0542 .
## Year2009     -0.11680    0.08442   -1.38    0.1667
## Year2010     -0.26397    0.08497   -3.11    0.0019 **
## Year2011     -0.17672    0.08398   -2.10    0.0355 *
## Year2012     -0.22945    0.08220   -2.79    0.0053 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.0743, Adjusted R-squared:  0.0614
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 148 weights are ~= 1. The remaining 1450 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.875   0.947   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          6.26e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1          1.018
## LastAuthorFemale  1.037 1          1.018
## Year              1.063 16          1.002

```

Residuals from first and last author



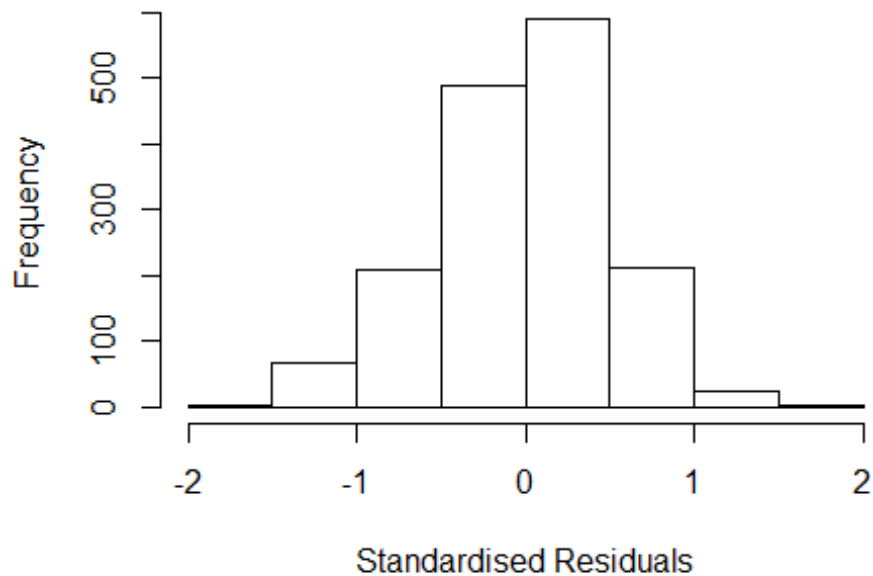
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6041 -0.3450 0.0227 0.3406 1.7563
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25080 0.07662 16.33 <2e-16 ***
## FirstAuthorFemale1 0.11091 0.03938 2.82 0.0049 **
## LastAuthorFemale1 0.10433 0.04005 2.60 0.0093 **
## Year1997 0.03241 0.11971 0.27 0.7867
## Year1998 0.03590 0.12120 0.30 0.7671
## Year1999 0.24238 0.12586 1.93 0.0543 .
## Year2000 -0.00845 0.12339 -0.07 0.9454
## Year2001 0.06854 0.11046 0.62 0.5351
## Year2002 0.00416 0.12796 0.03 0.9741
## Year2003 0.00548 0.11521 0.05 0.9621
## Year2004 0.09210 0.10065 0.92 0.3603
## Year2005 -0.02634 0.09287 -0.28 0.7768
```

```

## Year2006      -0.01856    0.09010   -0.21    0.8368
## Year2007      0.02092    0.08582    0.24    0.8075
## Year2008     -0.09147    0.08980   -1.02    0.3085
## Year2009     -0.04370    0.08659   -0.50    0.6139
## Year2010     -0.18679    0.08679   -2.15    0.0315 *
## Year2011     -0.08758    0.08425   -1.04    0.2987
## Year2012     -0.13618    0.08298   -1.64    0.1010
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.0357, Adjusted R-squared:  0.0247
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 146 weights are ~= 1. The remaining 1452 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.193  0.874  0.947  0.900  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.034 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
## -1.6276 -0.3517 0.0254 0.3471 1.7462
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25241 0.07637 16.40 <2e-16 ***
## FirstAuthorFemale1 0.12467 0.03940 3.16 0.0016 **
## Year1997 0.04252 0.11820 0.36 0.7191
## Year1998 0.03959 0.12050 0.33 0.7425
## Year1999 0.25056 0.12617 1.99 0.0472 *
## Year2000 -0.00533 0.12284 -0.04 0.9654
## Year2001 0.07939 0.11017 0.72 0.4712
## Year2002 0.00609 0.12780 0.05 0.9620
## Year2003 0.00997 0.11483 0.09 0.9309
## Year2004 0.10123 0.10043 1.01 0.3136
## Year2005 -0.01775 0.09235 -0.19 0.8477
## Year2006 -0.00582 0.08960 -0.06 0.9482
```

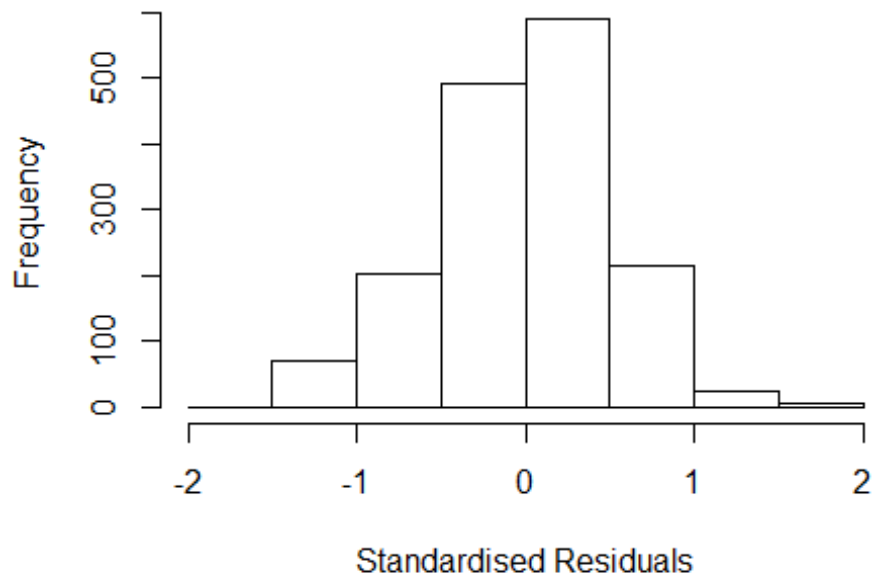


```

## Year2007          0.02938    0.08526    0.34    0.7305
## Year2008          -0.08084    0.08933   -0.90    0.3656
## Year2009          -0.03928    0.08642   -0.45    0.6496
## Year2010          -0.18011    0.08643   -2.08    0.0373 *
## Year2011          -0.08000    0.08410   -0.95    0.3416
## Year2012          -0.12762    0.08265   -1.54    0.1227
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.504
## Multiple R-squared:  0.0317, Adjusted R-squared:  0.0213
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1457 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.206  0.876  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      6.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.029 1          1.014
## Year            1.029 16          1.001

```

Residuals from last author



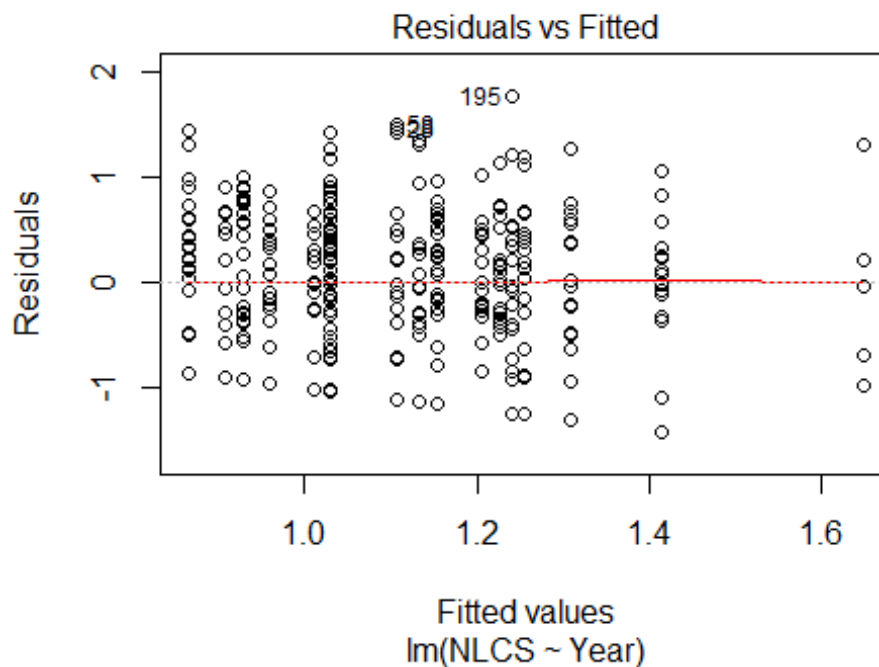
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4996 -0.3348 0.0227 0.3402 1.7968
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.271442 0.077021 16.51 <2e-16 ***
## LastAuthorFemale1 0.121219 0.039830 3.04 0.0024 **
## Year1997 0.016394 0.119584 0.14 0.8910
## Year1998 0.031051 0.122777 0.25 0.8004
## Year1999 0.228119 0.127482 1.79 0.0737 .
## Year2000 -0.006735 0.126091 -0.05 0.9574
## Year2001 0.053224 0.112077 0.47 0.6349
## Year2002 -0.006157 0.127312 -0.05 0.9614
## Year2003 -0.000232 0.117313 0.00 0.9984
## Year2004 0.078784 0.101267 0.78 0.4367
## Year2005 -0.031909 0.093508 -0.34 0.7330
## Year2006 -0.028033 0.090872 -0.31 0.7578
```

```

## Year2007      0.007276  0.086749  0.08  0.9332
## Year2008     -0.098101  0.090352 -1.09  0.2778
## Year2009     -0.051271  0.087392 -0.59  0.5575
## Year2010     -0.192947  0.087936 -2.19  0.0284 *
## Year2011     -0.095857  0.084859 -1.13  0.2588
## Year2012     -0.144364  0.083689 -1.72  0.0847 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.5
## Multiple R-squared:  0.0303, Adjusted R-squared:  0.0199
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 151 weights are ~= 1. The remaining 1447 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.170  0.873  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      6.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1598"
## [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
##   51  40  47  45  38  40  39  35  40  22  26  42  33  47  58
## 2011 2012
##   44  66
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   21  19  17  20  14  6  19  20  27  15  14  24  17  30  31
## 2011 2012

```

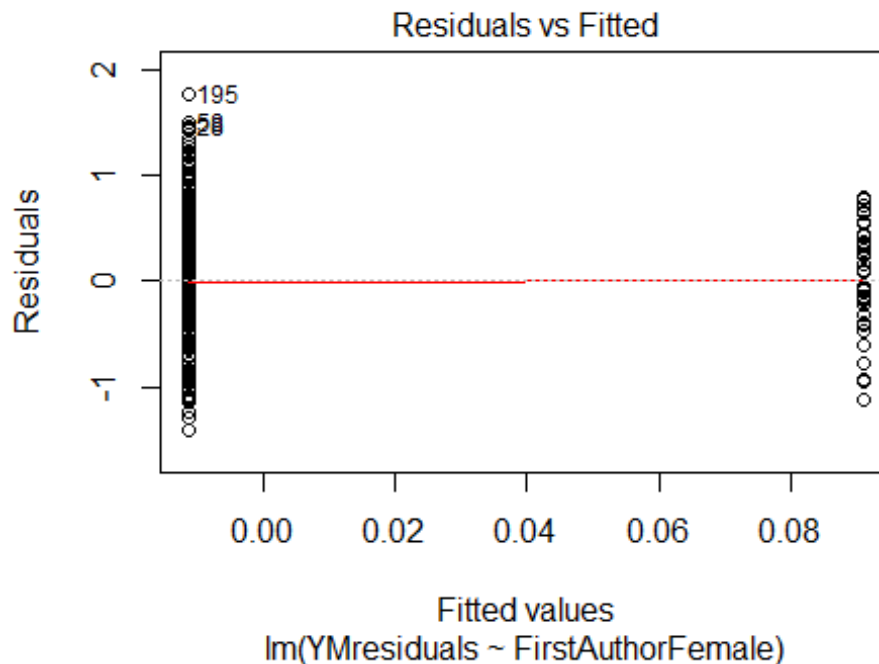
```
## 34 34
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 17 13 14 15 13 6 17 18 24 15 12 21 14 27 27
## 2011 2012
## 29 27
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 17, df = 16, p-value = 0.4
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3, df = 1, p-value = 0.09
## [1] "Female first author team size 2018 geometric mean: 3.1050186838948"
## [1] "Male first author team size 2018 geometric mean: 1.94121054959744"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

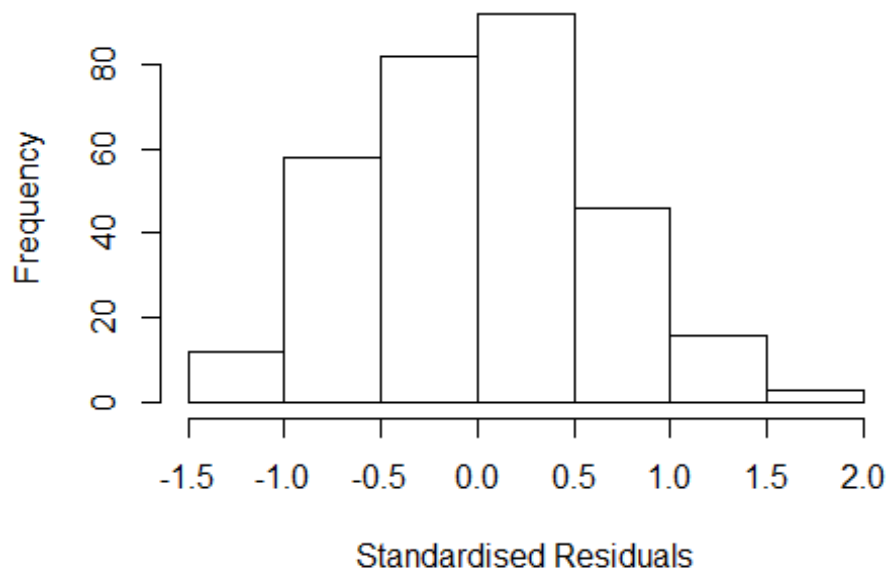
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 210, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.8039657955522"
## [1] "Male last author team size 2018 geometric mean: 2.02433144313075"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.2
## 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 1.421  1      1.192
## LastAuthorFemale  1.297  1      1.139
## UniqueAuthors    2.603  4      1.127
## Year              3.321 16      1.038
```

Residuals from first and last author and team size



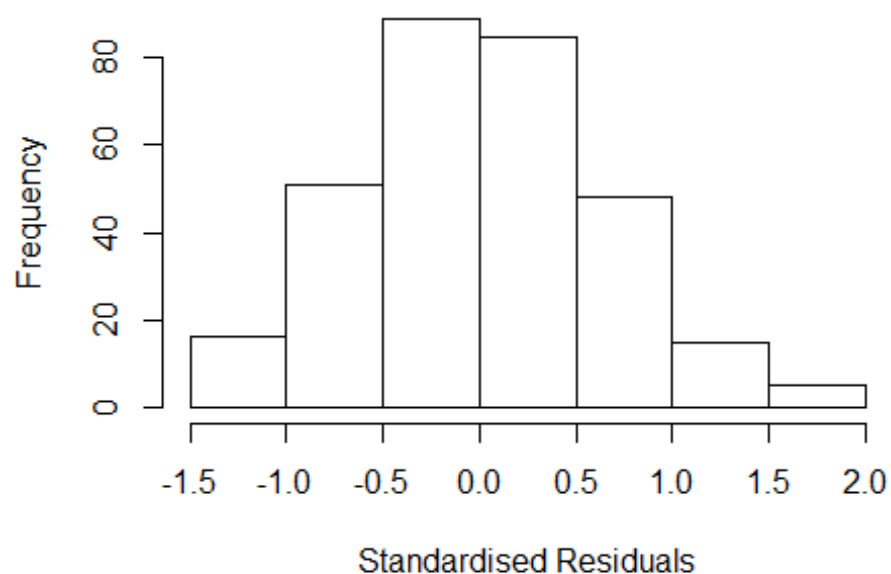
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4583 -0.4257 0.0217 0.3967 1.7980
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.83072 0.18640 4.46 1.2e-05 ***
## FirstAuthorFemale1 0.06100 0.10317 0.59 0.55481
## LastAuthorFemale1 -0.20649 0.15934 -1.30 0.19605
## UniqueAuthors2 0.34633 0.10237 3.38 0.00082 ***
## UniqueAuthors3 0.54509 0.12016 4.54 8.4e-06 ***
## UniqueAuthors4 0.46544 0.14664 3.17 0.00167 **
## UniqueAuthors5 0.52537 0.18252 2.88 0.00430 **
## Year1997 -0.21571 0.25702 -0.84 0.40202
## Year1998 0.29701 0.26126 1.14 0.25656
## Year1999 0.08250 0.25783 0.32 0.74921
```

```

## Year2000      -0.16258      0.30975      -0.52      0.60006
## Year2001      0.62253      0.37936      1.64      0.10190
## Year2002     -0.00913      0.24418      -0.04      0.97021
## Year2003      0.06965      0.21733      0.32      0.74882
## Year2004      0.05030      0.21893      0.23      0.81844
## Year2005      0.23954      0.25194      0.95      0.34251
## Year2006     -0.11483      0.21352      -0.54      0.59112
## Year2007      0.02156      0.22680      0.10      0.92435
## Year2008     -0.15668      0.25205      -0.62      0.53470
## Year2009     -0.42277      0.24066      -1.76      0.08004 .
## Year2010     -0.21965      0.22758      -0.97      0.33530
## Year2011     -0.06499      0.22306      -0.29      0.77100
## Year2012     -0.30201      0.22995      -1.31      0.19011
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.624
## Multiple R-squared:  0.171, Adjusted R-squared:  0.108
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 283 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.387  0.868  0.949  0.908  0.986  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.24e-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 1.255 1           1.120
## LastAuthorFemale  1.323 1           1.150
## Year              1.581 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
## -1.42229 -0.43446 -0.00992  0.45423  1.80665
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9873     0.2088   4.73 3.5e-06 ***
## FirstAuthorFemale1  0.1174     0.1056   1.11  0.27
## LastAuthorFemale1 -0.1782     0.1690  -1.05  0.29
## Year1997          -0.1013     0.2459  -0.41  0.68
## Year1998           0.3024     0.2777   1.09  0.28
## Year1999           0.2050     0.2590   0.79  0.43
## Year2000          -0.0560     0.3015  -0.19  0.85
## Year2001           0.5992     0.4171   1.44  0.15
## Year2002           0.1791     0.2734   0.66  0.51
## Year2003           0.1887     0.2373   0.80  0.43
## Year2004           0.1514     0.2492   0.61  0.54
## Year2005           0.4349     0.2640   1.65  0.10
```

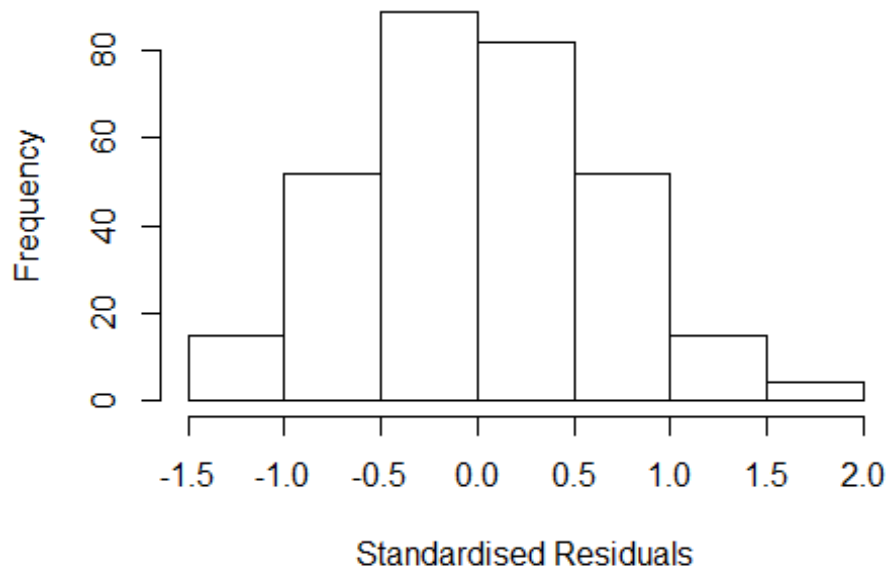


```

## Year2006          0.0767      0.2264      0.34      0.74
## Year2007          0.1701      0.2319      0.73      0.46
## Year2008          0.0258      0.2706      0.10      0.92
## Year2009         -0.1554      0.2470     -0.63      0.53
## Year2010         -0.0546      0.2748     -0.20      0.84
## Year2011          0.0792      0.2374      0.33      0.74
## Year2012         -0.1769      0.2725     -0.65      0.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.669
## Multiple R-squared:  0.0746, Adjusted R-squared:  0.0172
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 287 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.446  0.871  0.956  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      3.24e-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 1.189 1      1.091
## Year              1.189 16      1.005

```

Residuals from first author



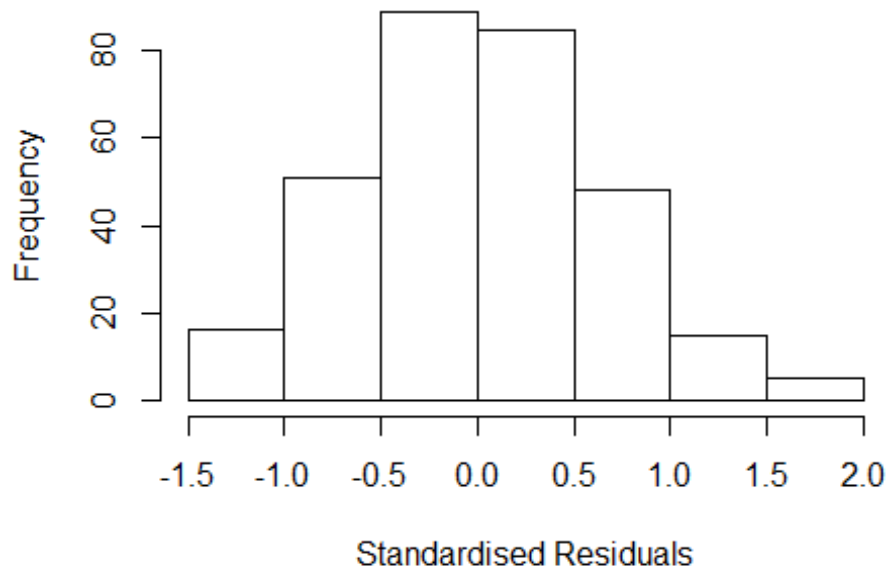
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.43022 -0.42910 -0.00642  0.46505  1.82658
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9587     0.2020   4.75 3.2e-06 ***
## FirstAuthorFemale1  0.0859     0.1059   0.81  0.418
## Year1997        -0.1133     0.2522  -0.45  0.654
## Year1998         0.3175     0.2768   1.15  0.252
## Year1999         0.2137     0.2588   0.83  0.410
## Year2000        -0.0376     0.2947  -0.13  0.899
## Year2001         0.6351     0.3878   1.64  0.103
## Year2002         0.2079     0.2687   0.77  0.440
## Year2003         0.2172     0.2323   0.94  0.351
## Year2004         0.1740     0.2474   0.70  0.483
## Year2005         0.4715     0.2564   1.84  0.067 .
## Year2006         0.0925     0.2252   0.41  0.682
```

```

## Year2007          0.2016      0.2244      0.90      0.370
## Year2008          0.0400      0.2682      0.15      0.882
## Year2009         -0.1363      0.2452     -0.56      0.579
## Year2010         -0.0248      0.2649     -0.09      0.926
## Year2011          0.0852      0.2362      0.36      0.719
## Year2012         -0.1616      0.2672     -0.60      0.546
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.666
## Multiple R-squared:  0.0714, Adjusted R-squared:  0.0171
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 284 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.432  0.868  0.954  0.911  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.24e-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 1.264 1      1.124
## Year            1.264 16      1.007

```

Residuals from last author



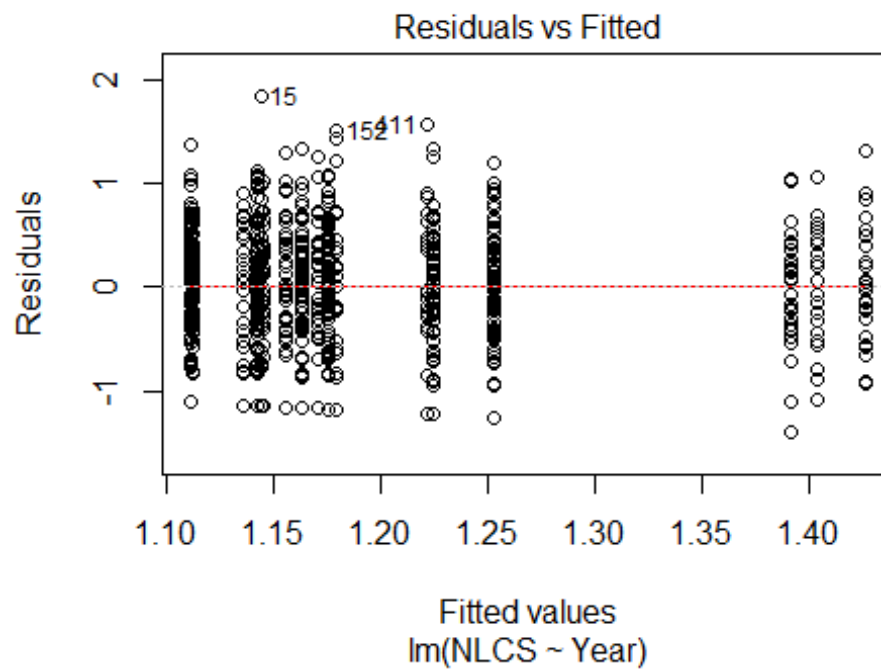
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.45212 -0.44361 -0.00219  0.46719  1.79461
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9905     0.2098   4.72 3.7e-06 ***
## LastAuthorFemale1 -0.1484     0.1665  -0.89  0.373
## Year1997       -0.1113     0.2473  -0.45  0.653
## Year1998        0.3075     0.2784   1.10  0.270
## Year1999        0.2139     0.2642   0.81  0.419
## Year2000       -0.0528     0.3060  -0.17  0.863
## Year2001        0.6147     0.4028   1.53  0.128
## Year2002        0.1767     0.2745   0.64  0.520
## Year2003        0.2103     0.2371   0.89  0.376
## Year2004        0.1546     0.2511   0.62  0.539
## Year2005        0.4616     0.2611   1.77  0.078 .
## Year2006        0.0804     0.2295   0.35  0.726
```

```

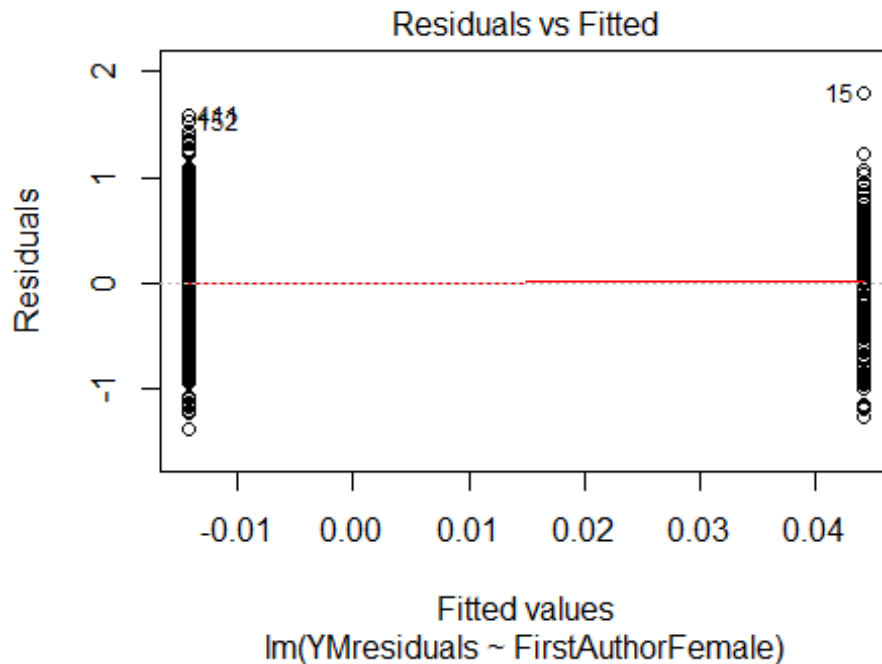
## Year2007          0.1779      0.2321      0.77      0.444
## Year2008          0.0282      0.2696      0.10      0.917
## Year2009         -0.1428      0.2499     -0.57      0.568
## Year2010         -0.0407      0.2751     -0.15      0.882
## Year2011          0.0825      0.2384      0.35      0.729
## Year2012         -0.1629      0.2744     -0.59      0.553
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.656
## Multiple R-squared:  0.0728, Adjusted R-squared:  0.0186
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 290 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.434  0.860  0.953  0.911  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.24e-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: 309"
## [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
##   77   68   62   52   61   60   66   58   47   69  117   95  155  150  146
## 2011 2012
##  160  181
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   27   22   22   23   24   31   27   30   41   59   48   91   88   88
## 2011 2012

```

```
## 98 103
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 22 17 21 19 20 23 25 24 25 27 46 38 74 77 69
## 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 = 29, df = 16, p-value = 0.02
```

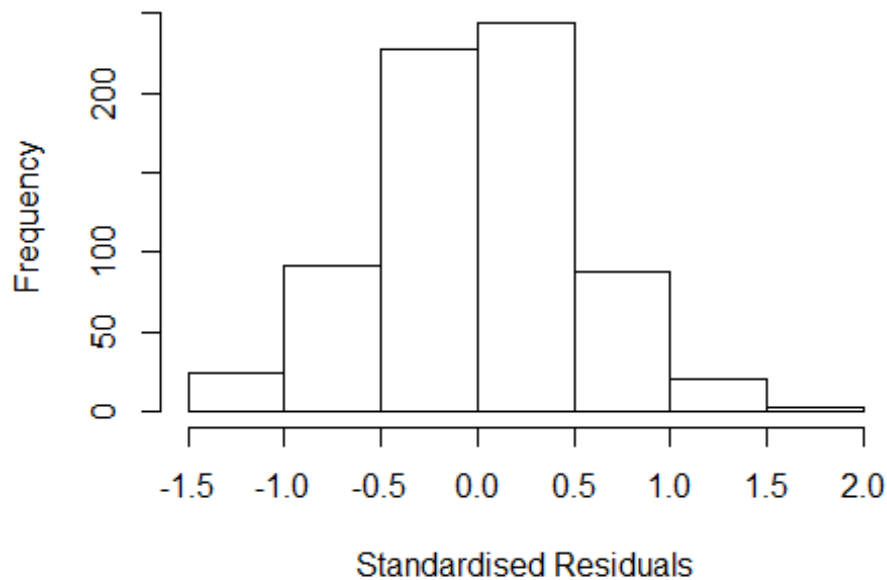


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.09792001677172"
## [1] "Male first author team size 2018 geometric mean: 2.59251705306673"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.62366796847238"
## [1] "Male last author team size 2018 geometric mean: 2.74745812845723"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 890, 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 1.282 1          1.132
## LastAuthorFemale  1.259 1          1.122
## UniqueAuthors    1.746 4          1.072
## Year              1.738 16         1.017
```

Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.396 -0.316 0.011 0.347 1.884
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.78277 0.16530 4.74 2.7e-06 ***
## FirstAuthorFemale1 0.01809 0.04933 0.37 0.714
## LastAuthorFemale1 0.09508 0.05239 1.81 0.070 .
## UniqueAuthors2 0.29843 0.06222 4.80 2.0e-06 ***
## UniqueAuthors3 0.38959 0.06315 6.17 1.2e-09 ***
## UniqueAuthors4 0.35292 0.07045 5.01 7.0e-07 ***
## UniqueAuthors5 0.51642 0.07487 6.90 1.2e-11 ***
## Year1997 0.26981 0.30416 0.89 0.375
## Year1998 0.32747 0.21902 1.50 0.135
## Year1999 0.00524 0.22445 0.02 0.981
```

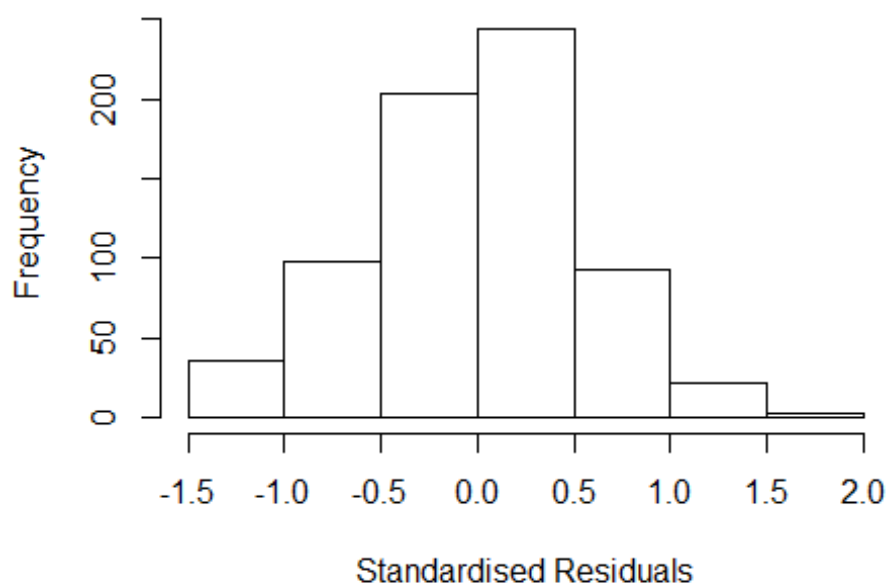


```

## Year2000      -0.05639    0.18810   -0.30    0.764
## Year2001      0.15984    0.21299    0.75    0.453
## Year2002      0.27993    0.18818    1.49    0.137
## Year2003      0.34348    0.20265    1.69    0.091 .
## Year2004      0.10041    0.20686    0.49    0.628
## Year2005      0.22385    0.19068    1.17    0.241
## Year2006      0.11616    0.17744    0.65    0.513
## Year2007      0.02010    0.16898    0.12    0.905
## Year2008      0.06962    0.16314    0.43    0.670
## Year2009      0.09665    0.17100    0.57    0.572
## Year2010      0.07302    0.16886    0.43    0.666
## Year2011      0.01205    0.16729    0.07    0.943
## Year2012      0.01983    0.16459    0.12    0.904
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.495
## Multiple R-squared:  0.12,   Adjusted R-squared:  0.091
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 57 weights are ~= 1. The remaining 639 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.116  0.864  0.949   0.898   0.988   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.44e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.113 1      1.055
## LastAuthorFemale  1.185 1      1.089
## Year              1.119 16      1.004

```

Residuals from first and last author



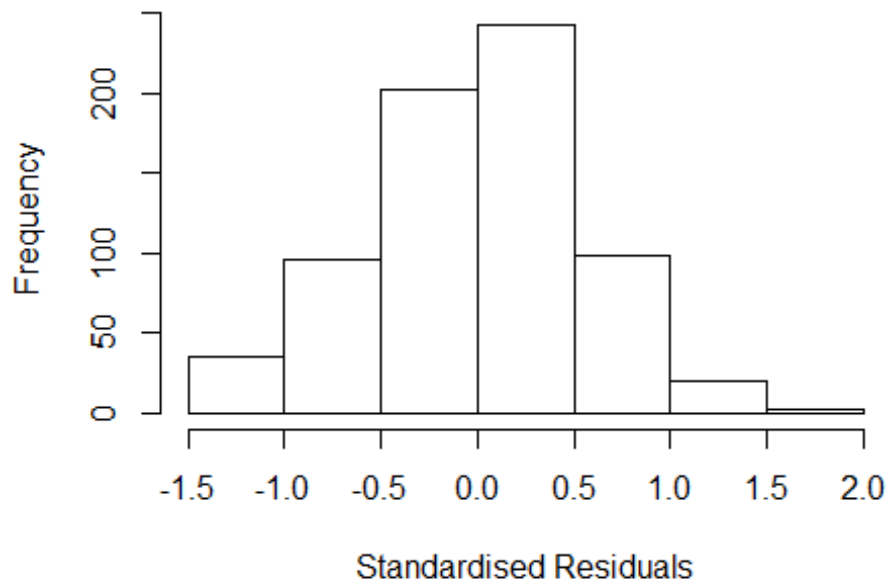
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3398 -0.3477 0.0399 0.3593 1.8260
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10431 0.17616 6.27 6.5e-10 ***
## FirstAuthorFemale1 0.05265 0.05020 1.05 0.29
## LastAuthorFemale1 0.08345 0.05533 1.51 0.13
## Year1997 0.19969 0.28711 0.70 0.49
## Year1998 0.26541 0.22570 1.18 0.24
## Year1999 -0.03677 0.23346 -0.16 0.87
## Year2000 -0.03645 0.21125 -0.17 0.86
## Year2001 0.09778 0.24091 0.41 0.68
## Year2002 0.23550 0.20369 1.16 0.25
## Year2003 0.28075 0.22100 1.27 0.20
## Year2004 0.04188 0.22399 0.19 0.85
## Year2005 0.12412 0.20617 0.60 0.55
```

```

## Year2006          0.07276      0.19401      0.38      0.71
## Year2007         -0.00701      0.18959     -0.04      0.97
## Year2008          0.02356      0.18211      0.13      0.90
## Year2009          0.05026      0.18651      0.27      0.79
## Year2010          0.06622      0.18861      0.35      0.73
## Year2011         -0.03041      0.18577     -0.16      0.87
## Year2012          0.00842      0.18565      0.05      0.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.0312, Adjusted R-squared:  0.00549
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 655 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.221  0.868  0.955  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      1.44e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.025 1      1.012
## Year              1.025 16      1.001

```

Residuals from first author



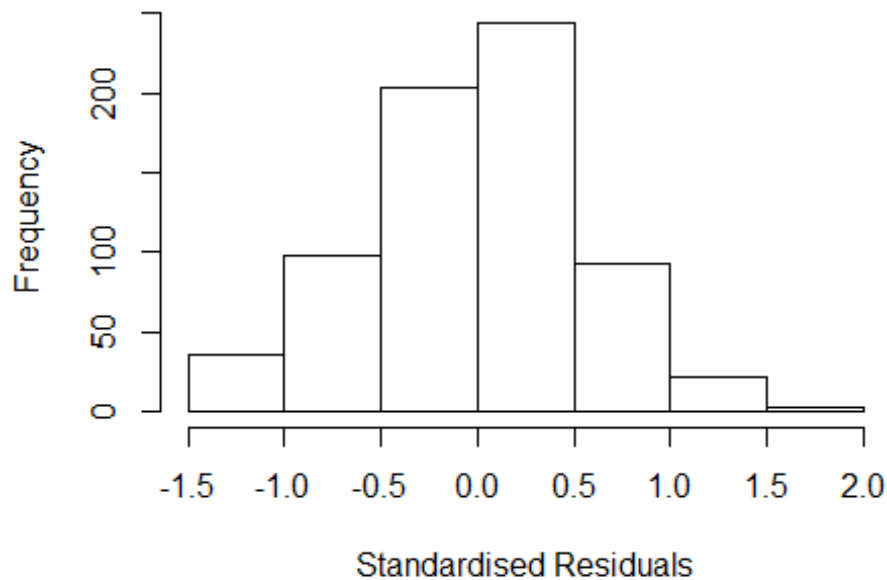
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3530 -0.3492 0.0485 0.3517 1.7952
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1102 0.1776 6.25 7.2e-10 ***
## FirstAuthorFemale1 0.0775 0.0476 1.63 0.10
## Year1997 0.1988 0.2862 0.69 0.49
## Year1998 0.2803 0.2255 1.24 0.21
## Year1999 -0.0354 0.2333 -0.15 0.88
## Year2000 -0.0188 0.2125 -0.09 0.93
## Year2001 0.0934 0.2399 0.39 0.70
## Year2002 0.2428 0.2046 1.19 0.24
## Year2003 0.2958 0.2197 1.35 0.18
## Year2004 0.0493 0.2251 0.22 0.83
## Year2005 0.1275 0.2073 0.62 0.54
## Year2006 0.0809 0.1948 0.42 0.68
```

```

## Year2007          -0.0094      0.1905   -0.05    0.96
## Year2008          0.0199      0.1833    0.11    0.91
## Year2009          0.0527      0.1875    0.28    0.78
## Year2010          0.0708      0.1893    0.37    0.71
## Year2011         -0.0229      0.1864   -0.12    0.90
## Year2012          0.0106      0.1867    0.06    0.95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.0282, Adjusted R-squared:  0.0038
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 656 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.238  0.868  0.955  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      1.44e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.088 1          1.043
## Year            1.088 16          1.003

```

Residuals from last author



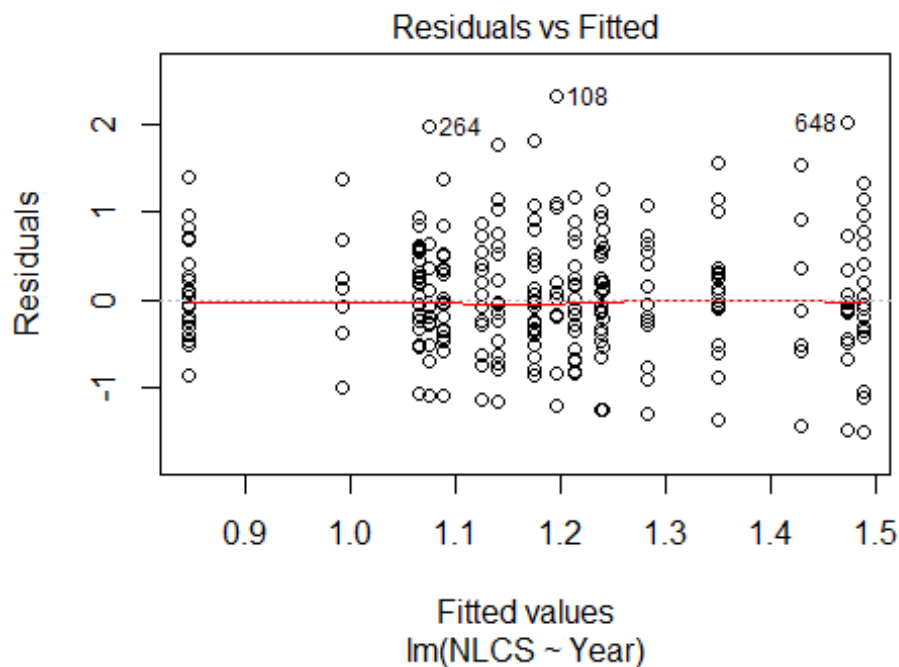
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3498 -0.3455 0.0358 0.3591 1.8657
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11728 0.17366 6.43 2.4e-10 ***
## LastAuthorFemale1 0.10267 0.05245 1.96 0.051 .
## Year1997 0.19190 0.28506 0.67 0.501
## Year1998 0.26599 0.22345 1.19 0.234
## Year1999 -0.04190 0.23208 -0.18 0.857
## Year2000 -0.04197 0.20928 -0.20 0.841
## Year2001 0.09947 0.24062 0.41 0.679
## Year2002 0.23256 0.20196 1.15 0.250
## Year2003 0.27872 0.21839 1.28 0.202
## Year2004 0.03216 0.22261 0.14 0.885
## Year2005 0.11887 0.20332 0.58 0.559
## Year2006 0.06887 0.19175 0.36 0.720
```

```

## Year2007          -0.01152      0.18806    -0.06      0.951
## Year2008           0.02041      0.18015      0.11      0.910
## Year2009           0.04757      0.18465      0.26      0.797
## Year2010           0.06030      0.18668      0.32      0.747
## Year2011          -0.03433      0.18379     -0.19      0.852
## Year2012           0.00814      0.18388      0.04      0.965
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.537
## Multiple R-squared:  0.0299, Adjusted R-squared:  0.00555
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 49 weights are ~= 1. The remaining 647 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.202  0.868  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      1.44e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 696"
## [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
##   39   53   37   36   43   37   29   35   36   45   54   55   44   35   45
## 2011 2012
##   36   42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   11   18   12    8   20   11   10   20   14   14   19   26   21   14   18
## 2011 2012

```

```
## 24 25
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 9 14 8 8 15 9 7 13 10 14 16 19 15 9 10
## 2011 2012
## 23 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 = 18, df = 16, p-value = 0.3
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.42, df = 1, p-value = 0.5

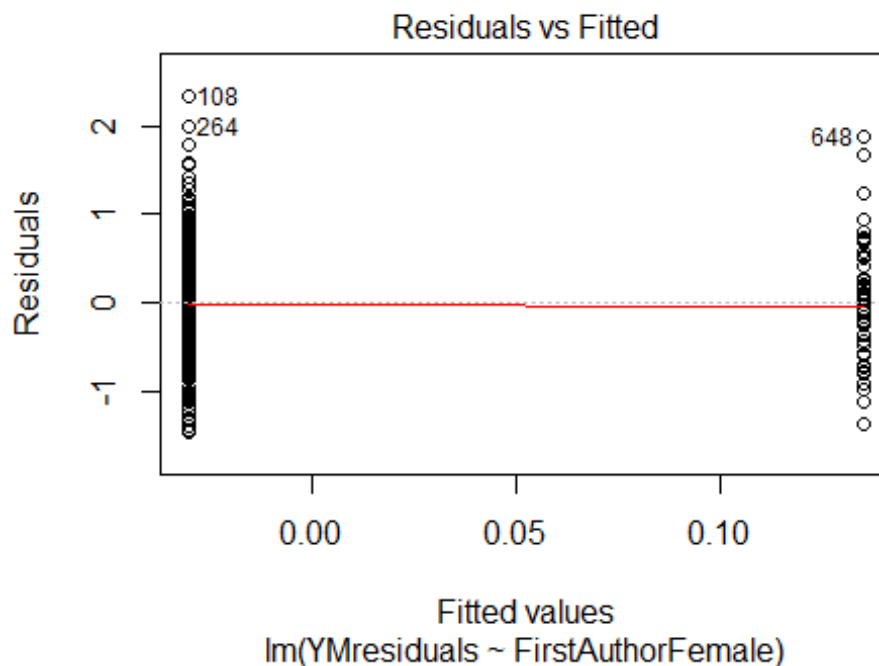
## [1] "Female first author team size 2018 geometric mean: 1.54221082540794"
## [1] "Male first author team size 2018 geometric mean: 1.99349609802557"

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



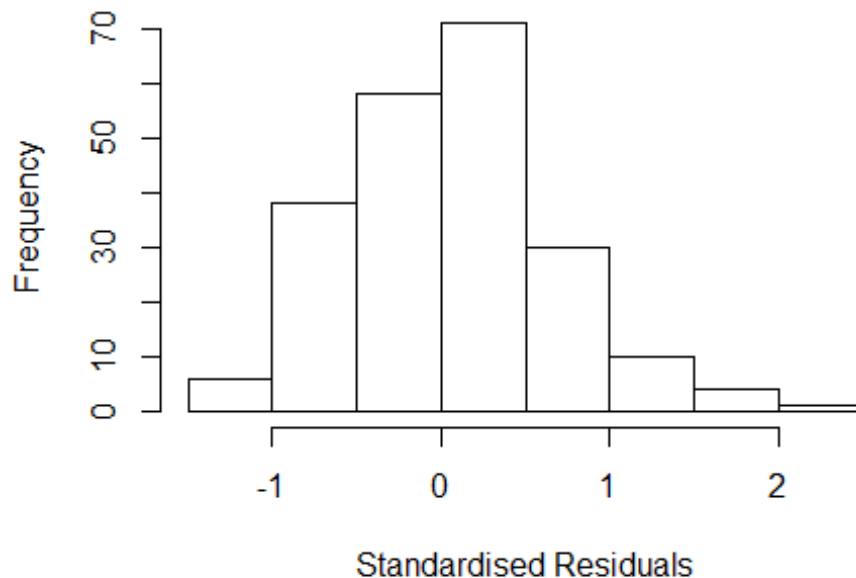
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 44, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.54221082540794"
## [1] "Male last author team size 2018 geometric mean: 1.99349609802557"

## Warning in wilcox.test.default(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 = 0.2
## 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 1.715  1      1.309
## LastAuthorFemale  1.396  1      1.181
## UniqueAuthors    3.376  4      1.164
## Year              4.647 16      1.049
```

Residuals from first and last author and team size



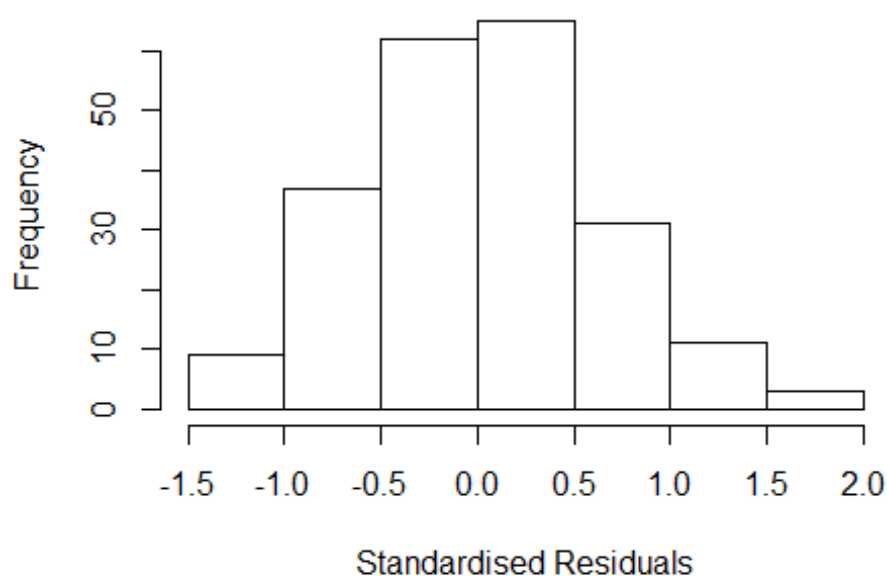
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2938 -0.4247 0.0208 0.3999 2.1589
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7037 0.3361 2.09 0.03758 *
## FirstAuthorFemale1 0.2008 0.1418 1.42 0.15825
## LastAuthorFemale1 0.1570 0.1458 1.08 0.28283
## UniqueAuthors2 0.4243 0.1208 3.51 0.00055 ***
## UniqueAuthors3 0.4650 0.1256 3.70 0.00028 ***
## UniqueAuthors4 0.1477 0.2700 0.55 0.58497
## UniqueAuthors5 0.4043 0.1685 2.40 0.01734 *
## Year1997 0.2752 0.3635 0.76 0.44991
## Year1998 -0.1136 0.4099 -0.28 0.78198
## Year1999 -0.0420 0.4102 -0.10 0.91854
```

```

## Year2000          -0.1023      0.3509   -0.29  0.77085
## Year2001           0.0241      0.3466    0.07  0.94463
## Year2002          -0.0257      0.4601   -0.06  0.95560
## Year2003           0.1251      0.3621    0.35  0.73014
## Year2004           0.1988      0.3443    0.58  0.56437
## Year2005           0.0950      0.3556    0.27  0.78961
## Year2006           0.0426      0.3359    0.13  0.89933
## Year2007          -0.1346      0.3372   -0.40  0.69019
## Year2008           0.0825      0.3567    0.23  0.81732
## Year2009           0.2843      0.4121    0.69  0.49107
## Year2010           0.0759      0.3651    0.21  0.83558
## Year2011           0.0579      0.3342    0.17  0.86259
## Year2012          -0.3644      0.3297   -1.11  0.27043
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.639
## Multiple R-squared:  0.154, Adjusted R-squared:  0.0586
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 200 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.231  0.889  0.955  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      4.59e-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 1.326 1      1.152
## LastAuthorFemale  1.437 1      1.199
## Year              1.733 16      1.017

```

Residuals from first and last author



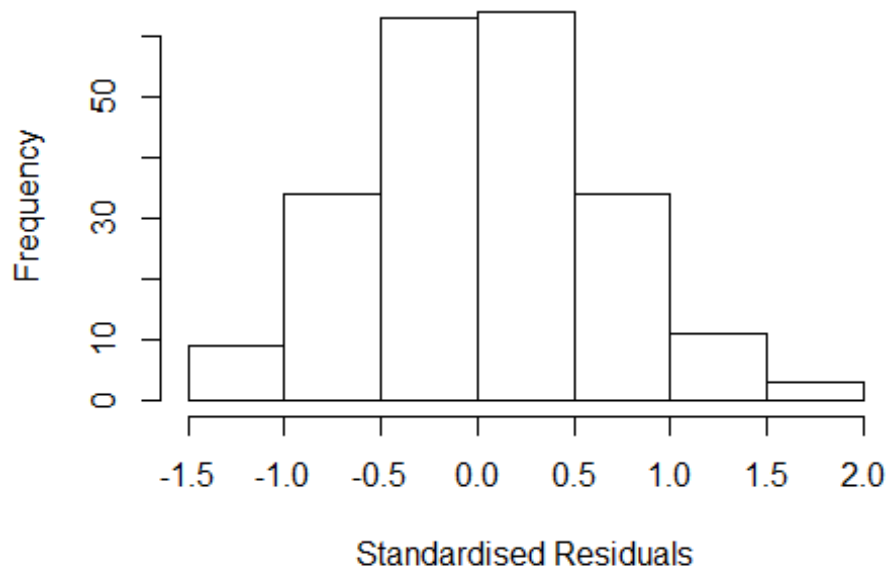
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.44228 -0.40448 0.00451 0.42572 1.94337
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0581 0.3167 3.34 0.001 ***
## FirstAuthorFemale1 0.2492 0.1348 1.85 0.066 .
## LastAuthorFemale1 0.1350 0.1567 0.86 0.390
## Year1997 0.1712 0.3639 0.47 0.639
## Year1998 -0.1945 0.4407 -0.44 0.659
## Year1999 -0.1944 0.4307 -0.45 0.652
## Year2000 -0.0935 0.3737 -0.25 0.803
## Year2001 -0.2295 0.3541 -0.65 0.518
## Year2002 -0.1057 0.4973 -0.21 0.832
## Year2003 0.1758 0.3669 0.48 0.632
## Year2004 0.2310 0.3635 0.64 0.526
## Year2005 -0.0174 0.3585 -0.05 0.961
```

```

## Year2006          0.0242      0.3438      0.07      0.944
## Year2007         -0.1166      0.3490     -0.33      0.739
## Year2008          0.0342      0.3691      0.09      0.926
## Year2009          0.2781      0.4237      0.66      0.512
## Year2010         -0.0492      0.3518     -0.14      0.889
## Year2011          0.0246      0.3414      0.07      0.943
## Year2012         -0.3630      0.3341     -1.09      0.279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.658
## Multiple R-squared:  0.0858, Adjusted R-squared:  0.00316
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 192 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.363  0.860  0.956  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      4.59e-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 1.298 1      1.139
## Year              1.298 16      1.008

```

Residuals from first author



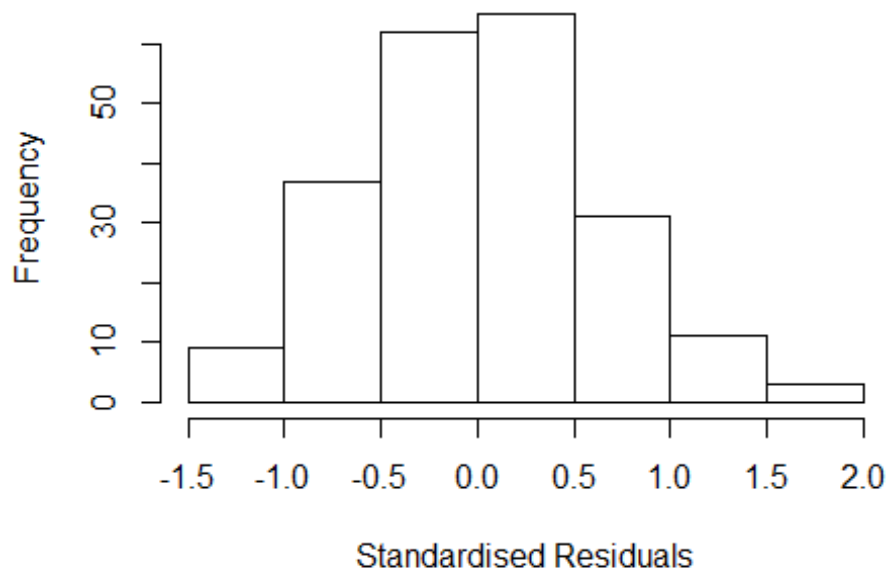
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3961 -0.4040 0.0136 0.4376 1.9363
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1120 0.3189 3.49 0.0006 ***
## FirstAuthorFemale1 0.2841 0.1301 2.18 0.0301 *
## Year1997 0.1266 0.3684 0.34 0.7315
## Year1998 -0.2266 0.4557 -0.50 0.6196
## Year1999 -0.2221 0.4340 -0.51 0.6094
## Year2000 -0.1403 0.3734 -0.38 0.7074
## Year2001 -0.2704 0.3600 -0.75 0.4534
## Year2002 -0.1678 0.4988 -0.34 0.7369
## Year2003 0.1341 0.3739 0.36 0.7202
## Year2004 0.1988 0.3703 0.54 0.5920
## Year2005 -0.0773 0.3542 -0.22 0.8274
## Year2006 -0.0234 0.3449 -0.07 0.9461
```

```

## Year2007          -0.1686      0.3523   -0.48   0.6327
## Year2008          -0.0249      0.3689   -0.07   0.9463
## Year2009           0.2202      0.4268    0.52   0.6065
## Year2010          -0.0665      0.3598   -0.18   0.8537
## Year2011          -0.0148      0.3469   -0.04   0.9659
## Year2012          -0.4070      0.3395   -1.20   0.2321
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.654
## Multiple R-squared:  0.0825, Adjusted R-squared:  0.00449
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 191 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.360  0.857   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      4.59e-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 1.31 1      1.145
## Year              1.31 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
## -1.3510 -0.4190 -0.0102 0.4116 2.1350
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0621 0.2998 3.54 0.00049 ***
## LastAuthorFemale1 0.2224 0.1458 1.53 0.12877
## Year1997 0.1824 0.3472 0.53 0.59996
## Year1998 -0.2134 0.4240 -0.50 0.61533
## Year1999 -0.1563 0.4198 -0.37 0.71012
## Year2000 -0.0934 0.3527 -0.26 0.79149
## Year2001 -0.2126 0.3383 -0.63 0.53048
## Year2002 -0.0557 0.4734 -0.12 0.90649
## Year2003 0.1829 0.3476 0.53 0.59947
## Year2004 0.2813 0.3547 0.79 0.42874
## Year2005 0.0626 0.3413 0.18 0.85457
## Year2006 0.0722 0.3306 0.22 0.82745
```

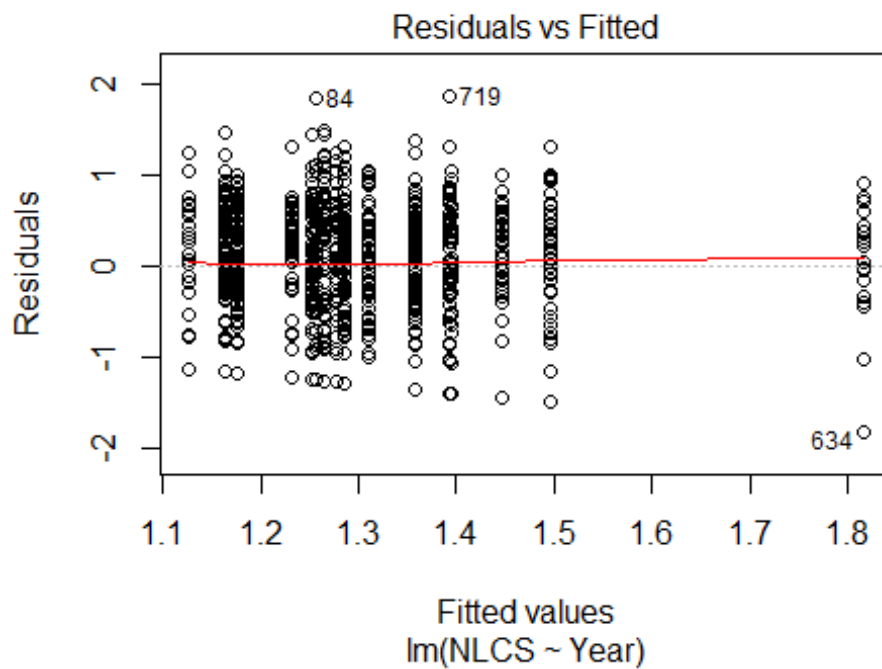


```

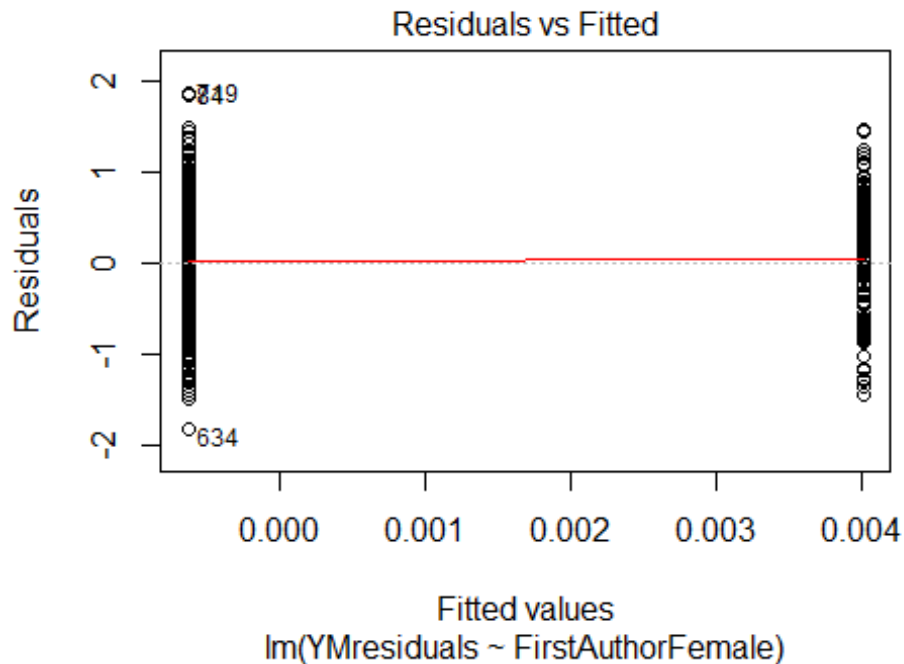
## Year2007          -0.0852      0.3349   -0.25  0.79942
## Year2008           0.0781      0.3481    0.22  0.82271
## Year2009           0.2889      0.4068    0.71  0.47844
## Year2010          -0.0176      0.3380   -0.05  0.95846
## Year2011           0.0246      0.3269    0.08  0.94011
## Year2012          -0.2907      0.3155   -0.92  0.35785
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.657
## Multiple R-squared:  0.0663, Adjusted R-squared:  -0.0131
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 196 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.269  0.859  0.959  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      4.59e-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: 218"
## [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
## 102 55 106 90 93 93 86 95 99 115 105 101 123 162 167
## 2011 2012
## 173 209
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 42 28 51 43 25 21 40 49 43 57 55 57 65 75 98
## 2011 2012

```

```
## 104 122
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 35 21 46 36 22 17 30 39 36 41 40 44 56 67 79
## 2011 2012
## 99 109
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 27, df = 16, p-value = 0.04
```

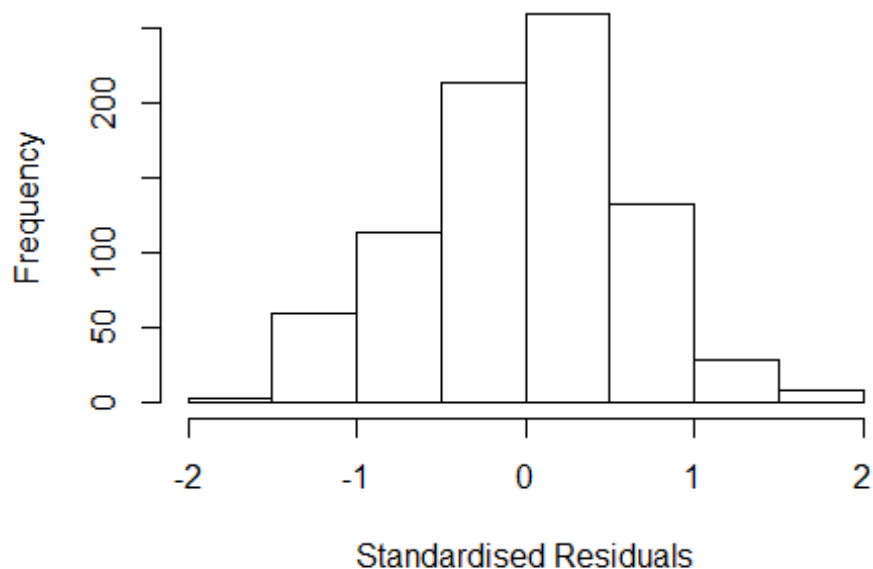


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4, df = 1, p-value = 0.05
```



```
## [1] "Female first author team size 2018 geometric mean: 2.69103449217768"
## [1] "Male first author team size 2018 geometric mean: 2.60374762026327"
##
## 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] "Female last author team size 2018 geometric mean: 2.81522372900289"
## [1] "Male last author team size 2018 geometric mean: 2.60001641378006"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, 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 1.107 1 1.052
## LastAuthorFemale 1.140 1 1.068
## UniqueAuthors 1.486 4 1.051
## Year 1.593 16 1.015
```

Residuals from first and last author and team size



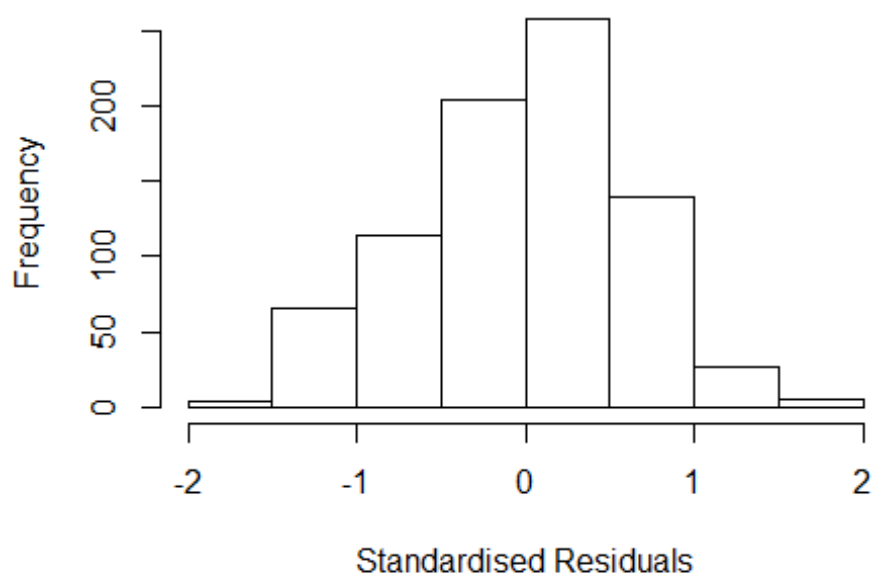
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5410 -0.4119 0.0322 0.4291 1.8801
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07769 0.13658 7.89 1.0e-14 ***
## FirstAuthorFemale1 -0.07462 0.08539 -0.87 0.38245
## LastAuthorFemale1 0.00112 0.09234 0.01 0.99030
## UniqueAuthors2 0.27811 0.07766 3.58 0.00036 ***
## UniqueAuthors3 0.38062 0.07759 4.91 1.1e-06 ***
## UniqueAuthors4 0.34198 0.09645 3.55 0.00041 ***
## UniqueAuthors5 0.46325 0.14687 3.15 0.00167 **
## Year1997 -0.27035 0.21020 -1.29 0.19878
## Year1998 -0.08061 0.20268 -0.40 0.69096
## Year1999 0.13589 0.19168 0.71 0.47857
```

```

## Year2000          0.13020    0.19389    0.67  0.50208
## Year2001          0.40062    0.18764    2.14  0.03306 *
## Year2002          0.01908    0.22776    0.08  0.93326
## Year2003          0.18521    0.17234    1.07  0.28285
## Year2004          0.04488    0.16506    0.27  0.78576
## Year2005         -0.03705    0.18426   -0.20  0.84068
## Year2006         -0.01782    0.17308   -0.10  0.91803
## Year2007         -0.07722    0.17227   -0.45  0.65408
## Year2008         -0.05788    0.16404   -0.35  0.72427
## Year2009         -0.11581    0.15619   -0.74  0.45863
## Year2010         -0.02676    0.15465   -0.17  0.86269
## Year2011         -0.16105    0.15216   -1.06  0.29018
## Year2012         -0.21658    0.15458   -1.40  0.16159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.617
## Multiple R-squared:  0.0771, Adjusted R-squared:  0.0515
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 747 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.333  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          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 1.073 1          1.036
## LastAuthorFemale  1.064 1          1.032
## Year              1.129 16          1.004

```

Residuals from first and last author



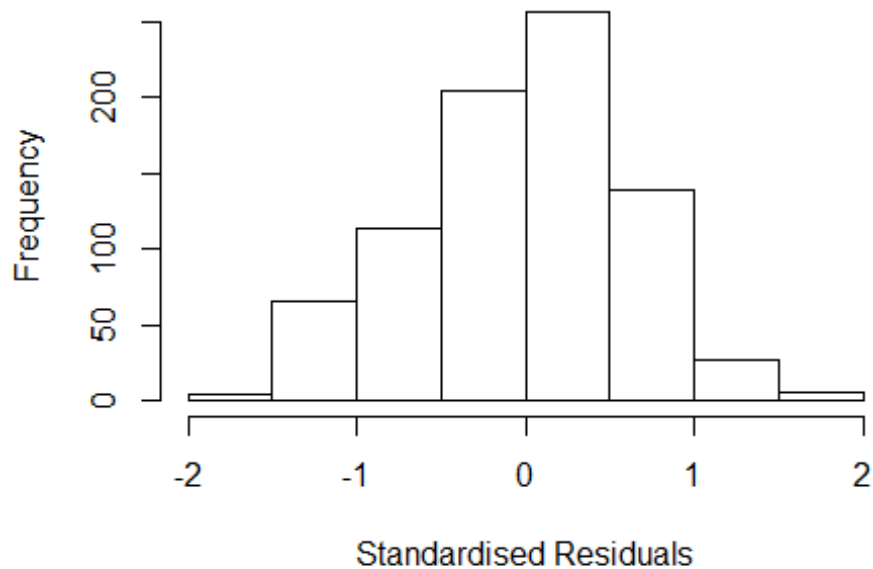
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7387 -0.4049 0.0337 0.4293 1.8766
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21736 0.12501 9.74 <2e-16 ***
## FirstAuthorFemale1 -0.02708 0.08449 -0.32 0.749
## LastAuthorFemale1 0.00943 0.09185 0.10 0.918
## Year1997 -0.19063 0.20666 -0.92 0.357
## Year1998 -0.00893 0.20207 -0.04 0.965
## Year1999 0.24953 0.18169 1.37 0.170
## Year2000 0.23698 0.18357 1.29 0.197
## Year2001 0.52137 0.17488 2.98 0.003 **
## Year2002 0.18125 0.19821 0.91 0.361
## Year2003 0.29650 0.15967 1.86 0.064 .
## Year2004 0.18837 0.15123 1.25 0.213
## Year2005 0.07538 0.17708 0.43 0.670
```

```

## Year2006          0.11090      0.16269      0.68      0.496
## Year2007          0.05050      0.16765      0.30      0.763
## Year2008          0.08170      0.14796      0.55      0.581
## Year2009          0.02376      0.14378      0.17      0.869
## Year2010          0.11193      0.13881      0.81      0.420
## Year2011         -0.01705      0.13693     -0.12      0.901
## Year2012         -0.05866      0.13903     -0.42      0.673
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.623
## Multiple R-squared:  0.0375, Adjusted R-squared:  0.0158
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 744 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.344  0.862  0.949  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      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 1.066 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
## -1.7388 -0.4053 0.0329 0.4288 1.8764
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21758 0.12497 9.74 <2e-16 ***
## FirstAuthorFemale1 -0.02637 0.08538 -0.31 0.758
## Year1997 -0.19018 0.20678 -0.92 0.358
## Year1998 -0.00877 0.20246 -0.04 0.965
## Year1999 0.25032 0.18172 1.38 0.169
## Year2000 0.23734 0.18357 1.29 0.196
## Year2001 0.52123 0.17484 2.98 0.003 **
## Year2002 0.18141 0.19833 0.91 0.361
## Year2003 0.29672 0.15970 1.86 0.064 .
## Year2004 0.18850 0.15129 1.25 0.213
## Year2005 0.07661 0.17671 0.43 0.665
## Year2006 0.11132 0.16283 0.68 0.494
```



```

## Year2007      0.05087    0.16782    0.30    0.762
## Year2008      0.08286    0.14792    0.56    0.576
## Year2009      0.02410    0.14390    0.17    0.867
## Year2010      0.11248    0.13900    0.81    0.419
## Year2011     -0.01635    0.13701   -0.12    0.905
## Year2012     -0.05815    0.13908   -0.42    0.676
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.622
## Multiple R-squared:  0.0375, Adjusted R-squared:  0.017
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 743 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.343  0.861  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      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 1.057 1      1.028
## Year      1.057 16      1.002

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.7358 -0.4058  0.0327  0.4290  1.8802

```

```

##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.21379    0.12383     9.80  <2e-16 ***
## LastAuthorFemale1 0.00646    0.09288     0.07   0.945
## Year1997       -0.18839    0.20578    -0.92   0.360
## Year1998       -0.00901    0.20121    -0.04   0.964
## Year1999        0.24964    0.18214     1.37   0.171
## Year2000        0.23804    0.18372     1.30   0.195
## Year2001        0.52202    0.17540     2.98   0.003 **
## Year2002        0.18057    0.19822     0.91   0.363
## Year2003        0.29630    0.15942     1.86   0.063 .
## Year2004        0.19201    0.14988     1.28   0.201
## Year2005        0.07722    0.17682     0.44   0.662
## Year2006        0.11382    0.16140     0.71   0.481
## Year2007        0.05094    0.16715     0.30   0.761
## Year2008        0.08309    0.14744     0.56   0.573
## Year2009        0.02350    0.14358     0.16   0.870
## Year2010        0.11135    0.13866     0.80   0.422
## Year2011       -0.01699    0.13662    -0.12   0.901
## Year2012       -0.05774    0.13864    -0.42   0.677
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.624
## Multiple R-squared:  0.0373, Adjusted R-squared:  0.0168
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 747 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.344  0.862  0.950  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           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: 817"
## [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
##    3   10    5    5    6    9    8   13    7   10   13   18   22   26   26
## 2011 2012
##   31   22
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    5    3    1    2    3    5    8    5    6    5   16   14   17   19
## 2011 2012
##   24   18
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    4    3    0    2    3    2    5    4    5    4   12   12   17   16
## 2011 2012
##   22   17
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.08707383067606"
## [1] "Male first author team size 2018 geometric mean: 1.68329952873076"

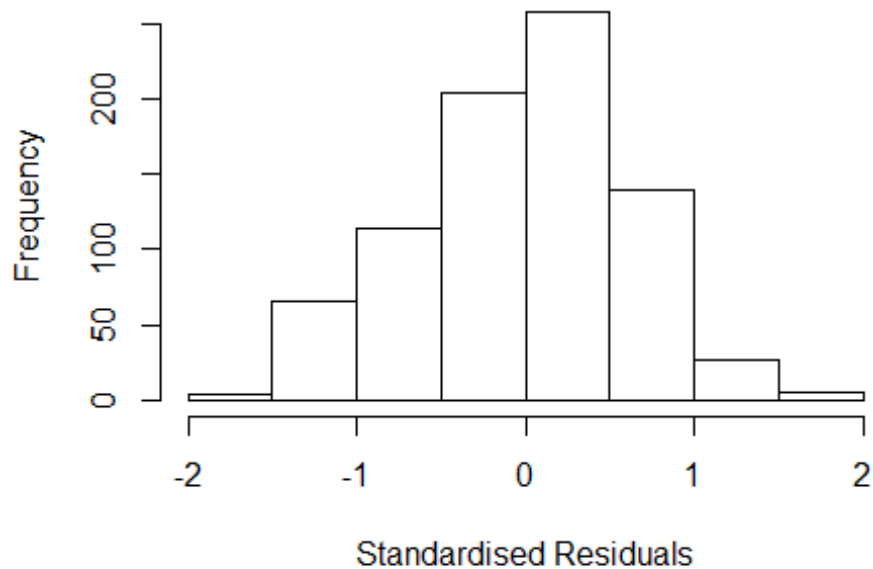
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 130, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.78179743628068"
## [1] "Male last author team size 2018 geometric mean: 1.76247267481325"

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

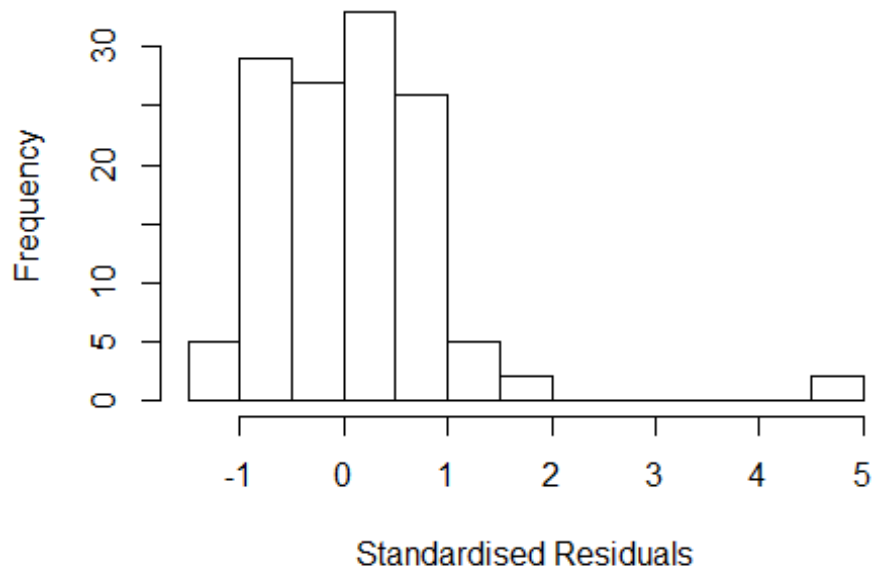
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 90, 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 12.936 1      3.597
## LastAuthorFemale  7.992 1      2.827
## UniqueAuthors    61.025 4      1.672
## Year              209.283 15     1.195
```

Residuals from first and last author and team size



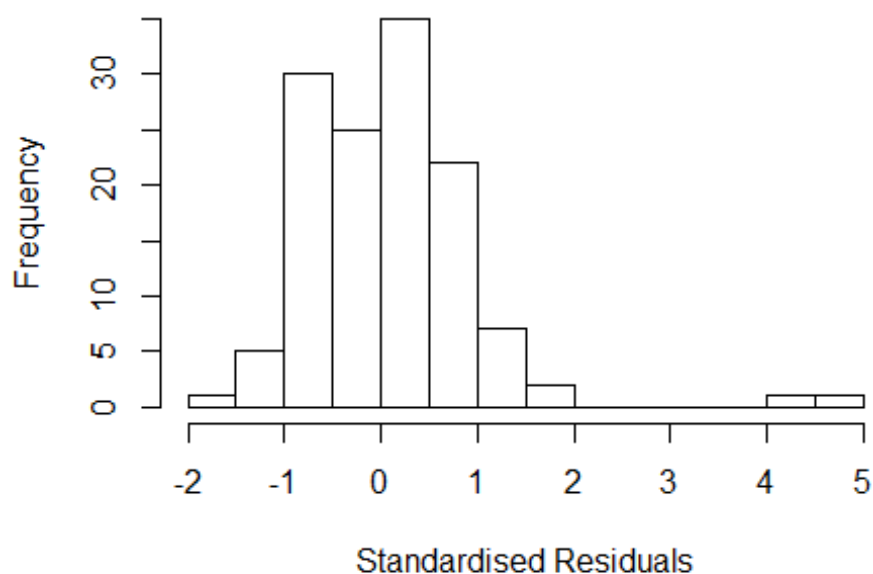
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 122 33847790037 5.209 2007    2216      1    4.501
## 146 45149134975 5.320 2008    2216      1    4.664
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4274 -0.5480  0.0428  0.5267  4.6635
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.322922   0.250398   5.28  6.7e-07 ***
## FirstAuthorFemale1 -0.000921   0.204022   0.00  0.99640
## LastAuthorFemale1  0.095223   0.243823   0.39  0.69691
## UniqueAuthors2     0.148000   0.175123   0.85  0.39993
## UniqueAuthors3     0.452766   0.198887   2.28  0.02481 *
## UniqueAuthors4     0.900013   0.243174   3.70  0.00034 ***
## UniqueAuthors5     1.081669   0.222916   4.85  4.2e-06 ***
## Year1997          -0.417919   0.381405  -1.10  0.27566
## Year1998          -0.842739   0.400530  -2.10  0.03772 *
```

```

## Year2000      -0.154922    0.494512   -0.31  0.75468
## Year2001      0.104503    0.850345    0.12  0.90242
## Year2002     -0.593922    0.666804   -0.89  0.37509
## Year2003     -0.502638    0.380982   -1.32  0.18988
## Year2004     -1.371783    0.257204   -5.33  5.4e-07 ***
## Year2005     -0.426764    0.493937   -0.86  0.38952
## Year2006     -0.900335    0.501603   -1.79  0.07549 .
## Year2007     -0.614509    0.326960   -1.88  0.06290 .
## Year2008     -0.760771    0.309988   -2.45  0.01573 *
## Year2009     -0.502518    0.269908   -1.86  0.06537 .
## Year2010     -0.567969    0.283562   -2.00  0.04771 *
## Year2011     -0.384567    0.283618   -1.36  0.17797
## Year2012     -0.589423    0.289584   -2.04  0.04428 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.778
## Multiple R-squared:  0.241, Adjusted R-squared:  0.0915
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(42,51) are outliers with |weight| = 0 ( < 0.00078);
## 11 weights are ~= 1. The remaining 116 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.558  0.905  0.954  0.933  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           7.75e-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 7.864 1 2.804
## LastAuthorFemale 3.555 1 1.886
## Year 15.011 15 1.094

```

Residuals from first and last author



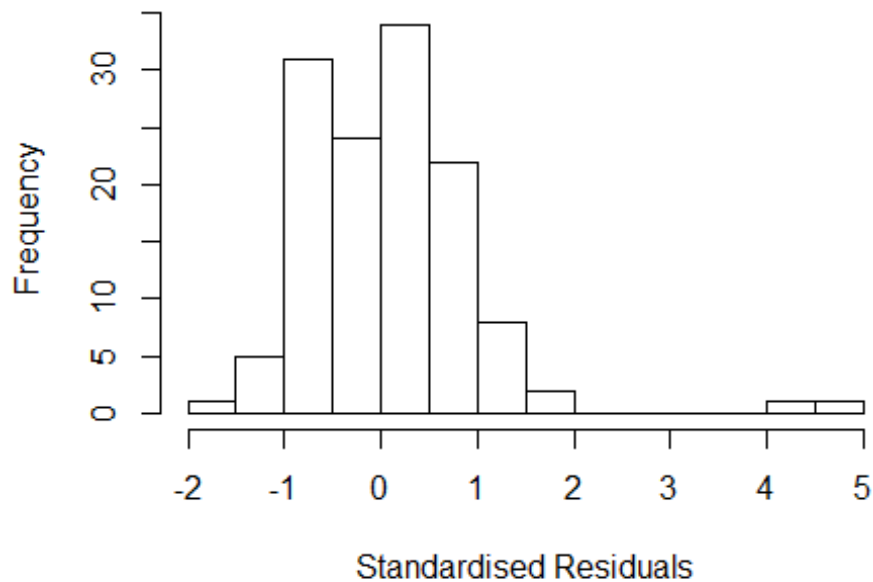
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 122 33847790037 5.209 2007      2216      1      4.252
## 146 45149134975 5.320 2008      2216      1      4.624
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.660 -0.561  0.075  0.548  4.624
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3813    0.2065   6.69 9.5e-10 ***
## FirstAuthorFemale1  0.0887    0.2065   0.43  0.668
## LastAuthorFemale1 -0.0709    0.2216  -0.32  0.750
## Year1997          -0.3469    0.3963  -0.88  0.383
## Year1998          -0.5579    0.4161  -1.34  0.183
## Year2000          -0.2133    0.4706  -0.45  0.651
## Year2001           0.2787    0.8254   0.34  0.736
## Year2002          -0.6523    0.6393  -1.02  0.310
## Year2003          -0.4684    0.3722  -1.26  0.211
## Year2004          -1.2445    0.2328  -5.35 4.9e-07 ***
## Year2005          -0.3111    0.4450  -0.70  0.486
```

```

## Year2006          -0.8202      0.4186   -1.96    0.053 .
## Year2007          -0.4242      0.3121   -1.36    0.177
## Year2008          -0.7033      0.3145   -2.24    0.027 *
## Year2009          -0.3293      0.2396   -1.37    0.172
## Year2010          -0.4212      0.2780   -1.52    0.133
## Year2011          -0.0768      0.2324   -0.33    0.742
## Year2012          -0.4129      0.2692   -1.53    0.128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.817
## Multiple R-squared:  0.114, Adjusted R-squared:  -0.0215
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(42,51) are outliers with |weight| = 0 ( < 0.00078);
## 9 weights are ~1. The remaining 118 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.614  0.890  0.953   0.929   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.75e-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 4.234 1          2.058
## Year              4.234 15          1.049

```


Residuals from first author



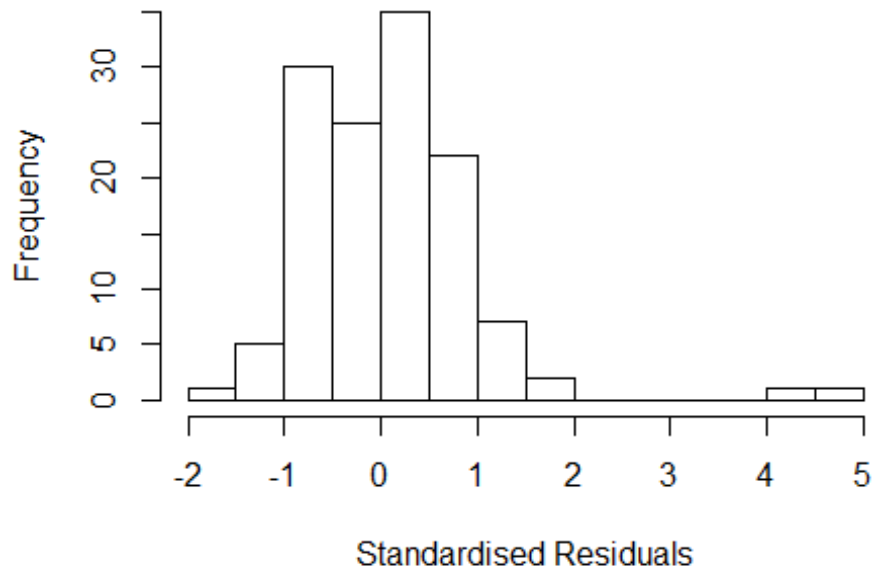
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 122 33847790037 5.209 2007      2216      1      4.252
## 146 45149134975 5.320 2008      2216      1      4.624
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6509 -0.5624  0.0666  0.5480  4.5962
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4254     0.1529   9.33 1.2e-15 ***
## FirstAuthorFemale1  0.0446     0.1529   0.29  0.771
## Year1997         -0.3978     0.3612  -1.10  0.273
## Year1998         -0.6010     0.3921  -1.53  0.128
## Year2000         -0.2574     0.4483  -0.57  0.567
## Year2001          0.2255     0.8056   0.28  0.780
## Year2002         -0.6964     0.6211  -1.12  0.265
## Year2003         -0.5372     0.2961  -1.81  0.072 .
## Year2004         -1.3017     0.1493  -8.72 2.9e-14 ***
## Year2005         -0.3434     0.4484  -0.77  0.445
## Year2006         -0.8630     0.3944  -2.19  0.031 *
```

```

## Year2007          -0.4718      0.3027   -1.56    0.122
## Year2008          -0.7462      0.2893   -2.58    0.011 *
## Year2009          -0.3780      0.2022   -1.87    0.064 .
## Year2010          -0.4726      0.2230   -2.12    0.036 *
## Year2011          -0.1331      0.1629   -0.82    0.416
## Year2012          -0.4665      0.2238   -2.08    0.039 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.833
## Multiple R-squared:  0.112, Adjusted R-squared:  -0.0149
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(42,51) are outliers with |weight| = 0 ( < 0.00078);
## 8 weights are ~= 1. The remaining 119 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.627  0.894  0.954  0.932  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.75e-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  1.9  1          1.378
## Year              1.9 15          1.022

```

Residuals from last author



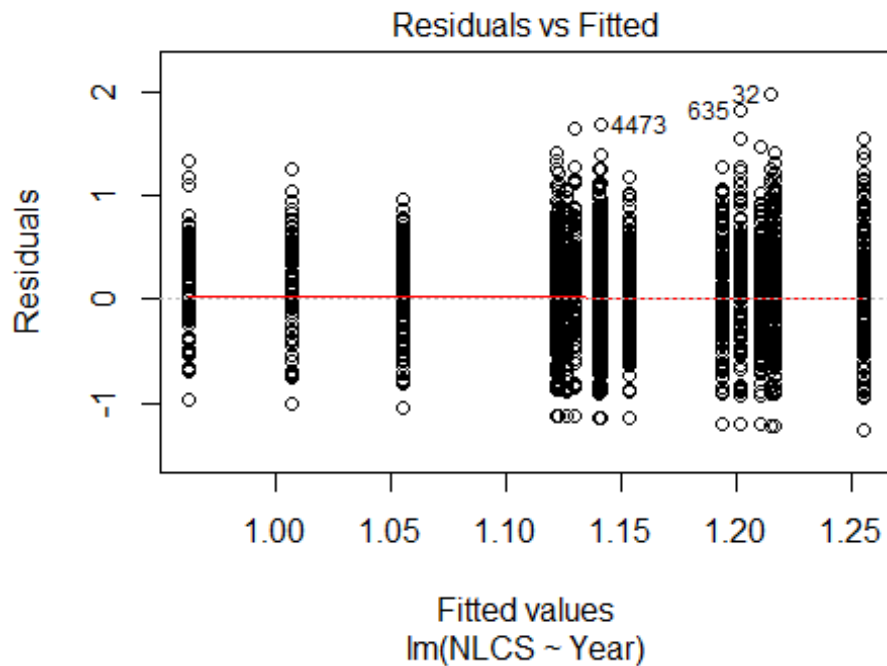
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 122 33847790037 5.209 2007    2216      1    4.252
## 146 45149134975 5.320 2008    2216      1    4.624
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6684 -0.5600  0.0452  0.5408  4.6319
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.47e+00  2.67e-08  5.50e+07 < 2e-16 ***
## LastAuthorFemale1 -1.58e-02  1.67e-01 -9.00e-02  0.925
## Year1997        -4.27e-01  3.44e-01 -1.24e+00  0.218
## Year1998        -6.47e-01  3.64e-01 -1.78e+00  0.078 .
## Year2000        -3.02e-01  4.24e-01 -7.10e-01  0.478
## Year2001         1.98e-01  8.02e-01  2.50e-01  0.805
## Year2002        -7.41e-01  6.11e-01 -1.21e+00  0.227
## Year2003        -5.53e-01  3.18e-01 -1.74e+00  0.085 .
## Year2004        -1.32e+00  1.51e-01 -8.70e+00 3.3e-14 ***
## Year2005        -3.73e-01  4.32e-01 -8.60e-01  0.390
## Year2006        -9.10e-01  3.67e-01 -2.48e+00  0.015 *
```

```

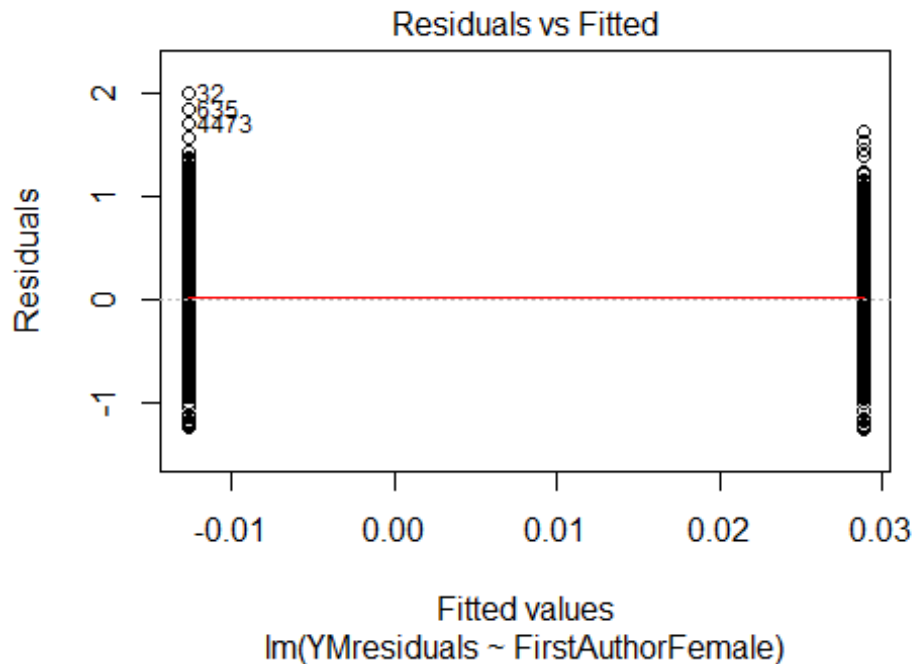
## Year2007          -5.01e-01    2.71e-01 -1.85e+00    0.068 .
## Year2008          -7.66e-01    3.18e-01 -2.41e+00    0.018 *
## Year2009          -4.01e-01    1.63e-01 -2.46e+00    0.015 *
## Year2010          -4.93e-01    2.14e-01 -2.30e+00    0.023 *
## Year2011          -1.61e-01    1.34e-01 -1.20e+00    0.232
## Year2012          -4.95e-01    2.03e-01 -2.44e+00    0.016 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.804
## Multiple R-squared:  0.113, Adjusted R-squared:  -0.0133
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(42,51) are outliers with |weight| = 0 ( < 0.00078);
## 9 weights are ~ = 1. The remaining 118 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.612  0.879   0.947   0.927   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      7.75e-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: 129"
## [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
## 339 330 243 268 289 369 277 271 292 305 478 437 429 324 293
## 2011 2012
## 322 345
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 155 154 133 146 121 101 149 165 174 198 281 267 260 220 178

```

```
## 2011 2012
## 205 228
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 138 126 117 123 104 90 125 134 151 169 225 236 222 191 158
## 2011 2012
## 173 204
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 61, df = 16, p-value = 3e-07
```

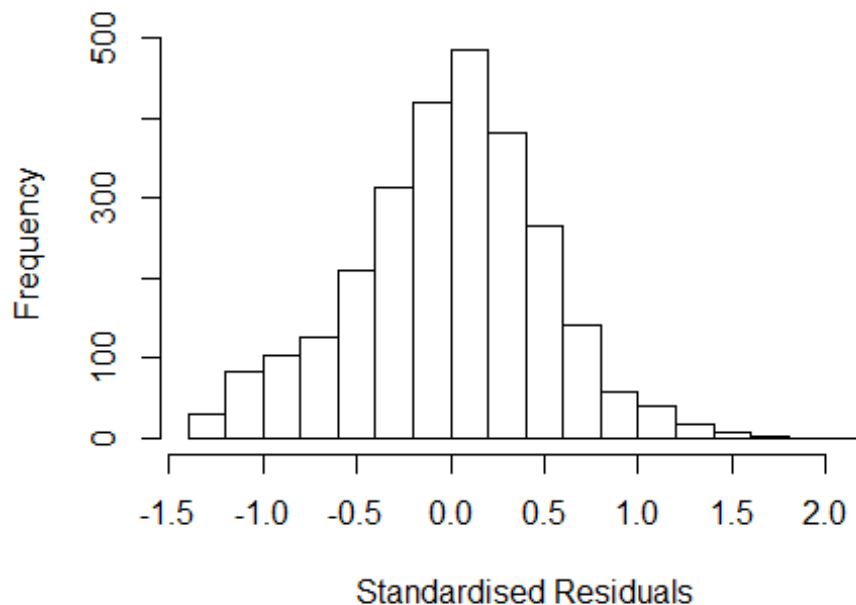


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 14, df = 1, p-value = 1e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 3.20418979142248"
## [1] "Male first author team size 2018 geometric mean: 2.90550004833612"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6400, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.49960807688049"
## [1] "Male last author team size 2018 geometric mean: 2.82585649920318"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6300, 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 1.105 1      1.051
## LastAuthorFemale  1.102 1      1.050
## UniqueAuthors     1.186 4      1.022
## Year              1.257 16     1.007
```

Residuals from first and last author and team size



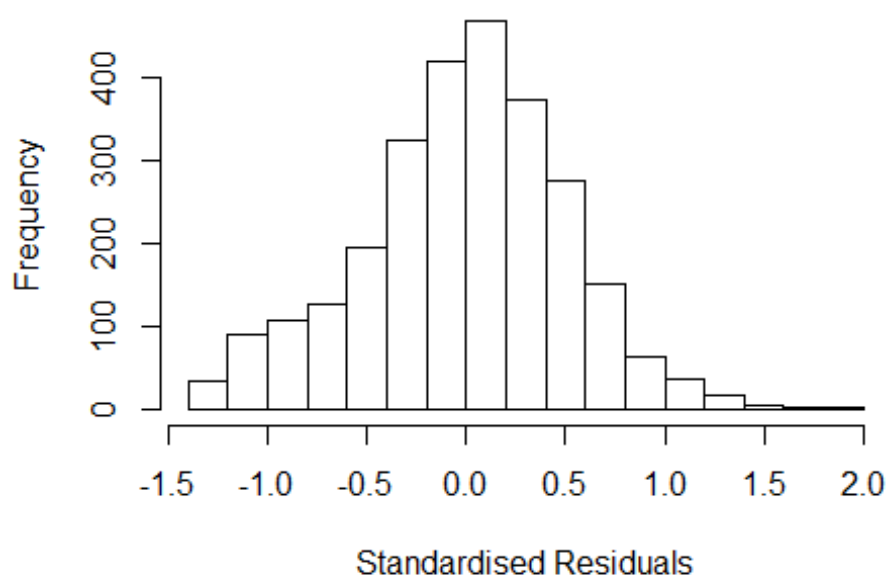
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.369 -0.319 0.022 0.319 2.056
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1357 0.0559 20.31 < 2e-16 ***
## FirstAuthorFemale1 0.0457 0.0213 2.15 0.03176 *
## LastAuthorFemale1 0.0250 0.0245 1.02 0.30858
## UniqueAuthors2 0.1250 0.0327 3.83 0.00013 ***
## UniqueAuthors3 0.1167 0.0335 3.48 0.00050 ***
## UniqueAuthors4 0.1687 0.0384 4.39 1.2e-05 ***
## UniqueAuthors5 0.2337 0.0359 6.51 8.9e-11 ***
## Year1997 -0.0361 0.0790 -0.46 0.64776
## Year1998 -0.0895 0.0735 -1.22 0.22369
## Year1999 -0.1222 0.0739 -1.65 0.09834 .
```

```

## Year2000          -0.1773      0.0821    -2.16   0.03103 *
## Year2001          -0.3306      0.0751    -4.40   1.1e-05 ***
## Year2002          -0.0905      0.0656    -1.38   0.16778
## Year2003          -0.2066      0.0683    -3.03   0.00250 **
## Year2004          -0.1238      0.0666    -1.86   0.06304 .
## Year2005          -0.1146      0.0603    -1.90   0.05771 .
## Year2006          -0.1726      0.0598    -2.88   0.00396 **
## Year2007          -0.1468      0.0590    -2.49   0.01295 *
## Year2008          -0.1425      0.0603    -2.36   0.01824 *
## Year2009          -0.0379      0.0614    -0.62   0.53715
## Year2010          -0.0929      0.0651    -1.43   0.15342
## Year2011          -0.0307      0.0665    -0.46   0.64421
## Year2012          -0.0367      0.0639    -0.57   0.56540
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.474
## Multiple R-squared:  0.041, Adjusted R-squared:  0.0331
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 222 weights are ~= 1. The remaining 2464 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.020  0.860  0.950  0.894  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.72e-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 1.086 1          1.042
## LastAuthorFemale 1.091 1          1.044
## Year              1.083 16          1.002

```


Residuals from first and last author



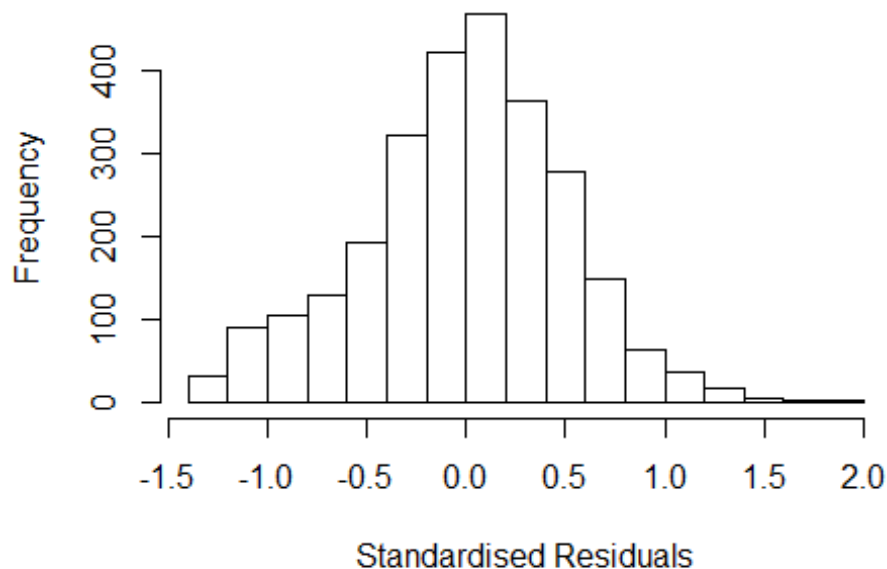
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3102 -0.3238 0.0191 0.3298 1.9589
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23306 0.05035 24.49 < 2e-16 ***
## FirstAuthorFemale1 0.05675 0.02130 2.67 0.0077 **
## LastAuthorFemale1 0.01586 0.02467 0.64 0.5205
## Year1997 -0.04171 0.08004 -0.52 0.6024
## Year1998 -0.08890 0.07301 -1.22 0.2235
## Year1999 -0.10889 0.07395 -1.47 0.1410
## Year2000 -0.17025 0.08268 -2.06 0.0396 *
## Year2001 -0.30651 0.07732 -3.96 7.6e-05 ***
## Year2002 -0.07571 0.06485 -1.17 0.2431
## Year2003 -0.19674 0.06771 -2.91 0.0037 **
## Year2004 -0.10332 0.06575 -1.57 0.1162
## Year2005 -0.09466 0.05995 -1.58 0.1145
```

```

## Year2006          -0.15411      0.05976      -2.58      0.0100 **
## Year2007          -0.11002      0.05827      -1.89      0.0591 .
## Year2008          -0.11465      0.05990      -1.91      0.0557 .
## Year2009          -0.02084      0.06047      -0.34      0.7304
## Year2010          -0.06791      0.06457      -1.05      0.2930
## Year2011           0.00458      0.06624       0.07      0.9449
## Year2012           0.00305      0.06372       0.05      0.9618
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.478
## Multiple R-squared:  0.0229, Adjusted R-squared:  0.0163
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 220 weights are ~= 1. The remaining 2466 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0551 0.8590 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      3.72e-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 1.041 1      1.020
## Year              1.041 16      1.001

```

Residuals from first author



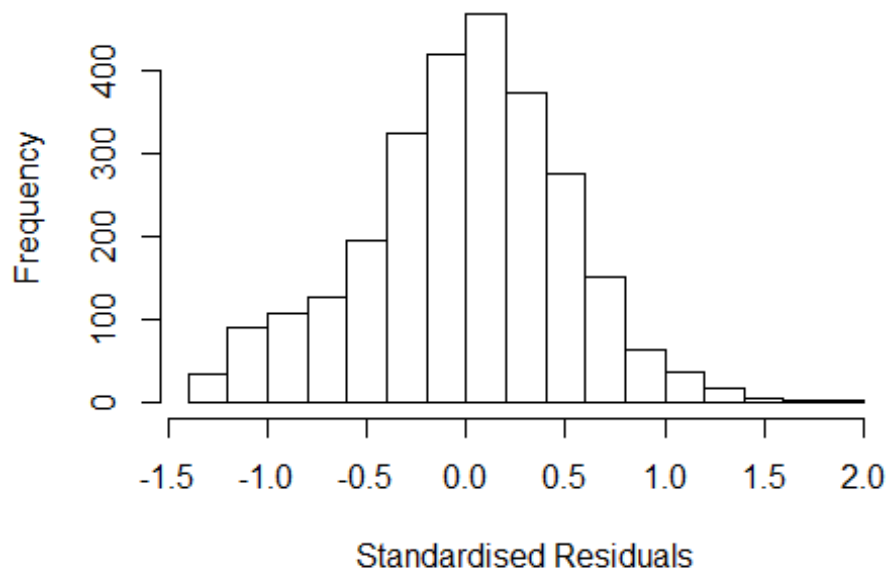
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30  -0.32   0.02   0.33   1.96
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.23488    0.05031   24.55 < 2e-16 ***
## FirstAuthorFemale1 0.06020    0.02086    2.89  0.0039 **
## Year1997      -0.04218    0.08007   -0.53  0.5984
## Year1998      -0.08799    0.07300   -1.21  0.2282
## Year1999      -0.10936    0.07396   -1.48  0.1394
## Year2000      -0.16967    0.08276   -2.05  0.0405 *
## Year2001      -0.30742    0.07739   -3.97 7.3e-05 ***
## Year2002      -0.07604    0.06493   -1.17  0.2417
## Year2003      -0.19722    0.06772   -2.91  0.0036 **
## Year2004      -0.10289    0.06567   -1.57  0.1173
## Year2005      -0.09418    0.05991   -1.57  0.1160
## Year2006      -0.15381    0.05975   -2.57  0.0101 *
```

```

## Year2007          -0.10969    0.05824   -1.88    0.0597 .
## Year2008          -0.11382    0.05987   -1.90    0.0574 .
## Year2009          -0.01972    0.06039   -0.33    0.7441
## Year2010          -0.06709    0.06448   -1.04    0.2982
## Year2011           0.00617    0.06608    0.09    0.9256
## Year2012           0.00428    0.06369    0.07    0.9465
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.478
## Multiple R-squared:  0.0227, Adjusted R-squared:  0.0165
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 227 weights are ~= 1. The remaining 2459 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0555 0.8590 0.9490 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.72e-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 1.044 1          1.022
## Year              1.044 16          1.001

```

Residuals from last author



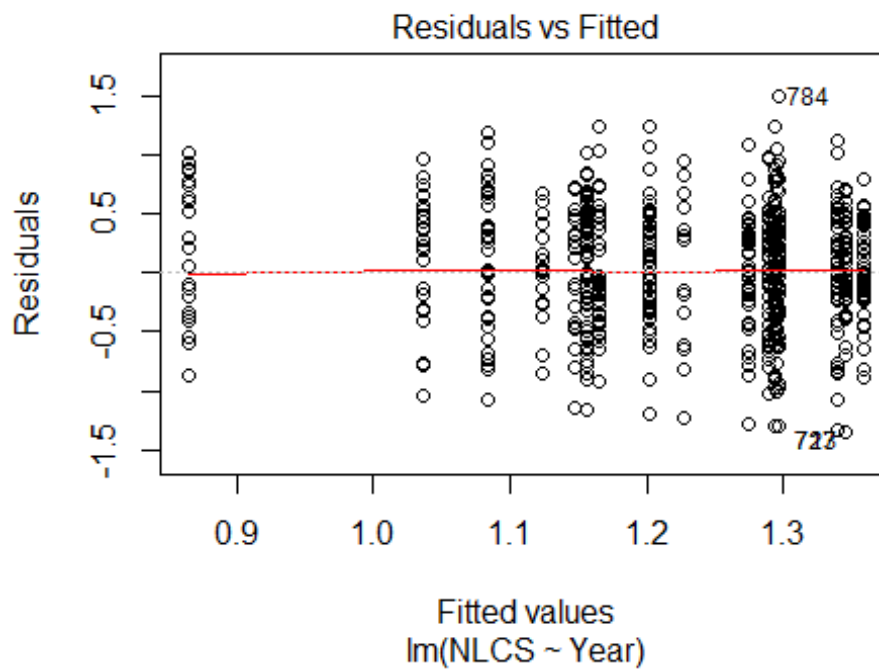
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2879 -0.3151  0.0269  0.3247  1.9505
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2415     0.0502   24.71  < 2e-16 ***
## LastAuthorFemale1  0.0318     0.0241    1.32  0.18808
## Year1997         -0.0381     0.0798   -0.48  0.63268
## Year1998         -0.0872     0.0732   -1.19  0.23376
## Year1999         -0.1021     0.0736   -1.39  0.16534
## Year2000         -0.1742     0.0827   -2.11  0.03535 *
## Year2001         -0.2994     0.0774   -3.87  0.00011 ***
## Year2002         -0.0752     0.0649   -1.16  0.24665
## Year2003         -0.1900     0.0676   -2.81  0.00497 **
## Year2004         -0.0973     0.0658   -1.48  0.13963
## Year2005         -0.0887     0.0599   -1.48  0.13897
## Year2006         -0.1461     0.0596   -2.45  0.01429 *
```

```

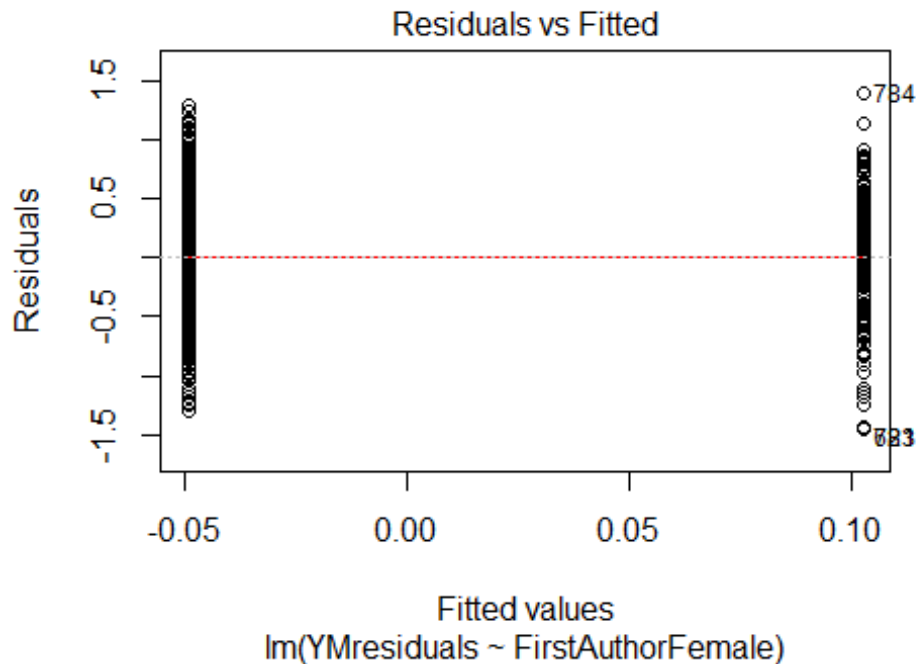
## Year2007          -0.1020      0.0583    -1.75   0.08018 .
## Year2008          -0.1014      0.0599    -1.69   0.09050 .
## Year2009          -0.0134      0.0604    -0.22   0.82486
## Year2010          -0.0595      0.0644    -0.92   0.35552
## Year2011           0.0146      0.0660     0.22   0.82465
## Year2012           0.0117      0.0636     0.18   0.85422
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.478
## Multiple R-squared:  0.0204, Adjusted R-squared:  0.0141
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 225 weights are ~= 1. The remaining 2461 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0583 0.8580 0.9490 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.72e-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: 2686"
## [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
##   35   34   44   30   38   37   47   42   59   57   51   78   62   50   57
## 2011 2012
##   63   68
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   22   21   29   21   25   14   40   35   47   49   41   65   54   43   50
## 2011 2012

```

```
## 53 59
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 22 21 28 21 24 14 32 34 43 45 38 63 51 38 44
## 2011 2012
## 51 55
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.06
```

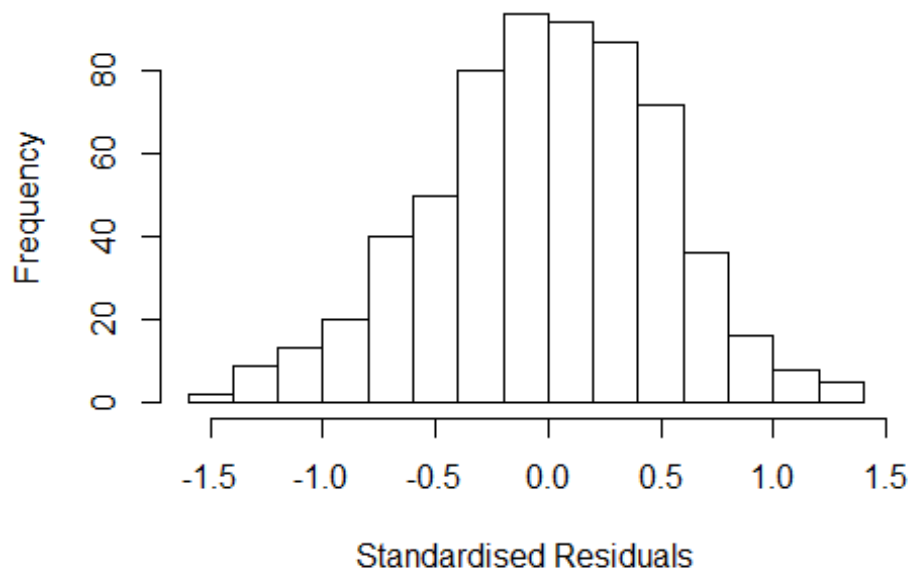


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.7, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 2.26412242084496"
## [1] "Male first author team size 2018 geometric mean: 1.66023670252734"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 890, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.19791614785318"
## [1] "Male last author team size 2018 geometric mean: 1.74819187201924"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 680, 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 1.608 1      1.268
## LastAuthorFemale  1.644 1      1.282
## UniqueAuthors    1.516 4      1.053
## Year              1.682 16     1.016
```


Residuals from first and last author and team size



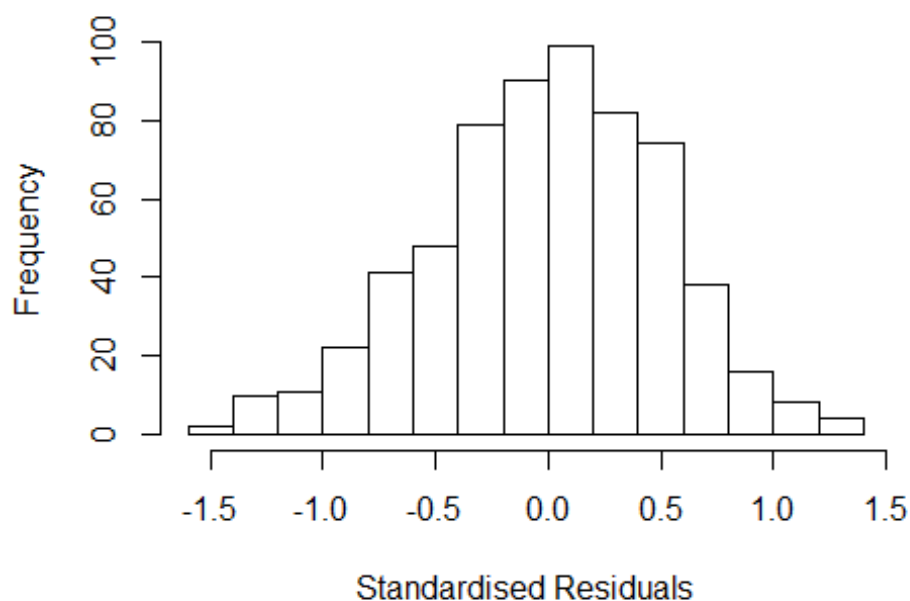
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5490 -0.3318 0.0115 0.3622 1.3621
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19736 0.15944 7.51 2.1e-13 ***
## FirstAuthorFemale1 0.15518 0.05177 3.00 0.0028 **
## LastAuthorFemale1 0.01199 0.05239 0.23 0.8191
## UniqueAuthors2 0.09528 0.04945 1.93 0.0545 .
## UniqueAuthors3 0.01628 0.07894 0.21 0.8367
## UniqueAuthors4 -0.10442 0.11594 -0.90 0.3681
## UniqueAuthors5 0.02491 0.24291 0.10 0.9184
## Year1997 -0.08987 0.19774 -0.45 0.6496
## Year1998 -0.46376 0.22278 -2.08 0.0378 *
## Year1999 -0.11211 0.18176 -0.62 0.5376
```

```

## Year2000      -0.18545    0.21856   -0.85    0.3965
## Year2001      0.00979    0.26068    0.04    0.9700
## Year2002     -0.23574    0.22549   -1.05    0.2962
## Year2003     -0.10618    0.18165   -0.58    0.5591
## Year2004      0.00701    0.17554    0.04    0.9682
## Year2005      0.01888    0.17504    0.11    0.9141
## Year2006      0.05434    0.17277    0.31    0.7532
## Year2007     -0.12522    0.17623   -0.71    0.4777
## Year2008      0.04999    0.17400    0.29    0.7740
## Year2009      0.08921    0.17124    0.52    0.6026
## Year2010      0.07033    0.17401    0.40    0.6862
## Year2011     -0.11698    0.17620   -0.66    0.5070
## Year2012      0.04079    0.17670    0.23    0.8175
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.494
## Multiple R-squared:  0.0854, Adjusted R-squared:  0.0519
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 574 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.305  0.863  0.946  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      1.60e-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 1.598 1      1.264
## LastAuthorFemale  1.630 1      1.277
## Year              1.203 16      1.006

```

Residuals from first and last author



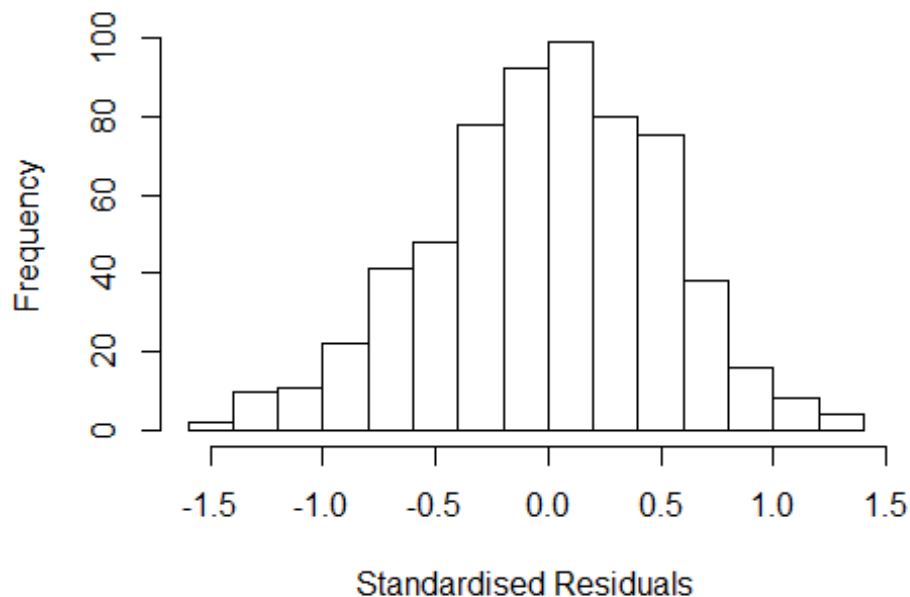
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.486 -0.340  0.011  0.355  1.326
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1978    0.1616   7.41 4.2e-13 ***
## FirstAuthorFemale1 0.1570    0.0524   3.00 0.0028 **
## LastAuthorFemale1  0.0155    0.0530   0.29 0.7702
## Year1997         -0.0727    0.1998  -0.36 0.7161
## Year1998         -0.4460    0.2258  -1.97 0.0487 *
## Year1999         -0.1067    0.1831  -0.58 0.5602
## Year2000         -0.1583    0.2190  -0.72 0.4702
## Year2001          0.0300    0.2637   0.11 0.9094
## Year2002         -0.2074    0.2283  -0.91 0.3639
## Year2003         -0.0967    0.1841  -0.52 0.5998
## Year2004          0.0330    0.1770   0.19 0.8520
## Year2005          0.0442    0.1761   0.25 0.8019
```

```

## Year2006          0.0830      0.1746      0.48      0.6349
## Year2007         -0.1043      0.1774     -0.59      0.5566
## Year2008          0.0816      0.1753      0.47      0.6418
## Year2009          0.1156      0.1733      0.67      0.5049
## Year2010          0.1008      0.1742      0.58      0.5628
## Year2011         -0.0754      0.1763     -0.43      0.6690
## Year2012          0.0667      0.1758      0.38      0.7046
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.078, Adjusted R-squared:  0.0506
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 578 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.335  0.860  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      1.60e-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



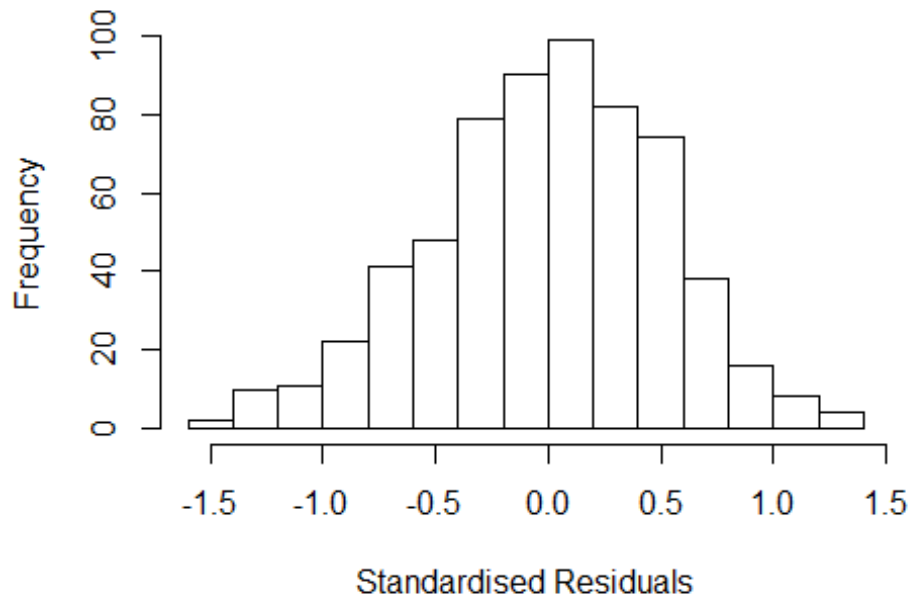
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4843 -0.3375 0.0123 0.3545 1.3289
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1994 0.1612 7.44 3.5e-13 ***
## FirstAuthorFemale1 0.1670 0.0435 3.84 0.00014 ***
## Year1997 -0.0721 0.1997 -0.36 0.71842
## Year1998 -0.4473 0.2255 -1.98 0.04771 *
## Year1999 -0.1069 0.1831 -0.58 0.55946
## Year2000 -0.1593 0.2187 -0.73 0.46662
## Year2001 0.0303 0.2640 0.11 0.90880
## Year2002 -0.2084 0.2281 -0.91 0.36117
## Year2003 -0.0977 0.1839 -0.53 0.59559
## Year2004 0.0320 0.1767 0.18 0.85629
## Year2005 0.0429 0.1759 0.24 0.80758
## Year2006 0.0836 0.1746 0.48 0.63222
```

```

## Year2007          -0.1033      0.1772   -0.58  0.55992
## Year2008           0.0840      0.1748    0.48  0.63116
## Year2009           0.1178      0.1730    0.68  0.49591
## Year2010           0.1016      0.1742    0.58  0.55979
## Year2011          -0.0748      0.1761   -0.42  0.67139
## Year2012           0.0667      0.1756    0.38  0.70413
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.078, Adjusted R-squared:  0.0521
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 580 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.334  0.858  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      1.60e-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 1.12 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
## -1.4376 -0.3442  0.0138  0.3531  1.3664
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2156     0.1591   7.64 8.5e-14 ***
## LastAuthorFemale1  0.1200     0.0443   2.71  0.0069 **
## Year1997         -0.0790     0.2007  -0.39  0.6940
## Year1998         -0.4330     0.2292  -1.89  0.0594 .
## Year1999         -0.1110     0.1814  -0.61  0.5408
## Year2000         -0.1348     0.2190  -0.62  0.5386
## Year2001          0.0161     0.2622   0.06  0.9511
## Year2002         -0.1929     0.2262  -0.85  0.3941
## Year2003         -0.1027     0.1835  -0.56  0.5761
## Year2004          0.0351     0.1745   0.20  0.8406
## Year2005          0.0558     0.1732   0.32  0.7474
## Year2006          0.0882     0.1714   0.51  0.6071
```

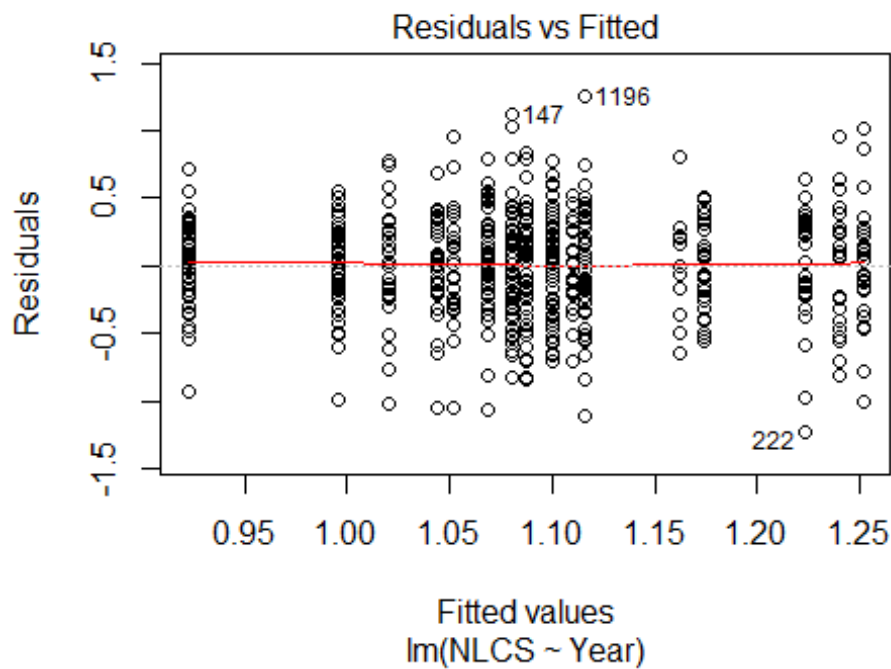
```

## Year2007          -0.1055      0.1758    -0.60    0.5486
## Year2008           0.0719      0.1730     0.42    0.6778
## Year2009           0.1020      0.1706     0.60    0.5501
## Year2010           0.0949      0.1713     0.55    0.5797
## Year2011          -0.0644      0.1750    -0.37    0.7129
## Year2012           0.0698      0.1749     0.40    0.6902
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.49
## Multiple R-squared:  0.0669, Adjusted R-squared:  0.0407
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 568 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.369  0.858  0.944  0.896  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      1.60e-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: 624"
## [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
##   54   48   38   36   51   56   51   49   46   51   55   72   77   80   78
## 2011 2012
##  109   93
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   24   19   13   28   26   26   31   27   26   35   42   45   48   52
## 2011 2012

```



```
## 59 57
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 21 19 17 12 24 23 25 26 24 21 30 35 42 37 45
## 2011 2012
## 52 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 = 15, df = 16, p-value = 0.5
```



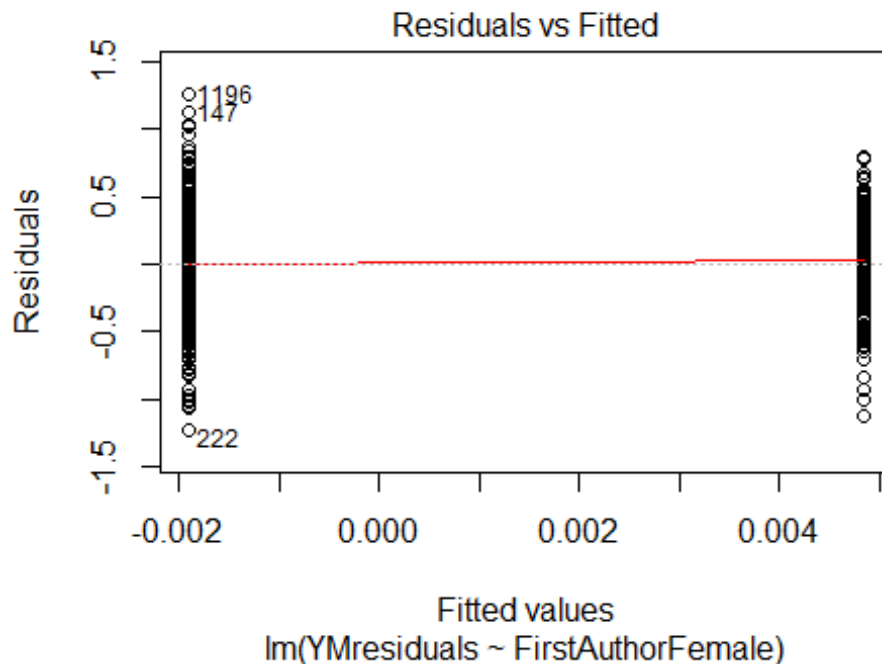
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.39, df = 1, p-value = 0.5

## [1] "Female first author team size 2018 geometric mean: 3.06127201857781"
## [1] "Male first author team size 2018 geometric mean: 3.28907369218098"

## Warning in wilcox.test.default(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 = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.25233988626749"
## [1] "Male last author team size 2018 geometric mean: 3.42623273336553"

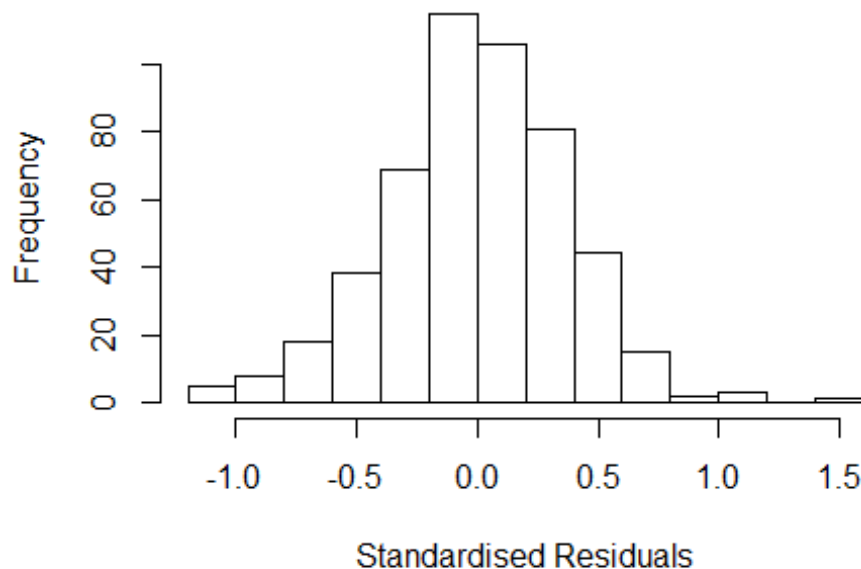
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "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	1.222	1	1.105
UniqueAuthors	2.426	4	1.117
Year	2.902	16	1.034

Residuals from first and last author and team size



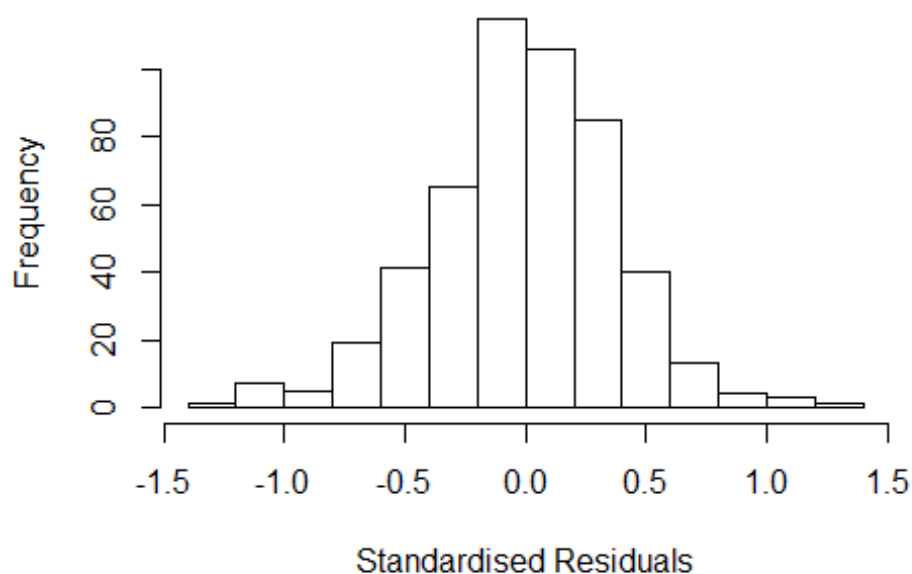
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.126441 -0.218438 -0.000442 0.232662 1.527391
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0194 0.1180 8.64 < 2e-16 ***
## FirstAuthorFemale1 0.0120 0.0375 0.32 0.74938
## LastAuthorFemale1 0.0495 0.0444 1.11 0.26612
## UniqueAuthors2 0.1788 0.0657 2.72 0.00675 **
## UniqueAuthors3 0.2288 0.0667 3.43 0.00066 ***
## UniqueAuthors4 0.1869 0.0690 2.71 0.00695 **
## UniqueAuthors5 0.2725 0.0692 3.94 9.5e-05 ***
## Year1997 -0.0952 0.1724 -0.55 0.58113
## Year1998 -0.2029 0.1350 -1.50 0.13347
## Year1999 -0.1147 0.1310 -0.88 0.38197
```

```

## Year2000          0.1070      0.1465      0.73  0.46530
## Year2001         -0.1083      0.1221     -0.89  0.37564
## Year2002          0.0680      0.1287      0.53  0.59714
## Year2003         -0.1263      0.1215     -1.04  0.29917
## Year2004         -0.1359      0.1323     -1.03  0.30482
## Year2005         -0.0287      0.1242     -0.23  0.81725
## Year2006         -0.1709      0.1219     -1.40  0.16166
## Year2007         -0.2871      0.1172     -2.45  0.01462 *
## Year2008         -0.2019      0.1117     -1.81  0.07138 .
## Year2009         -0.1169      0.1175     -0.99  0.32035
## Year2010         -0.1257      0.1120     -1.12  0.26198
## Year2011         -0.0893      0.1210     -0.74  0.46063
## Year2012         -0.1778      0.1162     -1.53  0.12672
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.34
## Multiple R-squared:  0.0934, Adjusted R-squared:  0.052
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 465 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0064 0.8570 0.9530 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          1.98e-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 1.107 1 1.052
## LastAuthorFemale 1.161 1 1.078
## Year 1.217 16 1.006

```

Residuals from first and last author



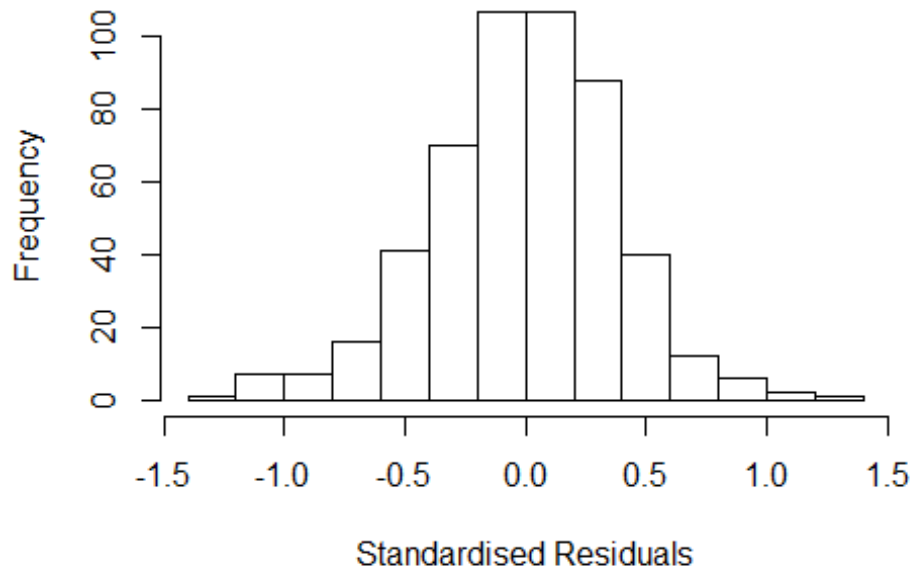
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26315 -0.23284 -0.00183 0.24438 1.32659
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1939 0.1134 10.53 <2e-16 ***
## FirstAuthorFemale1 0.0283 0.0379 0.75 0.456
## LastAuthorFemale1 0.0487 0.0452 1.08 0.282
## Year1997 -0.0874 0.1625 -0.54 0.591
## Year1998 -0.2212 0.1519 -1.46 0.146
## Year1999 -0.0999 0.1449 -0.69 0.491
## Year2000 0.0693 0.1405 0.49 0.622
## Year2001 -0.1101 0.1326 -0.83 0.407
## Year2002 0.0393 0.1404 0.28 0.780
## Year2003 -0.1370 0.1274 -1.08 0.283
## Year2004 -0.1511 0.1412 -1.07 0.285
## Year2005 -0.0423 0.1330 -0.32 0.751
```

```

## Year2006          -0.1662      0.1303   -1.28    0.203
## Year2007          -0.2813      0.1252   -2.25    0.025 *
## Year2008          -0.1804      0.1200   -1.50    0.133
## Year2009          -0.0867      0.1265   -0.69    0.493
## Year2010          -0.0954      0.1202   -0.79    0.427
## Year2011          -0.0781      0.1265   -0.62    0.538
## Year2012          -0.1515      0.1251   -1.21    0.227
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.0513, Adjusted R-squared:  0.0162
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 465 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.141  0.870  0.952  0.898  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.98e-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 1.08 1      1.039
## Year              1.08 16      1.002

```

Residuals from first author



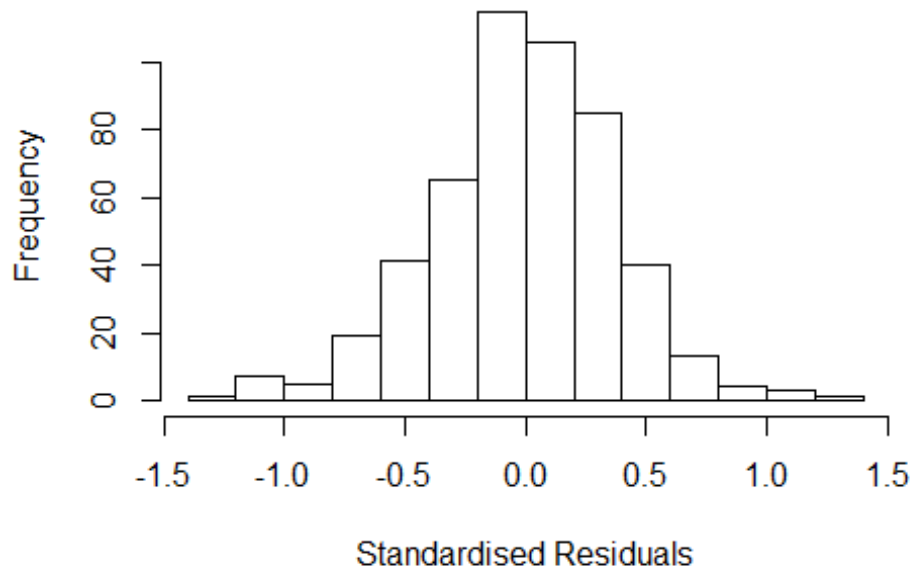
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26982 -0.22427 0.00607 0.24342 1.31228
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1995 0.1142 10.50 <2e-16 ***
## FirstAuthorFemale1 0.0319 0.0375 0.85 0.395
## Year1997 -0.0899 0.1631 -0.55 0.582
## Year1998 -0.2207 0.1528 -1.44 0.149
## Year1999 -0.1003 0.1456 -0.69 0.491
## Year2000 0.0703 0.1413 0.50 0.619
## Year2001 -0.1068 0.1335 -0.80 0.424
## Year2002 0.0322 0.1408 0.23 0.819
## Year2003 -0.1365 0.1285 -1.06 0.289
## Year2004 -0.1476 0.1415 -1.04 0.298
## Year2005 -0.0441 0.1335 -0.33 0.741
## Year2006 -0.1620 0.1311 -1.24 0.217
```

```

## Year2007          -0.2842      0.1262   -2.25    0.025 *
## Year2008          -0.1765      0.1213   -1.46    0.146
## Year2009          -0.0838      0.1276   -0.66    0.512
## Year2010          -0.0941      0.1215   -0.77    0.439
## Year2011          -0.0741      0.1280   -0.58    0.563
## Year2012          -0.1428      0.1267   -1.13    0.260
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.357
## Multiple R-squared:  0.0489, Adjusted R-squared:  0.0157
## 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.
##   0.148  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      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 1.132 1      1.064
## Year              1.132 16      1.004

```


Residuals from last author



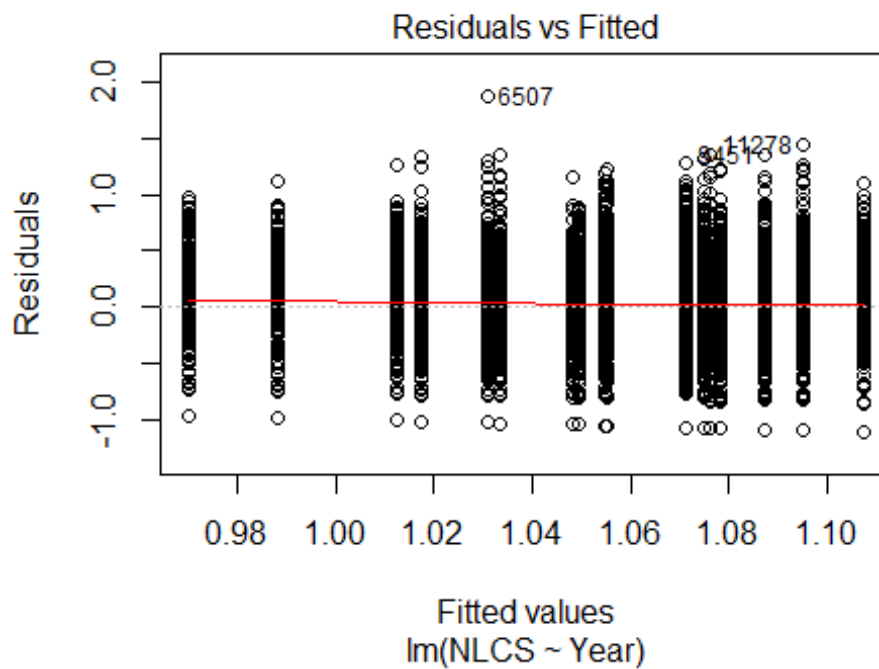
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26845 -0.23996 0.00473 0.24018 1.31804
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2023 0.1128 10.66 <2e-16 ***
## LastAuthorFemale1 0.0517 0.0447 1.16 0.247
## Year1997 -0.0896 0.1623 -0.55 0.581
## Year1998 -0.2221 0.1526 -1.46 0.146
## Year1999 -0.1015 0.1440 -0.70 0.481
## Year2000 0.0661 0.1414 0.47 0.640
## Year2001 -0.1155 0.1323 -0.87 0.383
## Year2002 0.0365 0.1409 0.26 0.796
## Year2003 -0.1377 0.1286 -1.07 0.285
## Year2004 -0.1534 0.1417 -1.08 0.280
## Year2005 -0.0392 0.1342 -0.29 0.770
## Year2006 -0.1701 0.1299 -1.31 0.191
```

```

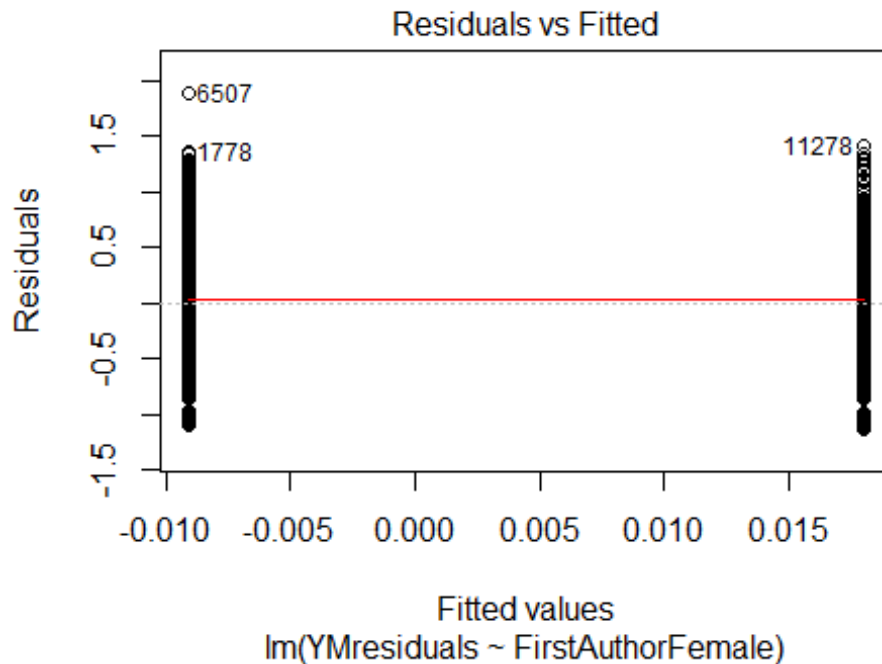
## Year2007          -0.2827      0.1256    -2.25      0.025 *
## Year2008          -0.1823      0.1206    -1.51      0.131
## Year2009          -0.0865      0.1274    -0.68      0.497
## Year2010          -0.0941      0.1211    -0.78      0.438
## Year2011          -0.0781      0.1271    -0.61      0.539
## Year2012          -0.1514      0.1261    -1.20      0.231
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.0502, Adjusted R-squared:  0.0171
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 466 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.152  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      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 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
##  473  513  524  484  527  545  516  544  473  535  591  644  693  689  718
## 2011 2012
##  736  806
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  241  259  309  287  235  203  351  373  322  381  413  466  493  476  535
## 2011 2012

```

```
## 533 613
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 220 229 268 257 204 189 316 326 285 336 370 419 439 410 470
## 2011 2012
## 473 559
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 62, df = 16, p-value = 2e-07
```

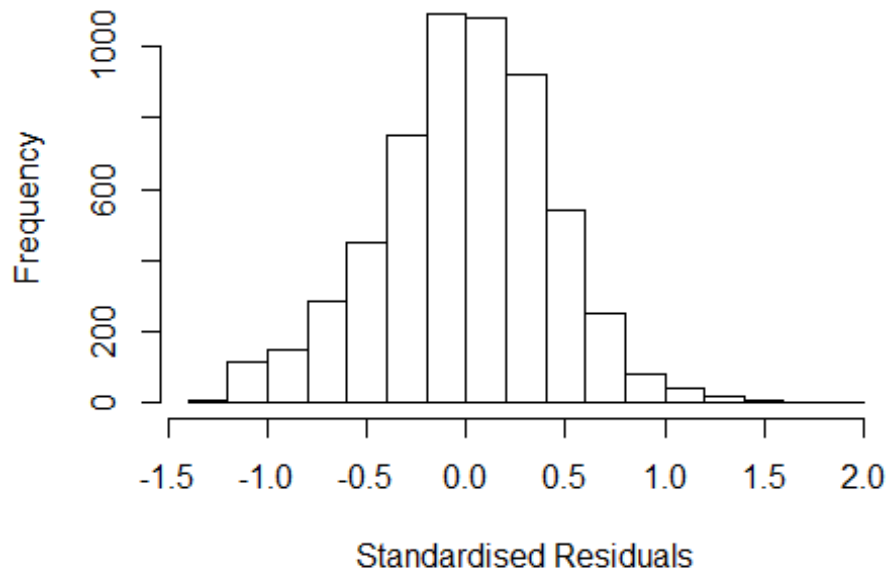


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.6, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.14265475087763"
## [1] "Male first author team size 2018 geometric mean: 3.04648658025126"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 22000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.95210671040931"
## [1] "Male last author team size 2018 geometric mean: 3.13085983657116"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, 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 1.061 1 1.030
## LastAuthorFemale 1.044 1 1.022
## UniqueAuthors 1.105 4 1.013
## Year 1.130 16 1.004
```

Residuals from first and last author and team size



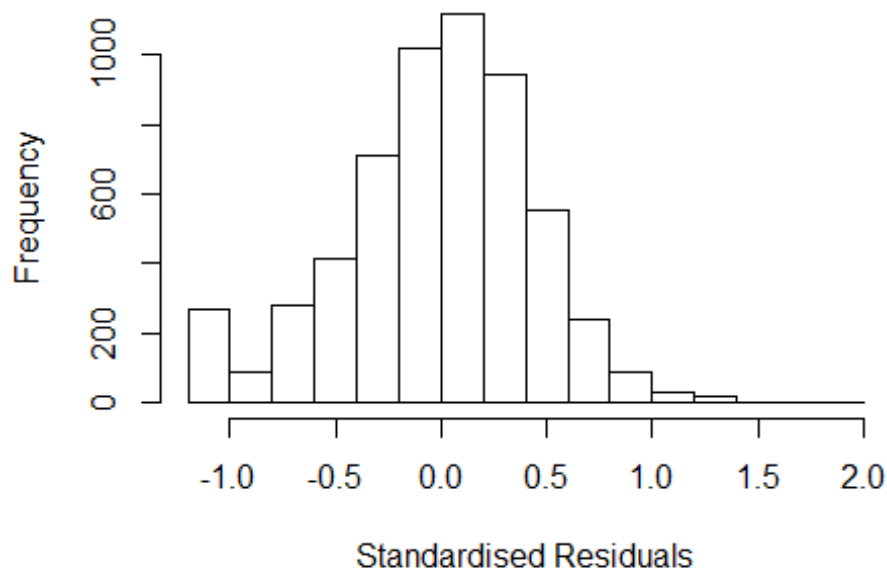
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.31535 -0.27334  0.00778  0.27972  1.85341
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8844    0.0372   23.77 < 2e-16 ***
## FirstAuthorFemale1  0.0112    0.0123    0.91  0.36213
## LastAuthorFemale1 -0.0112    0.0149   -0.75  0.45324
## UniqueAuthors2    0.2990    0.0219   13.67 < 2e-16 ***
## UniqueAuthors3    0.3093    0.0224   13.78 < 2e-16 ***
## UniqueAuthors4    0.3392    0.0242   14.00 < 2e-16 ***
## UniqueAuthors5    0.4310    0.0247   17.44 < 2e-16 ***
## Year1997         -0.0411    0.0461   -0.89  0.37299
## Year1998         -0.0603    0.0411   -1.47  0.14237
## Year1999         -0.0985    0.0425   -2.32  0.02054 *
```

```

## Year2000          -0.1443      0.0474   -3.04   0.00236 **
## Year2001          -0.1259      0.0496   -2.54   0.01119 *
## Year2002          -0.0825      0.0436   -1.89   0.05863 .
## Year2003          -0.0735      0.0391   -1.88   0.06032 .
## Year2004          -0.1448      0.0429   -3.37   0.00075 ***
## Year2005          -0.0759      0.0398   -1.91   0.05665 .
## Year2006          -0.1236      0.0406   -3.04   0.00235 **
## Year2007          -0.0656      0.0390   -1.68   0.09275 .
## Year2008          -0.0974      0.0380   -2.56   0.01048 *
## Year2009          -0.0868      0.0385   -2.25   0.02417 *
## Year2010          -0.1183      0.0387   -3.06   0.00223 **
## Year2011          -0.0889      0.0384   -2.32   0.02063 *
## Year2012          -0.1121      0.0376   -2.98   0.00289 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.406
## Multiple R-squared:  0.0886, Adjusted R-squared:  0.0851
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 527 weights are ~= 1. The remaining 5243 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0026 0.8620 0.9480 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.73e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1 1.018
## LastAuthorFemale 1.039 1 1.019
## Year 1.046 16 1.001

```

Residuals from first and last author



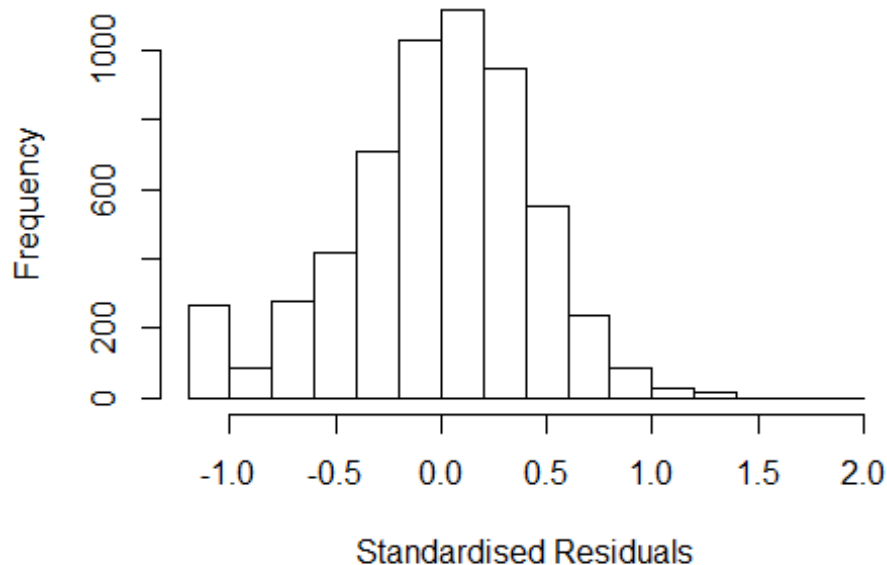
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1399 -0.2784 0.0174 0.2855 1.8869
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1242 0.0345 32.62 <2e-16 ***
## FirstAuthorFemale1 0.0360 0.0125 2.89 0.0039 **
## LastAuthorFemale1 -0.0203 0.0154 -1.32 0.1883
## Year1997 -0.0491 0.0479 -1.02 0.3060
## Year1998 -0.0455 0.0421 -1.08 0.2799
## Year1999 -0.0966 0.0441 -2.19 0.0286 *
## Year2000 -0.1364 0.0498 -2.74 0.0062 **
## Year2001 -0.1075 0.0530 -2.03 0.0426 *
## Year2002 -0.0761 0.0465 -1.63 0.1023
## Year2003 -0.0578 0.0407 -1.42 0.1553
## Year2004 -0.1177 0.0449 -2.62 0.0088 **
## Year2005 -0.0540 0.0409 -1.32 0.1868
```

```

## Year2006          -0.0889      0.0423   -2.10   0.0359 *
## Year2007          -0.0248      0.0406   -0.61   0.5409
## Year2008          -0.0559      0.0394   -1.42   0.1561
## Year2009          -0.0288      0.0396   -0.73   0.4678
## Year2010          -0.0641      0.0402   -1.59   0.1114
## Year2011          -0.0322      0.0397   -0.81   0.4169
## Year2012          -0.0482      0.0390   -1.24   0.2160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.00674,    Adjusted R-squared:  0.00363
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 499 weights are ~= 1. The remaining 5271 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0018 0.8590 0.9490 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      1.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.025 1      1.012
## Year              1.025 16      1.001

```


Residuals from first author



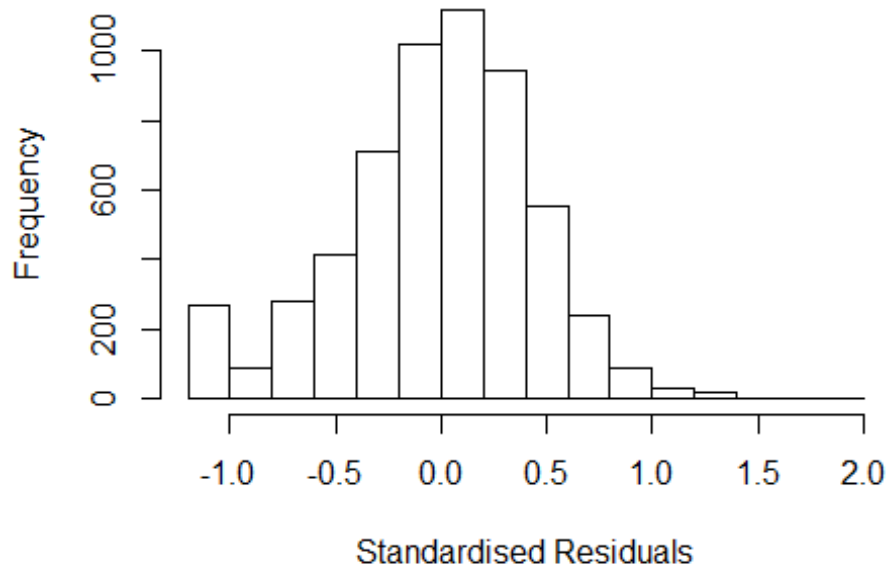
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1542 -0.2771  0.0168  0.2835  1.8693
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1215     0.0344   32.60  <2e-16 ***
## FirstAuthorFemale1  0.0327     0.0124    2.63  0.0085 **
## Year1997        -0.0495     0.0479   -1.03  0.3009
## Year1998        -0.0455     0.0421   -1.08  0.2794
## Year1999        -0.0961     0.0441   -2.18  0.0294 *
## Year2000        -0.1364     0.0498   -2.74  0.0061 **
## Year2001        -0.1069     0.0530   -2.02  0.0438 *
## Year2002        -0.0755     0.0465   -1.62  0.1043
## Year2003        -0.0571     0.0406   -1.40  0.1601
## Year2004        -0.1190     0.0449   -2.65  0.0080 **
## Year2005        -0.0541     0.0409   -1.32  0.1855
## Year2006        -0.0888     0.0423   -2.10  0.0359 *
```

```

## Year2007          -0.0255      0.0406   -0.63   0.5301
## Year2008          -0.0560      0.0394   -1.42   0.1550
## Year2009          -0.0294      0.0396   -0.74   0.4577
## Year2010          -0.0653      0.0402   -1.63   0.1041
## Year2011          -0.0330      0.0396   -0.83   0.4043
## Year2012          -0.0494      0.0389   -1.27   0.2042
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.00643,    Adjusted R-squared:  0.00349
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 490 weights are ~= 1. The remaining 5280 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0036 0.8610 0.9490 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      1.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.025 1          1.012
## Year            1.025 16          1.001

```

Residuals from last author



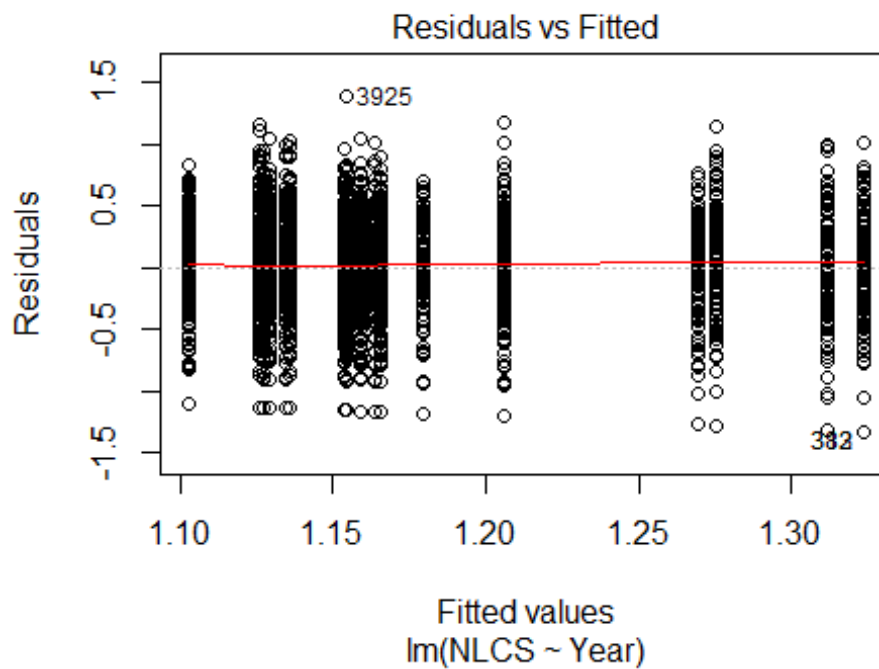
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1315 -0.2783 0.0159 0.2838 1.8674
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1315 0.0343 32.95 <2e-16 ***
## LastAuthorFemale1 -0.0124 0.0153 -0.81 0.417
## Year1997 -0.0500 0.0479 -1.04 0.297
## Year1998 -0.0451 0.0421 -1.07 0.284
## Year1999 -0.0949 0.0441 -2.15 0.032 *
## Year2000 -0.1344 0.0498 -2.70 0.007 **
## Year2001 -0.1077 0.0531 -2.03 0.042 *
## Year2002 -0.0744 0.0466 -1.60 0.110
## Year2003 -0.0551 0.0407 -1.35 0.176
## Year2004 -0.1151 0.0449 -2.57 0.010 *
## Year2005 -0.0503 0.0409 -1.23 0.219
## Year2006 -0.0844 0.0424 -1.99 0.047 *
```

```

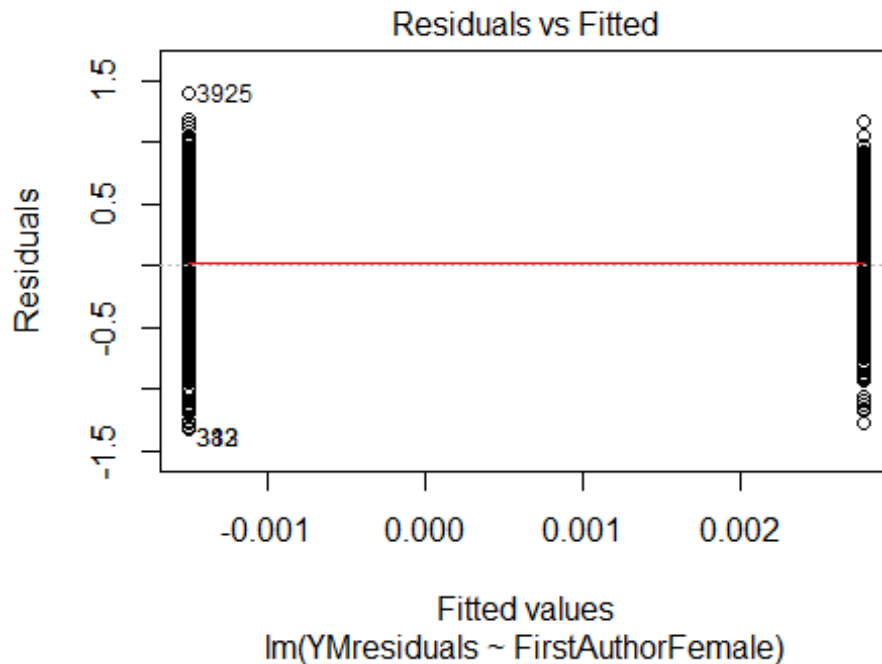
## Year2007          -0.0216      0.0406   -0.53    0.596
## Year2008          -0.0524      0.0394   -1.33    0.184
## Year2009          -0.0237      0.0396   -0.60    0.550
## Year2010          -0.0585      0.0403   -1.45    0.147
## Year2011          -0.0255      0.0397   -0.64    0.521
## Year2012          -0.0423      0.0390   -1.08    0.278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.00531,    Adjusted R-squared:  0.00237
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 504 weights are ~= 1. The remaining 5266 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0038 0.8600 0.9490 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      1.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5770"
## [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
## 251 270 231 276 261 264 253 270 291 318 351 363 405 381 388
## 2011 2012
## 433 450
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 125 140 120 158 128 110 137 183 180 203 222 227 255 237 257
## 2011 2012

```

```
## 279 305
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 102 114 98 138 111 98 121 150 155 166 180 201 216 202 212
## 2011 2012
## 243 267
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 56, df = 16, p-value = 3e-06
```

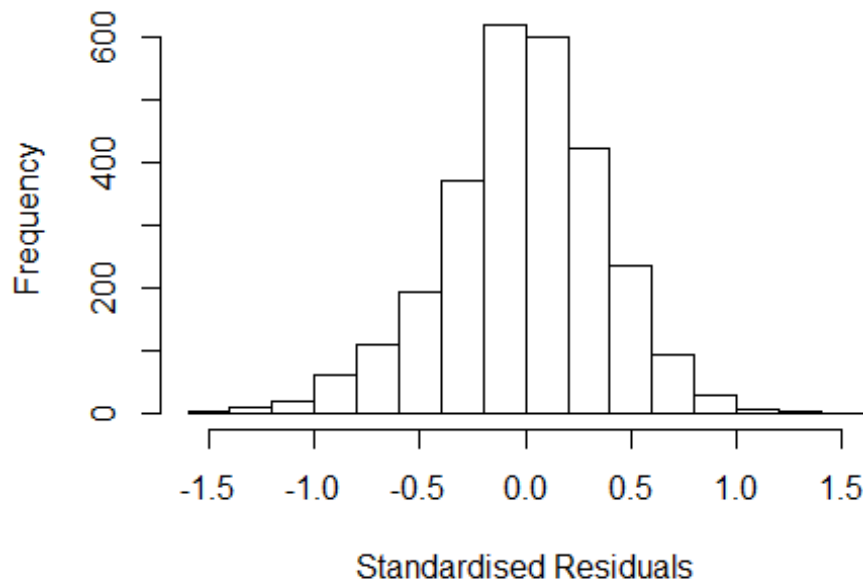


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.2, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 4.12137543181244"
## [1] "Male first author team size 2018 geometric mean: 3.63852007367249"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.72393820419443"
## [1] "Male last author team size 2018 geometric mean: 3.90402144291715"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7900, 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 1.050 1          1.025
## LastAuthorFemale  1.043 1          1.021
## UniqueAuthors    1.223 4          1.025
## Year              1.272 16         1.008
```

Residuals from first and last author and team size



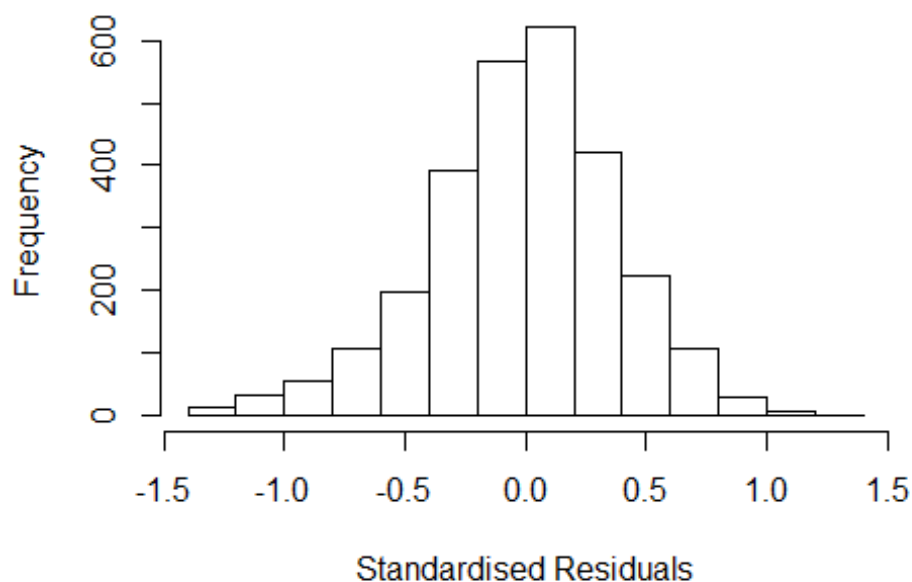
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.414019 -0.236988  0.000498  0.238493  1.407691
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17771    0.05879   20.03  < 2e-16 ***
## FirstAuthorFemale1 -0.00796    0.01504   -0.53  0.59692
## LastAuthorFemale1  0.01050    0.01866    0.56  0.57352
## UniqueAuthors2     0.21849    0.03942    5.54  3.2e-08 ***
## UniqueAuthors3     0.23376    0.03953    5.91  3.7e-09 ***
## UniqueAuthors4     0.21844    0.04045    5.40  7.2e-08 ***
## UniqueAuthors5     0.32019    0.03968    8.07  1.0e-15 ***
## Year1997          -0.02963    0.06275   -0.47  0.63684
## Year1998          -0.05690    0.06559   -0.87  0.38571
## Year1999          -0.16675    0.05839   -2.86  0.00432 **
```

```

## Year2000      -0.11218      0.05753      -1.95      0.05129      .
## Year2001      -0.17089      0.06582      -2.60      0.00947      **
## Year2002      -0.23721      0.05841      -4.06      5.0e-05      ***
## Year2003      -0.25978      0.05822      -4.46      8.5e-06      ***
## Year2004      -0.21975      0.05768      -3.81      0.00014      ***
## Year2005      -0.28389      0.05590      -5.08      4.1e-07      ***
## Year2006      -0.21516      0.05472      -3.93      8.6e-05      ***
## Year2007      -0.24773      0.05551      -4.46      8.4e-06      ***
## Year2008      -0.25855      0.05473      -4.72      2.4e-06      ***
## Year2009      -0.23419      0.05416      -4.32      1.6e-05      ***
## Year2010      -0.29495      0.05378      -5.48      4.5e-08      ***
## Year2011      -0.28738      0.05503      -5.22      1.9e-07      ***
## Year2012      -0.30187      0.05369      -5.62      2.1e-08      ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.0741, Adjusted R-squared:  0.0667
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 233 weights are ~= 1. The remaining 2541 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0836 0.8640 0.9520 0.8950 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          3.60e-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 1.044 1      1.022
## LastAuthorFemale 1.037 1      1.019
## Year 1.067 16      1.002

```


Residuals from first and last author



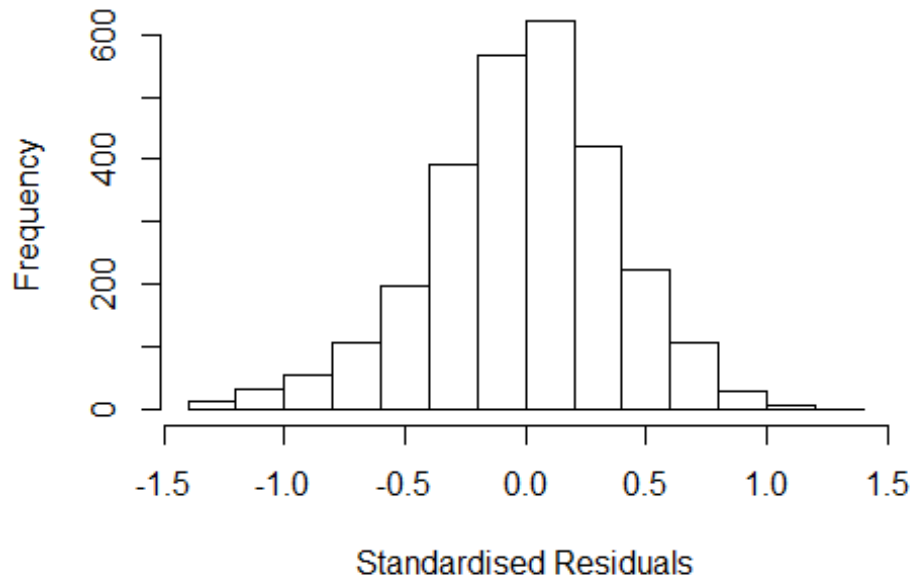
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3858 -0.2474 0.0074 0.2359 1.3780
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37947 0.05089 27.11 < 2e-16 ***
## FirstAuthorFemale1 0.00334 0.01529 0.22 0.82719
## LastAuthorFemale1 0.00297 0.01888 0.16 0.87524
## Year1997 -0.01671 0.06508 -0.26 0.79740
## Year1998 -0.04133 0.06586 -0.63 0.53032
## Year1999 -0.15504 0.06146 -2.52 0.01170 *
## Year2000 -0.09399 0.06028 -1.56 0.11905
## Year2001 -0.13244 0.06638 -2.00 0.04611 *
## Year2002 -0.22578 0.06048 -3.73 0.00019 ***
## Year2003 -0.24200 0.06024 -4.02 6.0e-05 ***
## Year2004 -0.19074 0.06003 -3.18 0.00150 **
## Year2005 -0.25900 0.05788 -4.47 8.0e-06 ***
```

```

## Year2006      -0.19762    0.05758   -3.43  0.00061 ***
## Year2007      -0.21541    0.05761   -3.74  0.00019 ***
## Year2008      -0.21928    0.05742   -3.82  0.00014 ***
## Year2009      -0.21073    0.05631   -3.74  0.00019 ***
## Year2010      -0.25742    0.05611   -4.59  4.7e-06 ***
## Year2011      -0.24317    0.05726   -4.25  2.2e-05 ***
## Year2012      -0.25403    0.05555   -4.57  5.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.362
## Multiple R-squared:  0.0368, Adjusted R-squared:  0.0305
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 230 weights are ~= 1. The remaining 2544 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.112  0.860  0.951  0.894  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.60e-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 1.037 1      1.018
## Year              1.037 16      1.001

```

Residuals from first author



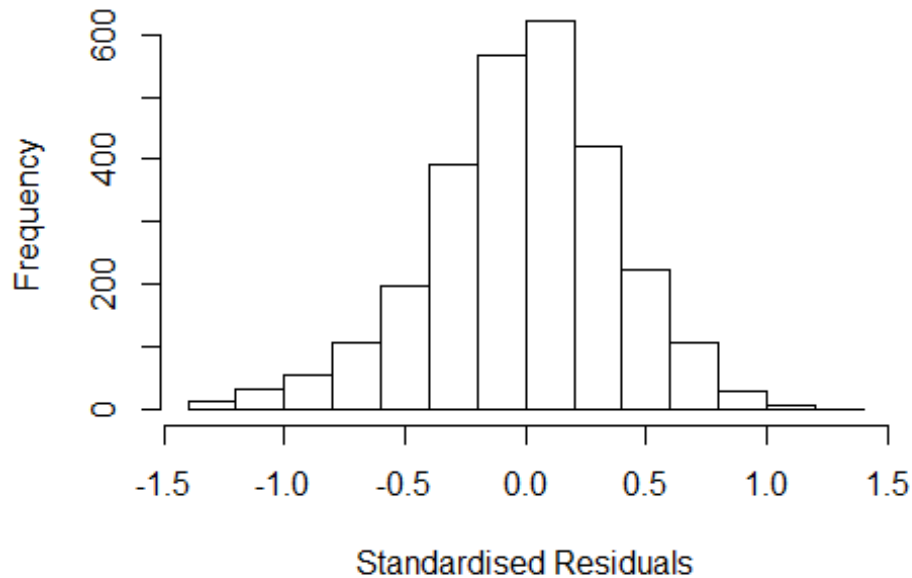
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.38340 -0.24608 0.00728 0.23590 1.38048
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37984 0.05081 27.16 < 2e-16 ***
## FirstAuthorFemale1 0.00356 0.01523 0.23 0.81516
## Year1997 -0.01666 0.06505 -0.26 0.79790
## Year1998 -0.04129 0.06584 -0.63 0.53058
## Year1999 -0.15504 0.06144 -2.52 0.01168 *
## Year2000 -0.09361 0.06009 -1.56 0.11939
## Year2001 -0.13245 0.06637 -2.00 0.04607 *
## Year2002 -0.22577 0.06045 -3.74 0.00019 ***
## Year2003 -0.24187 0.06020 -4.02 6.0e-05 ***
## Year2004 -0.19077 0.06001 -3.18 0.00149 **
## Year2005 -0.25892 0.05785 -4.48 7.9e-06 ***
## Year2006 -0.19760 0.05756 -3.43 0.00061 ***
```

```

## Year2007          -0.21532    0.05758   -3.74  0.00019 ***
## Year2008          -0.21915    0.05738   -3.82  0.00014 ***
## Year2009          -0.21067    0.05627   -3.74  0.00018 ***
## Year2010          -0.25720    0.05605   -4.59  4.7e-06 ***
## Year2011          -0.24303    0.05725   -4.25  2.3e-05 ***
## Year2012          -0.25379    0.05551   -4.57  5.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.362
## Multiple R-squared:  0.0368, Adjusted R-squared:  0.0309
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 230 weights are ~= 1. The remaining 2544 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.113  0.860  0.951  0.894  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.60e-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 1.03 1      1.015
## Year      1.03 16      1.001

```

Residuals from last author



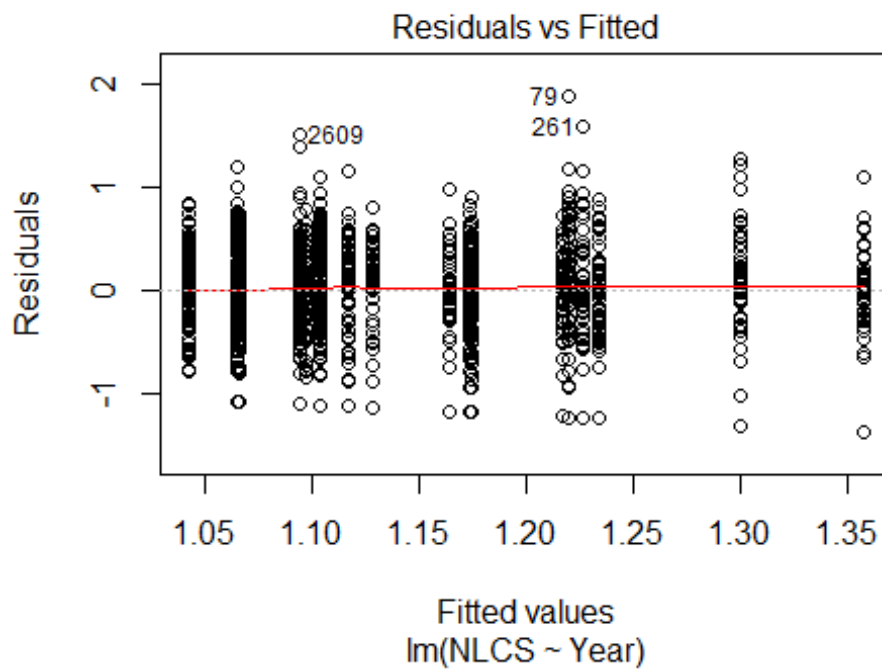
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.38375 -0.24695 0.00815 0.23614 1.37640
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.38040 0.05078 27.18 < 2e-16 ***
## LastAuthorFemale1 0.00335 0.01881 0.18 0.85880
## Year1997 -0.01688 0.06512 -0.26 0.79543
## Year1998 -0.04155 0.06589 -0.63 0.52840
## Year1999 -0.15484 0.06143 -2.52 0.01178 *
## Year2000 -0.09413 0.06031 -1.56 0.11871
## Year2001 -0.13251 0.06641 -2.00 0.04611 *
## Year2002 -0.22600 0.06050 -3.74 0.00019 ***
## Year2003 -0.24179 0.06022 -4.01 6.1e-05 ***
## Year2004 -0.19059 0.06005 -3.17 0.00152 **
## Year2005 -0.25871 0.05786 -4.47 8.1e-06 ***
## Year2006 -0.19738 0.05759 -3.43 0.00062 ***
```

```

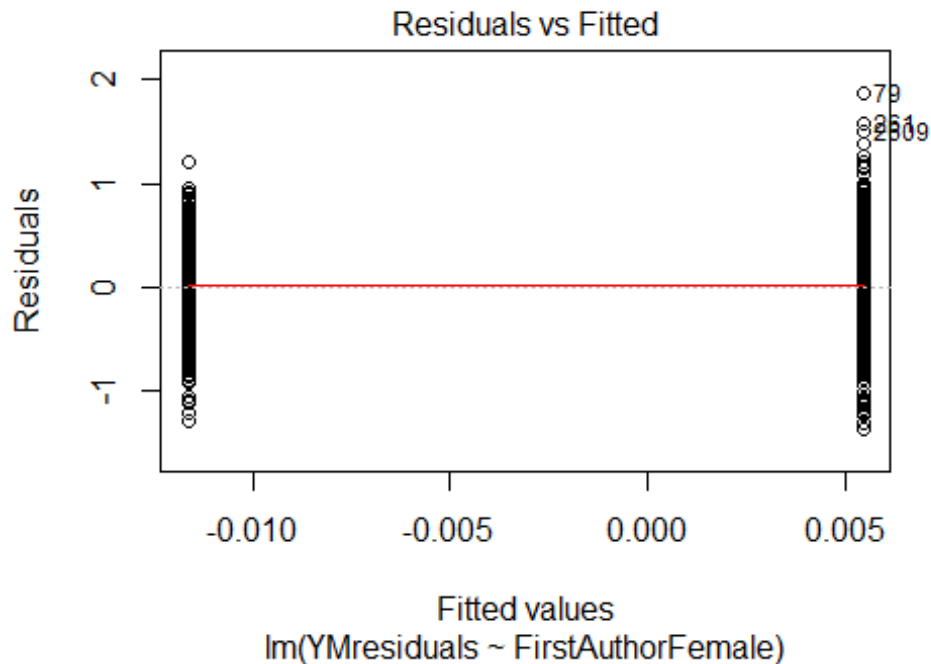
## Year2007          -0.21515      0.05760    -3.74  0.00019 ***
## Year2008          -0.21876      0.05732    -3.82  0.00014 ***
## Year2009          -0.21034      0.05627    -3.74  0.00019 ***
## Year2010          -0.25695      0.05607    -4.58  4.8e-06 ***
## Year2011          -0.24291      0.05726    -4.24  2.3e-05 ***
## Year2012          -0.25374      0.05554    -4.57  5.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.362
## Multiple R-squared:  0.0368, Adjusted R-squared:  0.0309
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 233 weights are ~= 1. The remaining 2541 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.113  0.860  0.951  0.894  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.60e-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: 2774"
## [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
## 125 126 113 95 105 125 101 100 103 136 144 162 219 274 259
## 2011 2012
## 260 295
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 64 51 51 37 46 49 48 52 60 72 79 84 127 146 130
## 2011 2012

```

```
## 135 163
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 39 41 30 43 46 44 39 49 59 54 63 102 126 105
## 2011 2012
## 119 137
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 96, df = 16, p-value = 2e-13
```

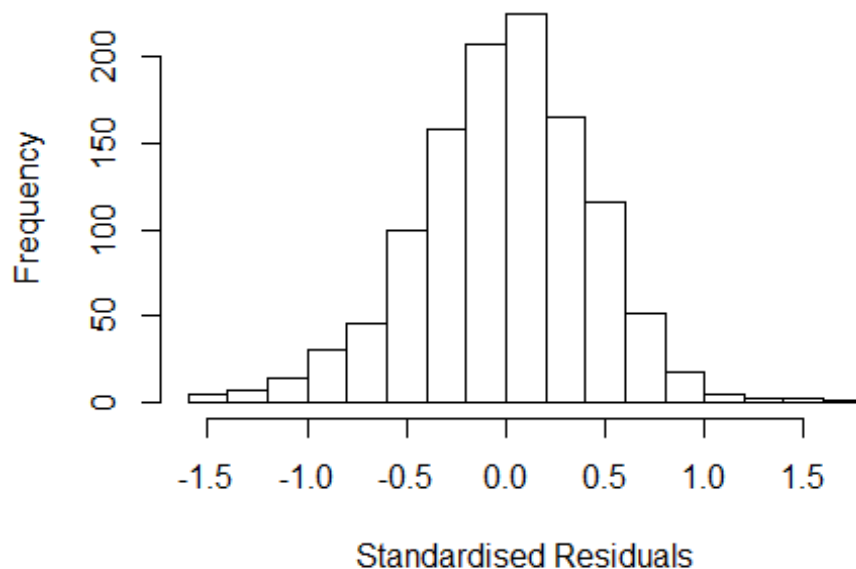


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 9e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.17293697526185"
## [1] "Male first author team size 2018 geometric mean: 3.38102198443253"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3700, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.11228931152141"
## [1] "Male last author team size 2018 geometric mean: 3.54018945962533"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.111 1      1.054
## LastAuthorFemale  1.089 1      1.043
## UniqueAuthors    1.493 4      1.051
## Year             1.549 16      1.014
```


Residuals from first and last author and team size



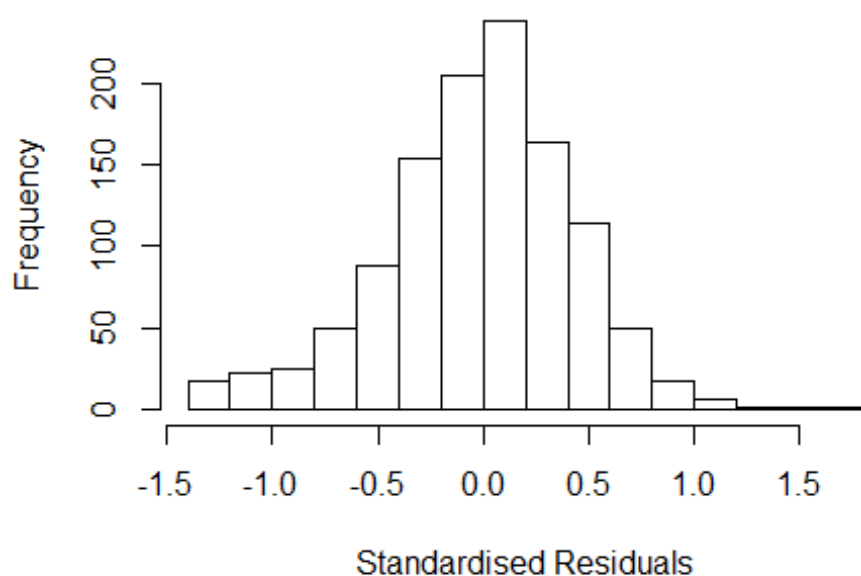
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.46812 -0.26445 0.00589 0.27375 1.67084
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1278 0.0978 11.53 < 2e-16 ***
## FirstAuthorFemale1 -0.0410 0.0272 -1.51 0.1316
## LastAuthorFemale1 -0.1198 0.0348 -3.44 0.0006 ***
## UniqueAuthors2 0.2954 0.0622 4.75 2.3e-06 ***
## UniqueAuthors3 0.2691 0.0625 4.30 1.8e-05 ***
## UniqueAuthors4 0.2603 0.0639 4.07 5.0e-05 ***
## UniqueAuthors5 0.3404 0.0612 5.57 3.3e-08 ***
## Year1997 -0.1591 0.1405 -1.13 0.2579
## Year1998 0.0167 0.1530 0.11 0.9130
## Year1999 -0.0486 0.1122 -0.43 0.6652
```

```

## Year2000          -0.1732      0.1176   -1.47    0.1412
## Year2001          -0.1990      0.1137   -1.75    0.0802 .
## Year2002          -0.2031      0.1085   -1.87    0.0615 .
## Year2003          -0.1661      0.1065   -1.56    0.1192
## Year2004          -0.1205      0.1102   -1.09    0.2744
## Year2005          -0.2658      0.1077   -2.47    0.0137 *
## Year2006          -0.1148      0.1048   -1.10    0.2737
## Year2007          -0.1777      0.1051   -1.69    0.0913 .
## Year2008          -0.3033      0.1029   -2.95    0.0033 **
## Year2009          -0.2565      0.1006   -2.55    0.0109 *
## Year2010          -0.2999      0.1030   -2.91    0.0037 **
## Year2011          -0.3110      0.1006   -3.09    0.0020 **
## Year2012          -0.2297      0.1003   -2.29    0.0223 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0905, Adjusted R-squared:  0.0728
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1034 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.059  0.862  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          8.70e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.056 1          1.028
## LastAuthorFemale  1.065 1          1.032
## Year              1.124 16          1.004

```

Residuals from first and last author



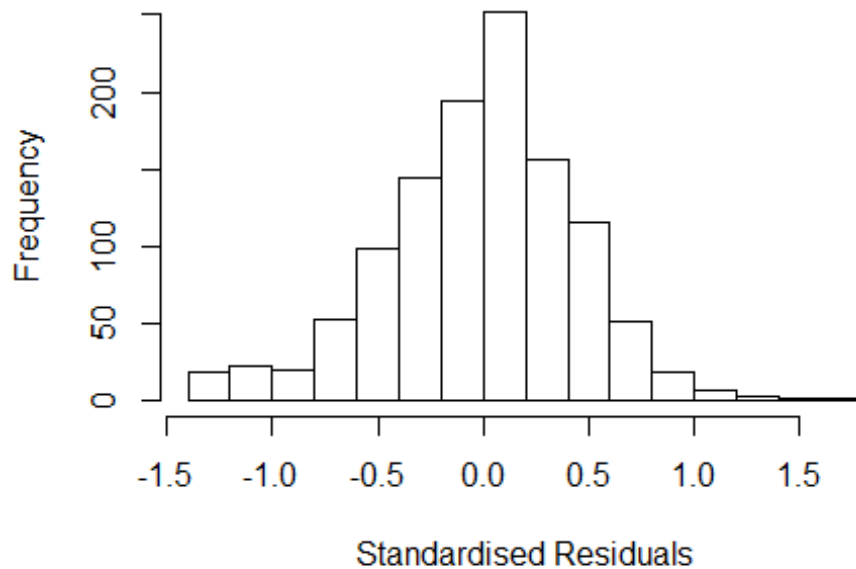
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.374 -0.273 0.012 0.262 1.780
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3137 0.1024 12.83 < 2e-16 ***
## FirstAuthorFemale1 -0.0294 0.0271 -1.08 0.27955
## LastAuthorFemale1 -0.1291 0.0352 -3.67 0.00026 ***
## Year1997 -0.0863 0.1470 -0.59 0.55731
## Year1998 0.0603 0.1562 0.39 0.69945
## Year1999 0.0380 0.1203 0.32 0.75209
## Year2000 -0.0971 0.1254 -0.77 0.43890
## Year2001 -0.1210 0.1248 -0.97 0.33236
## Year2002 -0.1286 0.1144 -1.12 0.26119
## Year2003 -0.0875 0.1152 -0.76 0.44771
## Year2004 -0.0547 0.1177 -0.46 0.64253
## Year2005 -0.1994 0.1147 -1.74 0.08236 .
```

```

## Year2006          -0.0192      0.1120   -0.17  0.86370
## Year2007          -0.0888      0.1138   -0.78  0.43578
## Year2008          -0.2237      0.1133   -1.98  0.04848 *
## Year2009          -0.1791      0.1085   -1.65  0.09912 .
## Year2010          -0.2110      0.1104   -1.91  0.05624 .
## Year2011          -0.2292      0.1092   -2.10  0.03599 *
## Year2012          -0.1482      0.1090   -1.36  0.17429
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0495, Adjusted R-squared:  0.0343
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1041 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0237 0.8650 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      8.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.059 1      1.029
## Year      1.059 16      1.002

```

Residuals from first author



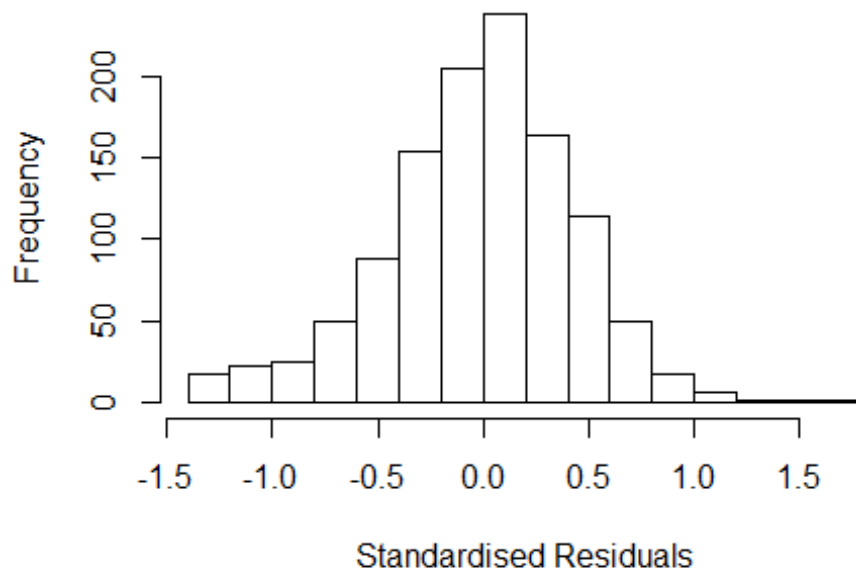
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3568 -0.2736 0.0176 0.2684 1.7950
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2990 0.1046 12.42 <2e-16 ***
## FirstAuthorFemale1 -0.0384 0.0273 -1.41 0.160
## Year1997 -0.0687 0.1498 -0.46 0.646
## Year1998 0.0578 0.1587 0.36 0.716
## Year1999 0.0354 0.1214 0.29 0.771
## Year2000 -0.1060 0.1278 -0.83 0.407
## Year2001 -0.1379 0.1267 -1.09 0.277
## Year2002 -0.1369 0.1158 -1.18 0.237
## Year2003 -0.0823 0.1180 -0.70 0.486
## Year2004 -0.0618 0.1204 -0.51 0.608
## Year2005 -0.2016 0.1171 -1.72 0.086 .
## Year2006 -0.0212 0.1143 -0.19 0.853
```

```

## Year2007          -0.0926      0.1163   -0.80    0.426
## Year2008          -0.2340      0.1154   -2.03    0.043 *
## Year2009          -0.1811      0.1110   -1.63    0.103
## Year2010          -0.2048      0.1123   -1.82    0.068 .
## Year2011          -0.2392      0.1118   -2.14    0.033 *
## Year2012          -0.1538      0.1114   -1.38    0.167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.414
## Multiple R-squared:  0.0383, Adjusted R-squared:  0.0239
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 105 weights are ~= 1. The remaining 1045 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0209 0.8600 0.9520 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      8.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.064 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
## -1.3641 -0.2703 0.0154 0.2670 1.7865
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3075 0.1027 12.73 < 2e-16 ***
## LastAuthorFemale1 -0.1323 0.0353 -3.75 0.00018 ***
## Year1997 -0.0878 0.1468 -0.60 0.54989
## Year1998 0.0566 0.1578 0.36 0.71995
## Year1999 0.0353 0.1209 0.29 0.77033
## Year2000 -0.0977 0.1261 -0.77 0.43876
## Year2001 -0.1237 0.1250 -0.99 0.32268
## Year2002 -0.1307 0.1150 -1.14 0.25581
## Year2003 -0.0879 0.1156 -0.76 0.44725
## Year2004 -0.0566 0.1180 -0.48 0.63149
## Year2005 -0.2012 0.1147 -1.75 0.07972 .
## Year2006 -0.0251 0.1121 -0.22 0.82310
```

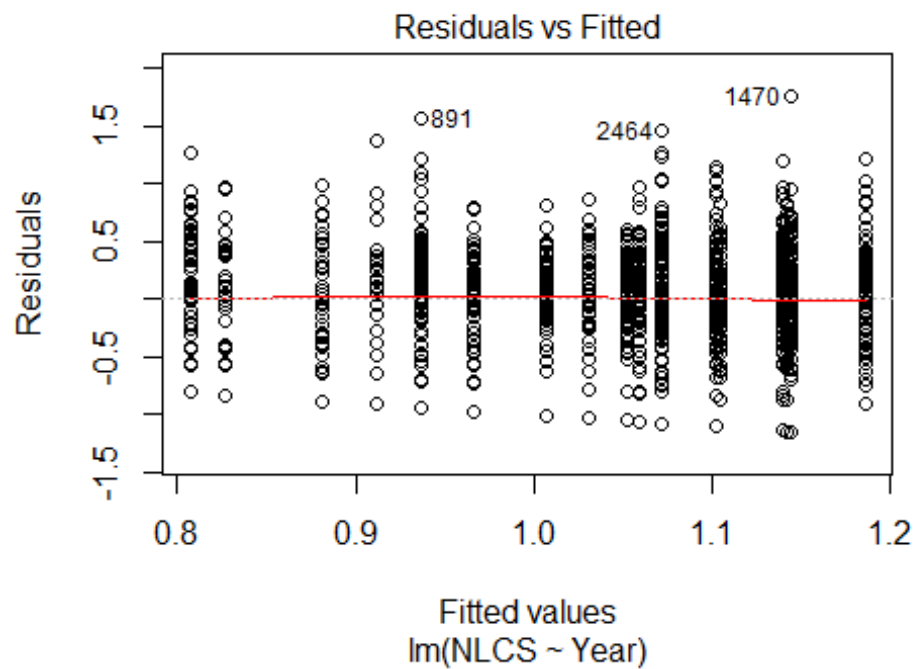
```

## Year2007          -0.0916      0.1139   -0.80   0.42166
## Year2008          -0.2252      0.1135   -1.98   0.04753 *
## Year2009          -0.1850      0.1085   -1.70   0.08854 .
## Year2010          -0.2160      0.1105   -1.95   0.05089 .
## Year2011          -0.2325      0.1094   -2.12   0.03383 *
## Year2012          -0.1498      0.1093   -1.37   0.17092
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0482, Adjusted R-squared:  0.034
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 1048 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0216 0.8650 0.9510 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      8.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1150"
## [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
## 111 107 110 86 156 129 143 116 125 113 131 112 125 164 156
## 2011 2012
## 191 182
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 40 29 57 53 76 42 100 71 75 71 85 74 90 108 115
## 2011 2012

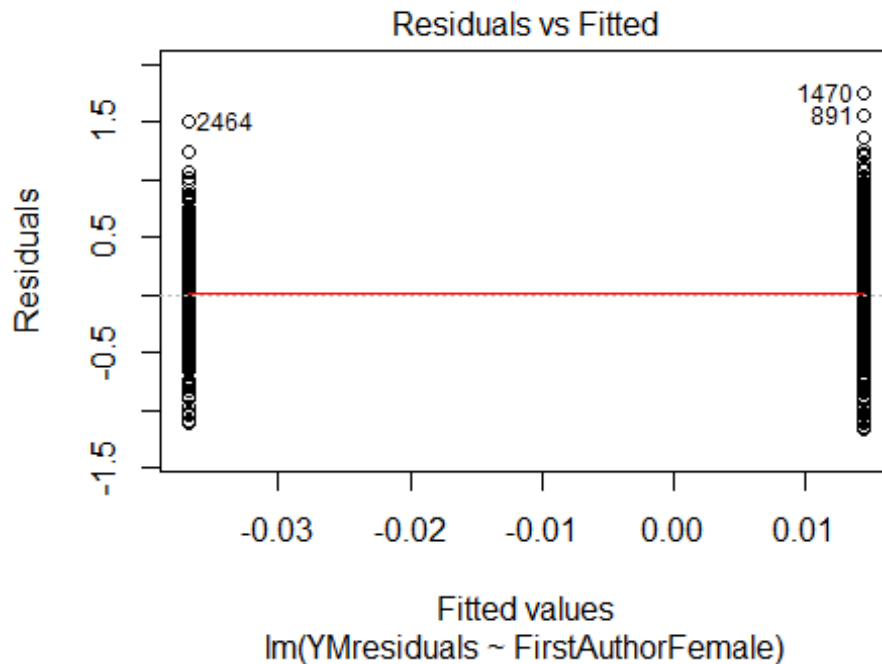
```



```
## 147 140
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 37 27 49 50 64 39 87 61 71 59 72 67 83 93 101
## 2011 2012
## 135 125
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 71, df = 16, p-value = 8e-09
```

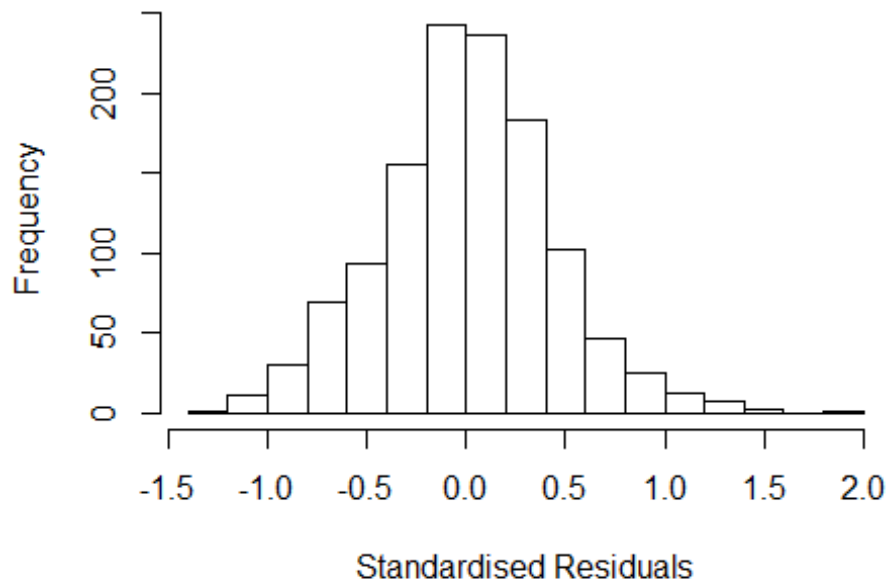


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.71, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 2.81634305384055"
## [1] "Male first author team size 2018 geometric mean: 2.55895627938071"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.58371407899625"
## [1] "Male last author team size 2018 geometric mean: 2.68000622077513"
##
## 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 1.139 1          1.067
## LastAuthorFemale  1.106 1          1.052
## UniqueAuthors    1.488 4          1.051
## Year             1.559 16          1.014
```

Residuals from first and last author and team size



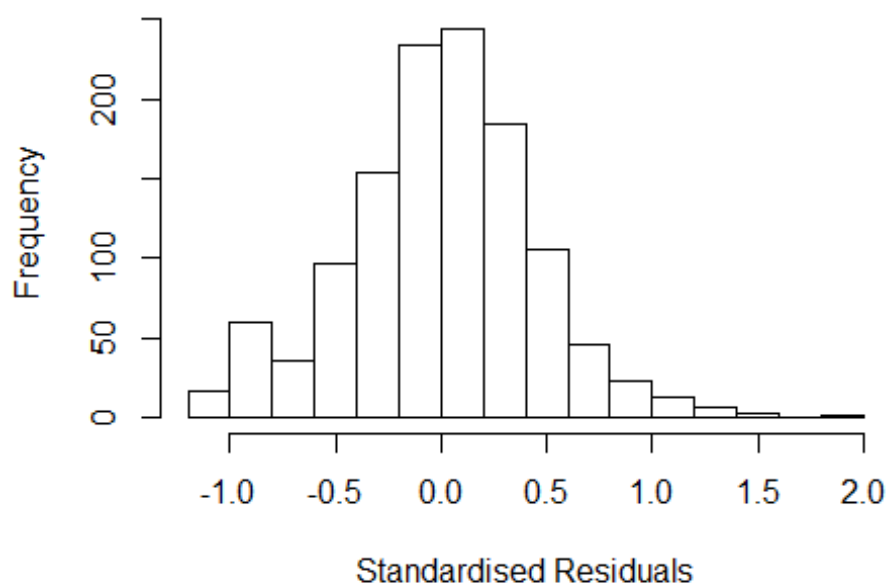
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22082 -0.27398  0.00814  0.26390  1.82381
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7306    0.0675   10.82 < 2e-16 ***
## FirstAuthorFemale1 -0.0525    0.0277   -1.89  0.0585 .
## LastAuthorFemale1 -0.0793    0.0339   -2.34  0.0195 *
## UniqueAuthors2     0.2138    0.0457    4.68 3.2e-06 ***
## UniqueAuthors3     0.2467    0.0467    5.28 1.5e-07 ***
## UniqueAuthors4     0.2434    0.0495    4.92 1.0e-06 ***
## UniqueAuthors5     0.3137    0.0517    6.06 1.8e-09 ***
## Year1997          0.1098    0.1339    0.82  0.4125
## Year1998          0.0757    0.0887    0.85  0.3936
## Year1999          0.1686    0.0892    1.89  0.0591 .
```

```

## Year2000          -0.0563      0.1018    -0.55    0.5805
## Year2001          -0.0534      0.1053    -0.51    0.6121
## Year2002           0.0610      0.0902     0.68    0.4991
## Year2003           0.1765      0.0768     2.30    0.0217 *
## Year2004           0.1440      0.0766     1.88    0.0603 .
## Year2005           0.1602      0.0805     1.99    0.0466 *
## Year2006           0.2131      0.0796     2.68    0.0075 **
## Year2007           0.2008      0.0778     2.58    0.0100 **
## Year2008           0.2429      0.0806     3.01    0.0027 **
## Year2009           0.2469      0.0751     3.29    0.0010 **
## Year2010           0.1886      0.0767     2.46    0.0141 *
## Year2011           0.1554      0.0771     2.02    0.0440 *
## Year2012           0.2365      0.0768     3.08    0.0021 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.114, Adjusted R-squared:  0.0977
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1104 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.003  0.856  0.949   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          8.20e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.075 1          1.037
## LastAuthorFemale  1.115 1          1.056
## Year              1.179 16          1.005

```

Residuals from first and last author



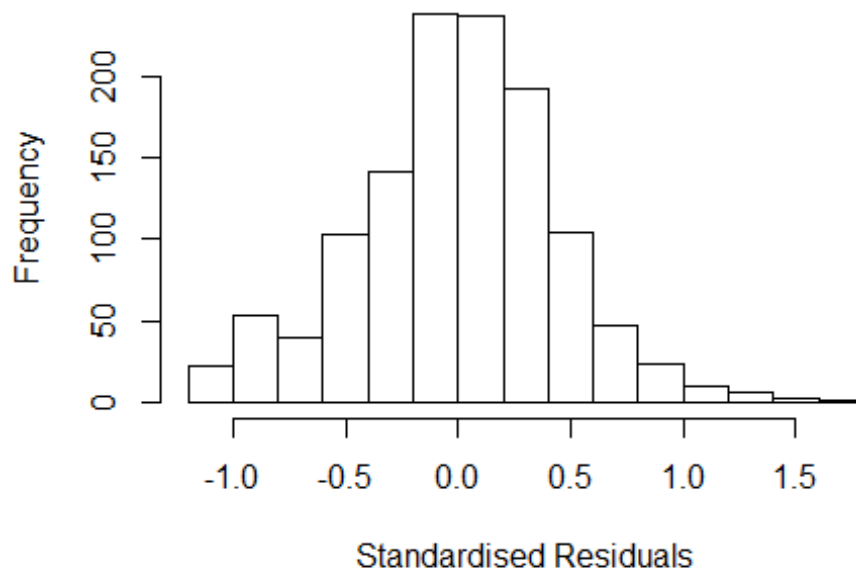
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20 -0.26 0.01 0.27 1.83
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8678 0.0753 11.52 < 2e-16 ***
## FirstAuthorFemale1 -0.0338 0.0278 -1.22 0.22378
## LastAuthorFemale1 -0.0790 0.0351 -2.25 0.02448 *
## Year1997 0.1053 0.1395 0.75 0.45049
## Year1998 0.1201 0.0981 1.22 0.22126
## Year1999 0.2227 0.0964 2.31 0.02103 *
## Year2000 -0.0533 0.1180 -0.45 0.65172
## Year2001 -0.0336 0.1154 -0.29 0.77098
## Year2002 0.1000 0.1060 0.94 0.34568
## Year2003 0.2419 0.0861 2.81 0.00503 **
## Year2004 0.2058 0.0868 2.37 0.01797 *
## Year2005 0.1977 0.0897 2.20 0.02777 *
```

```

## Year2006          0.2838      0.0883      3.21  0.00135 **
## Year2007          0.2552      0.0868      2.94  0.00335 **
## Year2008          0.3222      0.0890      3.62  0.00031 ***
## Year2009          0.3308      0.0838      3.95  8.3e-05 ***
## Year2010          0.2568      0.0854      3.01  0.00270 **
## Year2011          0.2183      0.0869      2.51  0.01213 *
## Year2012          0.3146      0.0850      3.70  0.00022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.0653, Adjusted R-squared:  0.0513
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1102 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0037 0.8490 0.9490 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.20e-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 1.074 1      1.036
## Year              1.074 16      1.002

```

Residuals from first author



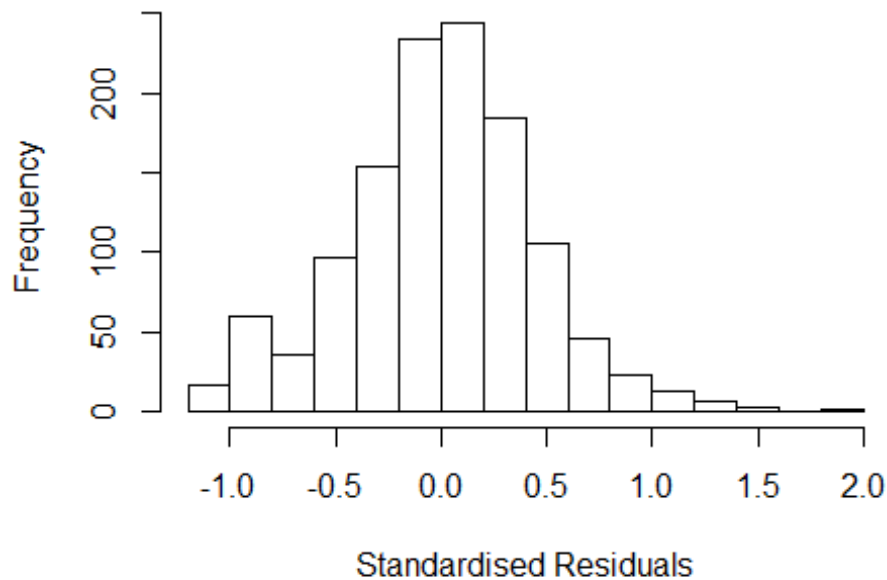
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1821 -0.2545 0.0124 0.2768 1.7522
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8510 0.0772 11.03 < 2e-16 ***
## FirstAuthorFemale1 -0.0465 0.0280 -1.66 0.09754 .
## Year1997 0.1101 0.1435 0.77 0.44296
## Year1998 0.1229 0.0996 1.23 0.21750
## Year1999 0.2232 0.0990 2.25 0.02434 *
## Year2000 -0.0445 0.1190 -0.37 0.70810
## Year2001 -0.0219 0.1169 -0.19 0.85129
## Year2002 0.1135 0.1082 1.05 0.29456
## Year2003 0.2452 0.0883 2.78 0.00557 **
## Year2004 0.2101 0.0893 2.35 0.01880 *
## Year2005 0.1978 0.0914 2.16 0.03068 *
## Year2006 0.2987 0.0901 3.31 0.00094 ***
```

```

## Year2007          0.2586      0.0886      2.92  0.00358 **
## Year2008          0.3322      0.0905      3.67  0.00025 ***
## Year2009          0.3311      0.0862      3.84  0.00013 ***
## Year2010          0.2631      0.0876      3.00  0.00273 **
## Year2011          0.2225      0.0889      2.50  0.01252 *
## Year2012          0.3203      0.0871      3.68  0.00025 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.404
## Multiple R-squared:  0.0598, Adjusted R-squared:  0.0465
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 1112 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0203 0.8530 0.9510 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      8.20e-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 1.11 1          1.054
## Year              1.11 16          1.003

```


Residuals from last author



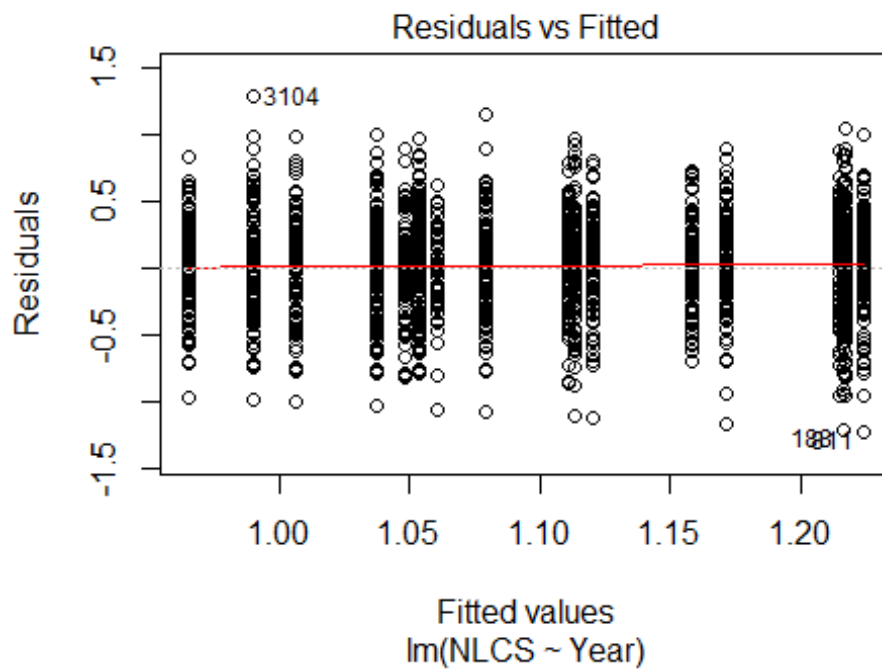
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18675 -0.25445 0.00432 0.27262 1.84868
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8675 0.0758 11.44 < 2e-16 ***
## LastAuthorFemale1 -0.0863 0.0352 -2.45 0.01431 *
## Year1997 0.0961 0.1393 0.69 0.49038
## Year1998 0.1138 0.0983 1.16 0.24712
## Year1999 0.2166 0.0965 2.24 0.02500 *
## Year2000 -0.0582 0.1184 -0.49 0.62337
## Year2001 -0.0394 0.1151 -0.34 0.73241
## Year2002 0.0903 0.1062 0.85 0.39514
## Year2003 0.2335 0.0861 2.71 0.00680 **
## Year2004 0.2019 0.0871 2.32 0.02069 *
## Year2005 0.1898 0.0897 2.12 0.03453 *
## Year2006 0.2721 0.0877 3.10 0.00197 **
```

```

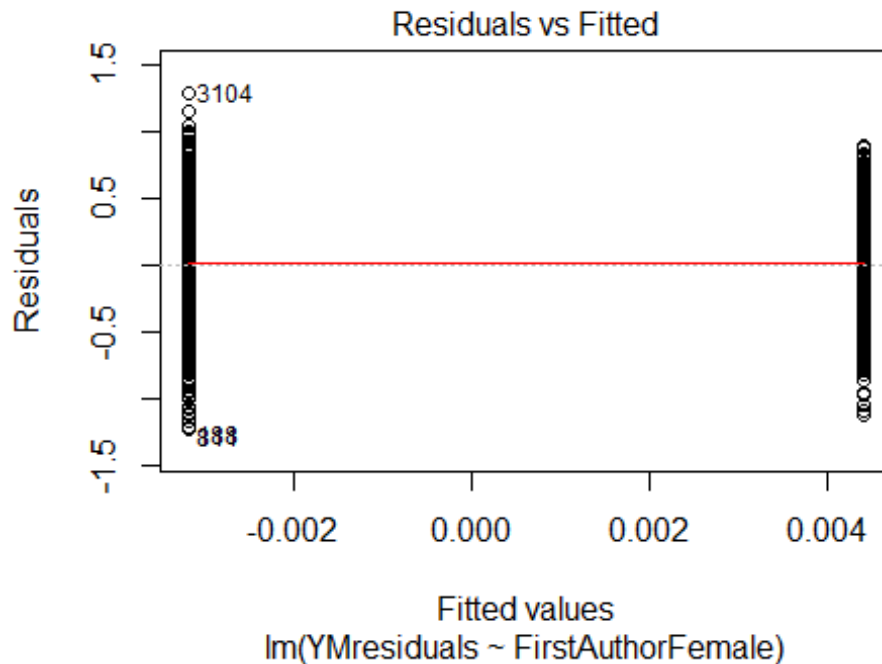
## Year2007          0.2497      0.0869      2.87  0.00411 **
## Year2008          0.3133      0.0888      3.53  0.00044 ***
## Year2009          0.3193      0.0835      3.82  0.00014 ***
## Year2010          0.2475      0.0854      2.90  0.00382 **
## Year2011          0.2087      0.0867      2.41  0.01618 *
## Year2012          0.3050      0.0848      3.59  0.00034 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.0641, Adjusted R-squared:  0.0509
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 117 weights are ~= 1. The remaining 1103 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0018 0.8550 0.9490 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.20e-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: 1220"
## [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
## 184 187 170 160 181 181 182 163 169 168 163 163 193 239 241
## 2011 2012
## 264 220
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 82 105 83 93 101 64 94 103 104 110 112 113 121 149 157
## 2011 2012

```

```
## 172 158
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 70 87 76 85 85 56 81 82 91 97 88 91 107 126 132
## 2011 2012
## 148 143
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```

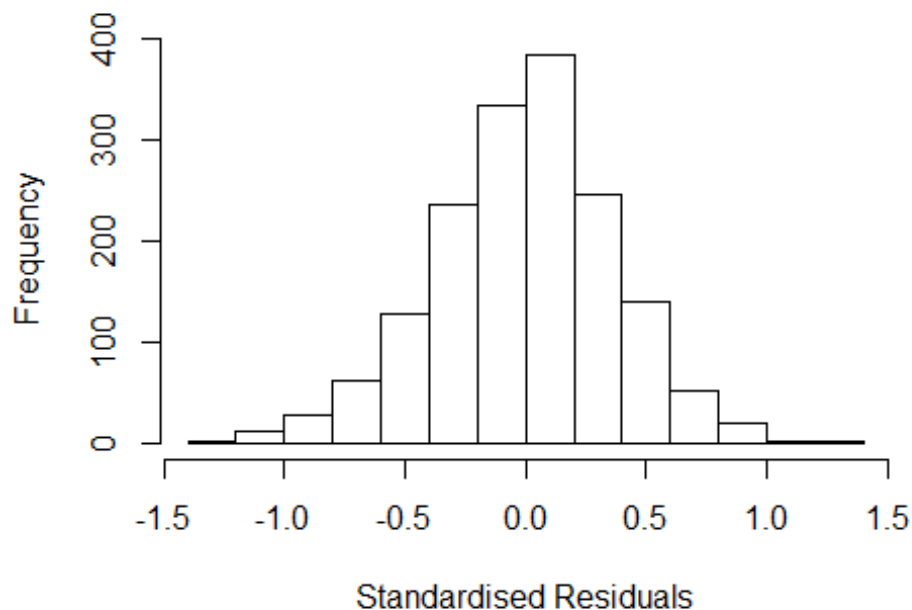


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 16, df = 1, p-value = 8e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.33827177957732"
## [1] "Male first author team size 2018 geometric mean: 3.88255396242911"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5200, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.18788347381356"
## [1] "Male last author team size 2018 geometric mean: 4.06823633667681"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4200, 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 1.062 1          1.031
## LastAuthorFemale  1.071 1          1.035
## UniqueAuthors    1.255 4          1.029
## Year              1.336 16         1.009
```

Residuals from first and last author and team size



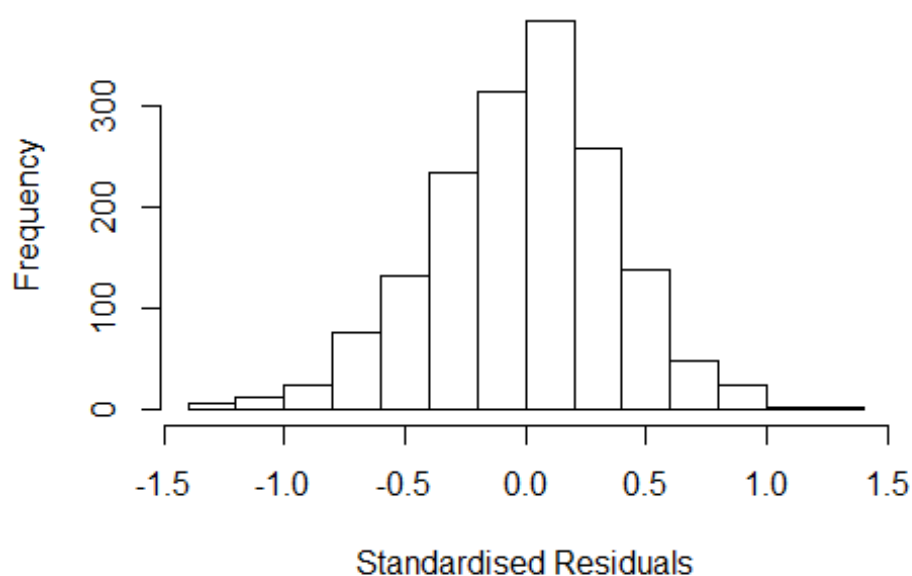
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30892 -0.23287  0.00664  0.23040  1.29236
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.01598    0.07577   13.41 < 2e-16 ***
## FirstAuthorFemale1 -0.00317    0.01898   -0.17  0.86720
## LastAuthorFemale1  0.01214    0.02206    0.55  0.58224
## UniqueAuthors2     0.20983    0.05717    3.67  0.00025 ***
## UniqueAuthors3     0.20241    0.05623    3.60  0.00033 ***
## UniqueAuthors4     0.27336    0.05653    4.84  1.5e-06 ***
## UniqueAuthors5     0.31386    0.05577    5.63  2.1e-08 ***
## Year1997          -0.00456    0.07335   -0.06  0.95043
## Year1998          -0.02734    0.06692   -0.41  0.68299
## Year1999          -0.00056    0.06958   -0.01  0.99358
```

```

## Year2000      -0.03306      0.06923      -0.48      0.63308
## Year2001      -0.14147      0.06553      -2.16      0.03101 *
## Year2002      -0.09700      0.06402      -1.52      0.12991
## Year2003      -0.12231      0.06631      -1.84      0.06528 .
## Year2004      -0.12935      0.06678      -1.94      0.05293 .
## Year2005      -0.24234      0.06210      -3.90      9.9e-05 ***
## Year2006      -0.11541      0.06442      -1.79      0.07341 .
## Year2007      -0.23015      0.06885      -3.34      0.00085 ***
## Year2008      -0.29662      0.06120      -4.85      1.4e-06 ***
## Year2009      -0.24817      0.06293      -3.94      8.4e-05 ***
## Year2010      -0.21983      0.06110      -3.60      0.00033 ***
## Year2011      -0.23142      0.06110      -3.79      0.00016 ***
## Year2012      -0.18580      0.06053      -3.07      0.00218 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.0901, Adjusted R-squared:  0.0777
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 153 weights are ~= 1. The remaining 1492 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.132  0.855   0.949   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          6.08e-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 1.048 1      1.024
## LastAuthorFemale  1.046 1      1.023
## Year              1.082 16      1.002

```

Residuals from first and last author



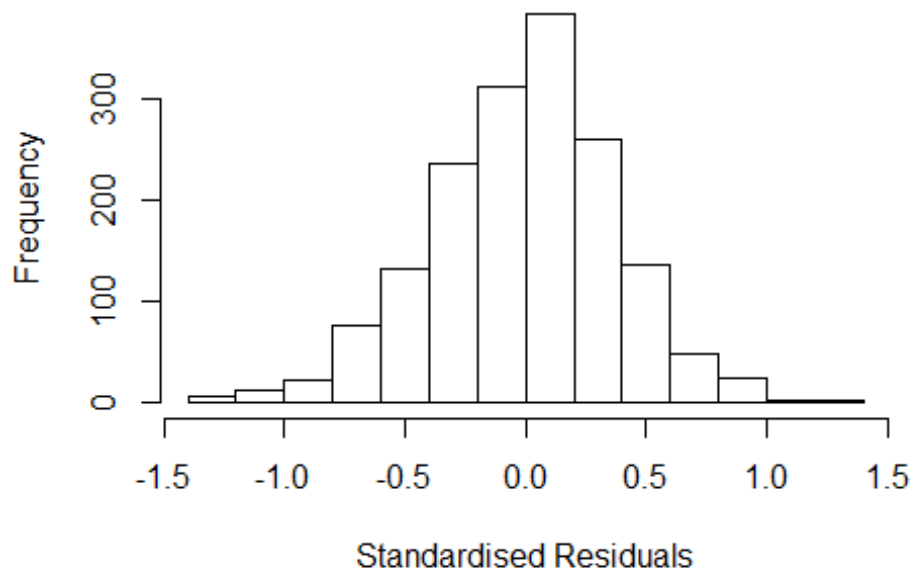
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2405 -0.2493  0.0181  0.2332  1.2723
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.239930   0.057036   21.74  < 2e-16 ***
## FirstAuthorFemale1  0.009233   0.019214    0.48  0.63090
## LastAuthorFemale1  0.008500   0.022115    0.38  0.70077
## Year1997        -0.017289   0.077459   -0.22  0.82340
## Year1998        -0.025930   0.070962   -0.37  0.71486
## Year1999         0.000607   0.073398    0.01  0.99340
## Year2000        -0.036178   0.071261   -0.51  0.61175
## Year2001        -0.139677   0.069730   -2.00  0.04533 *
## Year2002        -0.085707   0.068271   -1.26  0.20952
## Year2003        -0.103685   0.069558   -1.49  0.13625
## Year2004        -0.124372   0.069186   -1.80  0.07242 .
## Year2005        -0.216435   0.064873   -3.34  0.00087 ***
```

```

## Year2006      -0.107566    0.068088    -1.58    0.11434
## Year2007      -0.213860    0.072354    -2.96    0.00316 **
## Year2008      -0.294323    0.065832    -4.47    8.3e-06 ***
## Year2009      -0.242234    0.066812    -3.63    0.00030 ***
## Year2010      -0.199869    0.064377    -3.10    0.00194 **
## Year2011      -0.199685    0.064281    -3.11    0.00193 **
## Year2012      -0.160847    0.064293    -2.50    0.01246 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.0519, Adjusted R-squared:  0.0414
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 127 weights are ~= 1. The remaining 1518 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.189  0.866  0.953  0.899  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.08e-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 1.042 1      1.021
## Year      1.042 16      1.001

```


Residuals from first author



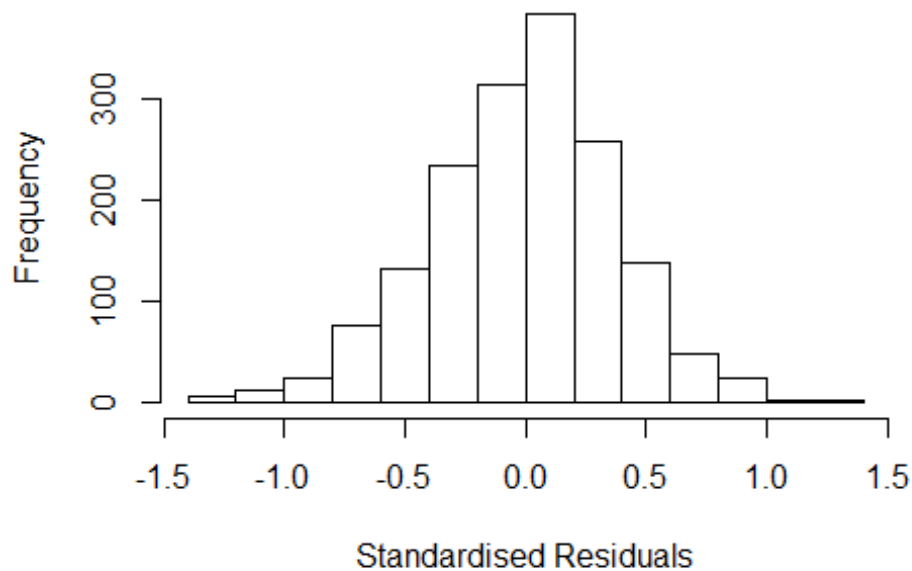
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2416 -0.2505  0.0164  0.2314  1.2700
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.24145    0.05677   21.87 < 2e-16 ***
## FirstAuthorFemale1 0.00978    0.01916    0.51  0.60997
## Year1997       -0.01680    0.07744   -0.22  0.82828
## Year1998       -0.02654    0.07095   -0.37  0.70838
## Year1999        0.00019    0.07342    0.00  0.99793
## Year2000       -0.03612    0.07130   -0.51  0.61251
## Year2001       -0.14007    0.06964   -2.01  0.04445 *
## Year2002       -0.08607    0.06829   -1.26  0.20770
## Year2003       -0.10388    0.06960   -1.49  0.13578
## Year2004       -0.12456    0.06926   -1.80  0.07227 .
## Year2005       -0.21655    0.06491   -3.34  0.00087 ***
## Year2006       -0.10659    0.06815   -1.56  0.11803
```

```

## Year2007          -0.21283    0.07250   -2.94  0.00338 **
## Year2008          -0.29340    0.06593   -4.45  9.2e-06 ***
## Year2009          -0.24145    0.06687   -3.61  0.00031 ***
## Year2010          -0.19897    0.06458   -3.08  0.00210 **
## Year2011          -0.19925    0.06440   -3.09  0.00201 **
## Year2012          -0.16066    0.06434   -2.50  0.01263 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.0519, Adjusted R-squared:  0.042
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1514 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.189  0.865  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      6.08e-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 1.041 1          1.020
## Year            1.041 16          1.001

```

Residuals from last author



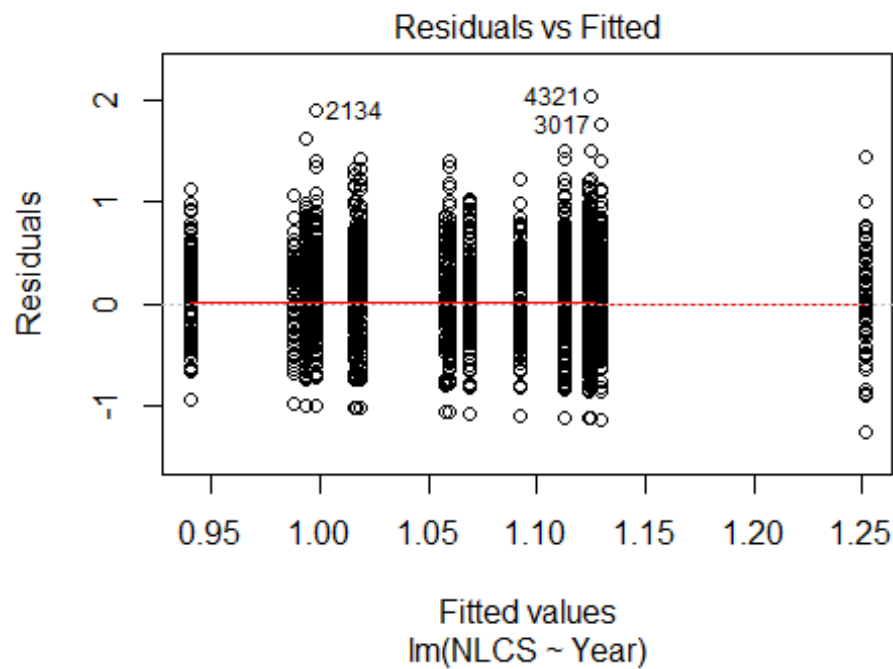
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2445 -0.2524 0.0192 0.2335 1.2675
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.24270 0.05667 21.93 < 2e-16 ***
## LastAuthorFemale1 0.00932 0.02205 0.42 0.67259
## Year1997 -0.01823 0.07736 -0.24 0.81373
## Year1998 -0.02604 0.07101 -0.37 0.71385
## Year1999 0.00179 0.07324 0.02 0.98047
## Year2000 -0.03551 0.07119 -0.50 0.61796
## Year2001 -0.13928 0.06978 -2.00 0.04610 *
## Year2002 -0.08549 0.06834 -1.25 0.21114
## Year2003 -0.10318 0.06948 -1.49 0.13774
## Year2004 -0.12265 0.06899 -1.78 0.07561 .
## Year2005 -0.21569 0.06487 -3.32 0.00090 ***
## Year2006 -0.10633 0.06802 -1.56 0.11817
```

```

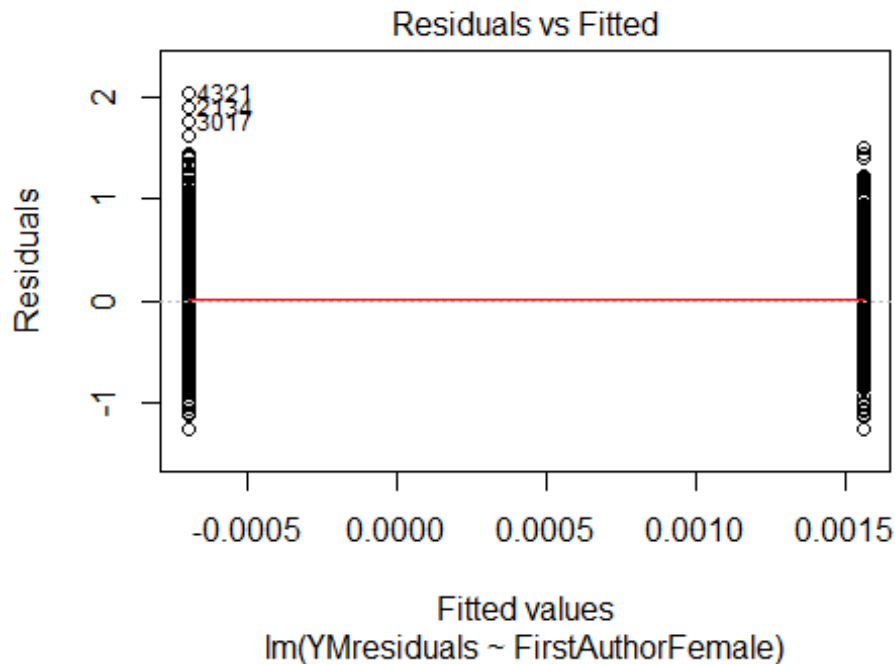
## Year2007          -0.21281      0.07221    -2.95  0.00325 **
## Year2008          -0.29252      0.06565    -4.46  8.9e-06 ***
## Year2009          -0.24021      0.06666    -3.60  0.00032 ***
## Year2010          -0.19810      0.06426    -3.08  0.00208 **
## Year2011          -0.19838      0.06423    -3.09  0.00205 **
## Year2012          -0.15887      0.06414    -2.48  0.01335 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.0518, Adjusted R-squared:  0.0418
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~ = 1. The remaining 1516 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.193  0.867  0.953  0.899  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.08e-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: 1645"
## [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
## 108 159 156 139 160 223 197 212 171 186 246 265 255 335 314
## 2011 2012
## 376 377
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 68 97 93 74 79 125 143 154 133 133 198 193 201 245 228
## 2011 2012

```

```
## 281 281
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 63 89 83 64 75 116 135 134 119 123 181 174 183 217 202
## 2011 2012
## 255 253
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 27, df = 16, p-value = 0.04
```

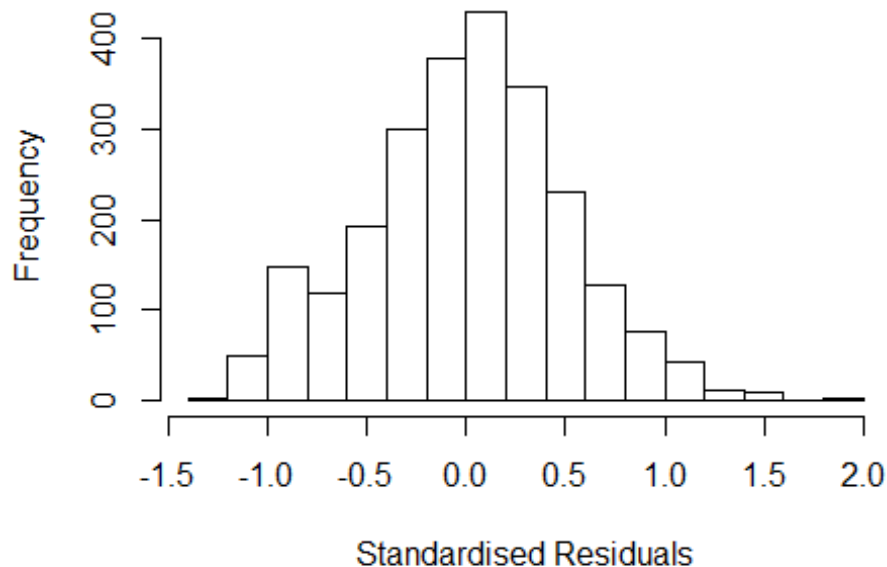


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.3e-05, df = 1, p-value = 1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.53503068559995"
## [1] "Male first author team size 2018 geometric mean: 2.51889088862333"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.54535095602229"
## [1] "Male last author team size 2018 geometric mean: 2.51736748904798"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8700, 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 1.159 1      1.076
## LastAuthorFemale  1.126 1      1.061
## UniqueAuthors     1.189 4      1.022
## Year              1.291 16      1.008
```

Residuals from first and last author and team size



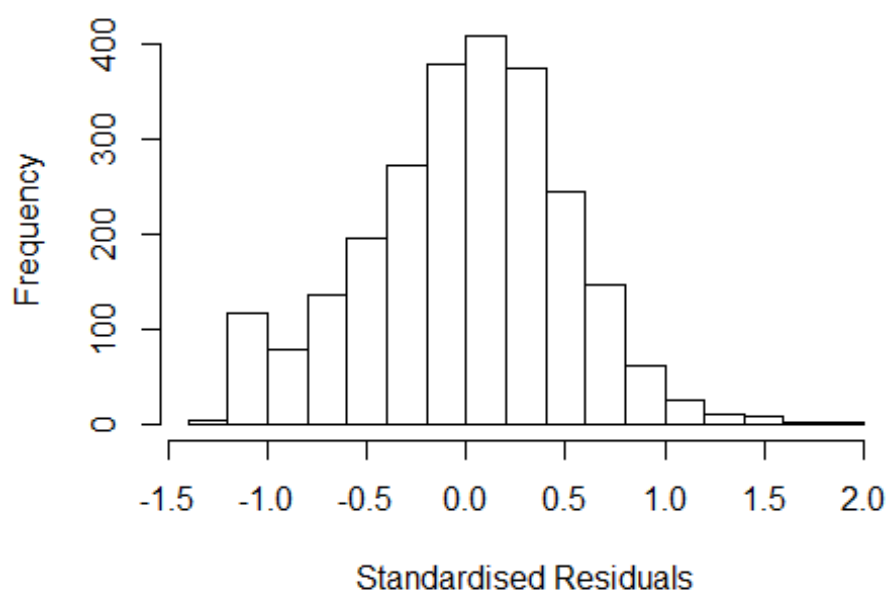
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3674 -0.3196 0.0215 0.3274 1.9026
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15e+00 7.57e-02 15.18 < 2e-16 ***
## FirstAuthorFemale1 -1.02e-02 2.39e-02 -0.43 0.67040
## LastAuthorFemale1 -6.41e-05 2.54e-02 0.00 0.99799
## UniqueAuthors2 1.70e-01 2.92e-02 5.82 6.5e-09 ***
## UniqueAuthors3 2.18e-01 3.05e-02 7.14 1.2e-12 ***
## UniqueAuthors4 1.74e-01 3.51e-02 4.96 7.5e-07 ***
## UniqueAuthors5 2.73e-01 3.43e-02 7.97 2.5e-15 ***
## Year1997 -2.39e-01 1.03e-01 -2.33 0.01977 *
## Year1998 -2.23e-01 8.97e-02 -2.48 0.01310 *
## Year1999 -2.87e-01 9.51e-02 -3.02 0.00259 **
```

```

## Year2000      -1.70e-01  1.02e-01  -1.67  0.09518 .
## Year2001      -2.89e-01  9.18e-02  -3.15  0.00165 **
## Year2002      -2.33e-01  9.51e-02  -2.45  0.01431 *
## Year2003      -2.59e-01  8.52e-02  -3.03  0.00244 **
## Year2004      -3.45e-01  9.06e-02  -3.80  0.00015 ***
## Year2005      -1.87e-01  8.63e-02  -2.17  0.02986 *
## Year2006      -2.89e-01  8.58e-02  -3.37  0.00075 ***
## Year2007      -1.98e-01  8.50e-02  -2.33  0.01963 *
## Year2008      -2.33e-01  8.41e-02  -2.77  0.00564 **
## Year2009      -1.69e-01  8.20e-02  -2.06  0.03928 *
## Year2010      -1.69e-01  8.31e-02  -2.03  0.04238 *
## Year2011      -1.84e-01  8.15e-02  -2.26  0.02385 *
## Year2012      -1.62e-01  8.21e-02  -1.98  0.04801 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.483
## Multiple R-squared:  0.053, Adjusted R-squared:  0.0444
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 211 weights are ~= 1. The remaining 2255 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0853 0.8590 0.9500 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.06e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.162 1 1.078
## LastAuthorFemale 1.127 1 1.062
## Year 1.123 16 1.004

```


Residuals from first and last author



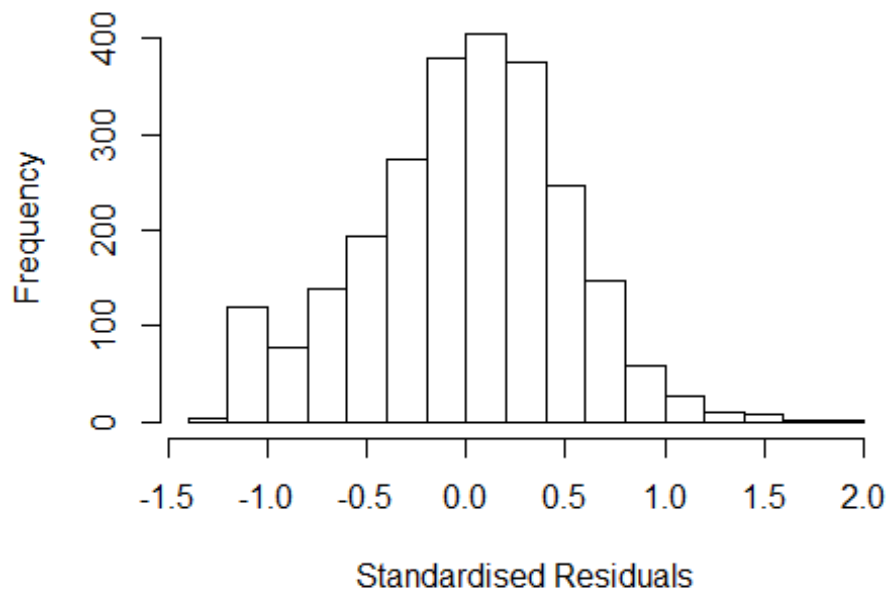
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.230 -0.329 0.026 0.337 1.920
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23044 0.07271 16.92 < 2e-16 ***
## FirstAuthorFemale1 0.00777 0.02442 0.32 0.75042
## LastAuthorFemale1 -0.00899 0.02600 -0.35 0.72947
## Year1997 -0.22224 0.10302 -2.16 0.03108 *
## Year1998 -0.20267 0.08867 -2.29 0.02237 *
## Year1999 -0.26793 0.09597 -2.79 0.00529 **
## Year2000 -0.13299 0.09885 -1.35 0.17863
## Year2001 -0.24270 0.09124 -2.66 0.00787 **
## Year2002 -0.21098 0.09291 -2.27 0.02325 *
## Year2003 -0.21704 0.08428 -2.58 0.01008 *
## Year2004 -0.31386 0.09040 -3.47 0.00053 ***
## Year2005 -0.13311 0.08365 -1.59 0.11166
```

```

## Year2006          -0.23978      0.08399      -2.85      0.00434 **
## Year2007          -0.14203      0.08228      -1.73      0.08446 .
## Year2008          -0.17761      0.08152      -2.18      0.02945 *
## Year2009          -0.10519      0.07924      -1.33      0.18448
## Year2010          -0.10635      0.07992      -1.33      0.18337
## Year2011          -0.12334      0.07921      -1.56      0.11959
## Year2012          -0.09793      0.07960      -1.23      0.21874
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.0173, Adjusted R-squared:  0.01
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 211 weights are ~= 1. The remaining 2255 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0852 0.8570 0.9490 0.8980 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



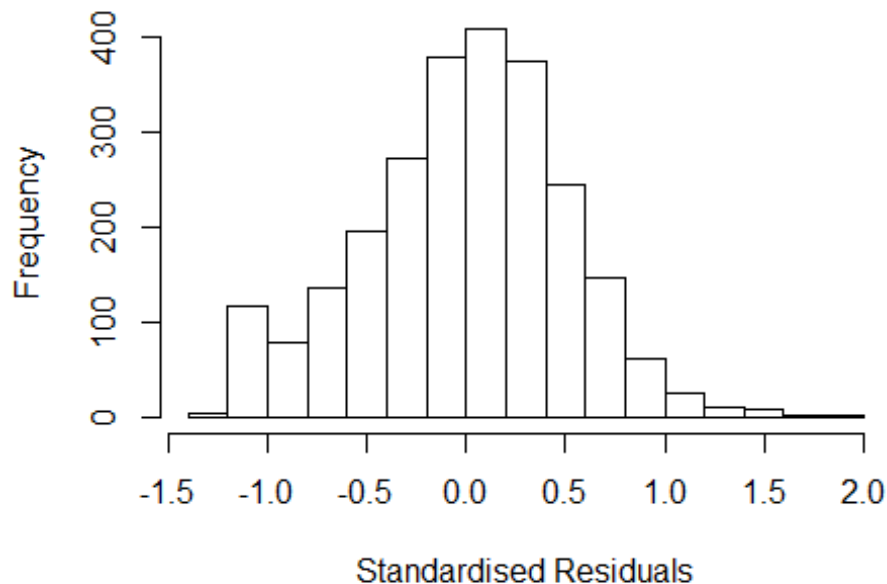
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2338 -0.3313 0.0243 0.3371 1.9128
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22932 0.07259 16.94 < 2e-16 ***
## FirstAuthorFemale1 0.00453 0.02375 0.19 0.84887
## Year1997 -0.22186 0.10303 -2.15 0.03139 *
## Year1998 -0.20303 0.08872 -2.29 0.02221 *
## Year1999 -0.26733 0.09599 -2.78 0.00539 **
## Year2000 -0.13254 0.09884 -1.34 0.18005
## Year2001 -0.24334 0.09123 -2.67 0.00769 **
## Year2002 -0.21038 0.09289 -2.26 0.02362 *
## Year2003 -0.21758 0.08423 -2.58 0.00985 **
## Year2004 -0.31375 0.09044 -3.47 0.00053 ***
## Year2005 -0.13282 0.08366 -1.59 0.11251
## Year2006 -0.24016 0.08399 -2.86 0.00428 **
```

```

## Year2007          -0.14222    0.08230   -1.73  0.08410 .
## Year2008          -0.17771    0.08153   -2.18  0.02937 *
## Year2009          -0.10539    0.07927   -1.33  0.18380
## Year2010          -0.10661    0.07995   -1.33  0.18246
## Year2011          -0.12343    0.07923   -1.56  0.11939
## Year2012          -0.09805    0.07963   -1.23  0.21831
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.0172, Adjusted R-squared:  0.0104
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 211 weights are ~= 1. The remaining 2255 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0885 0.8560 0.9490 0.8980 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



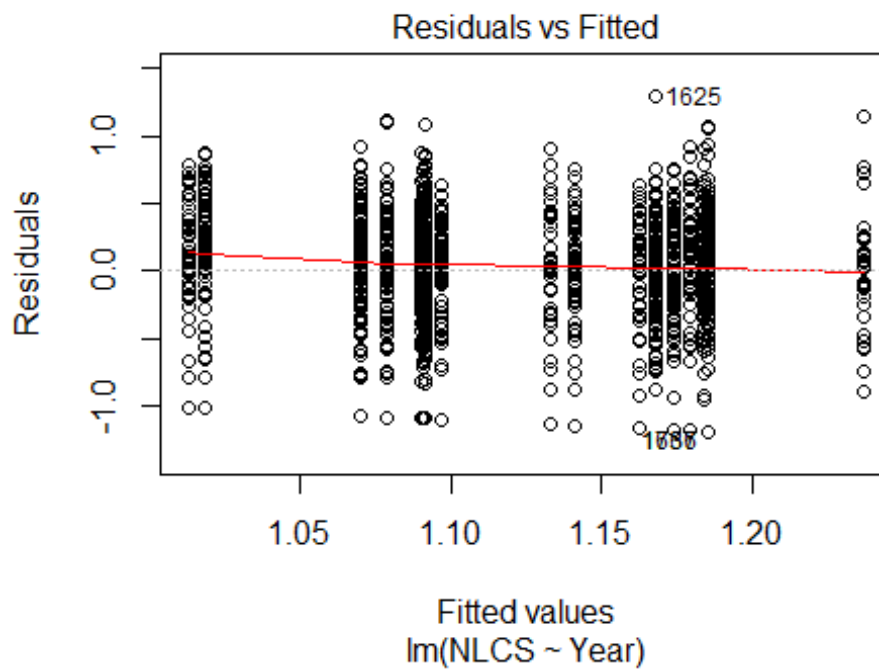
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2312 -0.3299 0.0244 0.3360 1.9157
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23122 0.07269 16.94 < 2e-16 ***
## LastAuthorFemale1 -0.00583 0.02523 -0.23 0.81732
## Year1997 -0.22220 0.10307 -2.16 0.03119 *
## Year1998 -0.20218 0.08866 -2.28 0.02268 *
## Year1999 -0.26802 0.09601 -2.79 0.00529 **
## Year2000 -0.13296 0.09885 -1.35 0.17872
## Year2001 -0.24250 0.09127 -2.66 0.00794 **
## Year2002 -0.21061 0.09296 -2.27 0.02356 *
## Year2003 -0.21692 0.08430 -2.57 0.01014 *
## Year2004 -0.31241 0.09026 -3.46 0.00055 ***
## Year2005 -0.13230 0.08369 -1.58 0.11406
## Year2006 -0.23904 0.08397 -2.85 0.00445 **
```

```

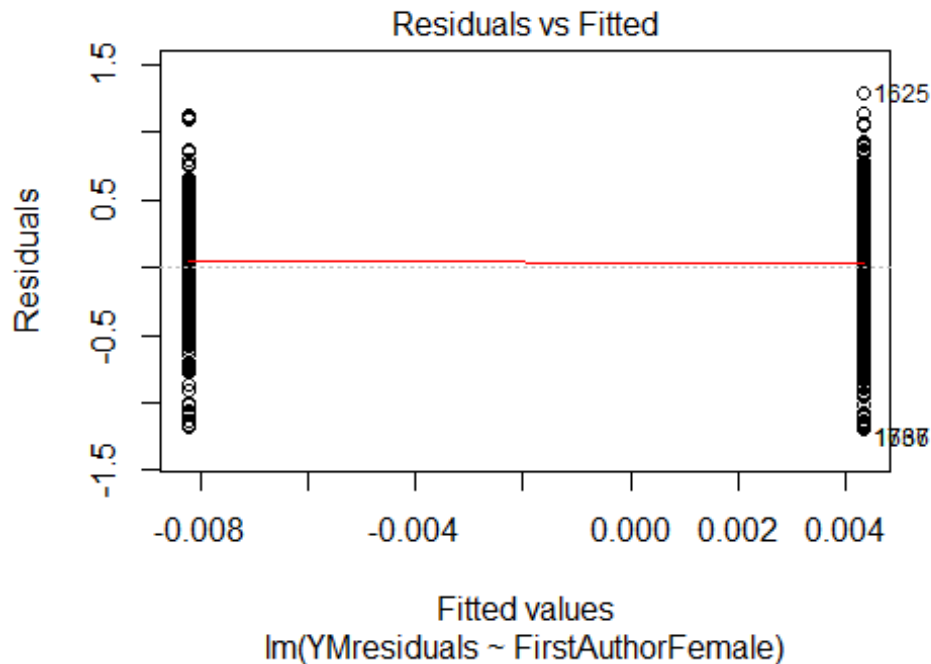
## Year2007          -0.14092      0.08225    -1.71  0.08677 .
## Year2008          -0.17626      0.08131    -2.17  0.03027 *
## Year2009          -0.10358      0.07907    -1.31  0.19034
## Year2010          -0.10582      0.07994    -1.32  0.18573
## Year2011          -0.12183      0.07901    -1.54  0.12322
## Year2012          -0.09678      0.07951    -1.22  0.22361
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.0172, Adjusted R-squared:  0.0104
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 209 weights are ~= 1. The remaining 2257 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0868 0.8560 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      4.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2466"
## [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
##   64   86   95   56  100   97  115  101   88  139  132  144  152  187  167
## 2011 2012
##  174  191
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   43   61   70   36   50   41   91   72   62  104   99  108  121  144  128
## 2011 2012

```

```
## 136 157
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 37 54 59 30 48 34 84 66 51 94 91 98 110 131 115
## 2011 2012
## 128 132
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 54, df = 16, p-value = 5e-06
```

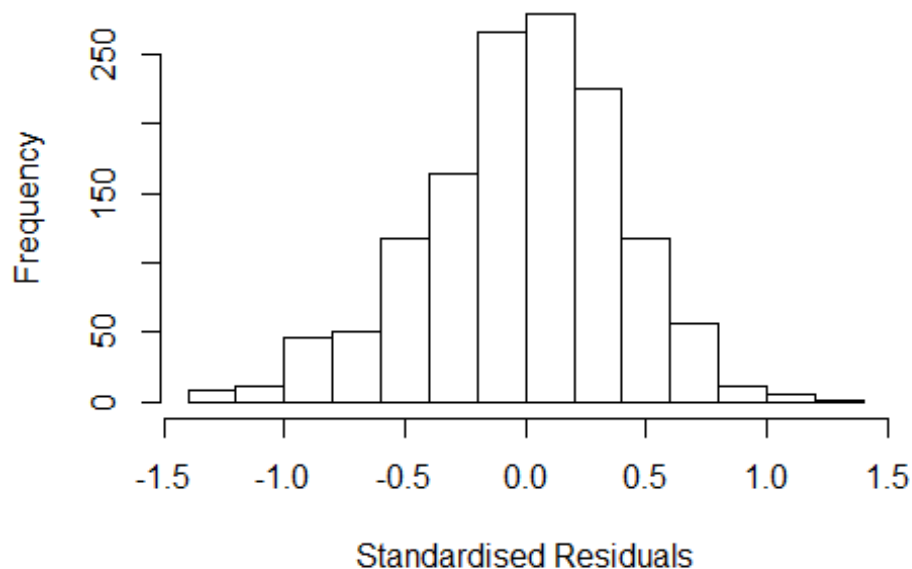


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.034, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 3.03617621083199"
## [1] "Male first author team size 2018 geometric mean: 2.9923759254253"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.06925204249115"
## [1] "Male last author team size 2018 geometric mean: 2.98584632586771"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, 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 1.160 1          1.077
## LastAuthorFemale  1.090 1          1.044
## UniqueAuthors    1.524 4          1.054
## Year              1.558 16         1.014
```


Residuals from first and last author and team size



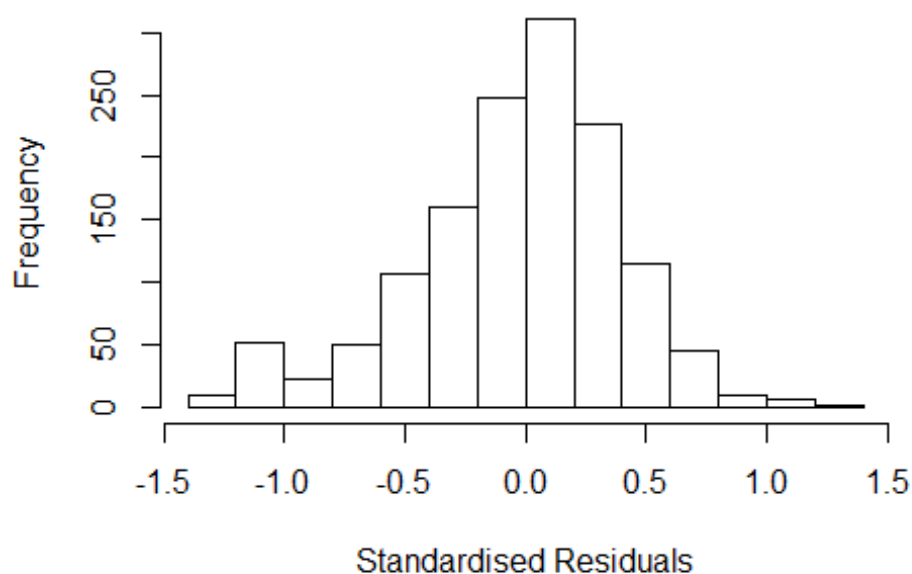
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2793 -0.2614  0.0142  0.2587  1.3284
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.963284   0.094020  10.25 < 2e-16 ***
## FirstAuthorFemale1 -0.017564   0.024788  -0.71  0.479
## LastAuthorFemale1  0.065418   0.029374   2.23  0.026 *
## UniqueAuthors2    0.277482   0.052756   5.26 1.7e-07 ***
## UniqueAuthors3    0.288117   0.053309   5.40 7.7e-08 ***
## UniqueAuthors4    0.311279   0.055970   5.56 3.2e-08 ***
## UniqueAuthors5    0.260307   0.058104   4.48 8.1e-06 ***
## Year1997          0.021705   0.105558   0.21  0.837
## Year1998         -0.014177   0.094444  -0.15  0.881
## Year1999          0.000523   0.105008   0.00  0.996
```

```

## Year2000      -0.172423    0.111107    -1.55    0.121
## Year2001      -0.026927    0.107838    -0.25    0.803
## Year2002      -0.116560    0.110440    -1.06    0.291
## Year2003      -0.047468    0.095810    -0.50    0.620
## Year2004      -0.091778    0.096154    -0.95    0.340
## Year2005      -0.002356    0.093512    -0.03    0.980
## Year2006      -0.103411    0.092594    -1.12    0.264
## Year2007      -0.040451    0.089875    -0.45    0.653
## Year2008      -0.045160    0.088953    -0.51    0.612
## Year2009      -0.036733    0.089326    -0.41    0.681
## Year2010      -0.112356    0.094332    -1.19    0.234
## Year2011      -0.127893    0.091088    -1.40    0.161
## Year2012      -0.158544    0.090014    -1.76    0.078 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.373
## Multiple R-squared:  0.0832, Adjusted R-squared:  0.0681
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1249 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.178  0.851  0.948  0.890  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          7.34e-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 1.096 1          1.047
## LastAuthorFemale  1.118 1          1.057
## Year              1.169 16          1.005

```

Residuals from first and last author



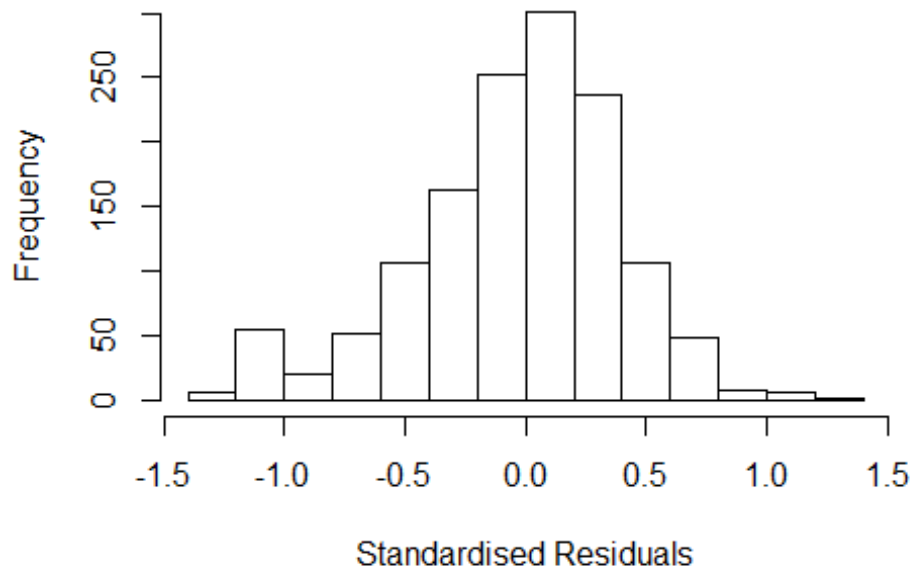
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2210 -0.2642  0.0221  0.2431  1.3017
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17769    0.07979   14.76  <2e-16 ***
## FirstAuthorFemale1  0.00513    0.02385    0.22   0.830
## LastAuthorFemale1  0.06238    0.02892    2.16   0.031 *
## Year1997         0.04146    0.10357    0.40   0.689
## Year1998         0.00360    0.08982    0.04   0.968
## Year1999         0.00904    0.10740    0.08   0.933
## Year2000        -0.11955    0.11445   -1.04   0.296
## Year2001        -0.01903    0.10955   -0.17   0.862
## Year2002        -0.04695    0.10994   -0.43   0.669
## Year2003        -0.04230    0.09099   -0.46   0.642
## Year2004        -0.08731    0.09610   -0.91   0.364
## Year2005         0.01864    0.08829    0.21   0.833
```

```

## Year2006      -0.08823    0.08900   -0.99    0.322
## Year2007      -0.00634    0.08692   -0.07    0.942
## Year2008      -0.01943    0.08586   -0.23    0.821
## Year2009      -0.00324    0.08563   -0.04    0.970
## Year2010      -0.06254    0.09156   -0.68    0.495
## Year2011      -0.09598    0.08822   -1.09    0.277
## Year2012      -0.11289    0.08713   -1.30    0.195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.0185, Adjusted R-squared:  0.00539
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1249 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.195  0.859   0.949   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      7.34e-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 1.072 1      1.035
## Year              1.072 16      1.002

```

Residuals from first author



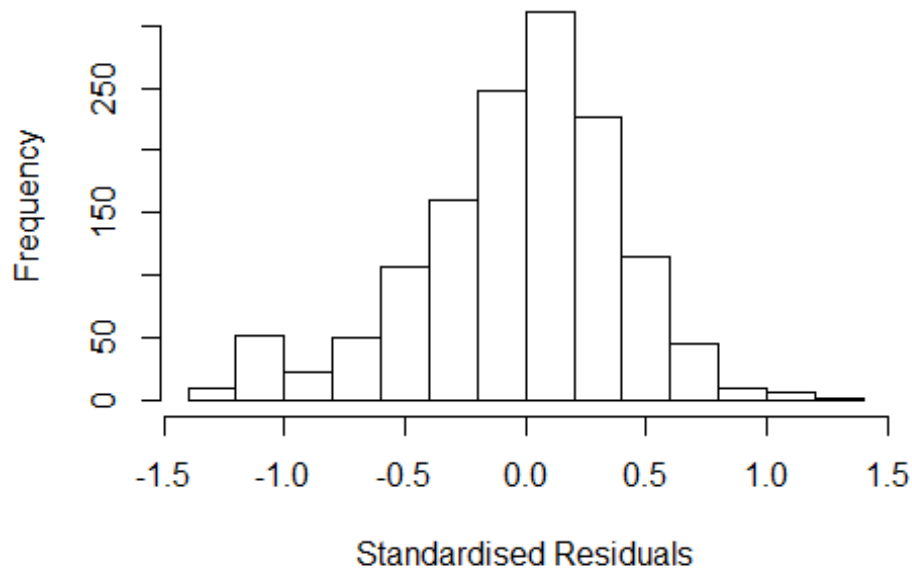
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2276 -0.2605 0.0216 0.2490 1.2906
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18540 0.08033 14.76 <2e-16 ***
## FirstAuthorFemale1 0.01540 0.02378 0.65 0.52
## Year1997 0.04144 0.10400 0.40 0.69
## Year1998 0.00768 0.09037 0.08 0.93
## Year1999 0.00241 0.10850 0.02 0.98
## Year2000 -0.11887 0.11433 -1.04 0.30
## Year2001 -0.02009 0.10822 -0.19 0.85
## Year2002 -0.05275 0.11007 -0.48 0.63
## Year2003 -0.04527 0.09188 -0.49 0.62
## Year2004 -0.08999 0.09715 -0.93 0.35
## Year2005 0.02685 0.08873 0.30 0.76
## Year2006 -0.08338 0.08943 -0.93 0.35
```

```

## Year2007          -0.00275    0.08781   -0.03    0.98
## Year2008          -0.01603    0.08666   -0.18    0.85
## Year2009          -0.00150    0.08614   -0.02    0.99
## Year2010          -0.05518    0.09219   -0.60    0.55
## Year2011          -0.09101    0.08896   -1.02    0.31
## Year2012          -0.11064    0.08763   -1.26    0.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.0144, Adjusted R-squared:  0.00197
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 128 weights are ~= 1. The remaining 1234 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.205  0.854  0.948  0.881  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.34e-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 1.097 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
## -1.2228 -0.2624  0.0232  0.2454  1.3001
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17844    0.07976   14.77  <2e-16 ***
## LastAuthorFemale1 0.06349    0.02867    2.21   0.027 *
## Year1997        0.04177    0.10365    0.40   0.687
## Year1998        0.00390    0.09002    0.04   0.965
## Year1999        0.00941    0.10743    0.09   0.930
## Year2000       -0.11934    0.11443   -1.04   0.297
## Year2001       -0.01913    0.10962   -0.17   0.862
## Year2002       -0.04677    0.11000   -0.43   0.671
## Year2003       -0.04161    0.09106   -0.46   0.648
## Year2004       -0.08628    0.09607   -0.90   0.369
## Year2005        0.01972    0.08836    0.22   0.823
## Year2006       -0.08748    0.08910   -0.98   0.326
```

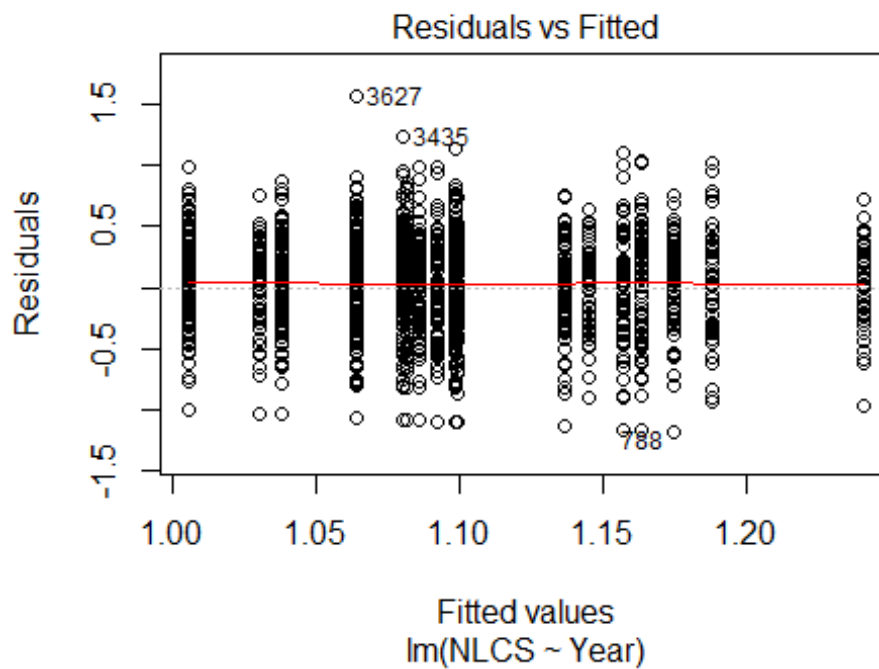
```

## Year2007          -0.00530      0.08700    -0.06      0.951
## Year2008          -0.01857      0.08596    -0.22      0.829
## Year2009          -0.00191      0.08563    -0.02      0.982
## Year2010          -0.06165      0.09175    -0.67      0.502
## Year2011          -0.09496      0.08814    -1.08      0.282
## Year2012          -0.11203      0.08739    -1.28      0.200
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.0185, Adjusted R-squared:  0.00609
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1246 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.196  0.858  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      7.34e-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: 1362"
## [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
## 229 170 135 131 162 178 159 139 148 166 208 210 234 275 251
## 2011 2012
## 272 302
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 81 60 48 49 64 62 69 83 67 88 129 127 134 152 152
## 2011 2012

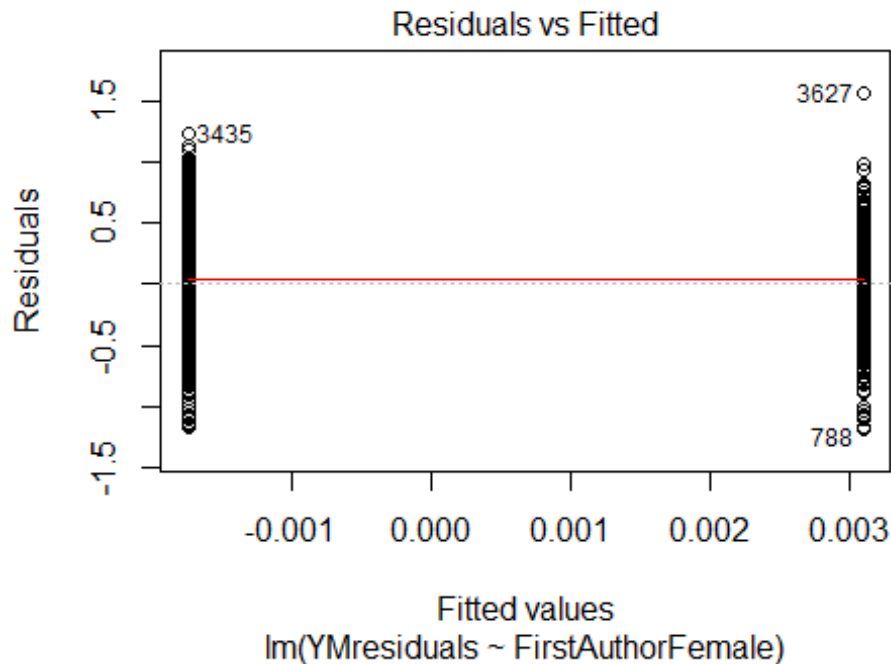
```



```
## 166 183
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 70 50 40 45 54 54 64 67 58 78 101 113 108 127 130
## 2011 2012
## 149 156
## [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
```

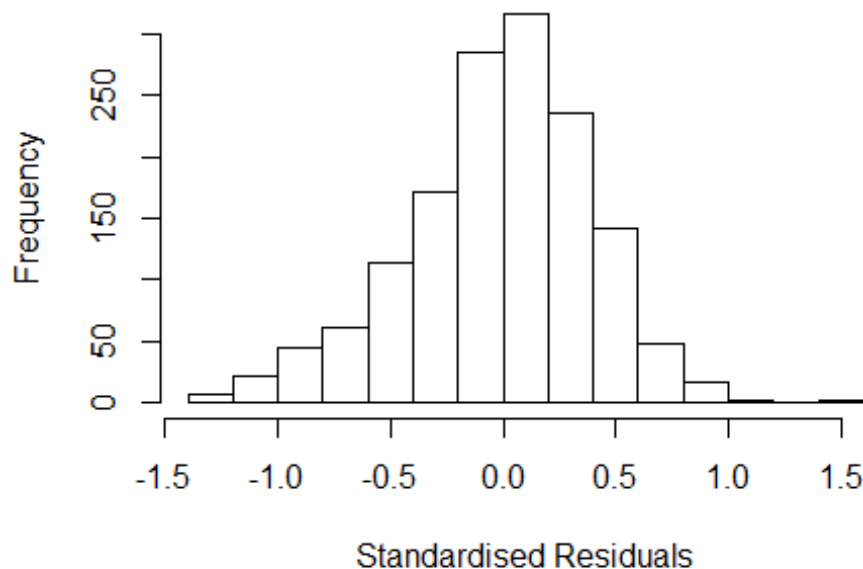


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.7, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 4.00548437469939"
## [1] "Male first author team size 2018 geometric mean: 3.57718315646925"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9300, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.89412823327522"
## [1] "Male last author team size 2018 geometric mean: 3.70643803323857"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6600, 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 1.120 1      1.058
## LastAuthorFemale  1.071 1      1.035
## UniqueAuthors     1.426 4      1.045
## Year               1.574 16     1.014
```

Residuals from first and last author and team size



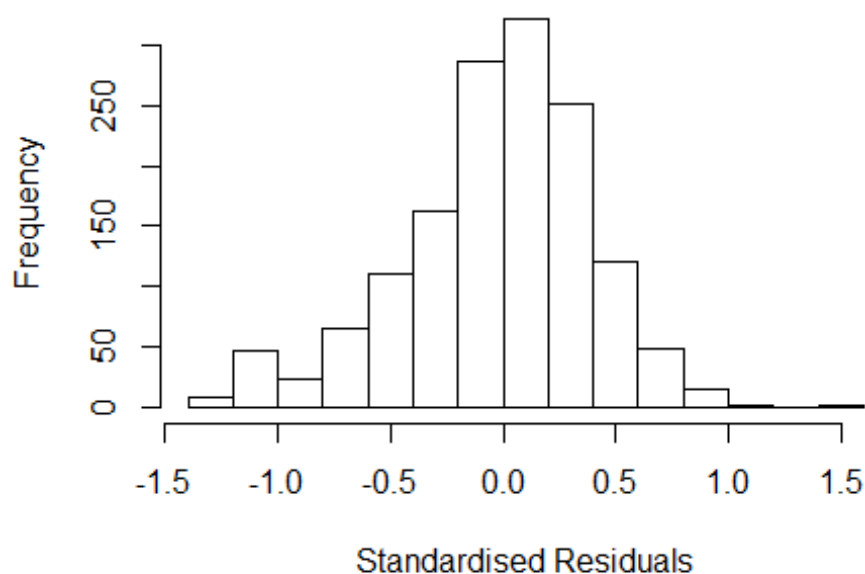
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3447 -0.2551 0.0168 0.2490 1.5774
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0449 0.0817 12.79 < 2e-16 ***
## FirstAuthorFemale1 -0.0140 0.0224 -0.62 0.53233
## LastAuthorFemale1 0.0367 0.0261 1.41 0.16004
## UniqueAuthors2 0.1756 0.0629 2.79 0.00530 **
## UniqueAuthors3 0.2360 0.0636 3.71 0.00021 ***
## UniqueAuthors4 0.2145 0.0645 3.33 0.00090 ***
## UniqueAuthors5 0.2985 0.0646 4.62 4.2e-06 ***
## Year1997 -0.0149 0.1022 -0.15 0.88383
## Year1998 -0.1228 0.0901 -1.36 0.17312
## Year1999 -0.0134 0.0784 -0.17 0.86417
```

```

## Year2000          0.0152      0.0890      0.17  0.86412
## Year2001         -0.1114      0.0796     -1.40  0.16173
## Year2002         -0.1361      0.0781     -1.74  0.08137 .
## Year2003         -0.1564      0.0778     -2.01  0.04460 *
## Year2004         -0.1962      0.0797     -2.46  0.01397 *
## Year2005         -0.1219      0.0729     -1.67  0.09469 .
## Year2006         -0.1496      0.0726     -2.06  0.03943 *
## Year2007         -0.2276      0.0767     -2.97  0.00307 **
## Year2008         -0.2141      0.0765     -2.80  0.00518 **
## Year2009         -0.1527      0.0711     -2.15  0.03187 *
## Year2010         -0.1426      0.0749     -1.90  0.05728 .
## Year2011         -0.1965      0.0730     -2.69  0.00722 **
## Year2012         -0.1714      0.0715     -2.40  0.01668 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.371
## Multiple R-squared:  0.0584, Adjusted R-squared:  0.044
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 130 weights are ~= 1. The remaining 1334 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0314 0.8570 0.9480 0.8900 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.83e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.116 1 1.056
## LastAuthorFemale 1.060 1 1.029
## Year 1.170 16 1.005

```

Residuals from first and last author



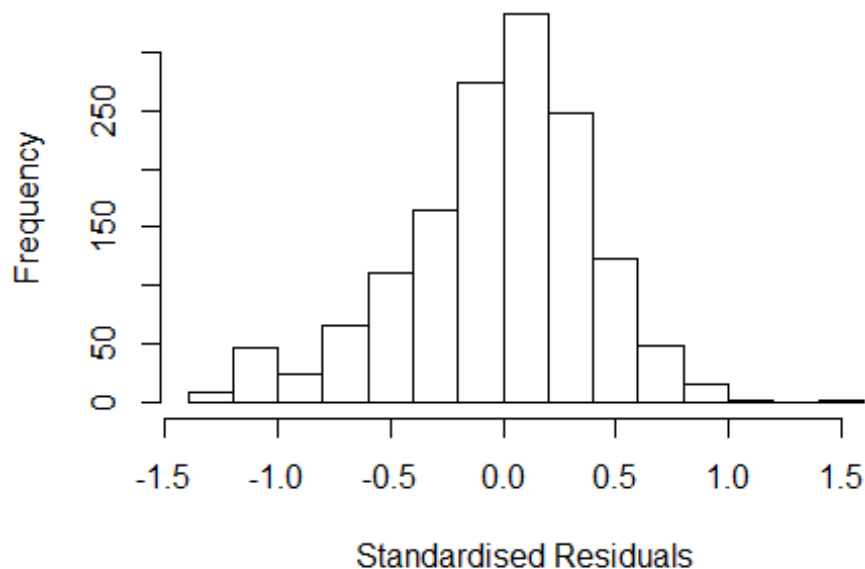
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2445 -0.2501 0.0163 0.2520 1.5277
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21467 0.06628 18.33 <2e-16 ***
## FirstAuthorFemale1 -0.00274 0.02252 -0.12 0.903
## LastAuthorFemale1 0.02442 0.02591 0.94 0.346
## Year1997 -0.02165 0.10017 -0.22 0.829
## Year1998 -0.13191 0.09539 -1.38 0.167
## Year1999 0.01864 0.08354 0.22 0.823
## Year2000 0.02981 0.08701 0.34 0.732
## Year2001 -0.07549 0.08324 -0.91 0.365
## Year2002 -0.09642 0.08040 -1.20 0.231
## Year2003 -0.13186 0.08091 -1.63 0.103
## Year2004 -0.16484 0.08315 -1.98 0.048 *
## Year2005 -0.07629 0.07502 -1.02 0.309
```

```

## Year2006      -0.10931    0.07566   -1.44    0.149
## Year2007      -0.17972    0.07971   -2.25    0.024 *
## Year2008      -0.16049    0.07773   -2.06    0.039 *
## Year2009      -0.09588    0.07373   -1.30    0.194
## Year2010      -0.08395    0.07616   -1.10    0.271
## Year2011      -0.14009    0.07585   -1.85    0.065 .
## Year2012      -0.11734    0.07431   -1.58    0.115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.0191, Adjusted R-squared:  0.00688
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 137 weights are ~= 1. The remaining 1327 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.054  0.853   0.948   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      6.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.108 1      1.053
## Year      1.108 16      1.003

```

Residuals from first author



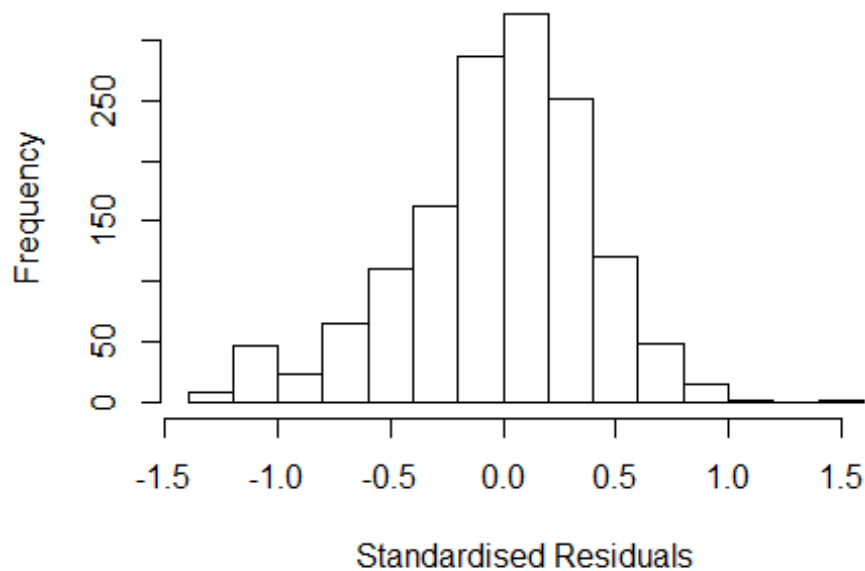
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2506 -0.2491  0.0159  0.2507  1.5442
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.21652    0.06621   18.37  <2e-16 ***
## FirstAuthorFemale1 -0.00108    0.02247   -0.05    0.962
## Year1997         -0.02135    0.10038   -0.21    0.832
## Year1998         -0.12759    0.09503   -1.34    0.180
## Year1999          0.01987    0.08341    0.24    0.812
## Year2000          0.03404    0.08647    0.39    0.694
## Year2001         -0.07294    0.08283   -0.88    0.379
## Year2002         -0.09448    0.08031   -1.18    0.240
## Year2003         -0.12829    0.08045   -1.59    0.111
## Year2004         -0.16177    0.08288   -1.95    0.051 .
## Year2005         -0.07427    0.07484   -0.99    0.321
## Year2006         -0.10694    0.07555   -1.42    0.157
```

```

## Year2007          -0.17841    0.07964   -2.24    0.025 *
## Year2008          -0.15490    0.07704   -2.01    0.045 *
## Year2009          -0.09383    0.07343   -1.28    0.202
## Year2010          -0.08188    0.07606   -1.08    0.282
## Year2011          -0.13564    0.07546   -1.80    0.072 .
## Year2012          -0.11392    0.07409   -1.54    0.124
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.0186, Adjusted R-squared:  0.00701
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1326 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0459 0.8540 0.9480 0.8860 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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

```


Residuals from last author



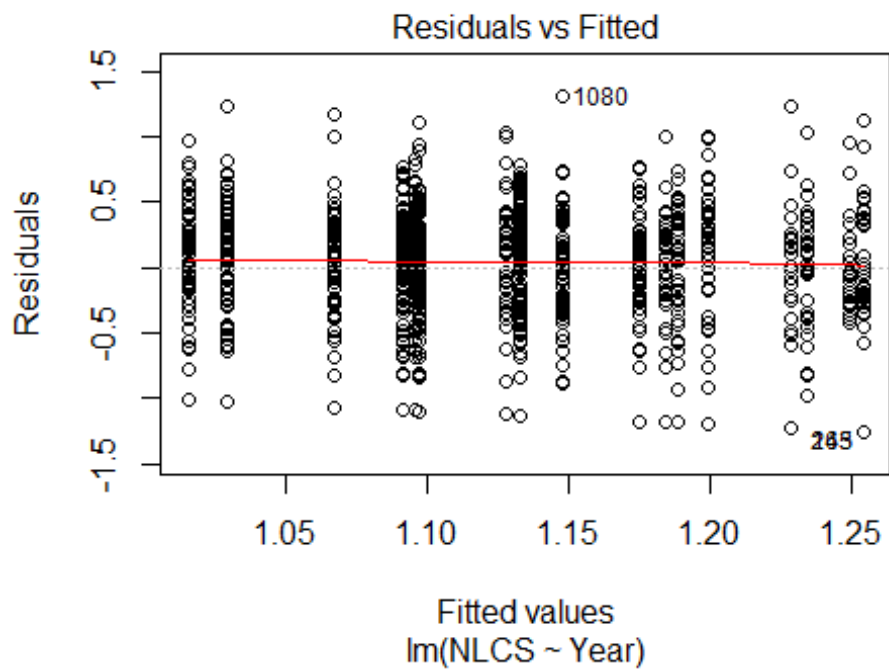
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2441 -0.2506 0.0161 0.2519 1.5263
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2143 0.0662 18.34 <2e-16 ***
## LastAuthorFemale1 0.0242 0.0258 0.94 0.350
## Year1997 -0.0217 0.1002 -0.22 0.829
## Year1998 -0.1324 0.0953 -1.39 0.165
## Year1999 0.0182 0.0834 0.22 0.827
## Year2000 0.0298 0.0870 0.34 0.732
## Year2001 -0.0760 0.0831 -0.92 0.360
## Year2002 -0.0968 0.0803 -1.21 0.228
## Year2003 -0.1321 0.0809 -1.63 0.102
## Year2004 -0.1650 0.0831 -1.98 0.047 *
## Year2005 -0.0768 0.0749 -1.03 0.305
## Year2006 -0.1101 0.0753 -1.46 0.144
```

```

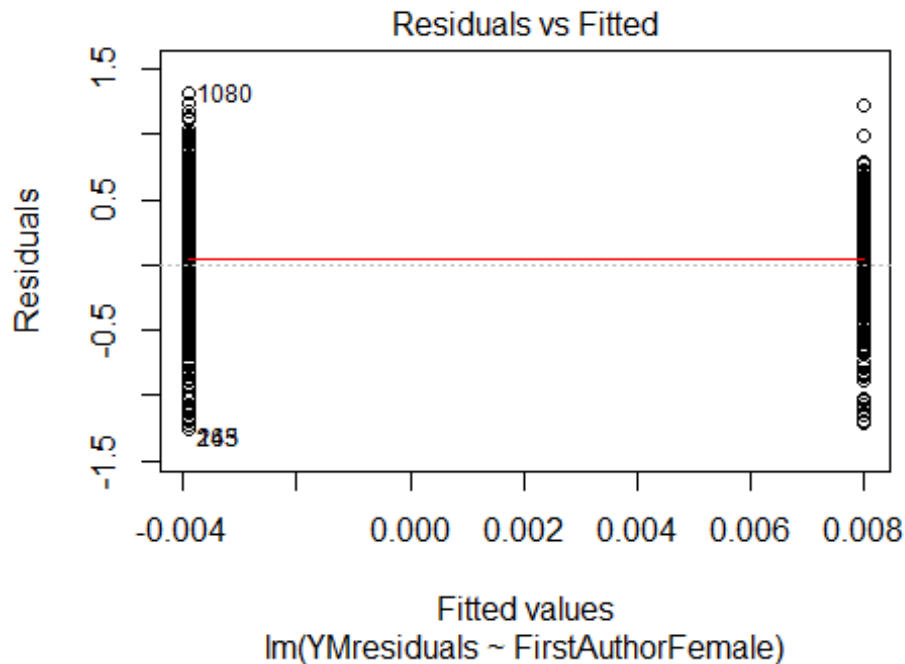
## Year2007          -0.1801      0.0797   -2.26    0.024 *
## Year2008          -0.1612      0.0773   -2.08    0.037 *
## Year2009          -0.0968      0.0731   -1.33    0.185
## Year2010          -0.0848      0.0755   -1.12    0.262
## Year2011          -0.1408      0.0756   -1.86    0.063 .
## Year2012          -0.1179      0.0741   -1.59    0.112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.0191, Adjusted R-squared:  0.00759
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1328 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.054  0.854  0.948  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      6.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1464"
## [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
## 125 100 98 83 108 96 88 106 84 112 138 117 147 214 198
## 2011 2012
## 250 250
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 43 33 25 23 43 32 39 58 36 56 68 59 80 113 109
## 2011 2012

```

```
## 129 136
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 37 30 19 19 41 28 37 41 33 47 52 50 65 98 85
## 2011 2012
## 118 111
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 29, df = 16, p-value = 0.02
```

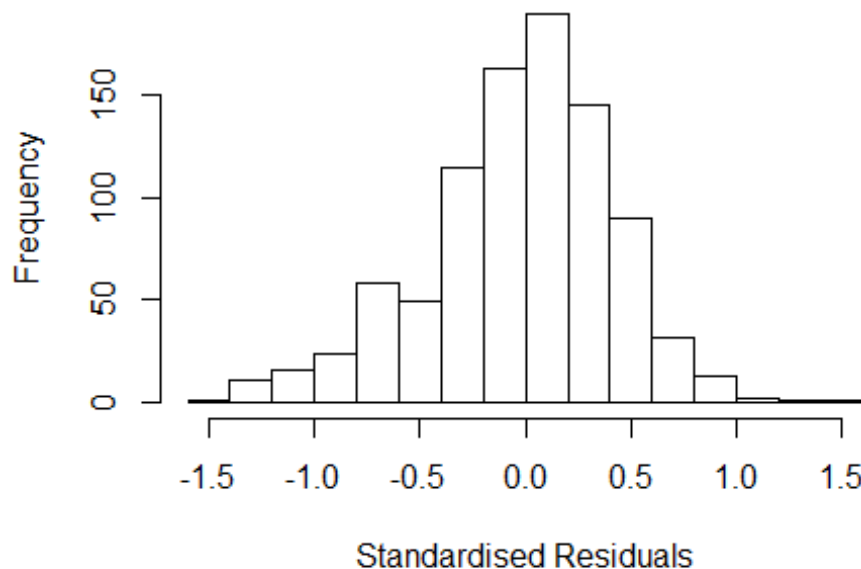


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 4.43952839987132"
## [1] "Male first author team size 2018 geometric mean: 3.47314240988111"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3500, p-value = 0.003
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.16567422765886"
## [1] "Male last author team size 2018 geometric mean: 3.712355967027"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.137 1      1.066
## LastAuthorFemale  1.182 1      1.087
## UniqueAuthors     1.789 4      1.075
## Year              2.120 16     1.024
```

Residuals from first and last author and team size



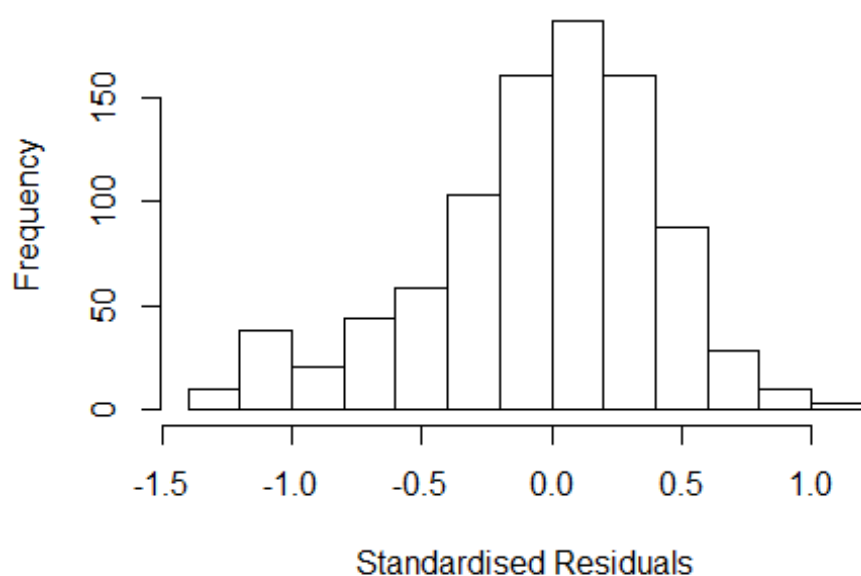
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4506 -0.2716 0.0214 0.2541 1.4238
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01660 0.12377 8.21 7.5e-16 ***
## FirstAuthorFemale1 -0.01590 0.03089 -0.51 0.60686
## LastAuthorFemale1 -0.06614 0.03623 -1.83 0.06824 .
## UniqueAuthors2 0.34340 0.08945 3.84 0.00013 ***
## UniqueAuthors3 0.37462 0.09061 4.13 3.9e-05 ***
## UniqueAuthors4 0.31348 0.09259 3.39 0.00074 ***
## UniqueAuthors5 0.43399 0.09125 4.76 2.3e-06 ***
## Year1997 -0.00529 0.13711 -0.04 0.96922
## Year1998 -0.16948 0.11909 -1.42 0.15505
## Year1999 -0.11875 0.12049 -0.99 0.32466
```

```

## Year2000      -0.08352    0.12942   -0.65   0.51886
## Year2001      -0.06610    0.13374   -0.49   0.62126
## Year2002      -0.09001    0.12570   -0.72   0.47415
## Year2003      -0.16030    0.11457   -1.40   0.16213
## Year2004      -0.21356    0.14141   -1.51   0.13133
## Year2005      -0.23283    0.11719   -1.99   0.04726 *
## Year2006      -0.21678    0.11833   -1.83   0.06728 .
## Year2007      -0.30530    0.12864   -2.37   0.01785 *
## Year2008      -0.27202    0.11978   -2.27   0.02339 *
## Year2009      -0.21265    0.10895   -1.95   0.05128 .
## Year2010      -0.23437    0.11555   -2.03   0.04282 *
## Year2011      -0.25122    0.10897   -2.31   0.02137 *
## Year2012      -0.17057    0.11056   -1.54   0.12326
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.39
## Multiple R-squared:  0.0959, Adjusted R-squared:  0.0735
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 831 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.137  0.843  0.949  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      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.140 1      1.068
## LastAuthorFemale  1.155 1      1.075
## Year              1.313 16      1.009

```

Residuals from first and last author



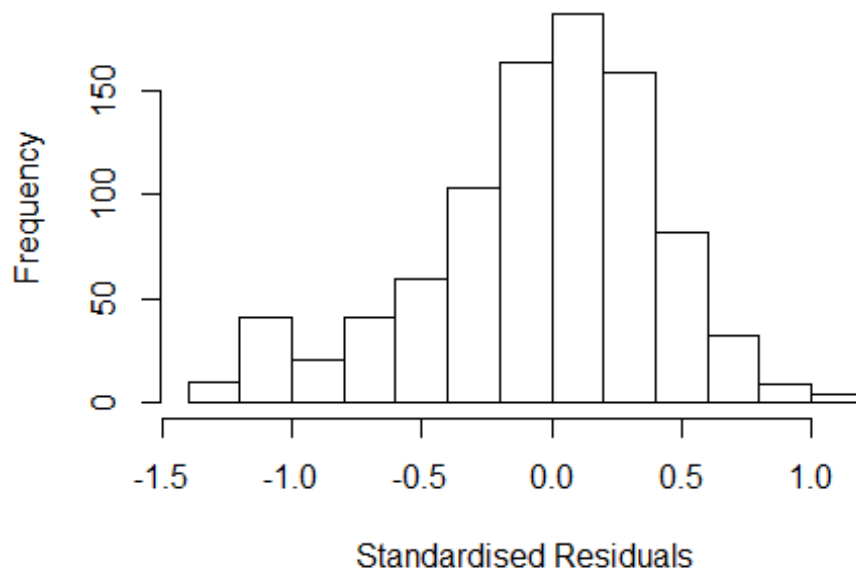
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2899 -0.2730  0.0284  0.2586  1.1035
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.28988    0.11731   11.00  <2e-16 ***
## FirstAuthorFemale1 -0.00750    0.03160   -0.24    0.812
## LastAuthorFemale1 -0.07209    0.03729   -1.93    0.054 .
## Year1997         -0.00151    0.14549   -0.01    0.992
## Year1998         -0.12588    0.12980   -0.97    0.332
## Year1999         -0.05373    0.13629   -0.39    0.693
## Year2000         -0.05784    0.13807   -0.42    0.675
## Year2001         -0.00544    0.14503   -0.04    0.970
## Year2002         -0.03216    0.13123   -0.25    0.806
## Year2003         -0.12241    0.12805   -0.96    0.339
## Year2004         -0.13678    0.15186   -0.90    0.368
## Year2005         -0.17720    0.12905   -1.37    0.170
```

```

## Year2006      -0.15742    0.13085   -1.20    0.229
## Year2007      -0.21316    0.14167   -1.50    0.133
## Year2008      -0.21906    0.13394   -1.64    0.102
## Year2009      -0.12183    0.12081   -1.01    0.313
## Year2010      -0.13650    0.12599   -1.08    0.279
## Year2011      -0.16670    0.12200   -1.37    0.172
## Year2012      -0.09209    0.12353   -0.75    0.456
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.0257, Adjusted R-squared:  0.00604
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 838 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.249  0.848  0.951  0.880  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      1.10e-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 1.149 1      1.072
## Year      1.149 16      1.004

```


Residuals from first author



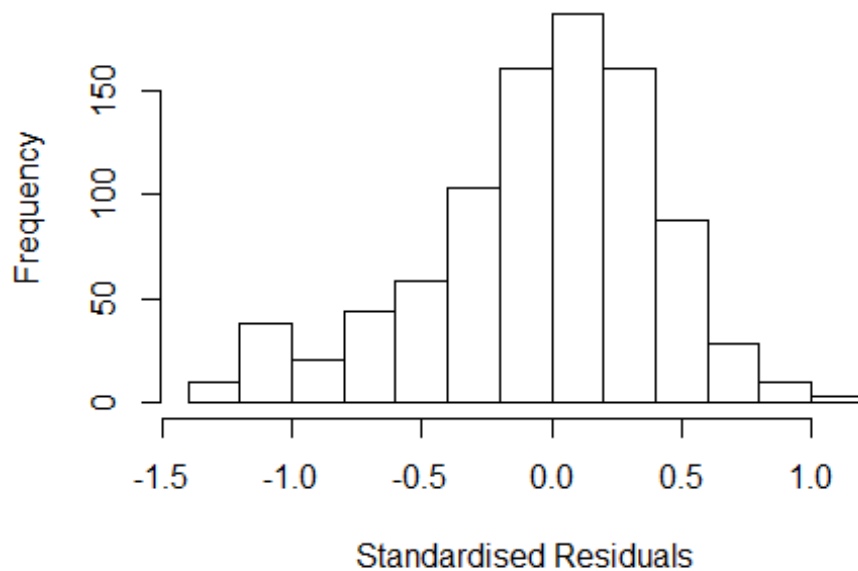
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2838 -0.2661 0.0236 0.2629 1.1075
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27503 0.11811 10.80 <2e-16 ***
## FirstAuthorFemale1 -0.01471 0.03210 -0.46 0.647
## Year1997 0.00881 0.14481 0.06 0.952
## Year1998 -0.13384 0.13197 -1.01 0.311
## Year1999 -0.05406 0.13648 -0.40 0.692
## Year2000 -0.05998 0.13908 -0.43 0.666
## Year2001 -0.01395 0.14795 -0.09 0.925
## Year2002 -0.02962 0.13251 -0.22 0.823
## Year2003 -0.12420 0.12949 -0.96 0.338
## Year2004 -0.13274 0.15290 -0.87 0.386
## Year2005 -0.16444 0.12996 -1.27 0.206
## Year2006 -0.14654 0.13202 -1.11 0.267
```

```

## Year2007          -0.20889    0.14274   -1.46    0.144
## Year2008          -0.22907    0.13466   -1.70    0.089 .
## Year2009          -0.12092    0.12222   -0.99    0.323
## Year2010          -0.12786    0.12691   -1.01    0.314
## Year2011          -0.16315    0.12359   -1.32    0.187
## Year2012          -0.08973    0.12471   -0.72    0.472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.0213, Adjusted R-squared:  0.00262
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 835 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.255  0.847  0.950  0.879  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.158 1          1.076
## Year              1.158 16          1.005

```

Residuals from last author



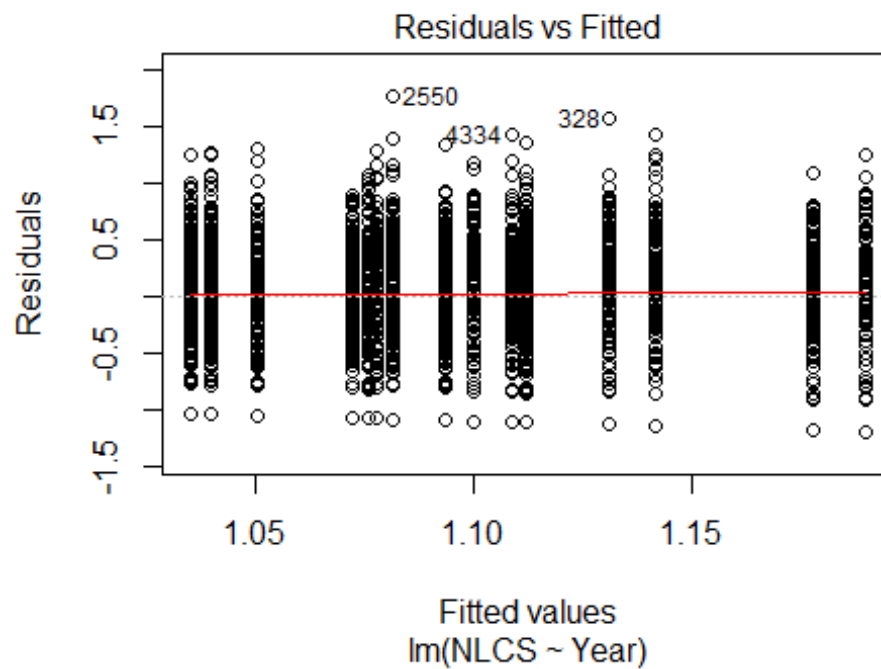
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2884 -0.2715 0.0255 0.2594 1.1070
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.288399 0.117120 11.00 <2e-16 ***
## LastAuthorFemale1 -0.073107 0.037670 -1.94 0.053 .
## Year1997 -0.000963 0.145508 -0.01 0.995
## Year1998 -0.125330 0.129883 -0.96 0.335
## Year1999 -0.053793 0.136265 -0.39 0.693
## Year2000 -0.057877 0.138074 -0.42 0.675
## Year2001 -0.006086 0.145070 -0.04 0.967
## Year2002 -0.032861 0.131362 -0.25 0.803
## Year2003 -0.123064 0.128019 -0.96 0.337
## Year2004 -0.137366 0.152073 -0.90 0.367
## Year2005 -0.177825 0.129088 -1.38 0.169
## Year2006 -0.159434 0.130644 -1.22 0.223
```

```

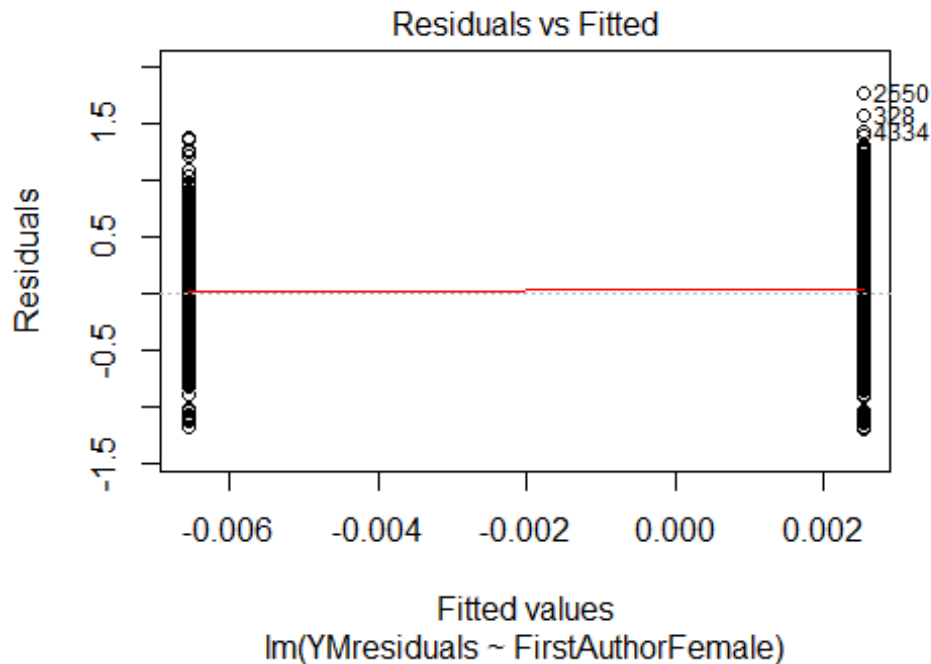
## Year2007          -0.212994    0.141703    -1.50     0.133
## Year2008          -0.219736    0.133877    -1.64     0.101
## Year2009          -0.123496    0.120515    -1.02     0.306
## Year2010          -0.138839    0.125492    -1.11     0.269
## Year2011          -0.167840    0.121901    -1.38     0.169
## Year2012          -0.092463    0.123635    -0.75     0.455
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.0256, Adjusted R-squared:  0.00701
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 836 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.250  0.850  0.951   0.879   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      1.10e-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: 911"
## [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
## 364 301 336 274 307 329 324 277 271 252 328 333 353 427 434
## 2011 2012
## 469 418
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 158 107 143 110 105 118 179 149 146 158 171 178 191 249 254
## 2011 2012

```

```
## 256 236
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 137 97 123 95 90 98 162 123 120 129 134 149 161 209 214
## 2011 2012
## 218 203
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 110, df = 16, p-value <2e-16
```

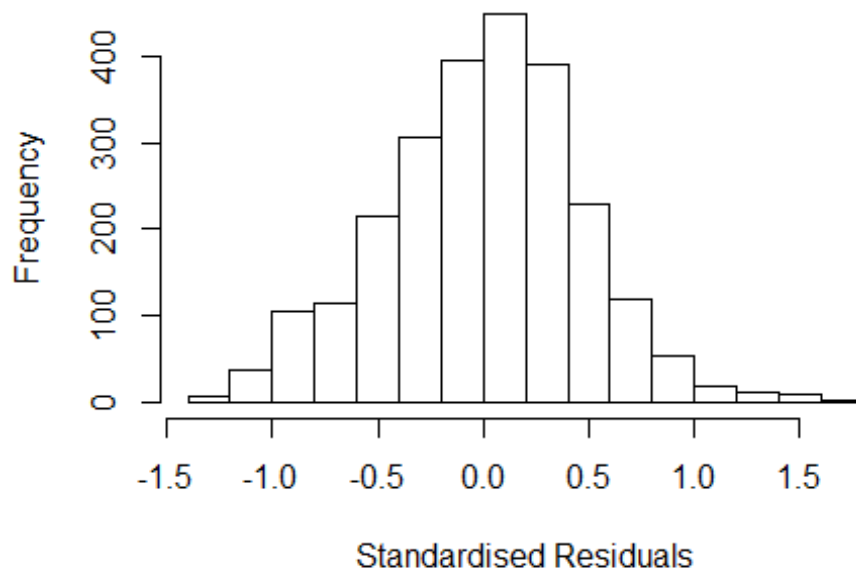


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 17, df = 1, p-value = 5e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 3.39690638723837"
## [1] "Male first author team size 2018 geometric mean: 2.93830753432811"
##
## 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: 3.20261553716955"
## [1] "Male last author team size 2018 geometric mean: 3.07437777081671"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4500, 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 1.045 1      1.022
## LastAuthorFemale  1.054 1      1.027
## UniqueAuthors    1.154 4      1.018
## Year              1.212 16     1.006
```

Residuals from first and last author and team size



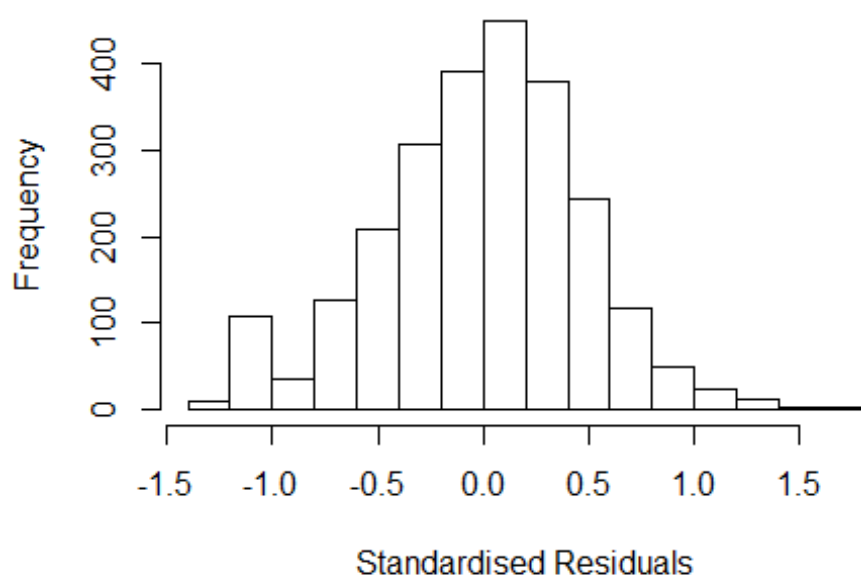
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2668 -0.3063  0.0207  0.2977  1.7137
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91178    0.06508   14.01  < 2e-16 ***
## FirstAuthorFemale1 -0.03325    0.02030   -1.64    0.102
## LastAuthorFemale1 -0.00132    0.02461   -0.05    0.957
## UniqueAuthors2     0.27509    0.03779    7.28 4.5e-13 ***
## UniqueAuthors3     0.28739    0.03806    7.55 6.1e-14 ***
## UniqueAuthors4     0.28327    0.04064    6.97 4.1e-12 ***
## UniqueAuthors5     0.32985    0.03971    8.31 < 2e-16 ***
## Year1997         -0.05934    0.08619   -0.69    0.491
## Year1998          0.00714    0.07622    0.09    0.925
## Year1999          0.07179    0.07978    0.90    0.368
```

```

## Year2000      -0.03310    0.08645   -0.38    0.702
## Year2001      -0.05252    0.07770   -0.68    0.499
## Year2002      -0.05457    0.06893   -0.79    0.429
## Year2003      -0.04737    0.07136   -0.66    0.507
## Year2004      -0.06767    0.07276   -0.93    0.352
## Year2005       0.04325    0.06621    0.65    0.514
## Year2006      -0.03629    0.06780   -0.54    0.593
## Year2007      -0.05345    0.06753   -0.79    0.429
## Year2008      -0.10384    0.06558   -1.58    0.113
## Year2009      -0.08367    0.06211   -1.35    0.178
## Year2010      -0.11247    0.06353   -1.77    0.077 .
## Year2011      -0.14026    0.06408   -2.19    0.029 *
## Year2012      -0.07420    0.06447   -1.15    0.250
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.448
## Multiple R-squared:  0.057, Adjusted R-squared:  0.0485
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 204 weights are ~= 1. The remaining 2258 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110  0.866  0.952  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      4.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "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  1.042 1      1.021
## Year              1.062 16      1.002

```


Residuals from first and last author



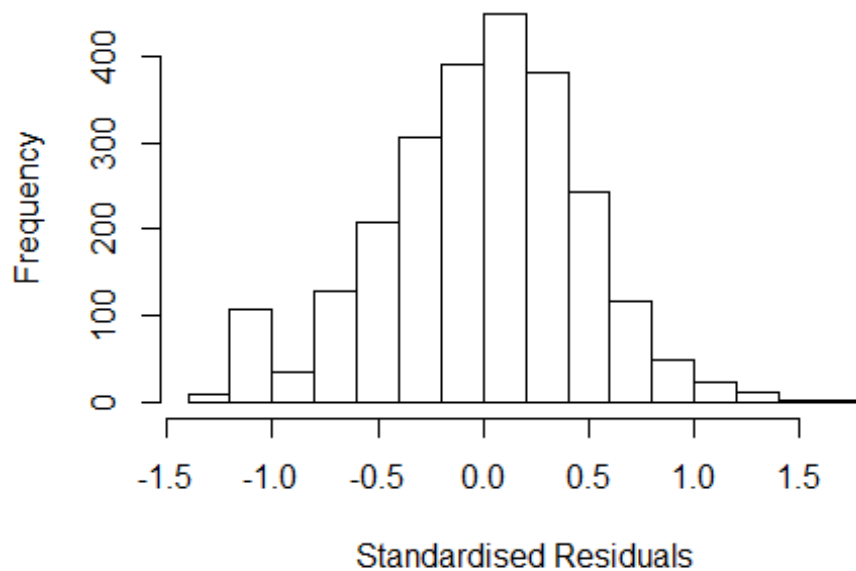
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2327 -0.3075 0.0224 0.3056 1.7625
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1459 0.0597 19.18 <2e-16 ***
## FirstAuthorFemale1 -0.0108 0.0204 -0.53 0.597
## LastAuthorFemale1 -0.0066 0.0249 -0.27 0.791
## Year1997 -0.0940 0.0883 -1.06 0.287
## Year1998 -0.0180 0.0790 -0.23 0.820
## Year1999 0.0868 0.0820 1.06 0.290
## Year2000 -0.0536 0.0894 -0.60 0.549
## Year2001 -0.0596 0.0831 -0.72 0.473
## Year2002 -0.0624 0.0727 -0.86 0.391
## Year2003 -0.0519 0.0731 -0.71 0.478
## Year2004 -0.0413 0.0765 -0.54 0.589
## Year2005 0.0304 0.0687 0.44 0.659
```

```

## Year2006          -0.0146      0.0709   -0.21    0.837
## Year2007          -0.0275      0.0701   -0.39    0.695
## Year2008          -0.0980      0.0683   -1.43    0.152
## Year2009          -0.0580      0.0647   -0.90    0.370
## Year2010          -0.0841      0.0661   -1.27    0.203
## Year2011          -0.1187      0.0669   -1.77    0.076 .
## Year2012          -0.0409      0.0676   -0.60    0.546
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.452
## Multiple R-squared:  0.0099, Adjusted R-squared:  0.00261
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 205 weights are ~= 1. The remaining 2257 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0944 0.8670 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      4.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1      1.012
## Year              1.023 16      1.001

```

Residuals from first author



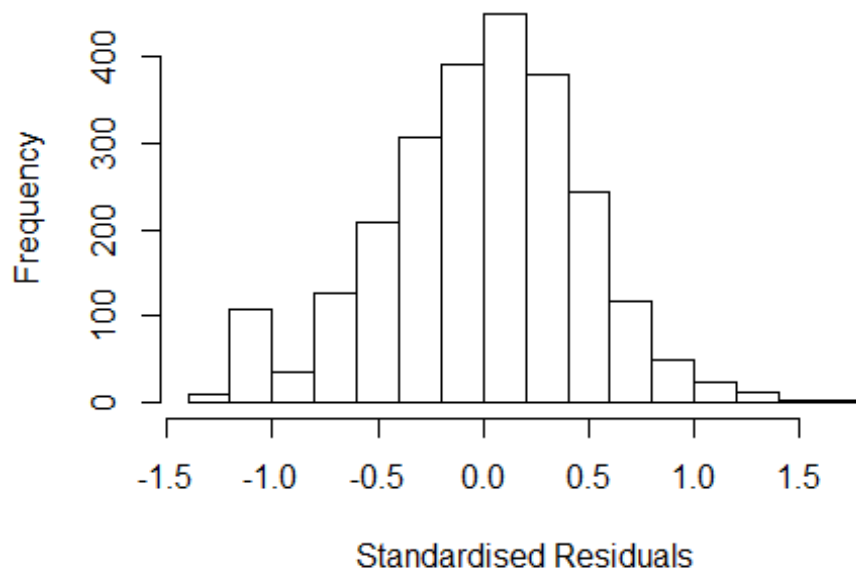
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2320 -0.3081 0.0226 0.3035 1.7632
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1458 0.0597 19.18 <2e-16 ***
## FirstAuthorFemale1 -0.0115 0.0205 -0.56 0.573
## Year1997 -0.0944 0.0881 -1.07 0.284
## Year1998 -0.0186 0.0789 -0.24 0.814
## Year1999 0.0863 0.0819 1.05 0.292
## Year2000 -0.0544 0.0891 -0.61 0.542
## Year2001 -0.0606 0.0830 -0.73 0.465
## Year2002 -0.0630 0.0726 -0.87 0.386
## Year2003 -0.0524 0.0730 -0.72 0.473
## Year2004 -0.0421 0.0762 -0.55 0.581
## Year2005 0.0294 0.0685 0.43 0.668
## Year2006 -0.0156 0.0706 -0.22 0.825
```

```

## Year2007          -0.0276      0.0701   -0.39    0.693
## Year2008          -0.0991      0.0680   -1.46    0.145
## Year2009          -0.0588      0.0645   -0.91    0.362
## Year2010          -0.0849      0.0659   -1.29    0.198
## Year2011          -0.1196      0.0668   -1.79    0.073 .
## Year2012          -0.0419      0.0674   -0.62    0.534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.452
## Multiple R-squared:  0.00986,    Adjusted R-squared:  0.00297
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 208 weights are ~= 1. The remaining 2254 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0945 0.8680 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      4.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



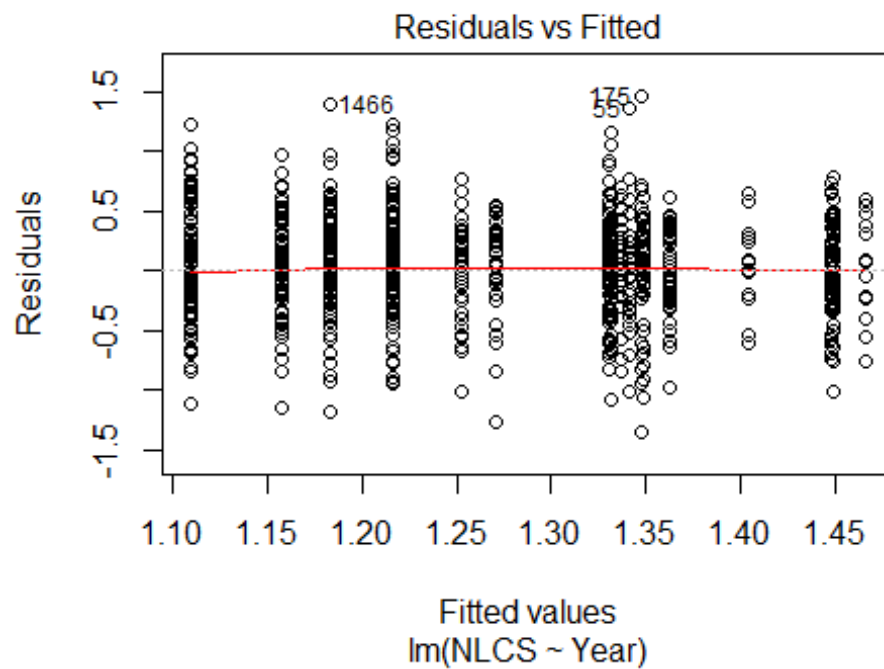
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2303 -0.3098 0.0234 0.3030 1.7650
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14400 0.05944 19.25 <2e-16 ***
## LastAuthorFemale1 -0.00848 0.02501 -0.34 0.735
## Year1997 -0.09316 0.08826 -1.06 0.291
## Year1998 -0.01825 0.07899 -0.23 0.817
## Year1999 0.08627 0.08197 1.05 0.293
## Year2000 -0.05379 0.08936 -0.60 0.547
## Year2001 -0.05919 0.08310 -0.71 0.476
## Year2002 -0.06301 0.07273 -0.87 0.386
## Year2003 -0.05284 0.07314 -0.72 0.470
## Year2004 -0.04125 0.07644 -0.54 0.590
## Year2005 0.02931 0.06868 0.43 0.670
## Year2006 -0.01536 0.07088 -0.22 0.828
```

```

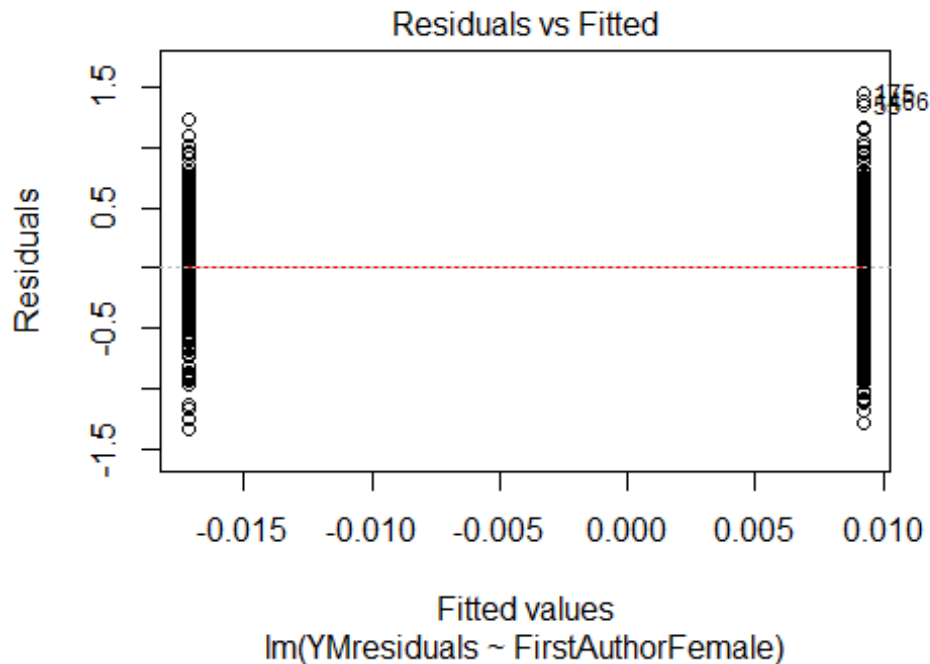
## Year2007          -0.02824      0.07014    -0.40      0.687
## Year2008          -0.09962      0.06841    -1.46      0.145
## Year2009          -0.05956      0.06468    -0.92      0.357
## Year2010          -0.08612      0.06602    -1.30      0.192
## Year2011          -0.12034      0.06691    -1.80      0.072 .
## Year2012          -0.04202      0.06766    -0.62      0.535
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.452
## Multiple R-squared:  0.00979,    Adjusted R-squared:  0.0029
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 205 weights are ~= 1. The remaining 2257 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0927 0.8660 0.9510 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      4.06e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2462"
## [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
##   59   37   53   57   56   47   52   53   60   55   62   72   86  122  142
## 2011 2012
##  136  145
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   23   14   23   35   39   16   30   33   45   42   40   48   57   92  100
## 2011 2012

```

```
## 98 104
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 20 14 18 31 31 13 24 27 42 40 36 43 39 80 88
## 2011 2012
## 86 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 = 22, df = 16, p-value = 0.1
```

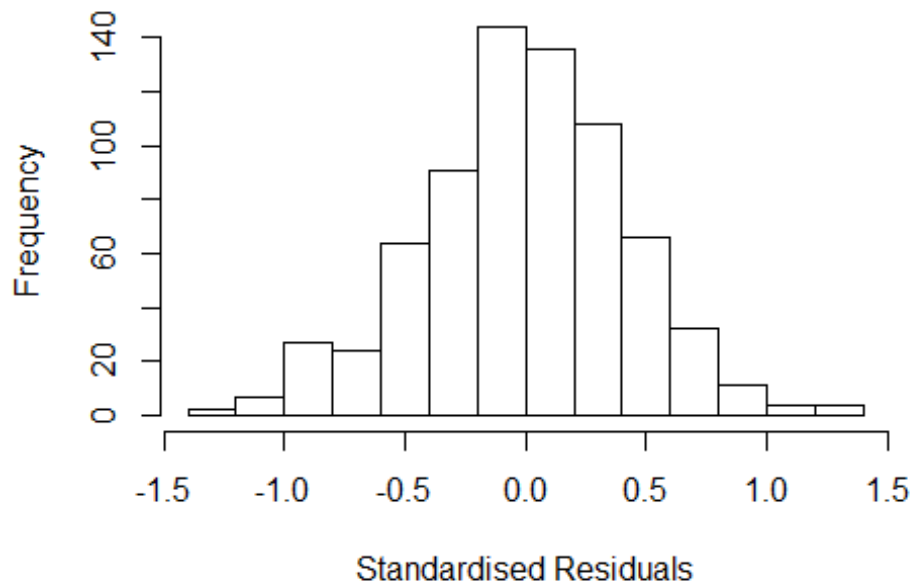


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.21, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 4.64806609396226"
## [1] "Male first author team size 2018 geometric mean: 3.57021013993251"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.06110252795315"
## [1] "Male last author team size 2018 geometric mean: 3.84577615862747"
##
## 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 1.095 1          1.046
## LastAuthorFemale  1.119 1          1.058
## UniqueAuthors    1.544 4          1.056
## Year              1.671 16         1.016
```


Residuals from first and last author and team size



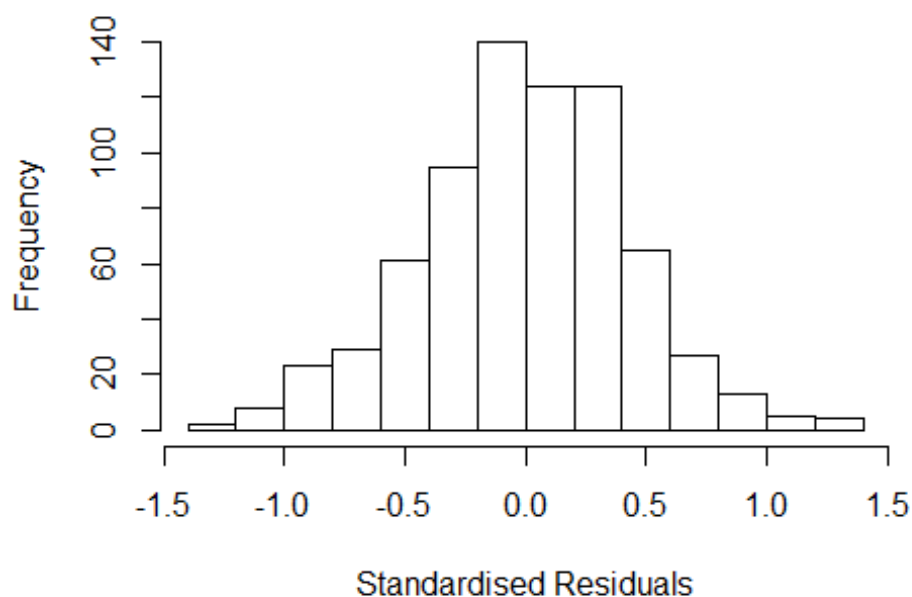
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.313405 -0.265030  0.000725  0.284371  1.347112
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.21945    0.11750   10.38  <2e-16 ***
## FirstAuthorFemale1 -0.03261    0.03478   -0.94   0.349
## LastAuthorFemale1  0.03050    0.03981    0.77   0.444
## UniqueAuthors2     0.09158    0.06079    1.51   0.132
## UniqueAuthors3    -0.00543    0.06469   -0.08   0.933
## UniqueAuthors4     0.11063    0.06861    1.61   0.107
## UniqueAuthors5     0.09872    0.05737    1.72   0.086 .
## Year1997           0.17142    0.16139    1.06   0.289
## Year1998           0.21419    0.13289    1.61   0.107
## Year1999           0.02785    0.13073    0.21   0.831
```

```

## Year2000      0.12612    0.12973    0.97    0.331
## Year2001      0.07621    0.15163    0.50    0.615
## Year2002      0.09337    0.13288    0.70    0.482
## Year2003     -0.03532    0.13190   -0.27    0.789
## Year2004      0.18923    0.12148    1.56    0.120
## Year2005      0.15554    0.12840    1.21    0.226
## Year2006      0.12071    0.12028    1.00    0.316
## Year2007     -0.02727    0.12132   -0.22    0.822
## Year2008      0.11660    0.13000    0.90    0.370
## Year2009     -0.08075    0.11993   -0.67    0.501
## Year2010     -0.17503    0.12152   -1.44    0.150
## Year2011     -0.09129    0.11647   -0.78    0.433
## Year2012     -0.05039    0.11939   -0.42    0.673
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.0815, Adjusted R-squared:  0.0525
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 662 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.261  0.870  0.952  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      1.39e-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 1.055 1      1.027
## LastAuthorFemale  1.101 1      1.049
## Year              1.159 16      1.005

```

Residuals from first and last author



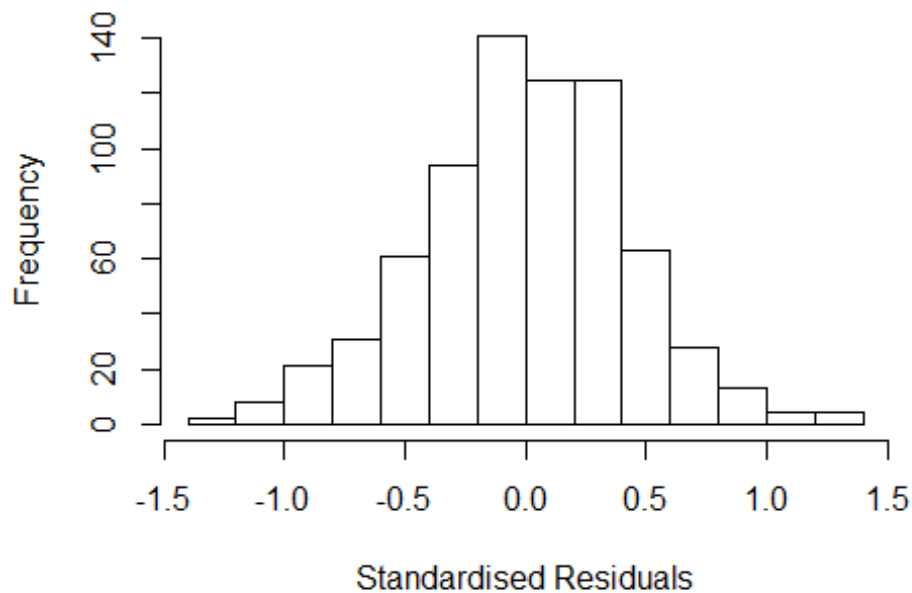
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30988 -0.26752 0.00377 0.28130 1.37859
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2850 0.1049 12.25 <2e-16 ***
## FirstAuthorFemale1 -0.0276 0.0343 -0.81 0.42
## LastAuthorFemale1 0.0247 0.0391 0.63 0.53
## Year1997 0.1879 0.1626 1.16 0.25
## Year1998 0.1948 0.1319 1.48 0.14
## Year1999 0.0249 0.1293 0.19 0.85
## Year2000 0.1339 0.1271 1.05 0.29
## Year2001 0.0482 0.1468 0.33 0.74
## Year2002 0.0877 0.1301 0.67 0.50
## Year2003 -0.0211 0.1315 -0.16 0.87
## Year2004 0.1869 0.1188 1.57 0.12
## Year2005 0.1578 0.1246 1.27 0.21
```

```

## Year2006          0.1011      0.1180      0.86      0.39
## Year2007         -0.0199      0.1219     -0.16      0.87
## Year2008          0.1135      0.1282      0.88      0.38
## Year2009         -0.0862      0.1183     -0.73      0.47
## Year2010         -0.1695      0.1198     -1.41      0.16
## Year2011         -0.0896      0.1150     -0.78      0.44
## Year2012         -0.0492      0.1174     -0.42      0.68
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0704, Adjusted R-squared:  0.0465
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 668 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.235  0.869  0.954  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      1.39e-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 1.055 1      1.027
## Year              1.055 16      1.002

```

Residuals from first author



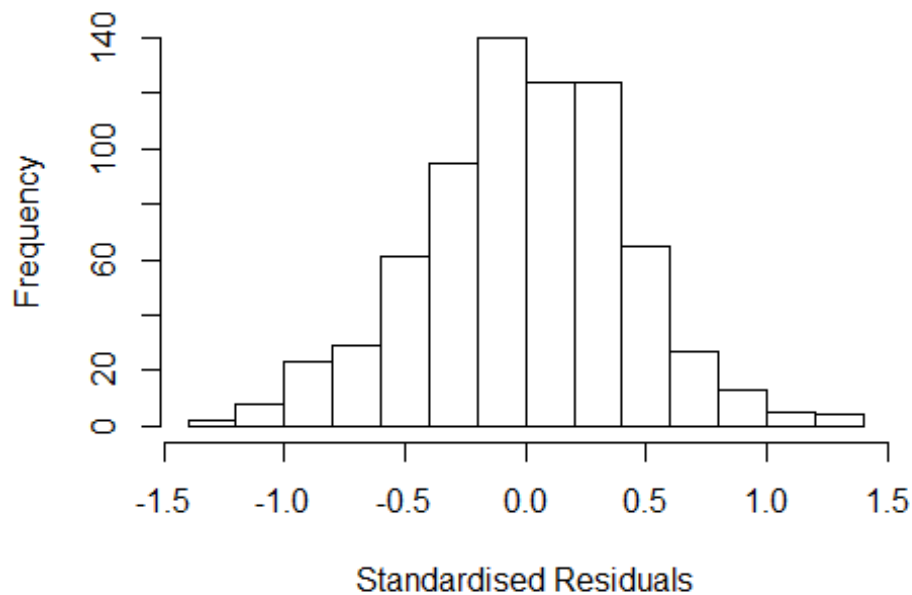
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.31372 -0.26070  0.00357  0.28103  1.37351
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2848    0.1049   12.25  <2e-16 ***
## FirstAuthorFemale1 -0.0260    0.0344   -0.76    0.45
## Year1997          0.1928    0.1607    1.20    0.23
## Year1998          0.1993    0.1312    1.52    0.13
## Year1999          0.0289    0.1289    0.22    0.82
## Year2000          0.1408    0.1265    1.11    0.27
## Year2001          0.0483    0.1468    0.33    0.74
## Year2002          0.0901    0.1303    0.69    0.49
## Year2003         -0.0158    0.1314   -0.12    0.90
## Year2004          0.1921    0.1186    1.62    0.11
## Year2005          0.1656    0.1232    1.34    0.18
## Year2006          0.1063    0.1176    0.90    0.37
```

```

## Year2007          -0.0152      0.1217   -0.12      0.90
## Year2008           0.1167      0.1282    0.91      0.36
## Year2009          -0.0827      0.1181   -0.70      0.48
## Year2010          -0.1626      0.1190   -1.37      0.17
## Year2011          -0.0843      0.1145   -0.74      0.46
## Year2012          -0.0426      0.1169   -0.36      0.72
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0698, Adjusted R-squared:  0.0473
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 665 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.239  0.869  0.954  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      1.39e-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 1.096 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
## -1.305318 -0.263076 -0.000547 0.276989 1.389893
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2816 0.1043 12.29 <2e-16 ***
## LastAuthorFemale1 0.0223 0.0392 0.57 0.57
## Year1997 0.1867 0.1625 1.15 0.25
## Year1998 0.1931 0.1317 1.47 0.14
## Year1999 0.0237 0.1287 0.18 0.85
## Year2000 0.1291 0.1272 1.01 0.31
## Year2001 0.0491 0.1468 0.33 0.74
## Year2002 0.0851 0.1297 0.66 0.51
## Year2003 -0.0292 0.1311 -0.22 0.82
## Year2004 0.1809 0.1184 1.53 0.13
## Year2005 0.1536 0.1242 1.24 0.22
## Year2006 0.0928 0.1172 0.79 0.43
```

```

## Year2007          -0.0252      0.1219    -0.21      0.84
## Year2008           0.1040      0.1271      0.82      0.41
## Year2009          -0.0917      0.1181    -0.78      0.44
## Year2010          -0.1756      0.1194    -1.47      0.14
## Year2011          -0.0975      0.1143    -0.85      0.39
## Year2012          -0.0568      0.1169    -0.49      0.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.0692, Adjusted R-squared:  0.0467
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 59 weights are ~= 1. The remaining 661 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.230  0.865   0.953   0.899   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.39e-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: 720"
## [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]"
##
## 2009 2010 2011 2012
##    2    3    4    2
##
## 2009 2010 2011 2012
##    2    1    3    2
##
## 2009 2010 2011 2012
##    2    1    2    2

```



```

## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 5.31329284591306"
## [1] "Male first author team size 2018 geometric mean: 5.94392195276313"
##
## Wilcoxon rank sum test
##
## 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: 3.80675409583932"
## [1] "Male last author team size 2018 geometric mean: 12.2474487139159"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0, 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: 7"
## [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
## 187 167 154 154 147 160 161 112 116 130 169 160 152 151 152
## 2011 2012
## 154 159
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 82 74 51 73 47 42 86 64 71 69 101 98 96 93 92
## 2011 2012
## 95 110
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 74 64 46 63 42 37 67 49 62 57 89 88 81 82 83
## 2011 2012
## 79 93
## [1] "Heteroscedasticity checks, confirming that there are problems with

```

these"

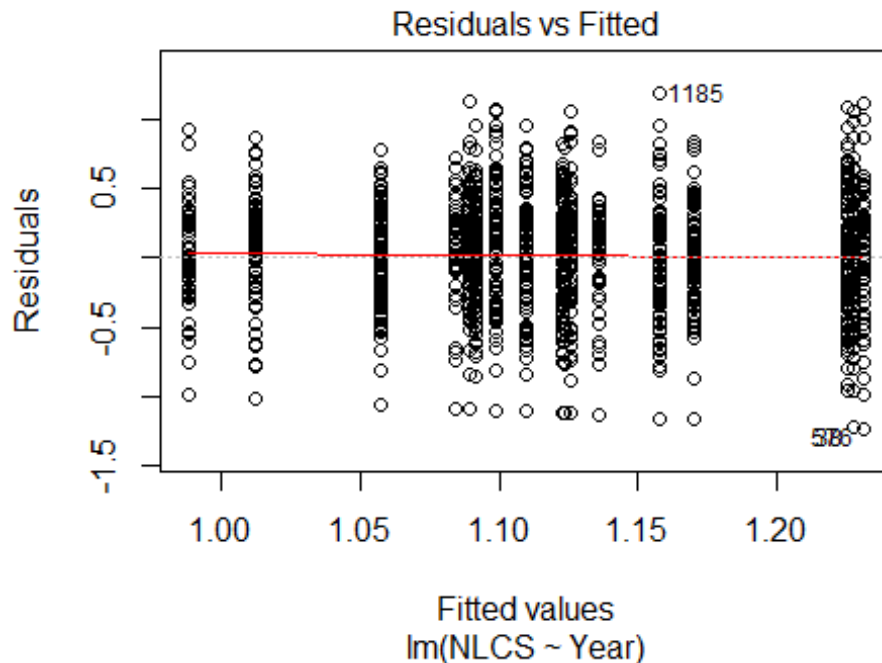
##

Bartlett test of homogeneity of variances

##

data: NLCS by Year

Bartlett's K-squared = 25, df = 16, p-value = 0.06



##

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: 3.75582990185108"

[1] "Male first author team size 2018 geometric mean: 3.85485405449255"

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

##

Wilcoxon rank sum test with continuity correction

##

data: FemaleTeamSizes2018 and MaleTeamSizes2018

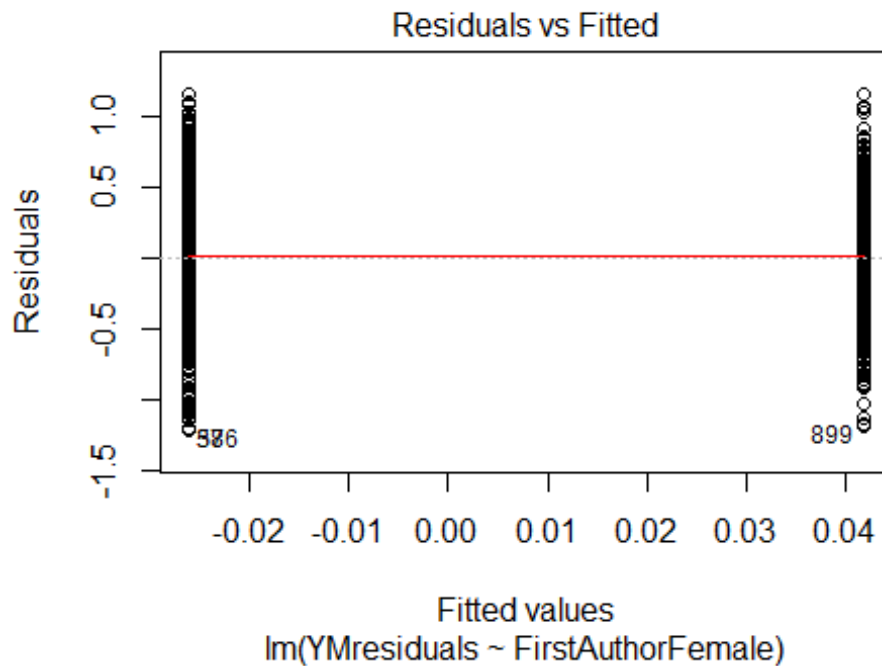
W = 280, p-value = 0.3

alternative hypothesis: true location shift is not equal to 0

##

```
## [1] "Female last author team size 2018 geometric mean: 3.96978171017282"
## [1] "Male last author team size 2018 geometric mean: 3.77148042355977"

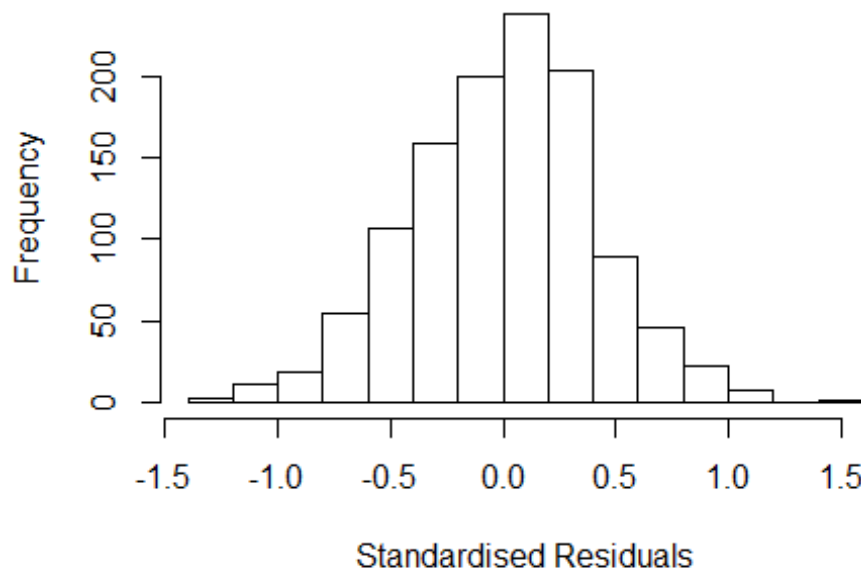
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 200, p-value = 0.8
## 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	1.081	1	1.040
LastAuthorFemale	1.117	1	1.057
UniqueAuthors	1.421	4	1.045
Year	1.581	16	1.014

Residuals from first and last author and team size



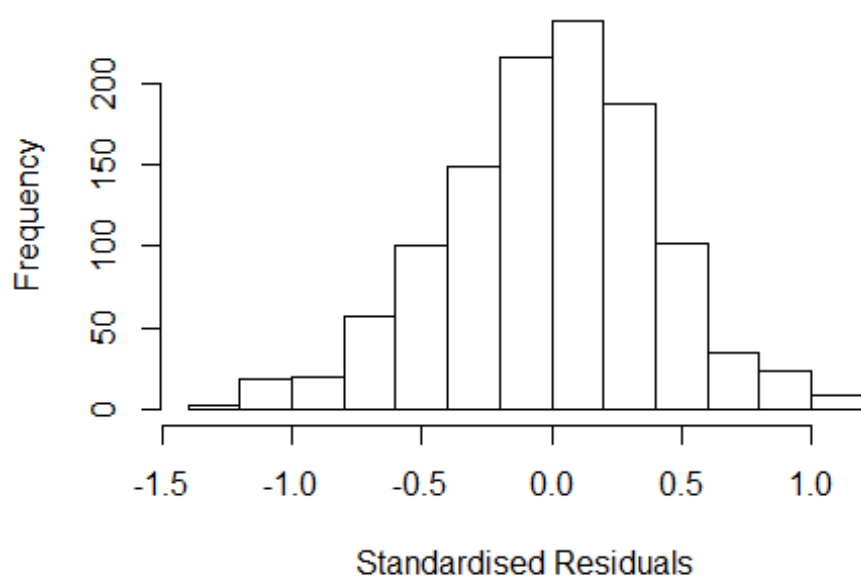
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2551 -0.2771  0.0191  0.2683  1.4179
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8711     0.1023   8.52 < 2e-16 ***
## FirstAuthorFemale1 0.0490     0.0251   1.95  0.05086 .
## LastAuthorFemale1 0.0386     0.0321   1.20  0.22967
## UniqueAuthors2    0.3376     0.0951   3.55  0.00040 ***
## UniqueAuthors3    0.3649     0.0937   3.89  0.00010 ***
## UniqueAuthors4    0.3839     0.0954   4.02  6.1e-05 ***
## UniqueAuthors5    0.4600     0.0933   4.93  9.6e-07 ***
## Year1997         -0.1101     0.0864  -1.27  0.20269
## Year1998         -0.0249     0.0944  -0.26  0.79189
## Year1999         -0.1316     0.0768  -1.71  0.08697 .
```

```

## Year2000          -0.0783      0.0820    -0.95   0.33986
## Year2001          -0.1170      0.0858    -1.36   0.17304
## Year2002          -0.1076      0.0762    -1.41   0.15836
## Year2003          -0.1154      0.0805    -1.43   0.15207
## Year2004          -0.2661      0.0784    -3.39   0.00071 ***
## Year2005          -0.2496      0.0750    -3.33   0.00090 ***
## Year2006          -0.2164      0.0679    -3.19   0.00148 **
## Year2007          -0.1713      0.0686    -2.50   0.01267 *
## Year2008          -0.1761      0.0736    -2.39   0.01686 *
## Year2009          -0.2324      0.0756    -3.07   0.00217 **
## Year2010          -0.1209      0.0695    -1.74   0.08240 .
## Year2011          -0.0354      0.0775    -0.46   0.64765
## Year2012          -0.1798      0.0757    -2.37   0.01777 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.0801, Adjusted R-squared:  0.0622
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1059 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.162  0.868  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          8.65e-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 1.085 1          1.041
## LastAuthorFemale  1.094 1          1.046
## Year              1.173 16          1.005

```

Residuals from first and last author



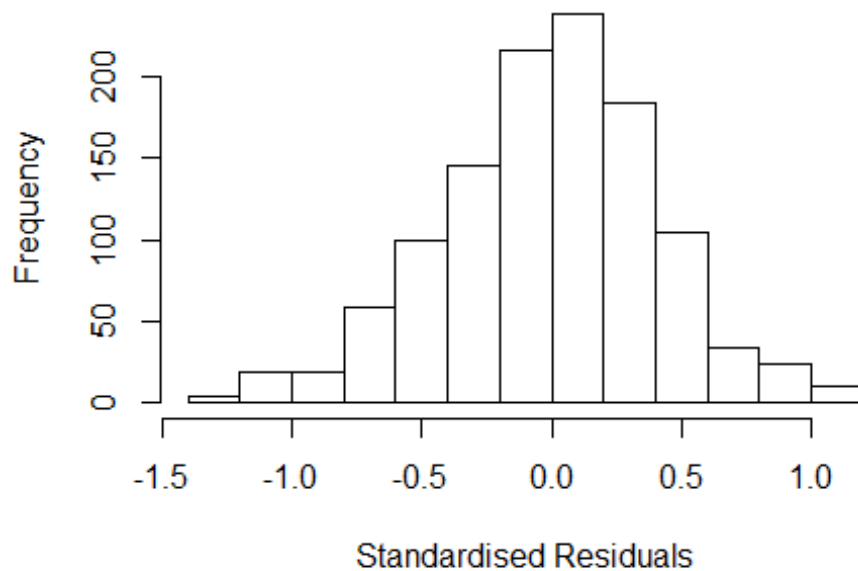
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20894 -0.27388  0.00893  0.26731  1.16368
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.208938   0.056382   21.44  <2e-16 ***
## FirstAuthorFemale1  0.063747   0.025513    2.50  0.0126 *
## LastAuthorFemale1  0.035260   0.033189    1.06  0.2883
## Year1997        -0.106828   0.084421   -1.27  0.2060
## Year1998        -0.006274   0.093802   -0.07  0.9467
## Year1999        -0.111334   0.076310   -1.46  0.1449
## Year2000        -0.079727   0.082155   -0.97  0.3320
## Year2001        -0.095384   0.092129   -1.04  0.3007
## Year2002        -0.073189   0.076673   -0.95  0.3400
## Year2003        -0.080503   0.079271   -1.02  0.3101
## Year2004        -0.221412   0.080336   -2.76  0.0059 **
## Year2005        -0.225328   0.074128   -3.04  0.0024 **
```

```

## Year2006      -0.171573    0.066932   -2.56    0.0105 *
## Year2007      -0.128554    0.068029   -1.89    0.0591 .
## Year2008      -0.126171    0.073181   -1.72    0.0850 .
## Year2009      -0.161617    0.074681   -2.16    0.0307 *
## Year2010      -0.062623    0.068796   -0.91    0.3629
## Year2011       0.000421    0.078549    0.01    0.9957
## Year2012      -0.140501    0.075949   -1.85    0.0646 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.0322, Adjusted R-squared:  0.0168
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 1049 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.322  0.859   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      8.65e-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 1.08 1      1.039
## Year              1.08 16      1.002

```

Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21337 -0.27042 0.00927 0.26205 1.16058
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21337 0.05633 21.54 <2e-16 ***
## FirstAuthorFemale1 0.06729 0.02545 2.64 0.0083 **
## Year1997 -0.10726 0.08435 -1.27 0.2038
## Year1998 -0.00636 0.09361 -0.07 0.9458
## Year1999 -0.10973 0.07636 -1.44 0.1510
## Year2000 -0.07941 0.08176 -0.97 0.3316
## Year2001 -0.09580 0.09189 -1.04 0.2974
## Year2002 -0.07404 0.07671 -0.97 0.3347
## Year2003 -0.07893 0.07931 -1.00 0.3198
## Year2004 -0.22205 0.08022 -2.77 0.0057 **
## Year2005 -0.22389 0.07412 -3.02 0.0026 **
## Year2006 -0.17197 0.06690 -2.57 0.0103 *
```

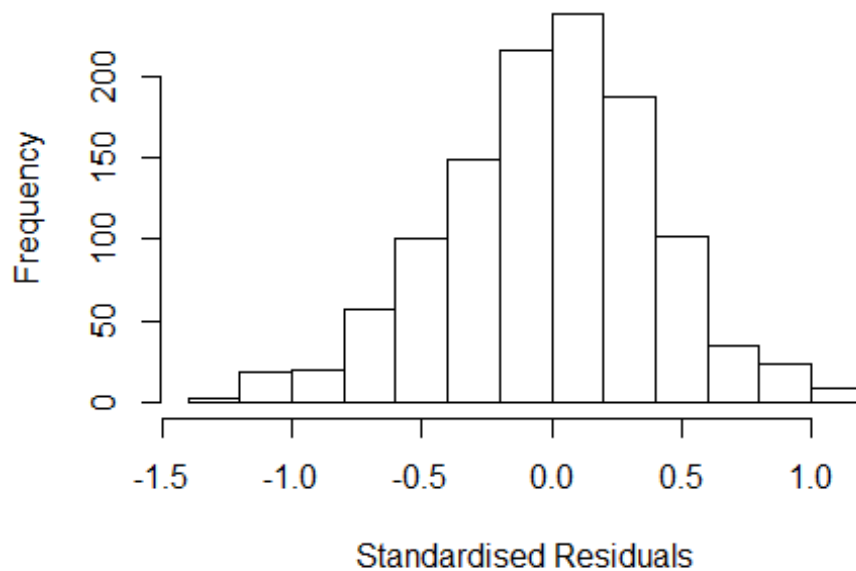


```

## Year2007          -0.12963    0.06798   -1.91    0.0568 .
## Year2008          -0.12873    0.07324   -1.76    0.0791 .
## Year2009          -0.16294    0.07474   -2.18    0.0295 *
## Year2010          -0.06146    0.06877   -0.89    0.3717
## Year2011           0.00661    0.07777    0.08    0.9323
## Year2012          -0.13617    0.07611   -1.79    0.0739 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.0312, Adjusted R-squared:  0.0167
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 111 weights are ~= 1. The remaining 1045 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.331  0.862  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      8.65e-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 1.086 1          1.042
## Year              1.086 16          1.003

```

Residuals from last author



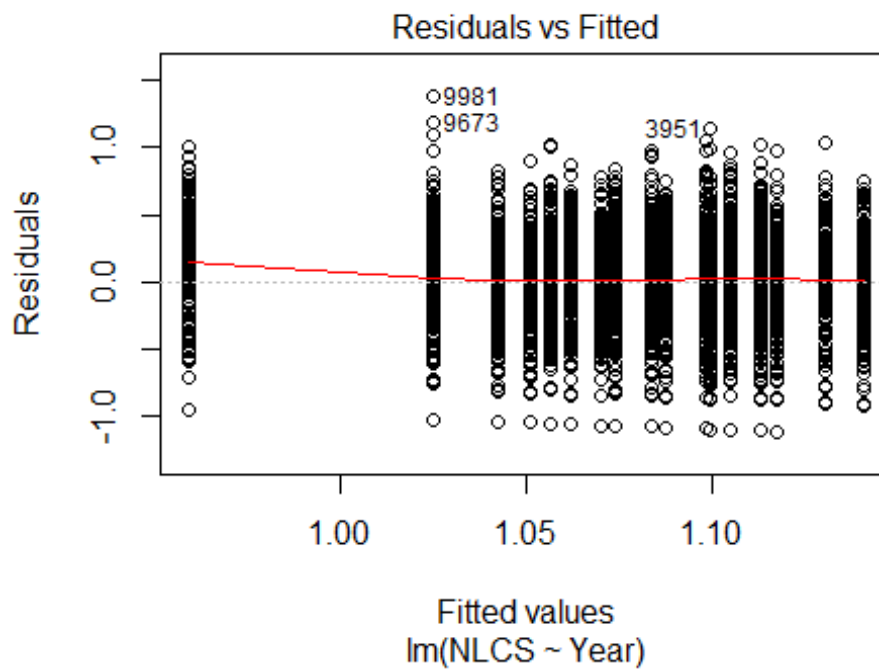
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2234 -0.2746  0.0106  0.2658  1.1756
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.21991    0.05650   21.59  <2e-16 ***
## LastAuthorFemale1 0.04497    0.03334    1.35  0.1777
## Year1997       -0.09882    0.08358   -1.18  0.2373
## Year1998        0.00354    0.09356    0.04  0.9699
## Year1999       -0.10174    0.07641   -1.33  0.1833
## Year2000       -0.07475    0.08326   -0.90  0.3695
## Year2001       -0.08727    0.09159   -0.95  0.3409
## Year2002       -0.05347    0.07629   -0.70  0.4835
## Year2003       -0.06789    0.07896   -0.86  0.3901
## Year2004       -0.20993    0.08038   -2.61  0.0091 **
## Year2005       -0.21670    0.07466   -2.90  0.0038 **
## Year2006       -0.15942    0.06723   -2.37  0.0179 *
```

```

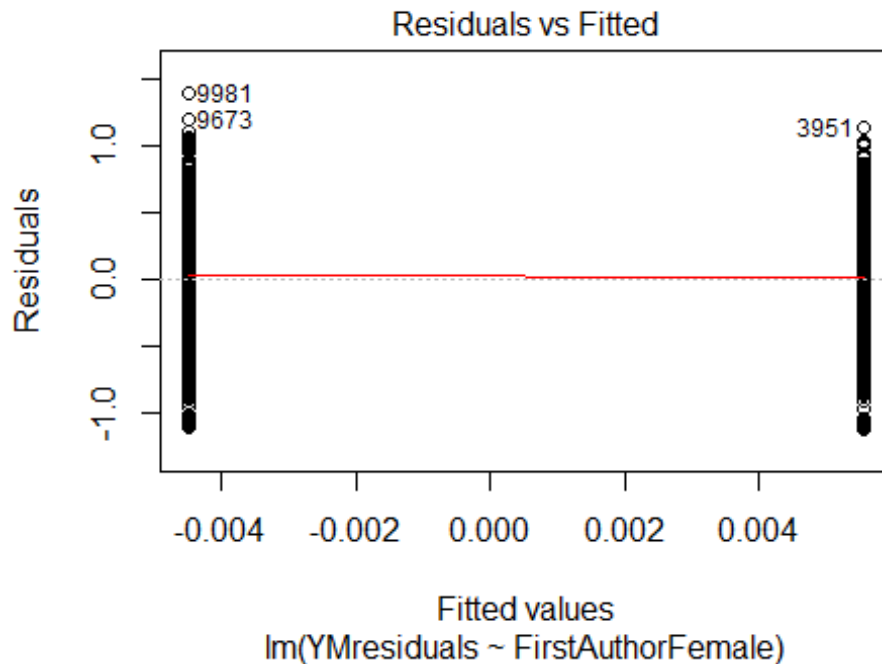
## Year2007          -0.11711      0.06807    -1.72    0.0856 .
## Year2008          -0.11098      0.07322    -1.52    0.1299
## Year2009          -0.13871      0.07434    -1.87    0.0623 .
## Year2010          -0.04861      0.06928    -0.70    0.4831
## Year2011           0.02198      0.07854     0.28    0.7796
## Year2012          -0.12066      0.07572    -1.59    0.1114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.0267, Adjusted R-squared:  0.0122
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 90 weights are ~= 1. The remaining 1066 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.315  0.864  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      8.65e-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: 1156"
## [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
## 471 441 452 426 530 460 504 332 368 440 416 456 418 422 472
## 2011 2012
## 461 464
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 265 263 243 250 231 131 345 213 224 278 267 316 283 279 333
## 2011 2012

```

```
## 338 343
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 238 231 206 205 198 107 292 170 192 222 218 254 236 246 281
## 2011 2012
## 273 294
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 170, df = 16, p-value <2e-16
```

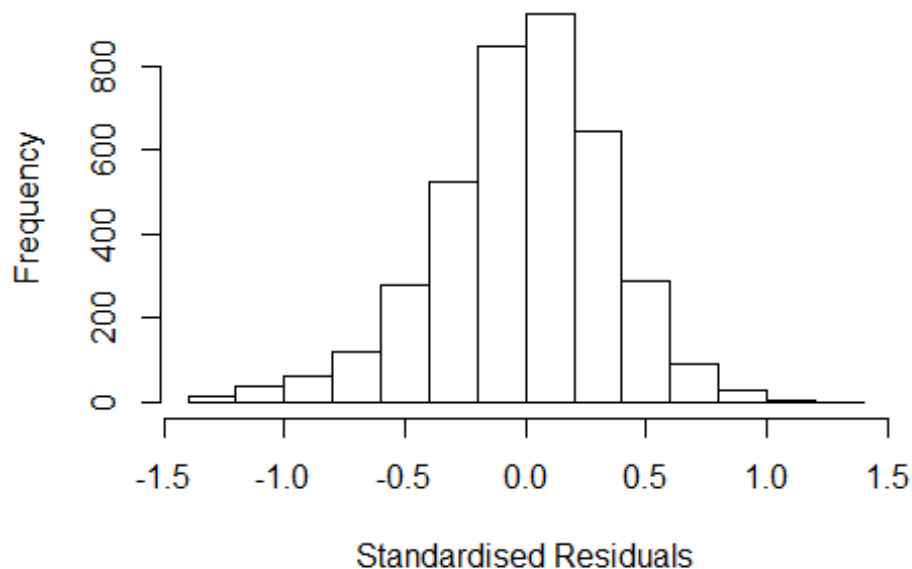


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 9e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 5.35991753733586"
## [1] "Male first author team size 2018 geometric mean: 4.87632824896086"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8200, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.1093477898696"
## [1] "Male last author team size 2018 geometric mean: 5.14514872085062"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6700, 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 1.039 1          1.019
## LastAuthorFemale  1.035 1          1.017
## UniqueAuthors    1.122 4          1.015
## Year             1.165 16          1.005
```

Residuals from first and last author and team size



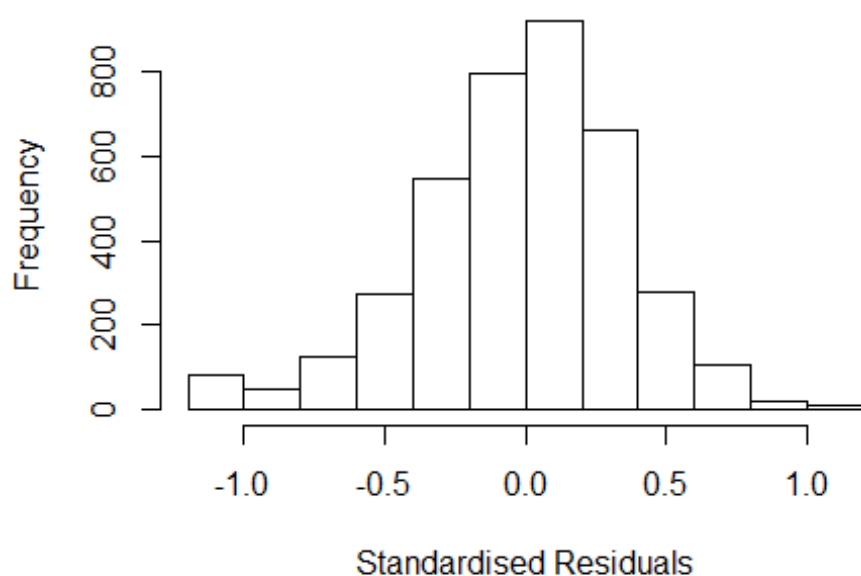
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22562 -0.22510 0.00829 0.22137 1.27903
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.930512 0.051660 18.01 < 2e-16 ***
## FirstAuthorFemale1 -0.008129 0.011485 -0.71 0.47914
## LastAuthorFemale1 0.000247 0.013401 0.02 0.98529
## UniqueAuthors2 0.135091 0.049346 2.74 0.00622 **
## UniqueAuthors3 0.166392 0.048763 3.41 0.00065 ***
## UniqueAuthors4 0.220465 0.048626 4.53 6e-06 ***
## UniqueAuthors5 0.295112 0.047743 6.18 7e-10 ***
## Year1997 -0.017037 0.039959 -0.43 0.66987
## Year1998 0.008174 0.035207 0.23 0.81643
## Year1999 0.004942 0.036108 0.14 0.89115
```

```

## Year2000      -0.014014    0.049466   -0.28   0.77696
## Year2001      0.002660    0.048057    0.06   0.95586
## Year2002     -0.042525    0.033206   -1.28   0.20040
## Year2003     -0.080389    0.036132   -2.22   0.02615 *
## Year2004     -0.088312    0.037617   -2.35   0.01894 *
## Year2005     -0.078778    0.034203   -2.30   0.02132 *
## Year2006     -0.065264    0.035740   -1.83   0.06792 .
## Year2007     -0.098787    0.035011   -2.82   0.00480 **
## Year2008     -0.056543    0.035937   -1.57   0.11571
## Year2009     -0.079345    0.035174   -2.26   0.02414 *
## Year2010     -0.079399    0.034857   -2.28   0.02279 *
## Year2011     -0.098308    0.034894   -2.82   0.00487 **
## Year2012     -0.134838    0.036569   -3.69   0.00023 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.331
## Multiple R-squared:  0.0544, Adjusted R-squared:  0.049
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 348 weights are ~= 1. The remaining 3515 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.102  0.864  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          2.59e-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 1.033 1          1.016
## LastAuthorFemale  1.031 1          1.015
## Year              1.060 16          1.002

```

Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1387 -0.2321 0.0118 0.2293 1.1937
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13021 0.02835 39.87 <2e-16 ***
## FirstAuthorFemale1 -0.00168 0.01168 -0.14 0.8857
## LastAuthorFemale1 -0.00483 0.01373 -0.35 0.7248
## Year1997 -0.00498 0.04014 -0.12 0.9012
## Year1998 0.00851 0.03645 0.23 0.8154
## Year1999 0.01411 0.03600 0.39 0.6951
## Year2000 -0.01729 0.05158 -0.34 0.7374
## Year2001 0.02547 0.04785 0.53 0.5946
## Year2002 -0.01170 0.03321 -0.35 0.7246
## Year2003 -0.04874 0.03660 -1.33 0.1830
## Year2004 -0.06911 0.03787 -1.82 0.0681 .
## Year2005 -0.05548 0.03486 -1.59 0.1116
```

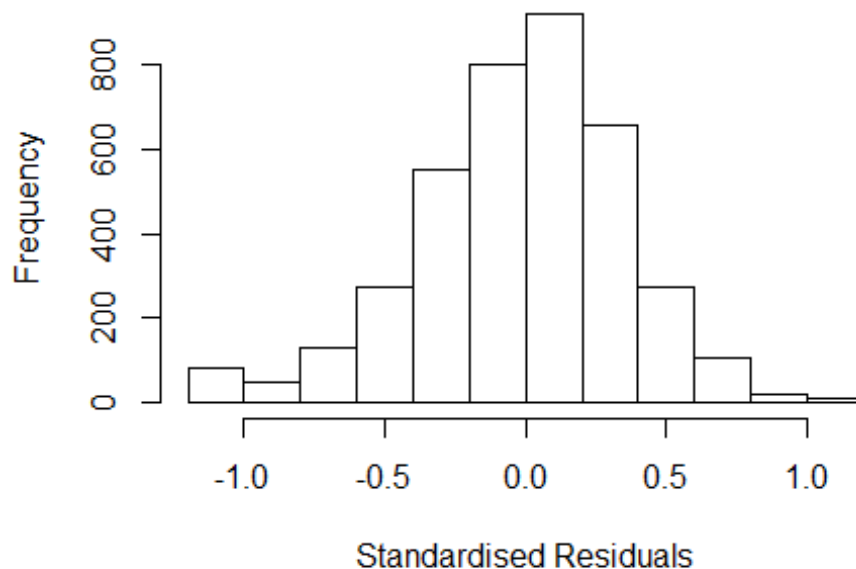


```

## Year2006          -0.03001    0.03508   -0.86    0.3923
## Year2007          -0.06679    0.03518   -1.90    0.0577 .
## Year2008          -0.02157    0.03594   -0.60    0.5484
## Year2009          -0.04807    0.03526   -1.36    0.1729
## Year2010          -0.03517    0.03472   -1.01    0.3111
## Year2011          -0.06003    0.03509   -1.71    0.0872 .
## Year2012          -0.10207    0.03682   -2.77    0.0056 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.338
## Multiple R-squared:  0.00893,    Adjusted R-squared:  0.00429
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 328 weights are ~= 1. The remaining 3535 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.185  0.869  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      2.59e-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 1.031 1      1.016
## Year              1.031 16      1.001

```

Residuals from first author



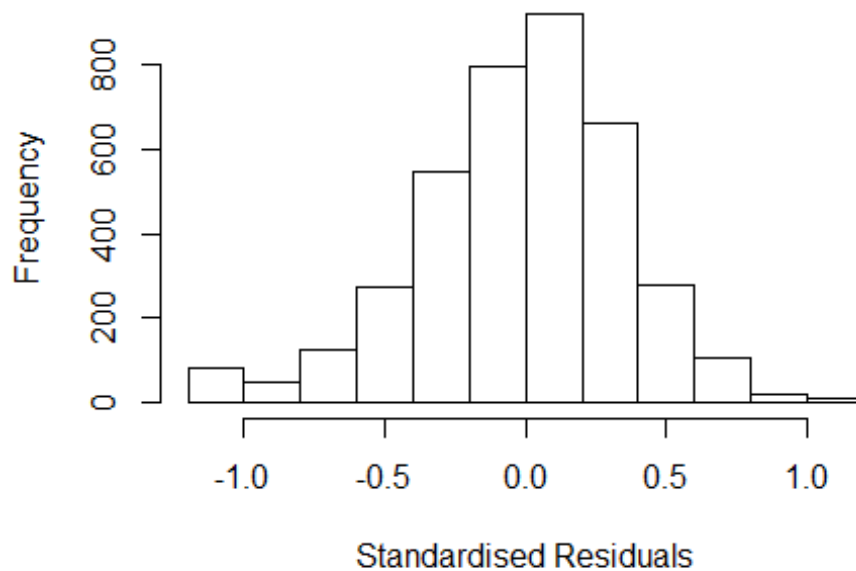
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1379 -0.2317 0.0117 0.2292 1.1901
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12922 0.02820 40.04 <2e-16 ***
## FirstAuthorFemale1 -0.00197 0.01168 -0.17 0.8658
## Year1997 -0.00496 0.04013 -0.12 0.9017
## Year1998 0.00871 0.03644 0.24 0.8110
## Year1999 0.01426 0.03599 0.40 0.6919
## Year2000 -0.01698 0.05157 -0.33 0.7419
## Year2001 0.02558 0.04782 0.53 0.5927
## Year2002 -0.01160 0.03319 -0.35 0.7267
## Year2003 -0.04866 0.03659 -1.33 0.1836
## Year2004 -0.06870 0.03782 -1.82 0.0694 .
## Year2005 -0.05563 0.03483 -1.60 0.1103
## Year2006 -0.02996 0.03506 -0.85 0.3928
```

```

## Year2007          -0.06685    0.03518   -1.90    0.0574 .
## Year2008          -0.02151    0.03591   -0.60    0.5492
## Year2009          -0.04817    0.03525   -1.37    0.1719
## Year2010          -0.03560    0.03466   -1.03    0.3045
## Year2011          -0.06019    0.03507   -1.72    0.0862 .
## Year2012          -0.10234    0.03679   -2.78    0.0054 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.338
## Multiple R-squared:  0.0089, Adjusted R-squared:  0.00452
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 320 weights are ~= 1. The remaining 3543 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.188  0.870  0.948  0.894  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.59e-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 1.03 1      1.015
## Year      1.03 16      1.001

```

Residuals from last author



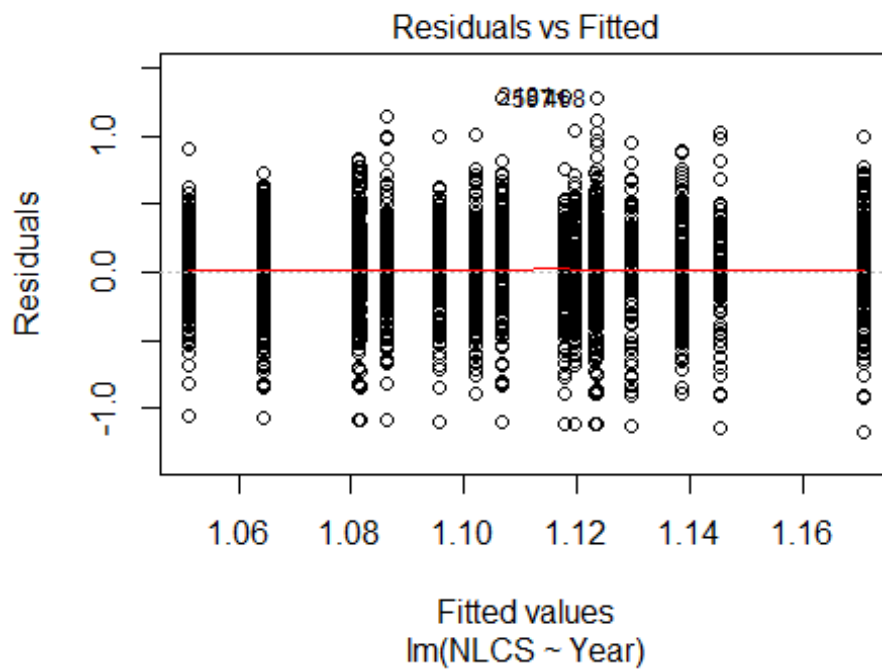
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1382 -0.2323 0.0123 0.2296 1.1947
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12967 0.02807 40.24 <2e-16 ***
## LastAuthorFemale1 -0.00497 0.01372 -0.36 0.7173
## Year1997 -0.00504 0.04014 -0.13 0.9002
## Year1998 0.00852 0.03646 0.23 0.8152
## Year1999 0.01397 0.03599 0.39 0.6980
## Year2000 -0.01729 0.05160 -0.34 0.7376
## Year2001 0.02530 0.04784 0.53 0.5969
## Year2002 -0.01195 0.03316 -0.36 0.7187
## Year2003 -0.04895 0.03657 -1.34 0.1808
## Year2004 -0.06927 0.03788 -1.83 0.0675 .
## Year2005 -0.05574 0.03479 -1.60 0.1092
## Year2006 -0.03027 0.03500 -0.86 0.3871
```

```

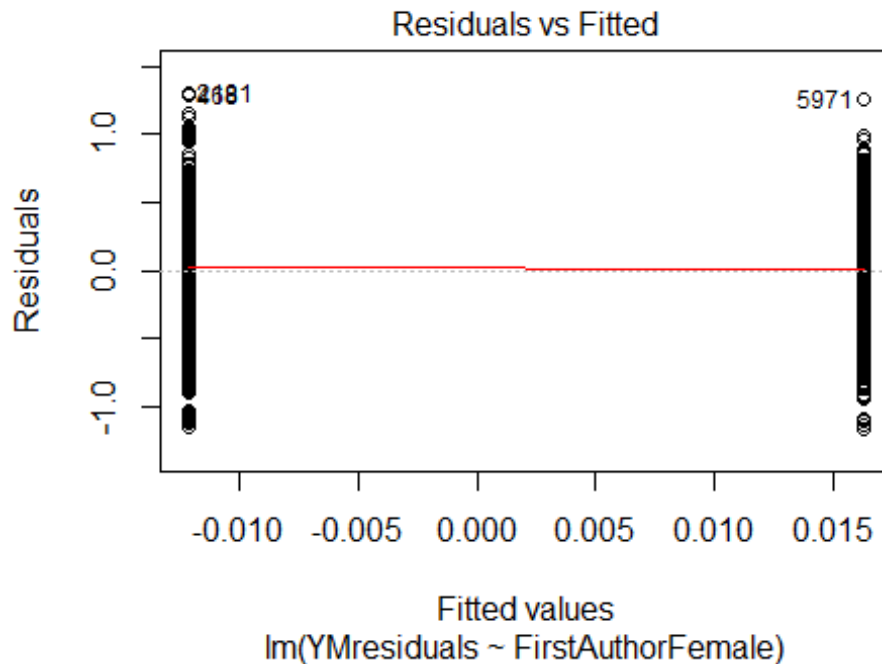
## Year2007          -0.06694      0.03516    -1.90    0.0570 .
## Year2008          -0.02176      0.03590    -0.61    0.5444
## Year2009          -0.04832      0.03520    -1.37    0.1699
## Year2010          -0.03536      0.03470    -1.02    0.3082
## Year2011          -0.06035      0.03501    -1.72    0.0848 .
## Year2012          -0.10240      0.03667    -2.79    0.0053 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.338
## Multiple R-squared:  0.00893,    Adjusted R-squared:  0.00455
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 329 weights are ~= 1. The remaining 3534 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.184  0.869  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      2.59e-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: 3863"
## [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
## 327 317 269 248 268 275 290 244 258 262 331 308 298 276 255
## 2011 2012
## 271 316
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 150 156 120 133 94 54 172 152 153 167 212 214 195 195 186
## 2011 2012

```

```
## 188 230
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 128 128 100 116 80 44 132 116 130 139 175 181 168 171 161
## 2011 2012
## 156 203
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 58, df = 16, p-value = 1e-06
```

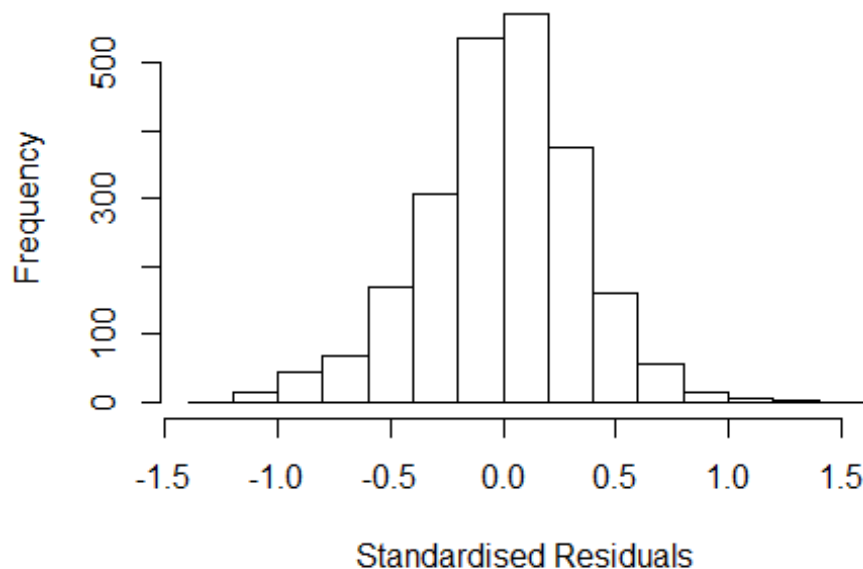


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 4.5452574426498"
## [1] "Male first author team size 2018 geometric mean: 4.23769209931312"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3200, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.08833475928907"
## [1] "Male last author team size 2018 geometric mean: 4.4874633511364"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "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  1.024 1      1.012
## UniqueAuthors    1.207 4      1.024
## 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
## -1.20791 -0.21168  0.00713  0.21339  1.43571
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9011    0.0823   10.96 < 2e-16 ***
## FirstAuthorFemale1  0.0165    0.0143    1.15  0.24909
## LastAuthorFemale1 -0.0110    0.0170   -0.65  0.51793
## UniqueAuthors2     0.2180    0.0793    2.75  0.00602 **
## UniqueAuthors3     0.2762    0.0789    3.50  0.00047 ***
## UniqueAuthors4     0.2748    0.0791    3.47  0.00052 ***
## UniqueAuthors5     0.3170    0.0787    4.03  5.9e-05 ***
## Year1997         -0.0199    0.0552   -0.36  0.71799
## Year1998         -0.0180    0.0515   -0.35  0.72610
## Year1999         -0.0709    0.0496   -1.43  0.15295
```

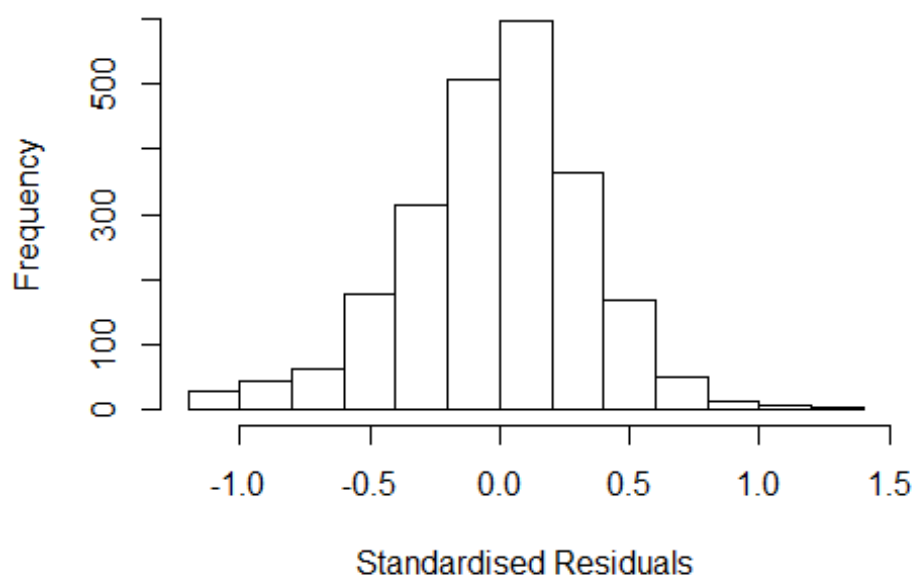


```

## Year2000          0.0141      0.0560      0.25  0.80150
## Year2001         -0.0276      0.0760     -0.36  0.71597
## Year2002         -0.0484      0.0433     -1.12  0.26393
## Year2003         -0.1112      0.0488     -2.28  0.02269 *
## Year2004         -0.1327      0.0452     -2.94  0.00335 **
## Year2005         -0.0626      0.0444     -1.41  0.15879
## Year2006         -0.0488      0.0426     -1.14  0.25248
## Year2007         -0.0807      0.0435     -1.85  0.06390 .
## Year2008         -0.1004      0.0429     -2.34  0.01927 *
## Year2009         -0.0781      0.0449     -1.74  0.08189 .
## Year2010         -0.0541      0.0446     -1.21  0.22479
## Year2011         -0.1048      0.0449     -2.33  0.01972 *
## Year2012         -0.0330      0.0441     -0.75  0.45429
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.0415, Adjusted R-squared:  0.0324
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 210 weights are ~= 1. The remaining 2118 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0058 0.8610 0.9500 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.30e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1      1.021
## LastAuthorFemale  1.017 1      1.009
## Year              1.058 16      1.002

```

Residuals from first and last author



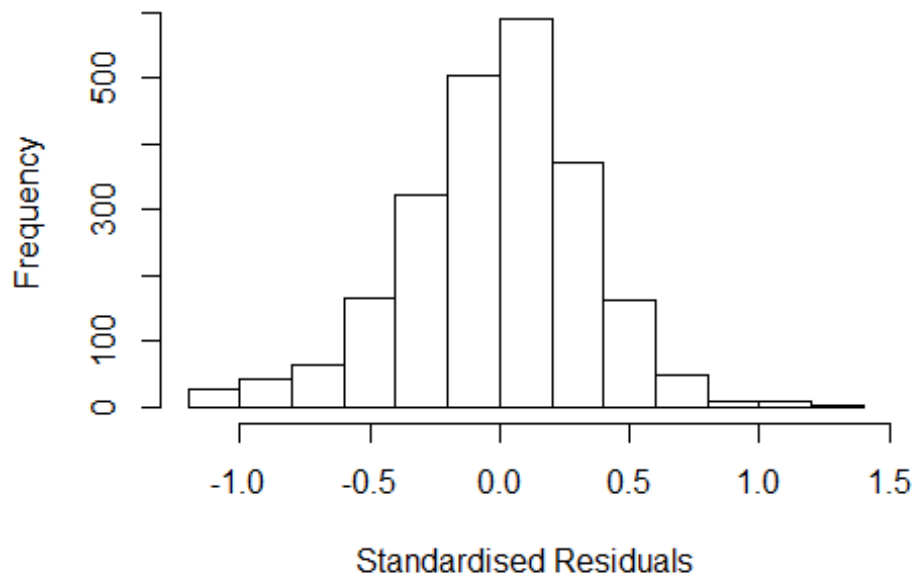
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19422 -0.21982 0.00961 0.20830 1.25945
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15354 0.03608 31.97 <2e-16 ***
## FirstAuthorFemale1 0.02222 0.01446 1.54 0.125
## LastAuthorFemale1 -0.01361 0.01704 -0.80 0.425
## Year1997 -0.00898 0.05542 -0.16 0.871
## Year1998 -0.00130 0.05104 -0.03 0.980
## Year1999 -0.06282 0.04938 -1.27 0.203
## Year2000 0.01846 0.05675 0.33 0.745
## Year2001 -0.00882 0.07546 -0.12 0.907
## Year2002 -0.03493 0.04356 -0.80 0.423
## Year2003 -0.10388 0.04907 -2.12 0.034 *
## Year2004 -0.11368 0.04582 -2.48 0.013 *
## Year2005 -0.04945 0.04448 -1.11 0.266
```

```

## Year2006      -0.03200    0.04261   -0.75    0.453
## Year2007      -0.06029    0.04372   -1.38    0.168
## Year2008      -0.08272    0.04350   -1.90    0.057 .
## Year2009      -0.05339    0.04444   -1.20    0.230
## Year2010      -0.02640    0.04454   -0.59    0.553
## Year2011      -0.07676    0.04532   -1.69    0.090 .
## Year2012      -0.00761    0.04463   -0.17    0.865
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.0119, Adjusted R-squared:  0.00424
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 193 weights are ~= 1. The remaining 2135 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0847 0.8600 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      4.30e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.041 1      1.020
## Year      1.041 16      1.001

```

Residuals from first author



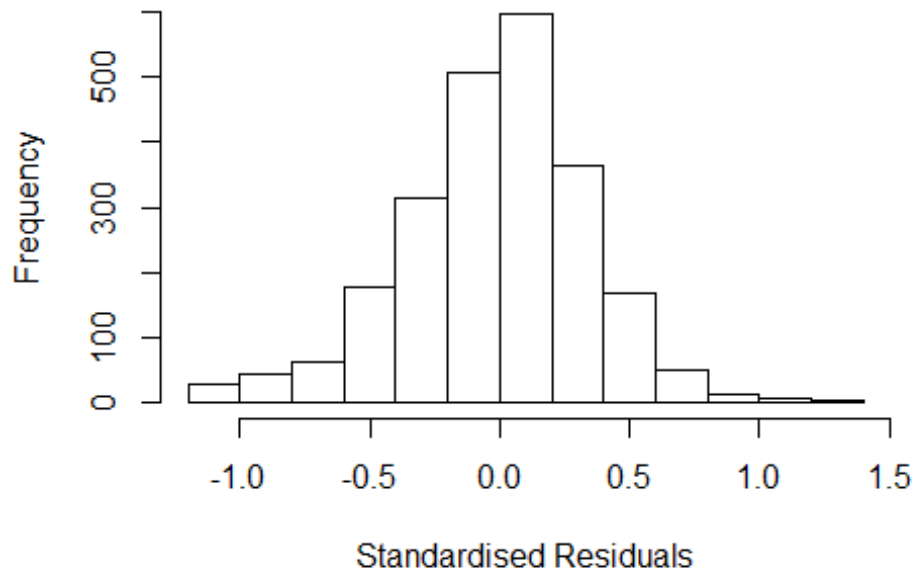
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19180 -0.21940 0.00884 0.20770 1.26259
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15082 0.03586 32.09 <2e-16 ***
## FirstAuthorFemale1 0.02203 0.01448 1.52 0.128
## Year1997 -0.00941 0.05548 -0.17 0.865
## Year1998 -0.00166 0.05108 -0.03 0.974
## Year1999 -0.06248 0.04937 -1.27 0.206
## Year2000 0.01895 0.05695 0.33 0.739
## Year2001 -0.00840 0.07555 -0.11 0.911
## Year2002 -0.03528 0.04353 -0.81 0.418
## Year2003 -0.10419 0.04908 -2.12 0.034 *
## Year2004 -0.11375 0.04591 -2.48 0.013 *
## Year2005 -0.04948 0.04454 -1.11 0.267
## Year2006 -0.03204 0.04263 -0.75 0.452
```

```

## Year2007          -0.06015    0.04371   -1.38    0.169
## Year2008          -0.08233    0.04348   -1.89    0.058 .
## Year2009          -0.05340    0.04443   -1.20    0.230
## Year2010          -0.02686    0.04450   -0.60    0.546
## Year2011          -0.07689    0.04529   -1.70    0.090 .
## Year2012          -0.00831    0.04462   -0.19    0.852
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.0117, Adjusted R-squared:  0.00438
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 194 weights are ~= 1. The remaining 2134 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0824 0.8610 0.9510 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      4.30e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.016 1          1.008
## Year            1.016 16          1.000

```

Residuals from last author



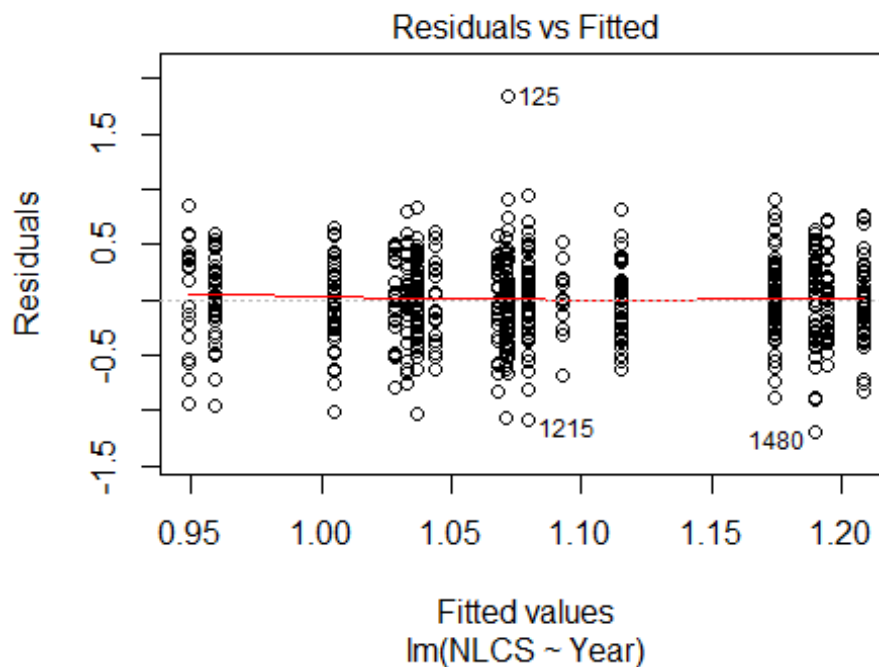
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1781 -0.2212 0.0105 0.2110 1.2533
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.158374 0.036130 32.06 <2e-16 ***
## LastAuthorFemale1 -0.013180 0.017060 -0.77 0.440
## Year1997 -0.007719 0.055494 -0.14 0.889
## Year1998 0.002548 0.050683 0.05 0.960
## Year1999 -0.059181 0.049299 -1.20 0.230
## Year2000 0.019770 0.056523 0.35 0.727
## Year2001 -0.006782 0.075479 -0.09 0.928
## Year2002 -0.028654 0.043236 -0.66 0.508
## Year2003 -0.099411 0.048906 -2.03 0.042 *
## Year2004 -0.109213 0.045900 -2.38 0.017 *
## Year2005 -0.043851 0.044188 -0.99 0.321
## Year2006 -0.027192 0.042434 -0.64 0.522
```

```

## Year2007          -0.055121    0.043419   -1.27    0.204
## Year2008          -0.077457    0.043347   -1.79    0.074 .
## Year2009          -0.047463    0.044113   -1.08    0.282
## Year2010          -0.019901    0.044258   -0.45    0.653
## Year2011          -0.069762    0.045020   -1.55    0.121
## Year2012          -0.000442    0.044165   -0.01    0.992
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.0109, Adjusted R-squared:  0.00364
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 199 weights are ~= 1. The remaining 2129 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0888 0.8590 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      4.30e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2328"
## [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
##   72   78   93   60   65   71   56   38   64   46   59   51   61   71   69
## 2011 2012
##   71   70
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   42   43   46   37   25   13   36   23   35   24   32   29   33   48   47
## 2011 2012

```

```
## 56 56
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 37 31 37 31 21 10 30 15 29 21 24 20 27 46 38
## 2011 2012
## 45 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 = 13, df = 16, p-value = 0.7
```

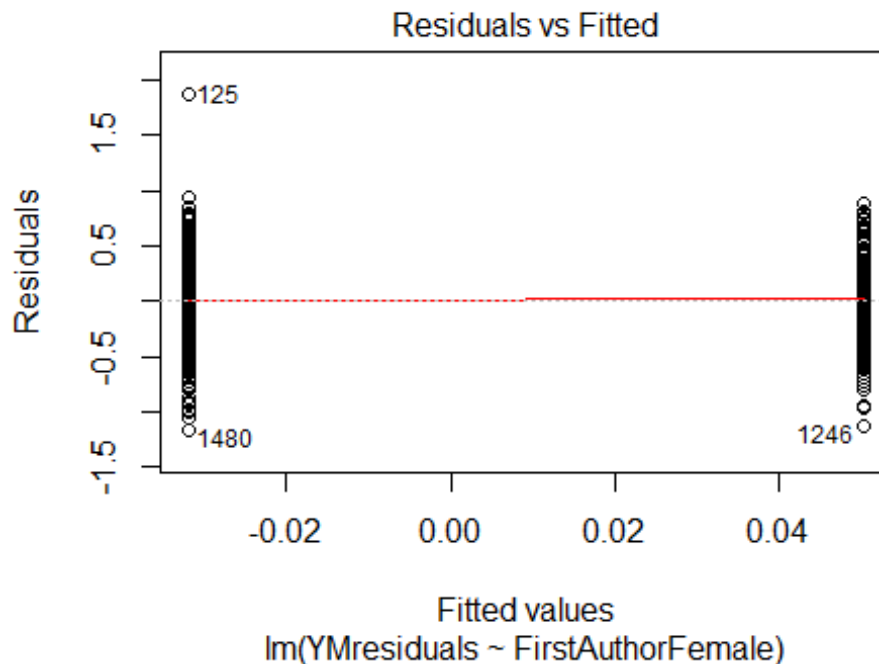


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.9, df = 1, p-value = 0.03
## [1] "Female first author team size 2018 geometric mean: 5.17035651715299"
## [1] "Male first author team size 2018 geometric mean: 4.37950134842466"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



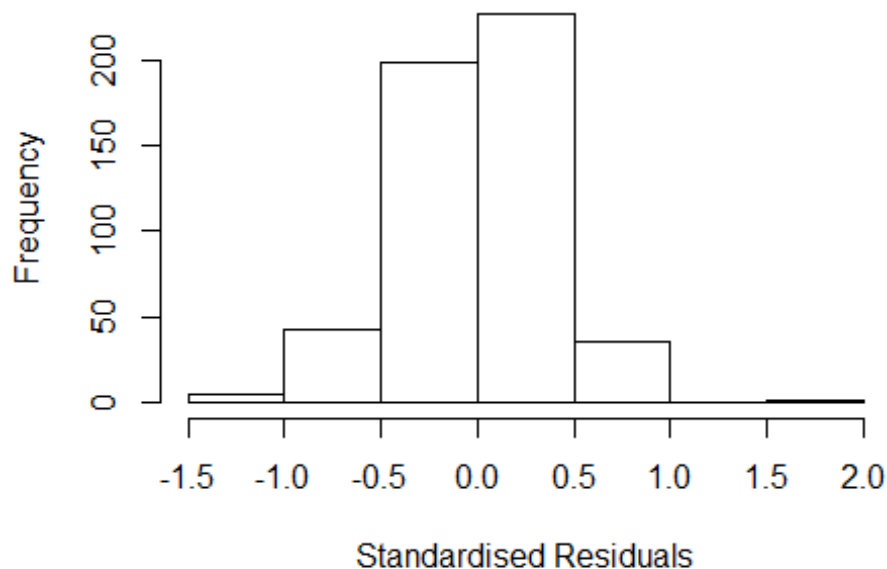
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.40643345387186"
## [1] "Male last author team size 2018 geometric mean: 4.72315381019964"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 99, 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 1.325  1      1.151
## LastAuthorFemale  1.227  1      1.108
## UniqueAuthors    2.816  4      1.138
## Year              3.269 16      1.038
```

Residuals from first and last author and team size



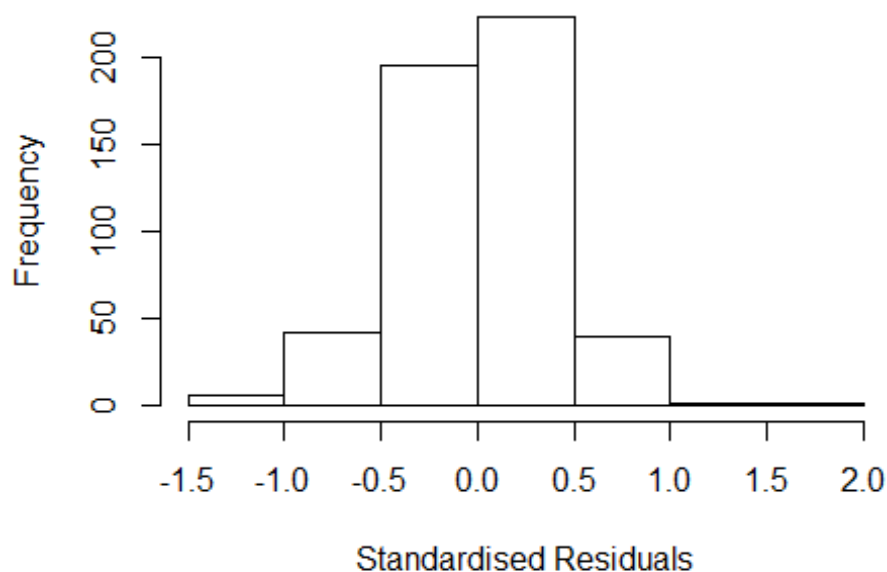
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.14891 -0.22353  0.00753  0.22026  1.94703
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0487    0.1052   9.97 < 2e-16 ***
## FirstAuthorFemale1  0.0734    0.0348   2.11  0.03550 *
## LastAuthorFemale1 -0.0121    0.0395  -0.31  0.75934
## UniqueAuthors2     0.1198    0.0921   1.30  0.19419
## UniqueAuthors3     0.2093    0.0943   2.22  0.02695 *
## UniqueAuthors4     0.1976    0.0967   2.04  0.04157 *
## UniqueAuthors5     0.3798    0.0929   4.09  5.1e-05 ***
## Year1997          -0.2813    0.0837  -3.36  0.00084 ***
## Year1998          -0.2797    0.0799  -3.50  0.00051 ***
## Year1999          -0.1915    0.0890  -2.15  0.03196 *
```

```

## Year2000          -0.2578      0.1005    -2.57  0.01056 *
## Year2001          -0.0116      0.1370    -0.08  0.93245
## Year2002          -0.2343      0.0953    -2.46  0.01427 *
## Year2003          -0.5466      0.1637    -3.34  0.00090 ***
## Year2004          -0.3807      0.0941    -4.05  6.1e-05 ***
## Year2005          -0.2715      0.0946    -2.87  0.00429 **
## Year2006          -0.2434      0.1031    -2.36  0.01860 *
## Year2007          -0.2577      0.0996    -2.59  0.00994 **
## Year2008          -0.1670      0.0928    -1.80  0.07274 .
## Year2009          -0.2904      0.0843    -3.44  0.00062 ***
## Year2010          -0.2391      0.0888    -2.69  0.00734 **
## Year2011          -0.1375      0.0826    -1.66  0.09665 .
## Year2012          -0.1090      0.0863    -1.26  0.20717
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.332
## Multiple R-squared:  0.192, Adjusted R-squared:  0.155
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 46 is an outlier with |weight| = 0 ( < 0.0002);
## 37 weights are ~= 1. The remaining 471 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.206  0.866  0.952   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          1.96e-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 1.198 1          1.095
## LastAuthorFemale  1.177 1          1.085
## Year              1.410 16          1.011

```

Residuals from first and last author



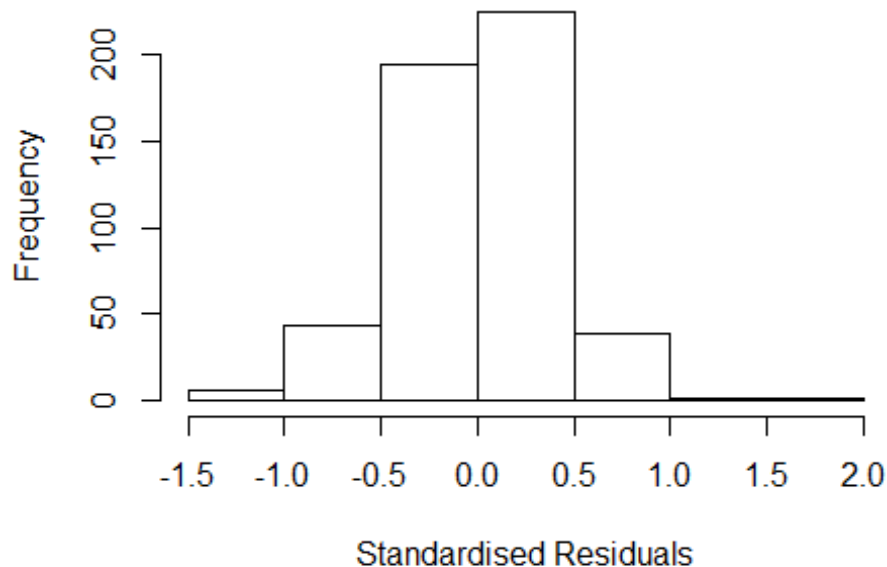
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1921 -0.2401 0.0128 0.2299 1.9609
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23313 0.06580 18.74 < 2e-16 ***
## FirstAuthorFemale1 0.12007 0.03506 3.42 0.00067 ***
## LastAuthorFemale1 -0.00995 0.04006 -0.25 0.80397
## Year1997 -0.28204 0.08252 -3.42 0.00068 ***
## Year1998 -0.27533 0.08613 -3.20 0.00148 **
## Year1999 -0.13597 0.08466 -1.61 0.10888
## Year2000 -0.24006 0.10162 -2.36 0.01855 *
## Year2001 -0.10138 0.10788 -0.94 0.34780
## Year2002 -0.23366 0.09904 -2.36 0.01871 *
## Year2003 -0.54940 0.17132 -3.21 0.00143 **
## Year2004 -0.32136 0.09967 -3.22 0.00135 **
## Year2005 -0.28370 0.10364 -2.74 0.00642 **
```

```

## Year2006      -0.25017    0.10198    -2.45    0.01451 *
## Year2007      -0.21375    0.10224    -2.09    0.03707 *
## Year2008      -0.12352    0.09080    -1.36    0.17432
## Year2009      -0.24722    0.08498    -2.91    0.00379 **
## Year2010      -0.19477    0.09591    -2.03    0.04282 *
## Year2011      -0.12568    0.08722    -1.44    0.15025
## Year2012      -0.04105    0.08706    -0.47    0.63748
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.351
## Multiple R-squared:  0.11,    Adjusted R-squared:  0.0771
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 46 is an outlier with |weight| = 0 ( < 0.0002);
## 42 weights are ~= 1. The remaining 466 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.225  0.862   0.953   0.899   0.986   0.999
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                  solve.tol          eps.outlier           eps.x
##           1.00e-07                1.00e-07           1.96e-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 1.201 1           1.096
## Year              1.201 16           1.006

```

Residuals from first author



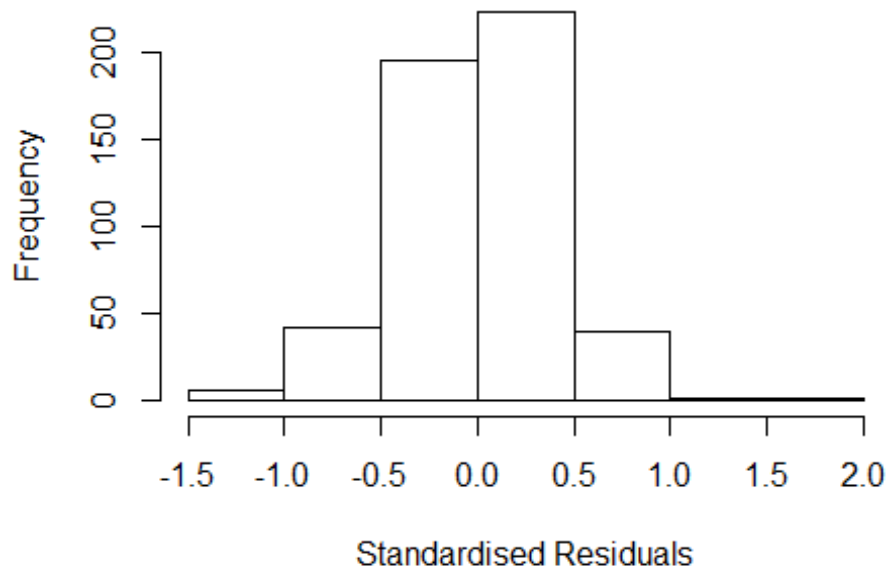
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1893 -0.2378 0.0147 0.2230 1.9639
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2307 0.0650 18.93 < 2e-16 ***
## FirstAuthorFemale1 0.1197 0.0351 3.41 0.00071 ***
## Year1997 -0.2826 0.0826 -3.42 0.00068 ***
## Year1998 -0.2749 0.0862 -3.19 0.00152 **
## Year1999 -0.1339 0.0839 -1.60 0.11135
## Year2000 -0.2395 0.1014 -2.36 0.01862 *
## Year2001 -0.1009 0.1082 -0.93 0.35164
## Year2002 -0.2343 0.0989 -2.37 0.01820 *
## Year2003 -0.5489 0.1720 -3.19 0.00151 **
## Year2004 -0.3209 0.0998 -3.22 0.00138 **
## Year2005 -0.2833 0.1034 -2.74 0.00637 **
## Year2006 -0.2495 0.1022 -2.44 0.01498 *
```

```

## Year2007          -0.2137      0.1023   -2.09  0.03728 *
## Year2008          -0.1218      0.0905   -1.35  0.17918
## Year2009          -0.2458      0.0847   -2.90  0.00389 **
## Year2010          -0.1946      0.0958   -2.03  0.04269 *
## Year2011          -0.1261      0.0872   -1.45  0.14901
## Year2012          -0.0414      0.0871   -0.47  0.63521
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.11,   Adjusted R-squared:  0.079
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 46 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.
##   0.225  0.862  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      1.96e-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 1.179 1          1.086
## Year              1.179 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
## -1.2398 -0.2492 0.0176 0.2350 1.9348
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26232 0.06519 19.36 < 2e-16 ***
## LastAuthorFemale1 -0.00123 0.04049 -0.03 0.97585
## Year1997 -0.28509 0.08161 -3.49 0.00052 ***
## Year1998 -0.26449 0.08661 -3.05 0.00238 **
## Year1999 -0.14436 0.08491 -1.70 0.08972 .
## Year2000 -0.22677 0.10467 -2.17 0.03075 *
## Year2001 -0.07164 0.09755 -0.73 0.46304
## Year2002 -0.22011 0.09669 -2.28 0.02325 *
## Year2003 -0.53223 0.17415 -3.06 0.00236 **
## Year2004 -0.30531 0.09897 -3.08 0.00215 **
## Year2005 -0.27583 0.10656 -2.59 0.00993 **
## Year2006 -0.21875 0.10581 -2.07 0.03922 *
```

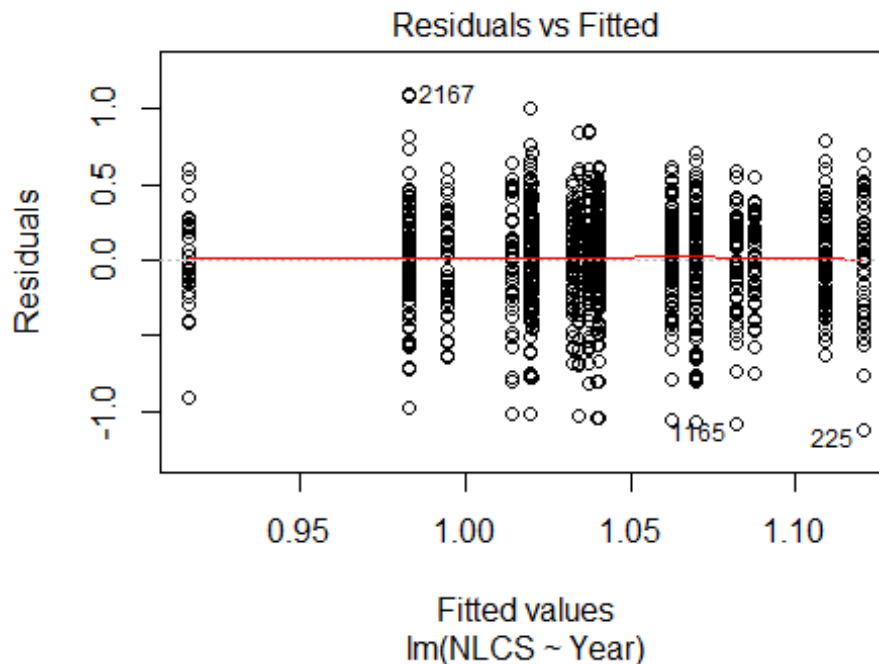


```

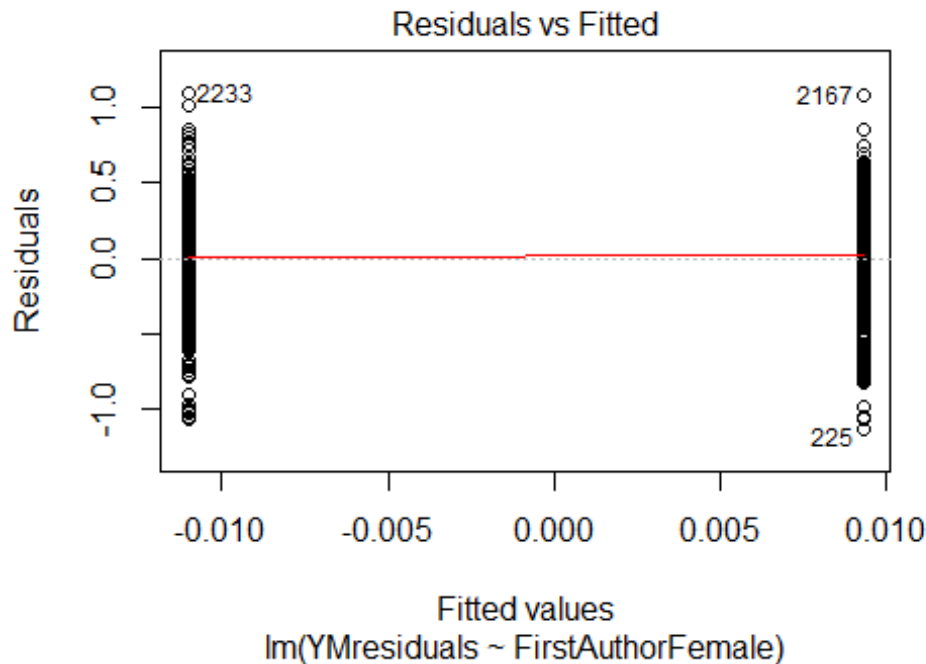
## Year2007          -0.22331      0.10228      -2.18   0.02948 *
## Year2008          -0.09312      0.09050      -1.03   0.30400
## Year2009          -0.22191      0.08685      -2.56   0.01092 *
## Year2010          -0.16289      0.09200      -1.77   0.07725 .
## Year2011          -0.09061      0.08679      -1.04   0.29699
## Year2012          -0.02254      0.08757      -0.26   0.79698
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.366
## Multiple R-squared:  0.0848, Adjusted R-squared:  0.0531
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 46 is an outlier with |weight| = 0 ( < 0.0002);
## 42 weights are ~= 1. The remaining 466 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.227  0.878  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          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 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
## 104  97  86  97  87  83  88  79  99  99  117  146  134  140  127
## 2011 2012
## 132 132
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 53  49  46  51  43  26  53  49  61  54  81  103  88  92  88

```

```
## 2011 2012
## 99 98
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 48 48 40 46 39 23 46 40 54 48 71 93 73 79 71
## 2011 2012
## 79 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 = 30, df = 16, p-value = 0.02
```



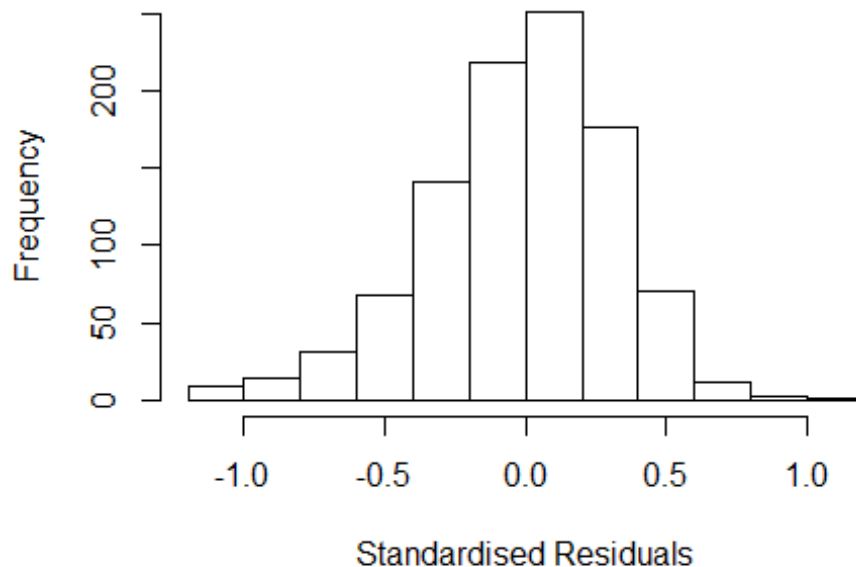
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.017, df = 1, p-value = 0.9
## [1] "Female first author team size 2018 geometric mean: 4.59914721516736"
## [1] "Male first author team size 2018 geometric mean: 5.41984626641916"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## 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] "Female last author team size 2018 geometric mean: 4.84764812180779"
## [1] "Male last author team size 2018 geometric mean: 5.03259009849932"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 520, 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	1.089	1	1.043
LastAuthorFemale	1.130	1	1.063
UniqueAuthors	1.385	4	1.042
Year	1.441	16	1.011

Residuals from first and last author and team size



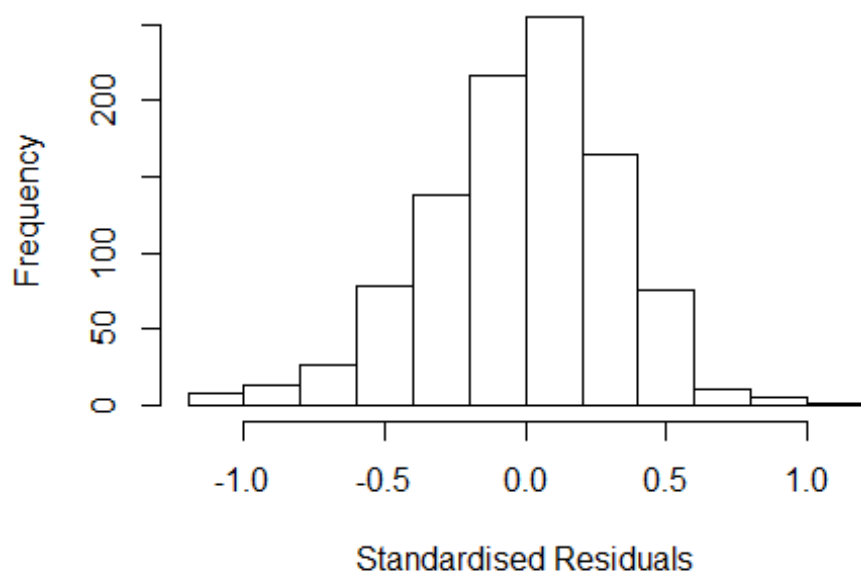
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1682 -0.2163  0.0125  0.2102  1.0364
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.988624   0.077119  12.82  <2e-16 ***
## FirstAuthorFemale1 -0.027320   0.021529  -1.27   0.205
## LastAuthorFemale1 -0.000591   0.029292  -0.02   0.984
## UniqueAuthors2     0.051384   0.068806   0.75   0.455
## UniqueAuthors3     0.043195   0.067818   0.64   0.524
## UniqueAuthors4     0.049556   0.067404   0.74   0.462
## UniqueAuthors5     0.154861   0.066413   2.33   0.020 *
## Year1997           0.077077   0.079401   0.97   0.332
## Year1998          -0.076998   0.073813  -1.04   0.297
## Year1999           0.030152   0.065092   0.46   0.643
```

```

## Year2000      0.013302    0.075993    0.18    0.861
## Year2001     -0.161124    0.079763   -2.02    0.044 *
## Year2002     -0.027686    0.074021   -0.37    0.708
## Year2003     -0.073898    0.066807   -1.11    0.269
## Year2004     -0.051501    0.065190   -0.79    0.430
## Year2005      0.048384    0.068144    0.71    0.478
## Year2006     -0.020152    0.061348   -0.33    0.743
## Year2007     -0.016031    0.060758   -0.26    0.792
## Year2008      0.028309    0.060946    0.46    0.642
## Year2009     -0.007233    0.061207   -0.12    0.906
## Year2010     -0.049141    0.067889   -0.72    0.469
## Year2011     -0.099871    0.065581   -1.52    0.128
## Year2012      0.052002    0.063888    0.81    0.416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.315
## Multiple R-squared:  0.0559, Adjusted R-squared:  0.0345
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 91 weights are ~= 1. The remaining 902 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.139  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          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 1.076 1          1.038
## LastAuthorFemale 1.063 1          1.031
## Year              1.144 16          1.004

```

Residuals from first and last author



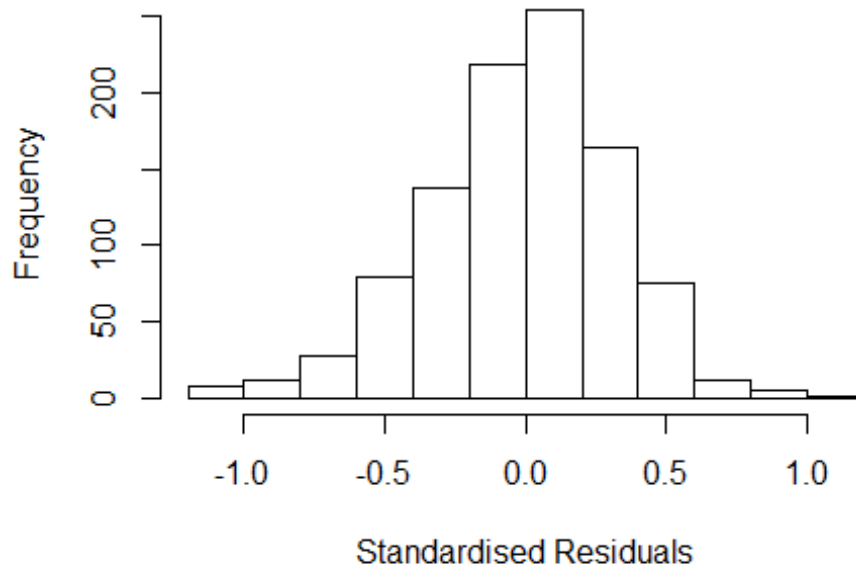
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1536 -0.2254  0.0126  0.2089  1.0936
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.07752    0.05397   19.97  <2e-16 ***
## FirstAuthorFemale1 -0.03536    0.02175   -1.63    0.104
## LastAuthorFemale1 -0.00887    0.02879   -0.31    0.758
## Year1997         0.07603    0.08179    0.93    0.353
## Year1998        -0.07460    0.07531   -0.99    0.322
## Year1999         0.03079    0.06646    0.46    0.643
## Year2000         0.00249    0.07903    0.03    0.975
## Year2001        -0.16797    0.07958   -2.11    0.035 *
## Year2002        -0.04140    0.07738   -0.54    0.593
## Year2003        -0.07861    0.06776   -1.16    0.246
## Year2004        -0.04321    0.06593   -0.66    0.512
## Year2005         0.04376    0.06993    0.63    0.532
```

```

## Year2006      -0.00390    0.06254   -0.06    0.950
## Year2007      -0.00222    0.06227   -0.04    0.972
## Year2008       0.04484    0.06207    0.72    0.470
## Year2009      -0.00241    0.06279   -0.04    0.969
## Year2010      -0.03454    0.07067   -0.49    0.625
## Year2011      -0.09116    0.06622   -1.38    0.169
## Year2012       0.04940    0.06548    0.75    0.451
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.318
## Multiple R-squared:  0.0292, Adjusted R-squared:  0.0112
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 909 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.161  0.868  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      1.01e-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 1.076 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
## -1.1525 -0.2242 0.0138 0.2108 1.0949
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07623 0.05384 19.99 <2e-16 ***
## FirstAuthorFemale1 -0.03547 0.02176 -1.63 0.103
## Year1997 0.07629 0.08185 0.93 0.351
## Year1998 -0.07483 0.07528 -0.99 0.320
## Year1999 0.03115 0.06649 0.47 0.640
## Year2000 0.00191 0.07914 0.02 0.981
## Year2001 -0.16915 0.07940 -2.13 0.033 *
## Year2002 -0.04199 0.07745 -0.54 0.588
## Year2003 -0.07922 0.06787 -1.17 0.243
## Year2004 -0.04247 0.06585 -0.64 0.519
## Year2005 0.04437 0.06989 0.63 0.526
## Year2006 -0.00399 0.06263 -0.06 0.949
```

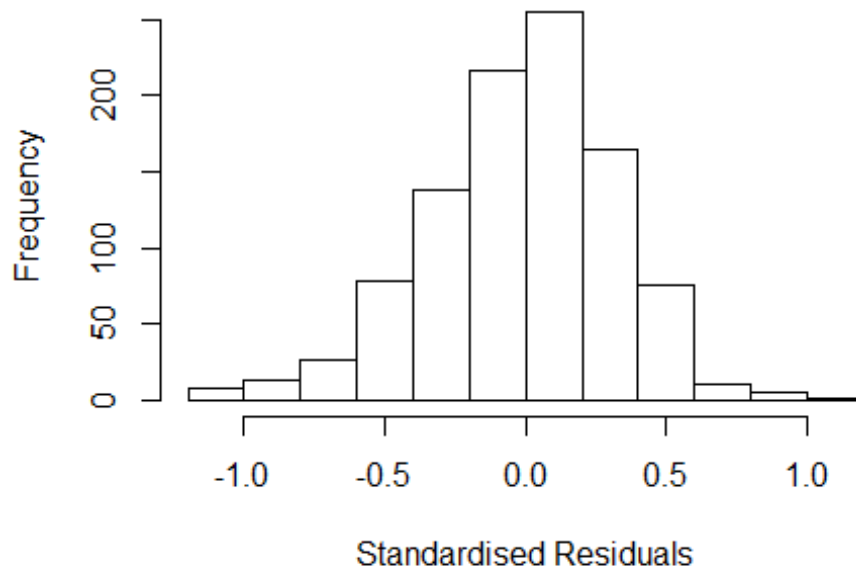


```

## Year2007      -0.00225    0.06233   -0.04    0.971
## Year2008      0.04452    0.06209    0.72    0.473
## Year2009     -0.00244    0.06284   -0.04    0.969
## Year2010     -0.03548    0.07053   -0.50    0.615
## Year2011     -0.09112    0.06628   -1.37    0.170
## Year2012      0.04900    0.06549    0.75    0.455
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.318
## Multiple R-squared:  0.0291, Adjusted R-squared:  0.0122
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 909 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.160  0.868  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      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 1.062 1      1.031
## Year      1.062 16      1.002

```

Residuals from last author



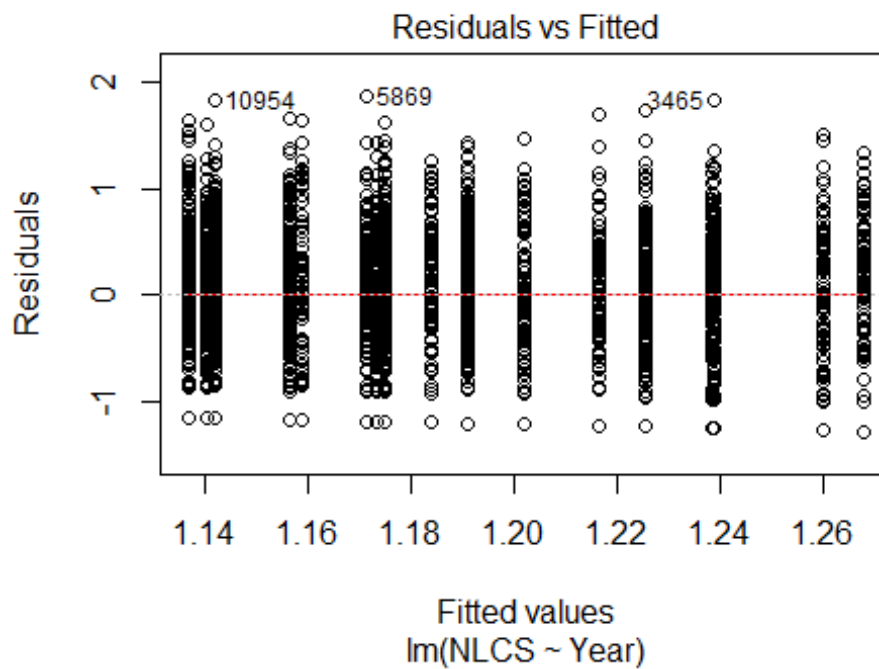
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1381 -0.2223  0.0151  0.2131  1.1138
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06e+00   5.20e-02  20.34  <2e-16 ***
## LastAuthorFemale1 -9.52e-03   2.91e-02  -0.33   0.743
## Year1997         8.03e-02   8.12e-02   0.99   0.323
## Year1998        -7.02e-02   7.49e-02  -0.94   0.349
## Year1999         3.96e-02   6.58e-02   0.60   0.548
## Year2000         9.51e-03   7.84e-02   0.12   0.903
## Year2001        -1.66e-01   7.98e-02  -2.07   0.038 *
## Year2002        -3.64e-02   7.71e-02  -0.47   0.637
## Year2003        -7.82e-02   6.67e-02  -1.17   0.241
## Year2004        -3.70e-02   6.54e-02  -0.57   0.571
## Year2005         4.07e-02   6.98e-02   0.58   0.560
## Year2006        -9.16e-04   6.23e-02  -0.01   0.988
```

```

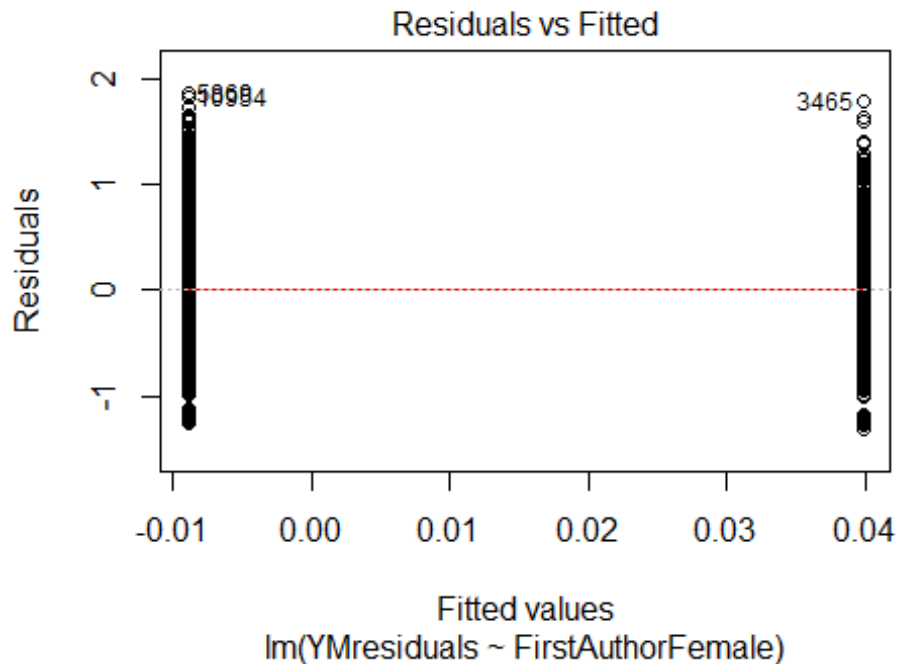
## Year2007          5.61e-03   6.16e-02   0.09   0.927
## Year2008          4.66e-02   6.20e-02   0.75   0.453
## Year2009          5.54e-05   6.28e-02   0.00   0.999
## Year2010         -2.96e-02   7.06e-02  -0.42   0.675
## Year2011         -9.16e-02   6.61e-02  -1.39   0.166
## Year2012          5.17e-02   6.52e-02   0.79   0.428
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.026, Adjusted R-squared:  0.00904
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 912 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.176  0.867  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      1.01e-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 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
##  442  390  417  418  415  396  347  388  451  463  521  523  690  689  700
## 2011 2012
##  774  791
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  113   80  104  110  113   81  101  129  156  189  206  206  284  280  304
## 2011 2012

```

```
## 368 387
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 92 62 85 85 98 70 78 106 130 146 162 165 217 227 239
## 2011 2012
## 297 315
## [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
```

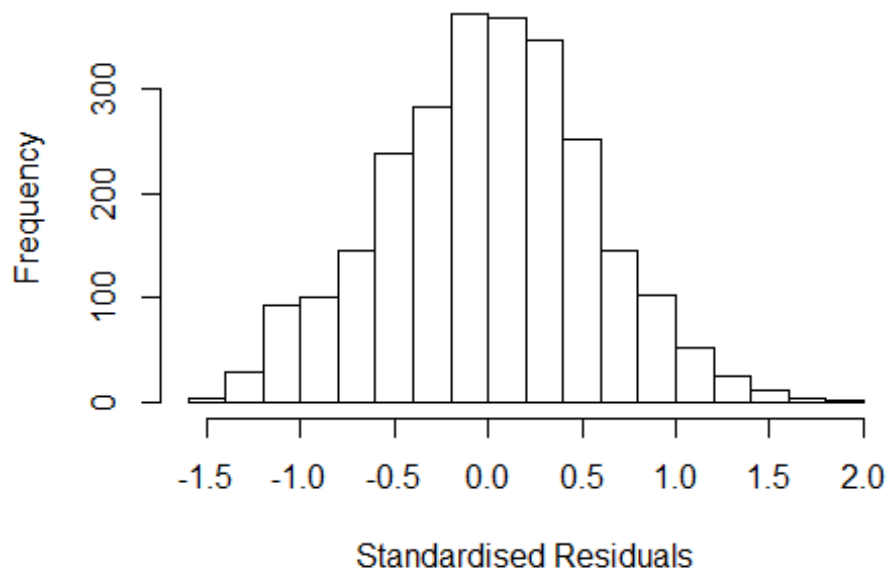


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.13, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 3.53845424312717"
## [1] "Male first author team size 2018 geometric mean: 3.25443419147113"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.13545048453186"
## [1] "Male last author team size 2018 geometric mean: 3.32894863587627"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7400, 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 1.062 1          1.030
## LastAuthorFemale  1.041 1          1.020
## UniqueAuthors    1.154 4          1.018
## Year              1.185 16         1.005
```

Residuals from first and last author and team size



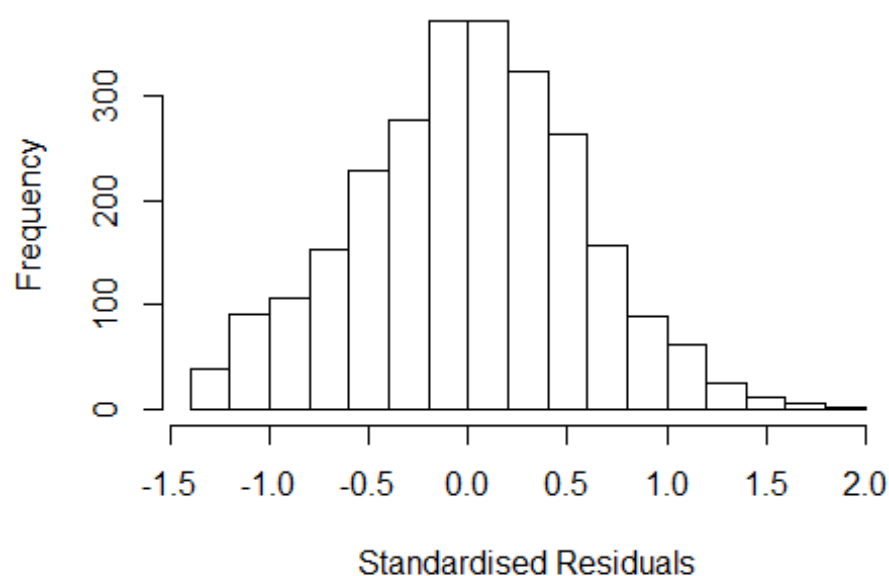
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4868 -0.3714 0.0118 0.3703 1.9302
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08424 0.07039 15.40 < 2e-16 ***
## FirstAuthorFemale1 0.03432 0.02979 1.15 0.24927
## LastAuthorFemale1 -0.00188 0.03532 -0.05 0.95765
## UniqueAuthors2 0.16157 0.04306 3.75 0.00018 ***
## UniqueAuthors3 0.08750 0.04510 1.94 0.05246 .
## UniqueAuthors4 0.21254 0.04697 4.52 6.3e-06 ***
## UniqueAuthors5 0.37551 0.05128 7.32 3.2e-13 ***
## Year1997 0.04649 0.10576 0.44 0.66030
## Year1998 0.05792 0.09396 0.62 0.53769
## Year1999 0.03769 0.09489 0.40 0.69122
```

```

## Year2000      -0.00732      0.09583      -0.08      0.93913
## Year2001      -0.11668      0.10709      -1.09      0.27602
## Year2002      -0.04657      0.09347      -0.50      0.61834
## Year2003      -0.00645      0.08944      -0.07      0.94250
## Year2004      -0.02984      0.08062      -0.37      0.71135
## Year2005      -0.08056      0.07473      -1.08      0.28112
## Year2006      -0.07935      0.07784      -1.02      0.30813
## Year2007      -0.07395      0.07575      -0.98      0.32902
## Year2008      -0.07020      0.07495      -0.94      0.34903
## Year2009      -0.10024      0.07490      -1.34      0.18091
## Year2010      -0.12759      0.07391      -1.73      0.08441 .
## Year2011      -0.07507      0.07210      -1.04      0.29793
## Year2012      -0.13619      0.07173      -1.90      0.05772 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.548
## Multiple R-squared:  0.039, Adjusted R-squared:  0.0307
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 239 weights are ~= 1. The remaining 2335 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.189  0.862  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      3.89e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.033 1      1.017
## LastAuthorFemale  1.033 1      1.016
## Year              1.066 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
## -1.3212 -0.3681  0.0101  0.3751  1.8554
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19246    0.06247   19.09  <2e-16 ***
## FirstAuthorFemale1 0.05866    0.02991    1.96    0.05 *
## LastAuthorFemale1 0.01716    0.03513    0.49    0.63
## Year1997         0.07006    0.10541    0.66    0.51
## Year1998         0.07985    0.09395    0.85    0.40
## Year1999         0.03028    0.09502    0.32    0.75
## Year2000         0.00972    0.09707    0.10    0.92
## Year2001        -0.11603    0.10704   -1.08    0.28
## Year2002        -0.01976    0.09413   -0.21    0.83
## Year2003         0.04322    0.08718    0.50    0.62
## Year2004         0.02088    0.08016    0.26    0.79
## Year2005        -0.03582    0.07325   -0.49    0.62
```

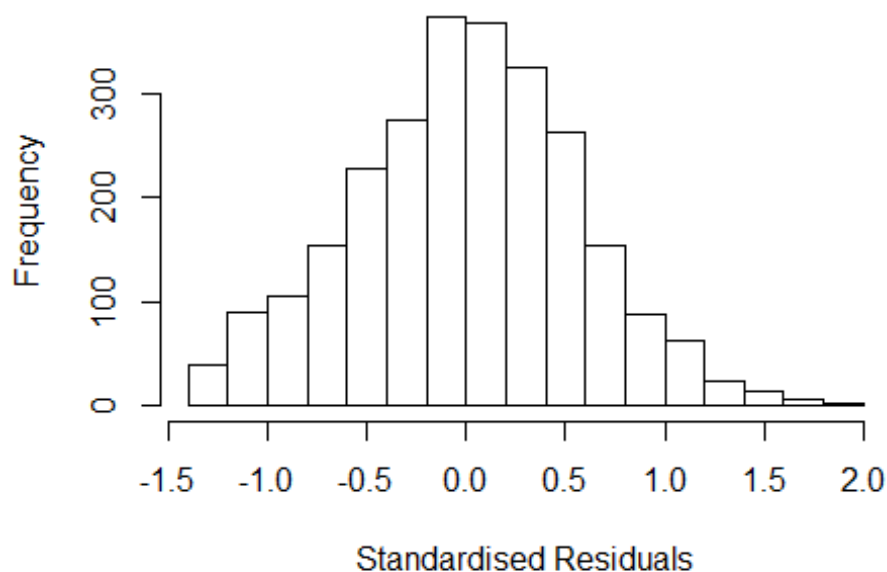


```

## Year2006      -0.05221    0.07663   -0.68    0.50
## Year2007      -0.01985    0.07363   -0.27    0.79
## Year2008      -0.03851    0.07388   -0.52    0.60
## Year2009      -0.04876    0.07369   -0.66    0.51
## Year2010      -0.07039    0.07295   -0.96    0.33
## Year2011      -0.01781    0.07046   -0.25    0.80
## Year2012      -0.07625    0.07001   -1.09    0.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.00739,    Adjusted R-squared:  0.000396
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 233 weights are ~= 1. The remaining 2341 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.242  0.866  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      3.89e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000      0
##      psi            subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.031 1      1.016
## 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
## -1.3255 -0.3680 0.0108 0.3740 1.8541
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19420 0.06214 19.22 <2e-16 ***
## FirstAuthorFemale1 0.05992 0.02993 2.00 0.045 *
## Year1997 0.07138 0.10510 0.68 0.497
## Year1998 0.07942 0.09378 0.85 0.397
## Year1999 0.02979 0.09492 0.31 0.754
## Year2000 0.00952 0.09701 0.10 0.922
## Year2001 -0.11540 0.10699 -1.08 0.281
## Year2002 -0.02061 0.09389 -0.22 0.826
## Year2003 0.04259 0.08707 0.49 0.625
## Year2004 0.02108 0.08009 0.26 0.792
## Year2005 -0.03649 0.07312 -0.50 0.618
## Year2006 -0.05253 0.07651 -0.69 0.492
```

```

## Year2007          -0.02031    0.07355   -0.28    0.782
## Year2008          -0.03817    0.07381   -0.52    0.605
## Year2009          -0.04894    0.07362   -0.66    0.506
## Year2010          -0.07011    0.07289   -0.96    0.336
## Year2011          -0.01734    0.07041   -0.25    0.806
## Year2012          -0.07593    0.06994   -1.09    0.278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.00729,    Adjusted R-squared:  0.000687
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 235 weights are ~= 1. The remaining 2339 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.242  0.866  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      3.89e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.03 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
## -1.28932 -0.37355  0.00839  0.38011  1.84393

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1966    0.0628   19.06 <2e-16 ***
## LastAuthorFemale1 0.0234    0.0351    0.67  0.51
## Year1997          0.0693    0.1057    0.66  0.51
## Year1998          0.0826    0.0943    0.88  0.38
## Year1999          0.0377    0.0947    0.40  0.69
## Year2000          0.0152    0.0974    0.16  0.88
## Year2001         -0.1162    0.1075   -1.08  0.28
## Year2002         -0.0127    0.0942   -0.14  0.89
## Year2003          0.0484    0.0873    0.55  0.58
## Year2004          0.0245    0.0802    0.31  0.76
## Year2005         -0.0284    0.0734   -0.39  0.70
## Year2006         -0.0453    0.0767   -0.59  0.55
## Year2007         -0.0125    0.0737   -0.17  0.86
## Year2008         -0.0343    0.0742   -0.46  0.64
## Year2009         -0.0408    0.0738   -0.55  0.58
## Year2010         -0.0636    0.0731   -0.87  0.38
## Year2011         -0.0126    0.0707   -0.18  0.86
## Year2012         -0.0706    0.0701   -1.01  0.31
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.0058, Adjusted R-squared:  -0.000813
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 242 weights are ~= 1. The remaining 2332 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.249  0.865  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      3.89e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2574"
## [1] ""

```

```

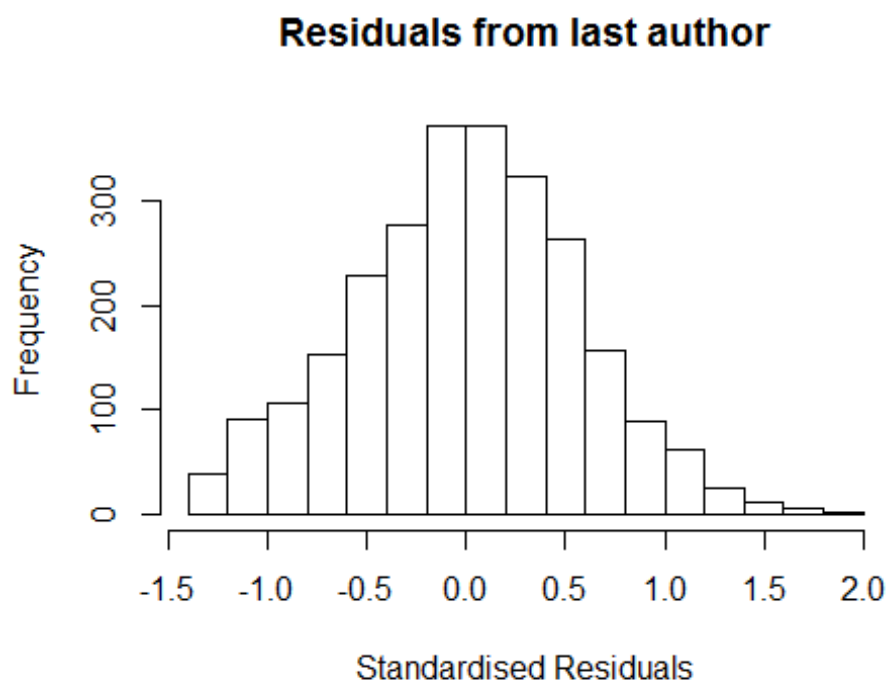
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2501"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 34 24 26 28 32 29 25 26 20 37 41 35 30 4 6
## 2011 2012
## 2 2
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 9 9 7 5 8 9 10 11 8 18 19 18 20 3 2
## 2011 2012
## 1 1
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 6 5 4 6 6 8 11 6 17 18 18 14 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: 4.89897948556636"
## [1] "Male first author team size 2018 geometric mean: 3.09411841113104"

## Warning in wilcox.test.default(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] "Female last author team size 2018 geometric mean: 4.89897948556636"
## [1] "Male last author team size 2018 geometric mean: 3.09411841113104"

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

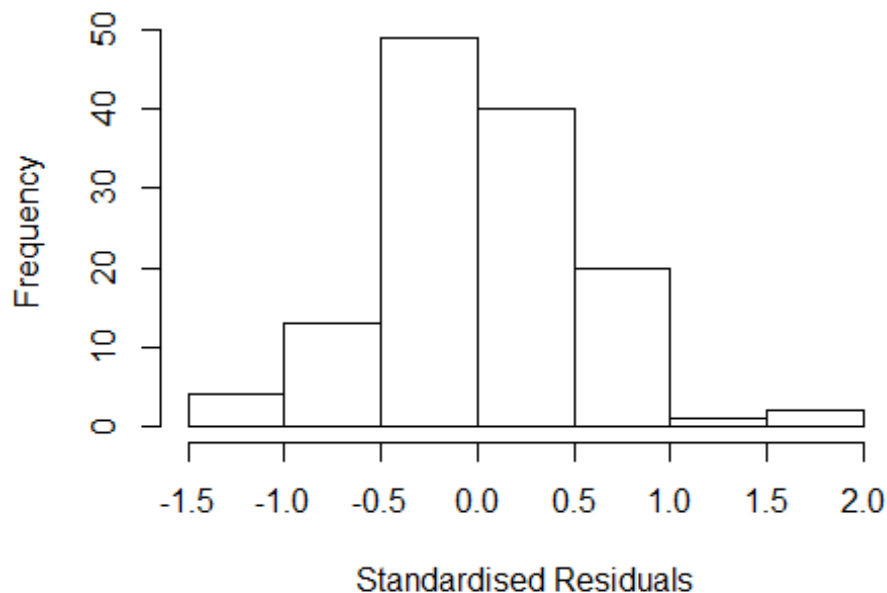
```



```
##
## 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"
##
```

	GVI	F	Df	$GVI^{1/(2 \cdot Df)}$
FirstAuthorFemale	7.452e-01	1		0.8633
LastAuthorFemale	5.606e+01	1		7.4873
UniqueAuthors	1.316e+16	4		103.4935
Year	5.007e+17	15		3.8903

Residuals from first and last author and team size



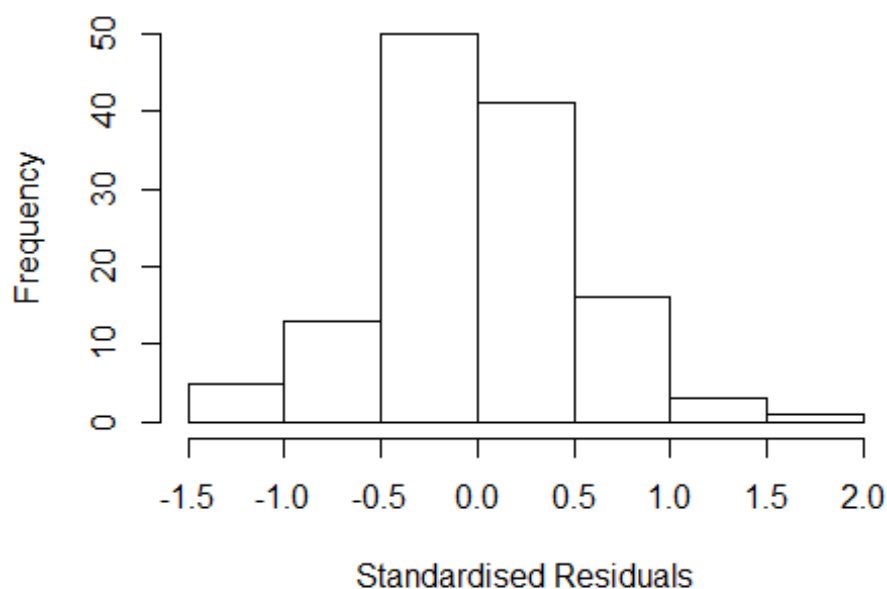
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.36e+00 -3.33e-01 -5.83e-16 3.30e-01 1.81e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0248 0.2229 4.60 1.2e-05 ***
## FirstAuthorFemale1 -0.0592 0.1658 -0.36 0.722
## LastAuthorFemale1 -0.1382 0.2240 -0.62 0.539
## UniqueAuthors2 0.3404 0.1981 1.72 0.089 .
## UniqueAuthors3 0.2349 0.1998 1.18 0.242
## UniqueAuthors4 0.3568 0.2167 1.65 0.103
## UniqueAuthors5 0.6041 0.2535 2.38 0.019 *
## Year1997 -0.1293 0.1711 -0.76 0.451
## Year1998 -0.2371 0.6170 -0.38 0.701
## Year1999 0.0278 0.5113 0.05 0.957
```

```

## Year2000          0.2455      0.2438      1.01      0.316
## Year2001         -0.3565      0.2102     -1.70      0.093 .
## Year2002          0.0575      0.4347      0.13      0.895
## Year2003         -0.2869      0.2183     -1.31      0.192
## Year2004         -0.1872      0.2295     -0.82      0.416
## Year2005         -0.0115      0.1999     -0.06      0.954
## Year2006         -0.2253      0.1841     -1.22      0.224
## Year2007         -0.1675      0.2066     -0.81      0.419
## Year2008         -0.3836      0.1722     -2.23      0.028 *
## Year2009         -0.1723      0.2024     -0.85      0.396
## Year2010          0.1063      0.2955      0.36      0.720
## Year2012         -0.2760      0.2410     -1.15      0.255
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.518
## Multiple R-squared:  0.188, Adjusted R-squared:  0.0282
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 116 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.872  0.958  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      7.75e-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 41.63 1      6.452
## LastAuthorFemale 231.06 1      15.201
## Year              667.81 15      1.242

```


Residuals from first and last author



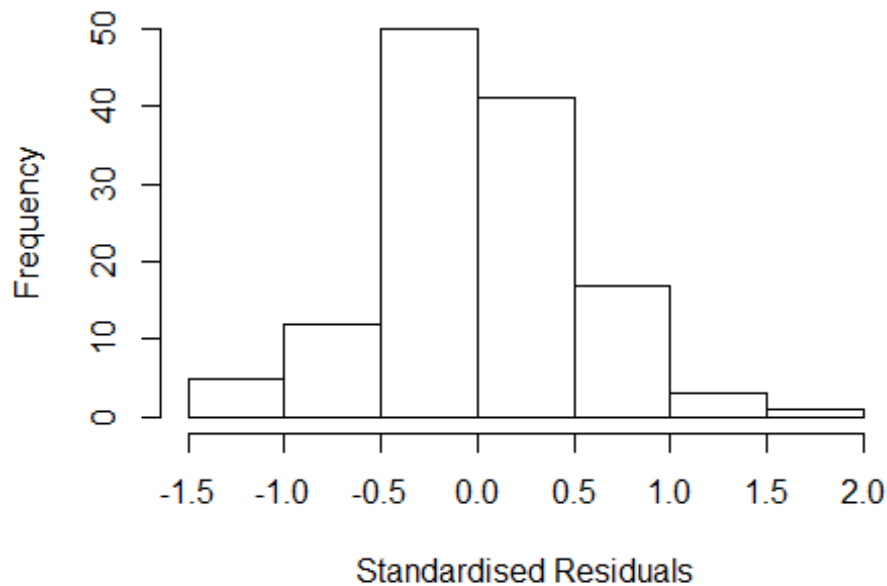
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.479180 -0.329712 -0.000536 0.285612 1.513600
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28688 0.12597 10.22 <2e-16 ***
## FirstAuthorFemale1 0.00164 0.17349 0.01 0.9925
## LastAuthorFemale1 -0.23408 0.26171 -0.89 0.3730
## Year1997 -0.06590 0.15750 -0.42 0.6765
## Year1998 -0.37376 0.46051 -0.81 0.4187
## Year1999 0.10987 0.50046 0.22 0.8266
## Year2000 0.23883 0.26257 0.91 0.3650
## Year2001 -0.40168 0.21366 -1.88 0.0627 .
## Year2002 0.19065 0.50095 0.38 0.7042
## Year2003 -0.31098 0.21337 -1.46 0.1478
## Year2004 -0.17023 0.23208 -0.73 0.4648
## Year2005 0.05450 0.20044 0.27 0.7862
```

```

## Year2006      -0.22341    0.18417   -1.21    0.2277
## Year2007      -0.15516    0.18296   -0.85    0.3982
## Year2008      -0.39805    0.14353   -2.77    0.0065 **
## Year2009      -0.17721    0.15805   -1.12    0.2646
## Year2010       0.54420    0.22060    2.47    0.0152 *
## Year2012      -0.10180    0.22060   -0.46    0.6454
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.499
## Multiple R-squared:  0.141, Adjusted R-squared:  0.00952
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 117 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.338  0.837   0.954   0.897   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      7.75e-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 14.66 1          3.828
## Year              14.66 15          1.094

```

Residuals from first author



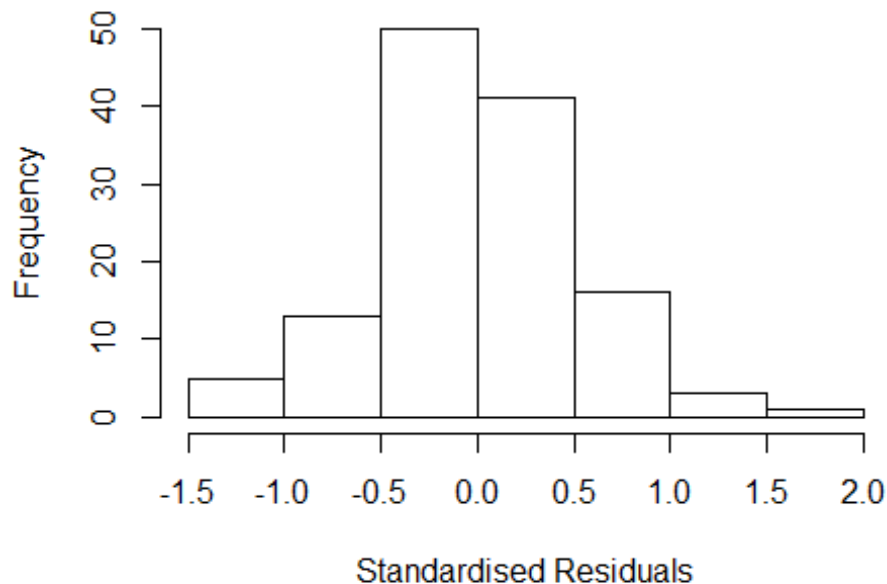
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.42723 -0.31915 -0.00308 0.28125 1.53426
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22223 0.12401 9.86 <2e-16 ***
## FirstAuthorFemale1 -0.04108 0.16364 -0.25 0.8022
## Year1997 -0.00171 0.15610 -0.01 0.9913
## Year1998 -0.30266 0.44931 -0.67 0.5020
## Year1999 0.19659 0.51102 0.38 0.7012
## Year2000 0.28236 0.28399 0.99 0.3223
## Year2001 -0.36792 0.20043 -1.84 0.0691 .
## Year2002 0.24608 0.47182 0.52 0.6030
## Year2003 -0.25216 0.21861 -1.15 0.2512
## Year2004 -0.09782 0.22843 -0.43 0.6693
## Year2005 0.12485 0.18847 0.66 0.5091
## Year2006 -0.17448 0.17951 -0.97 0.3332
```

```

## Year2007          -0.09082    0.17076   -0.53    0.5959
## Year2008          -0.32606    0.13601   -2.40    0.0182 *
## Year2009          -0.09119    0.14405   -0.63    0.5280
## Year2010           0.37477    0.12401    3.02    0.0031 **
## Year2012          -0.27123    0.12401   -2.19    0.0308 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.528
## Multiple R-squared:  0.128, Adjusted R-squared:  0.00353
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 116 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.378  0.867   0.959   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      7.75e-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 44.32  1          6.658
## Year            44.32 15          1.135

```

Residuals from last author



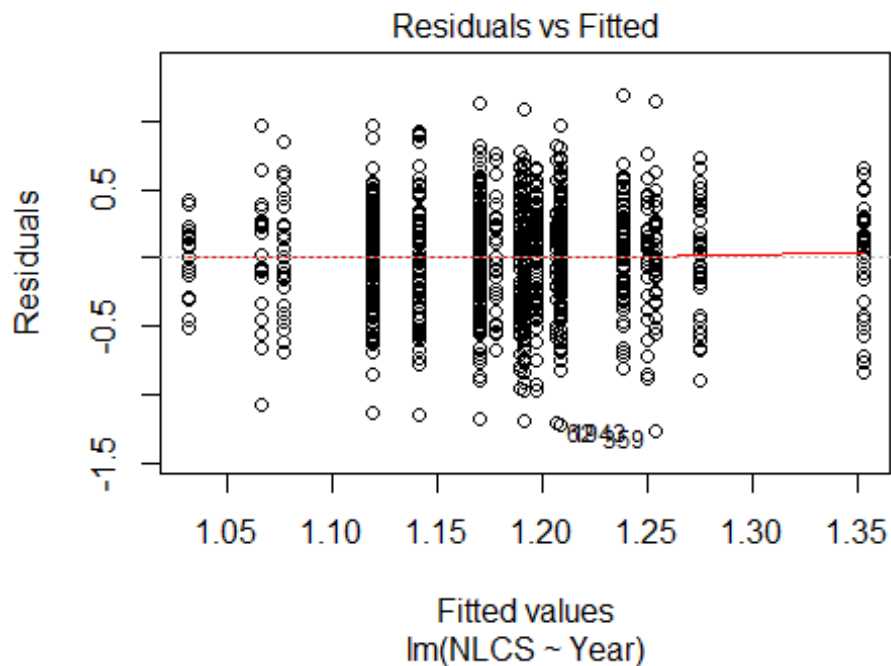
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.477388 -0.329658 -0.000388 0.285080 1.517942
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2872 0.1244 10.34 <2e-16 ***
## LastAuthorFemale1 -0.2333 0.2340 -1.00 0.3209
## Year1997 -0.0662 0.1564 -0.42 0.6731
## Year1998 -0.3737 0.4624 -0.81 0.4206
## Year1999 0.1069 0.4681 0.23 0.8198
## Year2000 0.2385 0.2617 0.91 0.3642
## Year2001 -0.4010 0.1912 -2.10 0.0383 *
## Year2002 0.1902 0.4838 0.39 0.6949
## Year2003 -0.3110 0.2133 -1.46 0.1476
## Year2004 -0.1694 0.2256 -0.75 0.4542
## Year2005 0.0548 0.1994 0.27 0.7842
## Year2006 -0.2231 0.1792 -1.25 0.2157
```

```

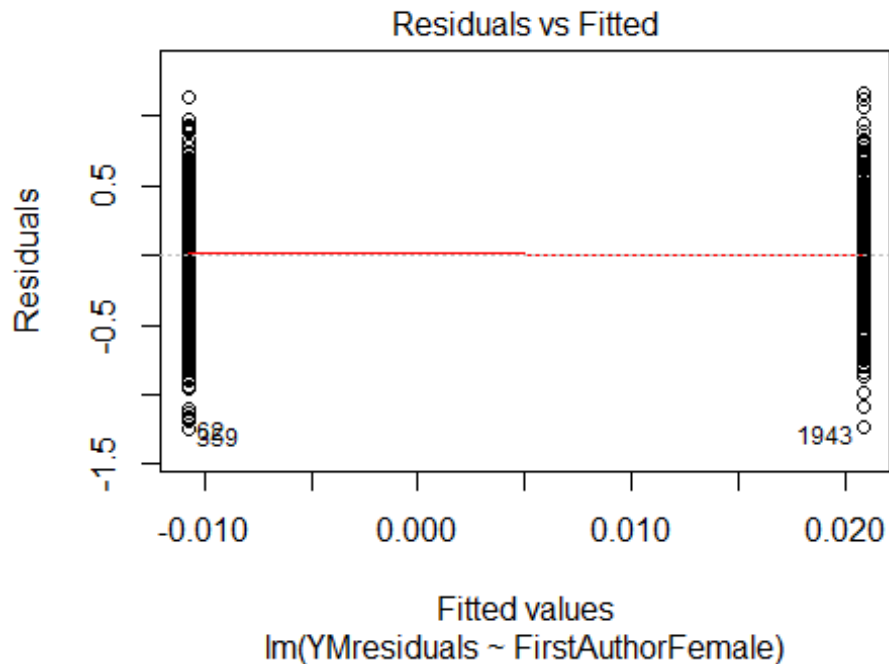
## Year2007          -0.1554      0.1830    -0.85    0.3977
## Year2008          -0.3981      0.1436    -2.77    0.0065 **
## Year2009          -0.1767      0.1439    -1.23    0.2222
## Year2010           0.5431      0.1711     3.17    0.0019 **
## Year2012          -0.1029      0.1711    -0.60    0.5489
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.141, Adjusted R-squared:  0.0185
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 117 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.332  0.837  0.954  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      7.75e-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: 129"
## [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
##   69   60   55   48   77   66   97   78   73  132  123  126  128  168  229
## 2011 2012
##  223  238
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   28   22   21   25   35   29   40   30   34   65   78   73   69  112  150
## 2011 2012
##  144  152

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    24    16    17    20    28    26    29    22    27    53    64    57    58    92   129
## 2011 2012
##   109  120
## [1] "Heteroscedasticity checks, confirming that there are problems with
##      these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6
```

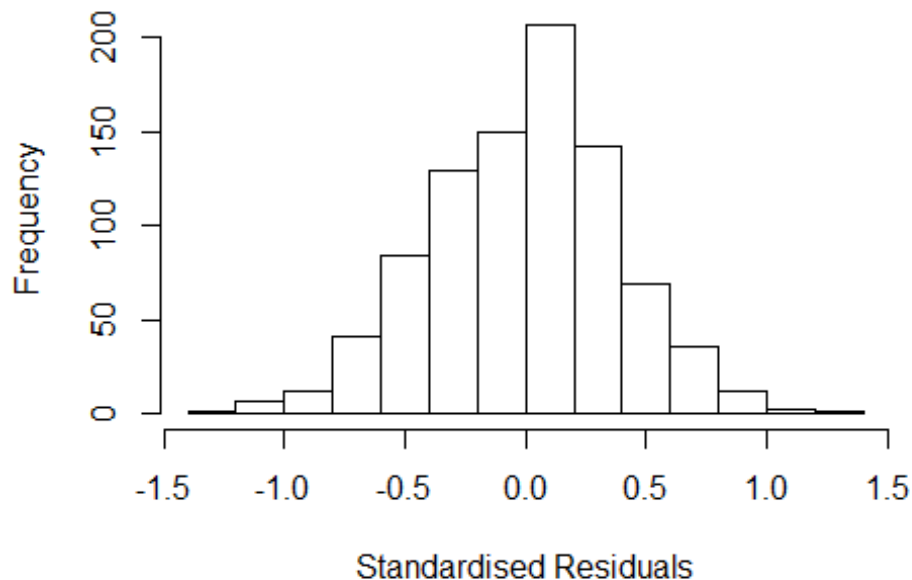


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.19, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 5.0141709144952"
## [1] "Male first author team size 2018 geometric mean: 4.3561082257935"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 710, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.76949760425359"
## [1] "Male last author team size 2018 geometric mean: 4.43714029393916"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 660, 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 1.146 1      1.070
## LastAuthorFemale  1.087 1      1.043
## UniqueAuthors    1.593 4      1.060
## Year              1.888 16      1.020
```


Residuals from first and last author and team size



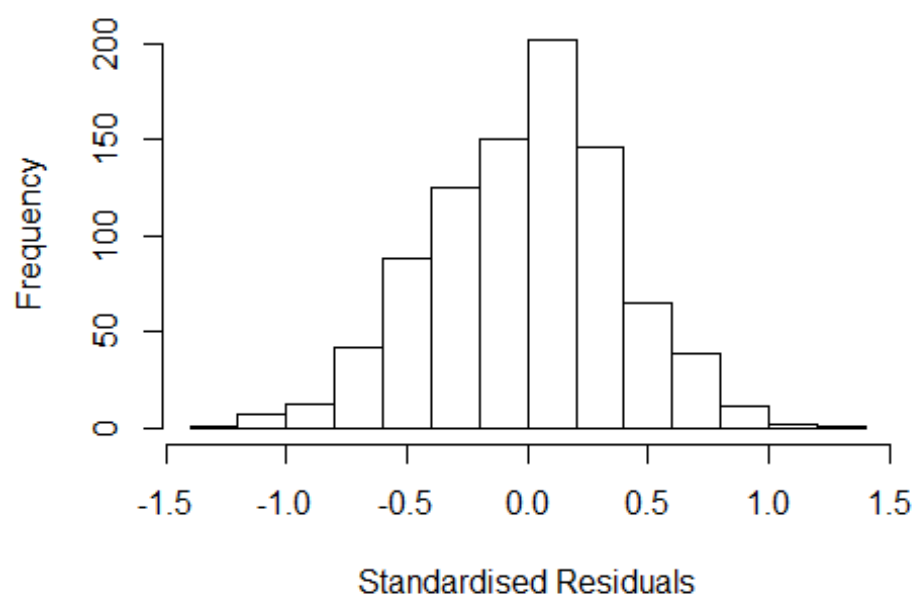
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3411 -0.2694  0.0232  0.2474  1.2465
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.051317   0.118277   8.89   <2e-16 ***
## FirstAuthorFemale1 -0.003348   0.029159  -0.11   0.9086
## LastAuthorFemale1  0.091731   0.029767   3.08   0.0021 **
## UniqueAuthors2     0.130494   0.113678   1.15   0.2513
## UniqueAuthors3     0.140949   0.112536   1.25   0.2107
## UniqueAuthors4     0.215759   0.112931   1.91   0.0564 .
## UniqueAuthors5     0.184766   0.111785   1.65   0.0987 .
## Year1997          -0.138212   0.111190  -1.24   0.2142
## Year1998          -0.218950   0.127948  -1.71   0.0874 .
## Year1999          -0.188132   0.135631  -1.39   0.1658
```

```

## Year2000      0.013278    0.108270    0.12    0.9024
## Year2001      0.037019    0.121574    0.30    0.7608
## Year2002     -0.044984    0.120166   -0.37    0.7082
## Year2003      0.085596    0.142026    0.60    0.5469
## Year2004     -0.013638    0.112470   -0.12    0.9035
## Year2005     -0.000614    0.100200   -0.01    0.9951
## Year2006     -0.088162    0.103770   -0.85    0.3958
## Year2007     -0.015975    0.106283   -0.15    0.8806
## Year2008     -0.015391    0.104336   -0.15    0.8828
## Year2009     -0.106499    0.097448   -1.09    0.2748
## Year2010     -0.053139    0.092209   -0.58    0.5646
## Year2011     -0.037447    0.093711   -0.40    0.6895
## Year2012     -0.090958    0.093651   -0.97    0.3317
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared:  0.0416, Adjusted R-squared:  0.0173
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 816 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.874  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          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 1.137 1          1.066
## LastAuthorFemale  1.084 1          1.041
## Year              1.191 16          1.005

```

Residuals from first and last author



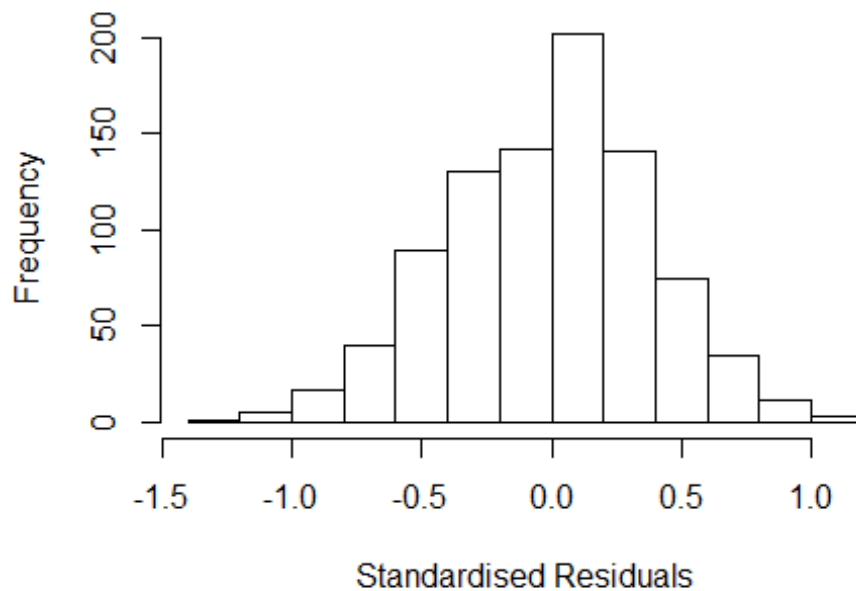
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.313 -0.277  0.023  0.260  1.215
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.187023   0.087517   13.56  <2e-16 ***
## FirstAuthorFemale1 -0.000684   0.029465  -0.02   0.9815
## LastAuthorFemale1  0.097196   0.029793   3.26   0.0011 **
## Year1997        -0.138917   0.111660  -1.24   0.2138
## Year1998        -0.239852   0.129984  -1.85   0.0653 .
## Year1999        -0.152661   0.134516  -1.13   0.2567
## Year2000         0.028821   0.112073   0.26   0.7971
## Year2001         0.055700   0.122393   0.46   0.6492
## Year2002        -0.009034   0.123428  -0.07   0.9417
## Year2003         0.109693   0.142353   0.77   0.4412
## Year2004         0.016056   0.114297   0.14   0.8883
## Year2005         0.018733   0.103791   0.18   0.8568
```

```

## Year2006      -0.067672    0.108561   -0.62    0.5332
## Year2007      0.007496    0.107749    0.07    0.9446
## Year2008      0.018686    0.103980    0.18    0.8574
## Year2009     -0.072099    0.098297   -0.73    0.4635
## Year2010     -0.018853    0.093714   -0.20    0.8406
## Year2011     -0.007600    0.096444   -0.08    0.9372
## Year2012     -0.054253    0.095419   -0.57    0.5698
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.0305, Adjusted R-squared:  0.0105
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 819 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.230  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      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 1.125 1      1.061
## Year              1.125 16      1.004

```

Residuals from first author



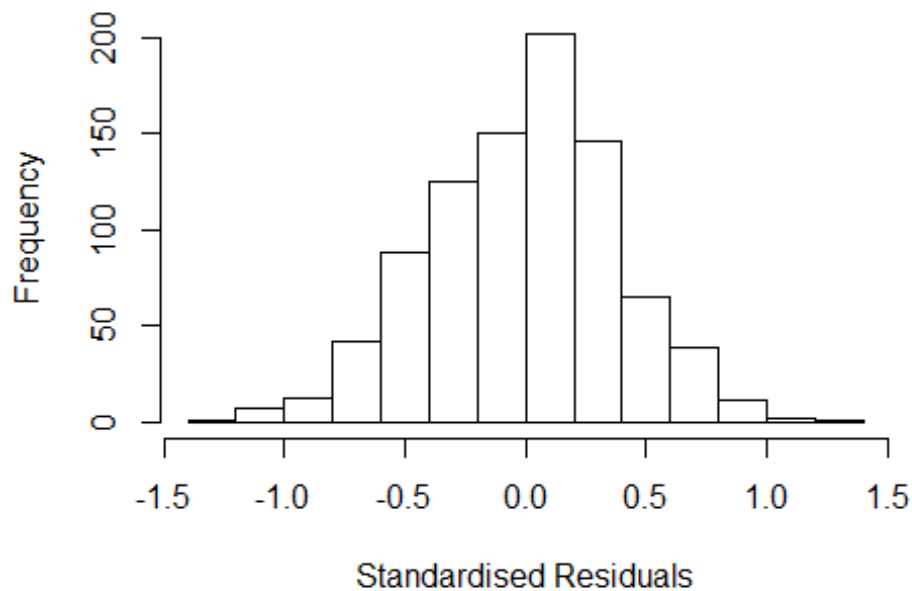
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2503 -0.2706 0.0218 0.2566 1.1885
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.194157 0.087597 13.63 <2e-16 ***
## FirstAuthorFemale1 0.012661 0.029509 0.43 0.67
## Year1997 -0.126634 0.108075 -1.17 0.24
## Year1998 -0.214684 0.132460 -1.62 0.11
## Year1999 -0.158691 0.136873 -1.16 0.25
## Year2000 0.056106 0.112714 0.50 0.62
## Year2001 0.058689 0.123215 0.48 0.63
## Year2002 0.000372 0.125054 0.00 1.00
## Year2003 0.133108 0.140181 0.95 0.34
## Year2004 0.029831 0.115904 0.26 0.80
## Year2005 0.025991 0.103657 0.25 0.80
## Year2006 -0.053442 0.109082 -0.49 0.62
```

```

## Year2007          0.016027    0.107956    0.15      0.88
## Year2008          0.024656    0.104549    0.24      0.81
## Year2009         -0.053978    0.098088   -0.55      0.58
## Year2010         -0.001657    0.093423   -0.02      0.99
## Year2011          0.011035    0.096273    0.11      0.91
## Year2012         -0.041622    0.095035   -0.44      0.66
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.02,   Adjusted R-squared:  0.000905
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 825 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.289  0.870  0.951  0.908  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.069 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
## -1.3127 -0.2771 0.0233 0.2594 1.2145
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18667 0.08777 13.52 <2e-16 ***
## LastAuthorFemale1 0.09699 0.02959 3.28 0.0011 **
## Year1997 -0.13879 0.11139 -1.25 0.2131
## Year1998 -0.24002 0.12973 -1.85 0.0646 .
## Year1999 -0.15221 0.13440 -1.13 0.2577
## Year2000 0.02904 0.11197 0.26 0.7954
## Year2001 0.05564 0.12234 0.45 0.6494
## Year2002 -0.00894 0.12298 -0.07 0.9420
## Year2003 0.10946 0.14165 0.77 0.4399
## Year2004 0.01624 0.11370 0.14 0.8865
## Year2005 0.01862 0.10331 0.18 0.8570
## Year2006 -0.06733 0.10829 -0.62 0.5342
```

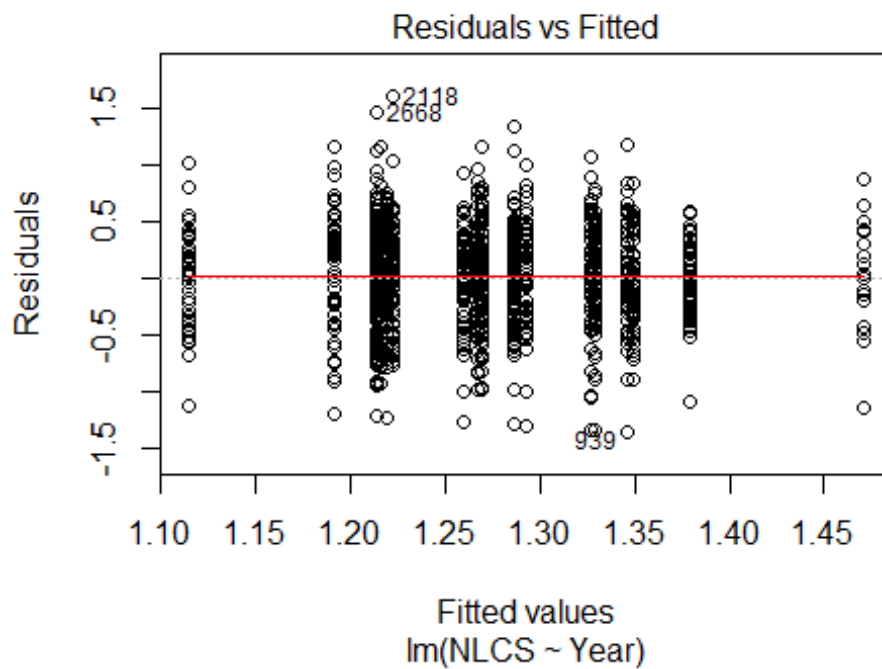
```

## Year2007      0.00730    0.10648    0.07    0.9453
## Year2008      0.01887    0.10303    0.18    0.8547
## Year2009     -0.07207    0.09742   -0.74    0.4596
## Year2010     -0.01866    0.09356   -0.20    0.8420
## Year2011     -0.00743    0.09576   -0.08    0.9382
## Year2012     -0.05421    0.09501   -0.57    0.5684
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.0304, Adjusted R-squared:  0.0115
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 819 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.236  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      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: 891"
## [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
## 138 121 130 127 139 100 112 134 144 170 154 176 155 178 219
## 2011 2012
## 229 222
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 52 48 58 53 36 19 66 56 53 65 73 71 67 97 133
## 2011 2012

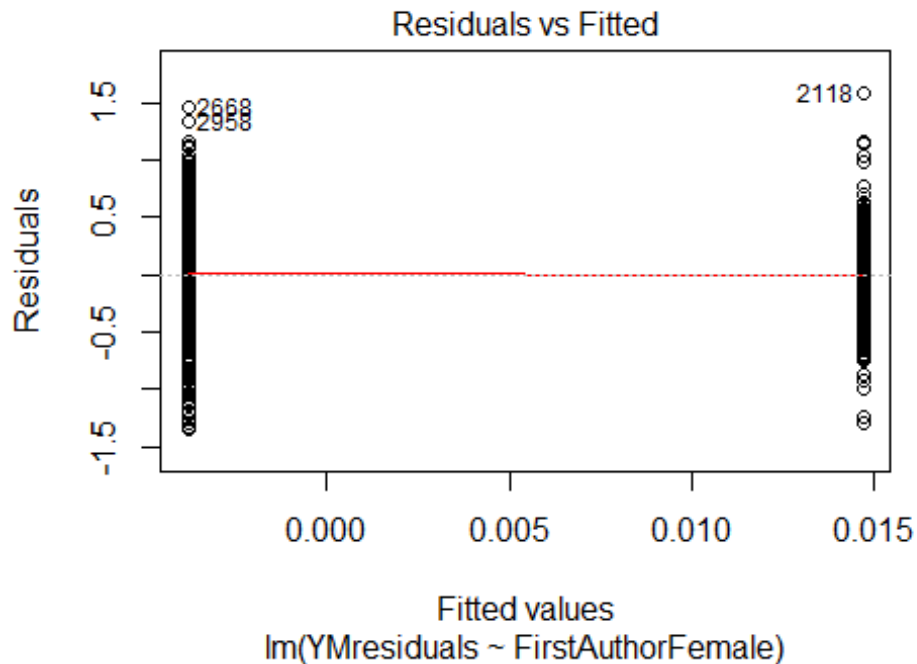
```



```
## 146 130
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 40 40 40 43 29 13 53 44 47 58 53 56 54 84 115
## 2011 2012
## 118 108
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 31, df = 16, p-value = 0.01
```

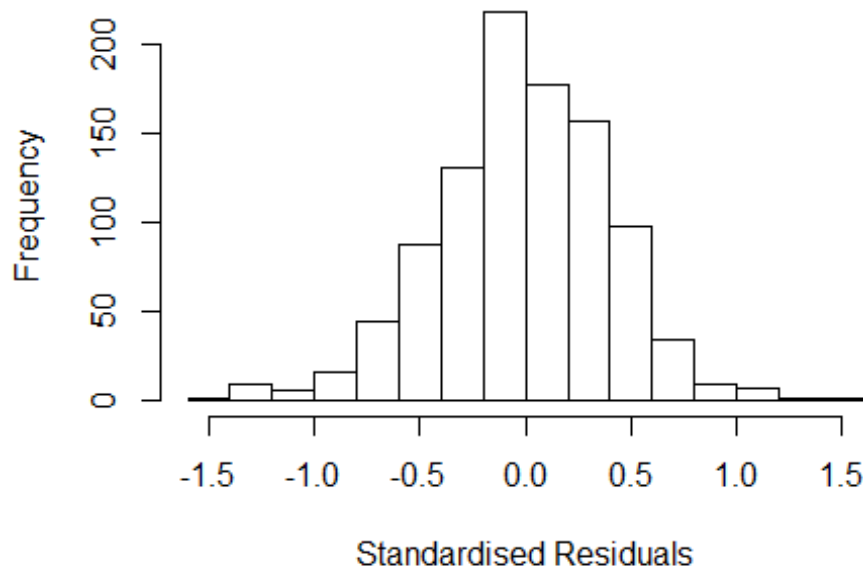


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.6, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 3.29306232378281"
## [1] "Male first author team size 2018 geometric mean: 2.91733797146519"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 470, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.29160805333235"
## [1] "Male last author team size 2018 geometric mean: 2.94570872299801"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 310, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.121 1      1.059
## LastAuthorFemale  1.218 1      1.104
## UniqueAuthors     1.472 4      1.050
## Year              1.631 16     1.015
```

Residuals from first and last author and team size



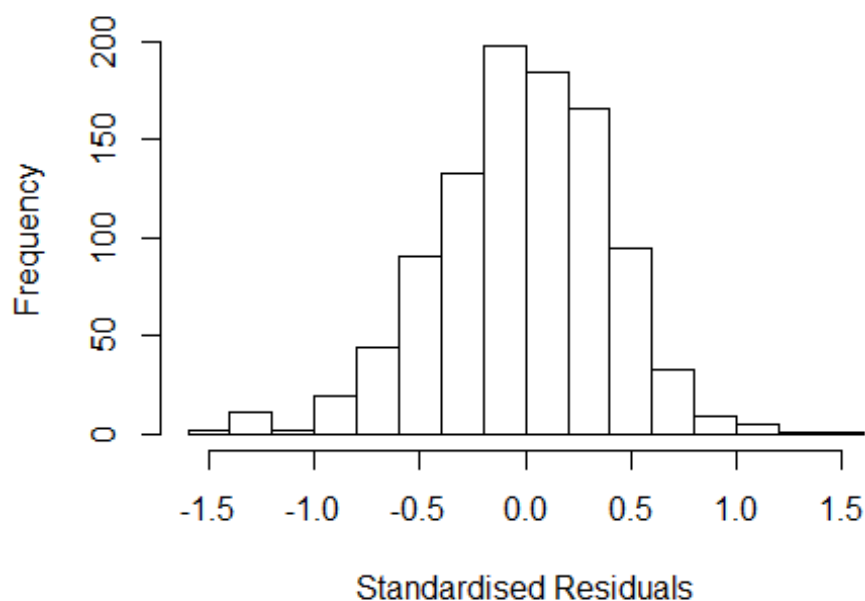
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.44688 -0.24735 -0.00732 0.26399 1.48793
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20448 0.09520 12.65 < 2e-16 ***
## FirstAuthorFemale1 -0.01636 0.03031 -0.54 0.58950
## LastAuthorFemale1 0.05629 0.04017 1.40 0.16149
## UniqueAuthors2 0.21843 0.08873 2.46 0.01400 *
## UniqueAuthors3 0.23804 0.08909 2.67 0.00767 **
## UniqueAuthors4 0.25506 0.09086 2.81 0.00510 **
## UniqueAuthors5 0.25321 0.09062 2.79 0.00531 **
## Year1997 -0.25266 0.07746 -3.26 0.00114 **
## Year1998 -0.18976 0.10311 -1.84 0.06603 .
## Year1999 0.00436 0.07912 0.06 0.95603
```

```

## Year2000      -0.25351    0.08819   -2.87  0.00414 **
## Year2001      -0.09280    0.12393   -0.75  0.45414
## Year2002      -0.12349    0.07018   -1.76  0.07880 .
## Year2003      -0.09351    0.10082   -0.93  0.35394
## Year2004      -0.10018    0.07627   -1.31  0.18936
## Year2005      -0.15058    0.07472   -2.02  0.04416 *
## Year2006      -0.17387    0.07652   -2.27  0.02330 *
## Year2007      -0.03284    0.07011   -0.47  0.63958
## Year2008      -0.25223    0.07300   -3.46  0.00057 ***
## Year2009      -0.11482    0.06851   -1.68  0.09404 .
## Year2010      -0.20222    0.06196   -3.26  0.00114 **
## Year2011      -0.24284    0.06600   -3.68  0.00025 ***
## Year2012      -0.19087    0.05943   -3.21  0.00136 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.0538, Adjusted R-squared:  0.0324
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 103 weights are ~= 1. The remaining 892 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.111  0.866  0.949   0.898  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          1.01e-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 1.105 1          1.051
## LastAuthorFemale  1.157 1          1.076
## Year              1.202 16          1.006

```

Residuals from first and last author



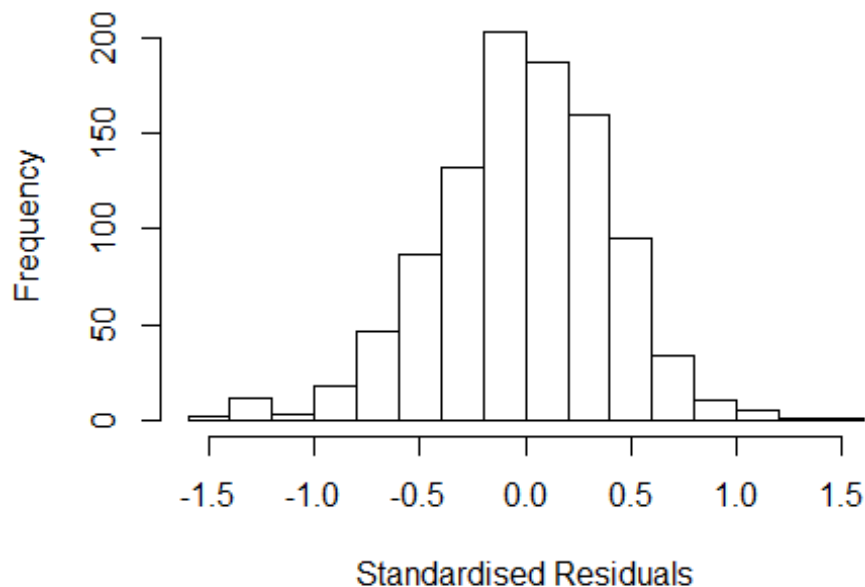
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41258 -0.25070 -0.00214  0.26429  1.47764
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.42435    0.04964   28.69 < 2e-16 ***
## FirstAuthorFemale1 -0.00317    0.03057   -0.10  0.91746
## LastAuthorFemale1  0.06670    0.03977    1.68  0.09382 .
## Year1997        -0.25360    0.07896   -3.21  0.00136 **
## Year1998        -0.21421    0.10727   -2.00  0.04610 *
## Year1999        -0.01177    0.08081   -0.15  0.88418
## Year2000        -0.27081    0.09165   -2.95  0.00320 **
## Year2001        -0.10074    0.13402   -0.75  0.45242
## Year2002        -0.11202    0.07008   -1.60  0.11027
## Year2003        -0.09005    0.09685   -0.93  0.35269
## Year2004        -0.09147    0.07708   -1.19  0.23563
## Year2005        -0.15720    0.07362   -2.14  0.03299 *
```

```

## Year2006          -0.16000    0.07591   -2.11  0.03530 *
## Year2007          -0.03332    0.06874   -0.48  0.62799
## Year2008          -0.24087    0.07313   -3.29  0.00102 **
## Year2009          -0.10847    0.06861   -1.58  0.11422
## Year2010          -0.19241    0.06236   -3.09  0.00209 **
## Year2011          -0.23399    0.06579   -3.56  0.00039 ***
## Year2012          -0.17957    0.05905   -3.04  0.00242 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.0375, Adjusted R-squared:  0.0197
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 888 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.121  0.866  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      1.01e-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 1.071 1      1.035
## Year              1.071 16      1.002

```

Residuals from first author



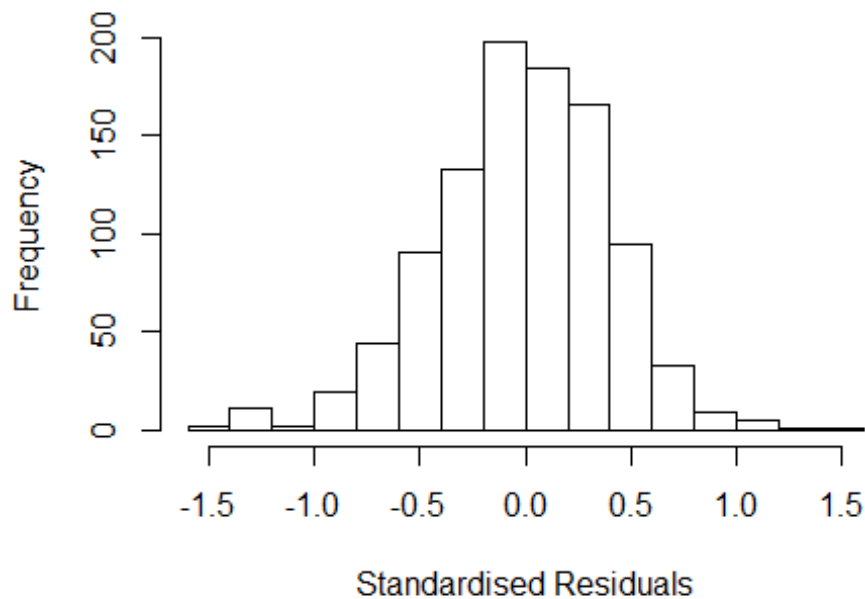
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.41832 -0.25760 -0.00432 0.27547 1.47140
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.42472 0.04925 28.93 < 2e-16 ***
## FirstAuthorFemale1 0.00521 0.03010 0.17 0.86257
## Year1997 -0.25217 0.07877 -3.20 0.00141 **
## Year1998 -0.21193 0.10631 -1.99 0.04649 *
## Year1999 -0.00639 0.08067 -0.08 0.93683
## Year2000 -0.26647 0.09202 -2.90 0.00387 **
## Year2001 -0.09473 0.13237 -0.72 0.47438
## Year2002 -0.11328 0.07015 -1.61 0.10666
## Year2003 -0.08342 0.09743 -0.86 0.39209
## Year2004 -0.08759 0.07708 -1.14 0.25610
## Year2005 -0.15540 0.07276 -2.14 0.03295 *
## Year2006 -0.14555 0.07539 -1.93 0.05381 .
```

```

## Year2007          -0.02776    0.06818   -0.41  0.68393
## Year2008          -0.23607    0.07338   -3.22  0.00134 **
## Year2009          -0.09852    0.06789   -1.45  0.14706
## Year2010          -0.17906    0.06118   -2.93  0.00350 **
## Year2011          -0.22812    0.06540   -3.49  0.00051 ***
## Year2012          -0.17292    0.05880   -2.94  0.00335 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.0345, Adjusted R-squared:  0.0177
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 887 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.129  0.867  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      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 1.118 1          1.057
## Year            1.118 16          1.003

```


Residuals from last author



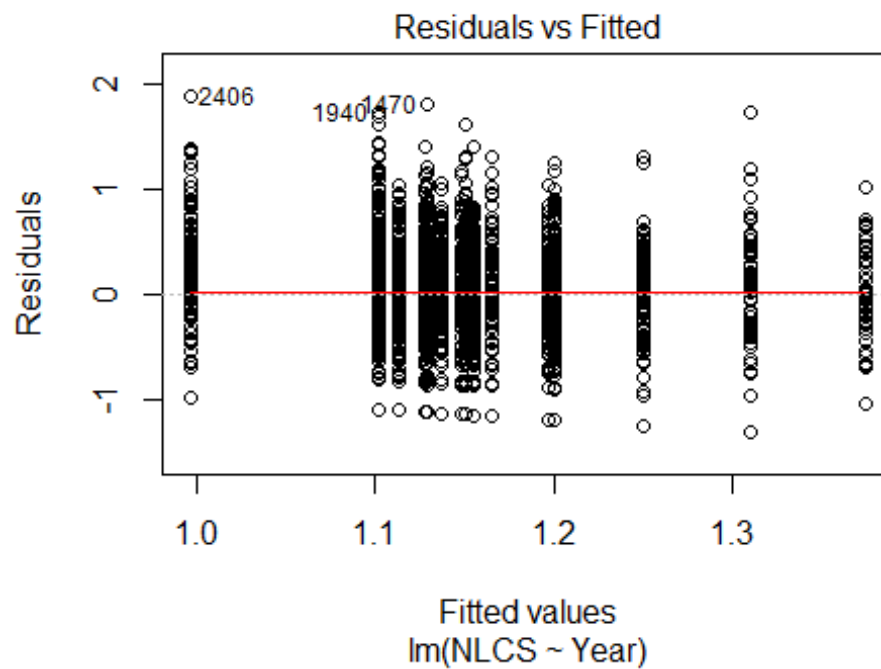
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.41189 -0.25199 -0.00227 0.26521 1.47827
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4239 0.0495 28.76 < 2e-16 ***
## LastAuthorFemale1 0.0661 0.0391 1.69 0.09146 .
## Year1997 -0.2540 0.0785 -3.23 0.00126 **
## Year1998 -0.2142 0.1072 -2.00 0.04603 *
## Year1999 -0.0120 0.0808 -0.15 0.88173
## Year2000 -0.2707 0.0917 -2.95 0.00322 **
## Year2001 -0.1005 0.1337 -0.75 0.45242
## Year2002 -0.1123 0.0701 -1.60 0.10933
## Year2003 -0.0905 0.0966 -0.94 0.34896
## Year2004 -0.0918 0.0769 -1.19 0.23317
## Year2005 -0.1576 0.0734 -2.15 0.03204 *
## Year2006 -0.1600 0.0759 -2.11 0.03536 *
```

```

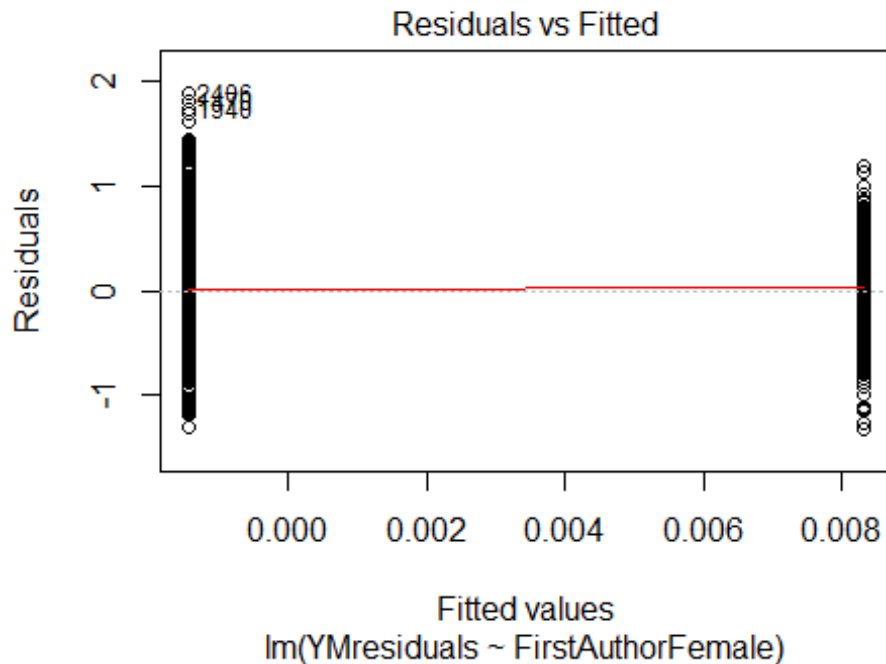
## Year2007          -0.0337      0.0688   -0.49  0.62420
## Year2008          -0.2414      0.0725   -3.33  0.00090 ***
## Year2009          -0.1087      0.0685   -1.59  0.11271
## Year2010          -0.1925      0.0624   -3.09  0.00208 **
## Year2011          -0.2342      0.0656   -3.57  0.00038 ***
## Year2012          -0.1798      0.0590   -3.05  0.00239 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.0374, Adjusted R-squared:  0.0207
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 888 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.123  0.866  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      1.01e-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: 995"
## [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
## 299 238 217 238 352 340 432 291 292 356 368 414 491 568 549
## 2011 2012
## 599 600
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 69 62 85 50 104 98 167 103 95 150 163 172 222 257 267
## 2011 2012

```

```
## 301 289
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 56 45 63 39 81 73 128 78 79 118 123 131 168 197 188
## 2011 2012
## 224 196
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 140, df = 16, p-value <2e-16
```

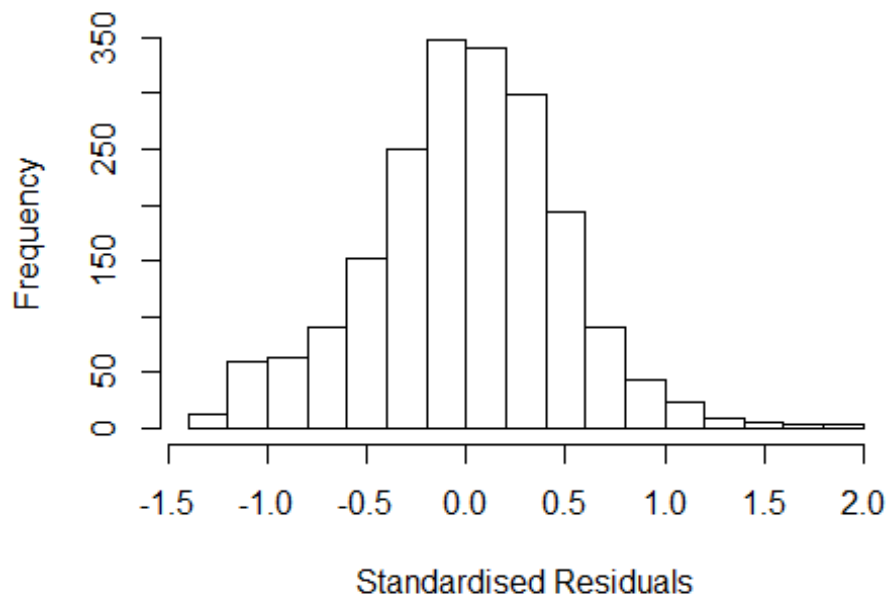


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 1e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 3.21555083141559"
## [1] "Male first author team size 2018 geometric mean: 2.85489118045489"
##
## 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] "Female last author team size 2018 geometric mean: 3.70194331678325"
## [1] "Male last author team size 2018 geometric mean: 2.82042755173234"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.061 1          1.030
## LastAuthorFemale  1.066 1          1.033
## UniqueAuthors     1.280 4          1.031
## Year              1.386 16         1.010
```

Residuals from first and last author and team size



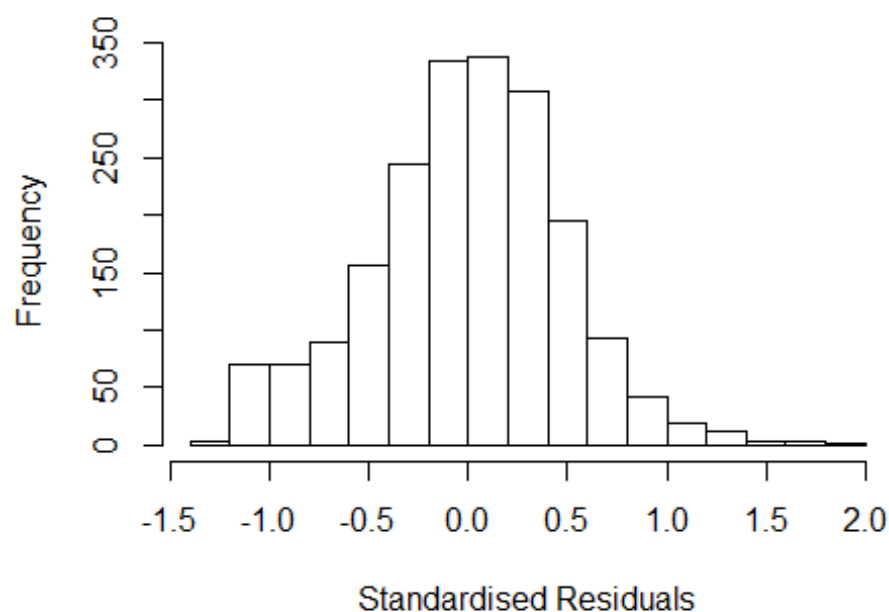
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34715 -0.29328 0.00894 0.31607 1.97598
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2268 0.0901 13.62 < 2e-16 ***
## FirstAuthorFemale1 0.0148 0.0279 0.53 0.59665
## LastAuthorFemale1 0.0162 0.0400 0.41 0.68534
## UniqueAuthors2 0.0658 0.0493 1.33 0.18218
## UniqueAuthors3 0.1203 0.0499 2.41 0.01611 *
## UniqueAuthors4 0.1374 0.0518 2.65 0.00809 **
## UniqueAuthors5 0.2051 0.0520 3.95 8.2e-05 ***
## Year1997 -0.2008 0.1271 -1.58 0.11430
## Year1998 -0.0529 0.1000 -0.53 0.59714
## Year1999 0.0928 0.0988 0.94 0.34773
```

```

## Year2000          -0.2034      0.1037    -1.96   0.05003 .
## Year2001          -0.3202      0.1336    -2.40   0.01664 *
## Year2002          -0.3896      0.1032    -3.78   0.00016 ***
## Year2003          -0.1298      0.1075    -1.21   0.22762
## Year2004          -0.1651      0.0975    -1.69   0.09072 .
## Year2005          -0.1898      0.0912    -2.08   0.03764 *
## Year2006          -0.1467      0.0939    -1.56   0.11814
## Year2007          -0.1349      0.0900    -1.50   0.13409
## Year2008          -0.1889      0.0875    -2.16   0.03093 *
## Year2009          -0.1739      0.0880    -1.98   0.04816 *
## Year2010          -0.2192      0.0878    -2.50   0.01256 *
## Year2011          -0.2146      0.0855    -2.51   0.01213 *
## Year2012          -0.2769      0.0878    -3.15   0.00164 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.452
## Multiple R-squared:  0.0424, Adjusted R-squared:  0.0317
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 177 weights are ~= 1. The remaining 1810 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0165 0.8620 0.9500 0.8920 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          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 1.043 1 1.021
## LastAuthorFemale 1.061 1 1.030
## Year 1.093 16 1.003

```

Residuals from first and last author



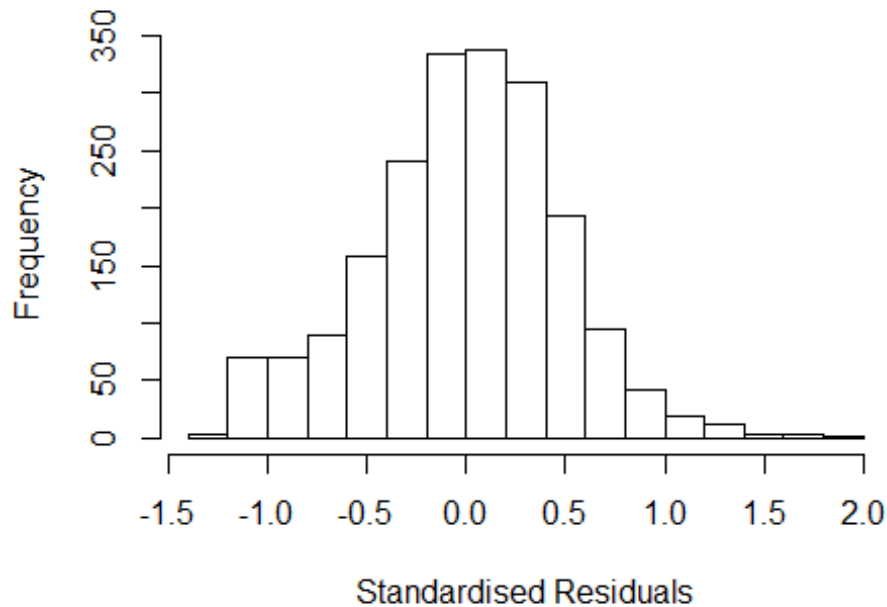
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3440 -0.2945  0.0146  0.3056  1.9276
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3075     0.0835   15.66 < 2e-16 ***
## FirstAuthorFemale1  0.0240     0.0279    0.86  0.38810
## LastAuthorFemale1  0.0124     0.0408    0.30  0.76094
## Year1997          -0.1852     0.1270   -1.46  0.14486
## Year1998          -0.0474     0.1022   -0.46  0.64255
## Year1999           0.1030     0.1032    1.00  0.31835
## Year2000          -0.2017     0.1054   -1.91  0.05583 .
## Year2001          -0.2694     0.1341   -2.01  0.04470 *
## Year2002          -0.3562     0.1057   -3.37  0.00077 ***
## Year2003          -0.1345     0.1097   -1.23  0.22038
## Year2004          -0.1322     0.1001   -1.32  0.18674
## Year2005          -0.1693     0.0933   -1.81  0.06979 .
```

```

## Year2006          -0.1232      0.0960   -1.28   0.19966
## Year2007          -0.1093      0.0919   -1.19   0.23440
## Year2008          -0.1598      0.0896   -1.78   0.07476 .
## Year2009          -0.1399      0.0901   -1.55   0.12043
## Year2010          -0.1862      0.0898   -2.07   0.03829 *
## Year2011          -0.1872      0.0878   -2.13   0.03311 *
## Year2012          -0.2385      0.0898   -2.66   0.00799 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.454
## Multiple R-squared:  0.0279, Adjusted R-squared:  0.0191
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 174 weights are ~= 1. The remaining 1813 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0323 0.8610 0.9520 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.036 1      1.018
## Year      1.036 16      1.001

```


Residuals from first author



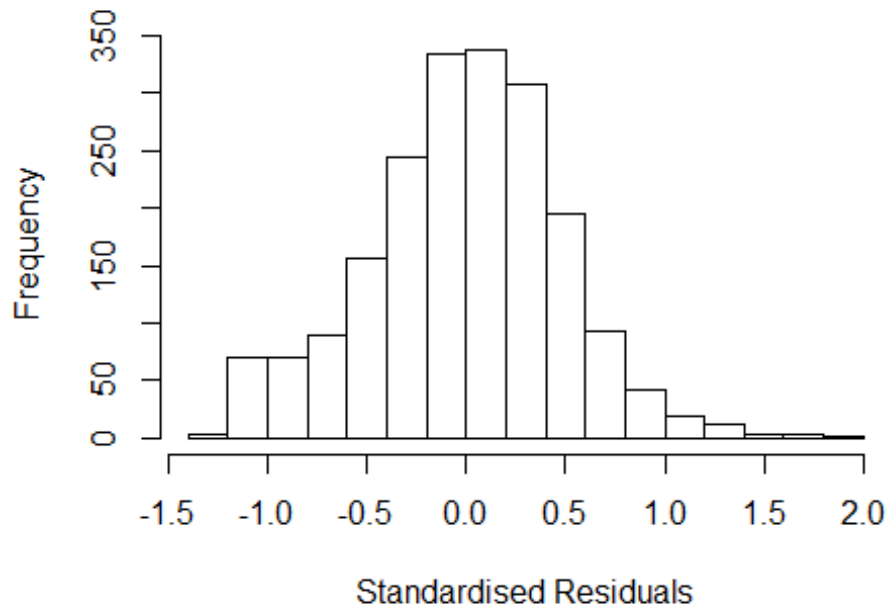
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3334 -0.2956  0.0135  0.3071  1.9277
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3084     0.0836   15.65 < 2e-16 ***
## FirstAuthorFemale1  0.0250     0.0278    0.90  0.36848
## Year1997         -0.1848     0.1270   -1.46  0.14568
## Year1998         -0.0475     0.1023   -0.46  0.64272
## Year1999          0.1029     0.1032    1.00  0.31904
## Year2000         -0.2020     0.1055   -1.91  0.05570 .
## Year2001         -0.2699     0.1342   -2.01  0.04448 *
## Year2002         -0.3571     0.1055   -3.38  0.00073 ***
## Year2003         -0.1346     0.1098   -1.22  0.22072
## Year2004         -0.1331     0.1001   -1.33  0.18385
## Year2005         -0.1691     0.0934   -1.81  0.07053 .
## Year2006         -0.1228     0.0961   -1.28  0.20141
```

```

## Year2007          -0.1091      0.0920   -1.19  0.23567
## Year2008          -0.1597      0.0898   -1.78  0.07538 .
## Year2009          -0.1395      0.0901   -1.55  0.12184
## Year2010          -0.1858      0.0899   -2.07  0.03888 *
## Year2011          -0.1869      0.0879   -2.13  0.03354 *
## Year2012          -0.2381      0.0899   -2.65  0.00816 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.454
## Multiple R-squared:  0.0279, Adjusted R-squared:  0.0195
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 172 weights are ~= 1. The remaining 1815 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0325 0.8600 0.9520 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.055 1      1.027
## Year      1.055 16      1.002

```

Residuals from last author



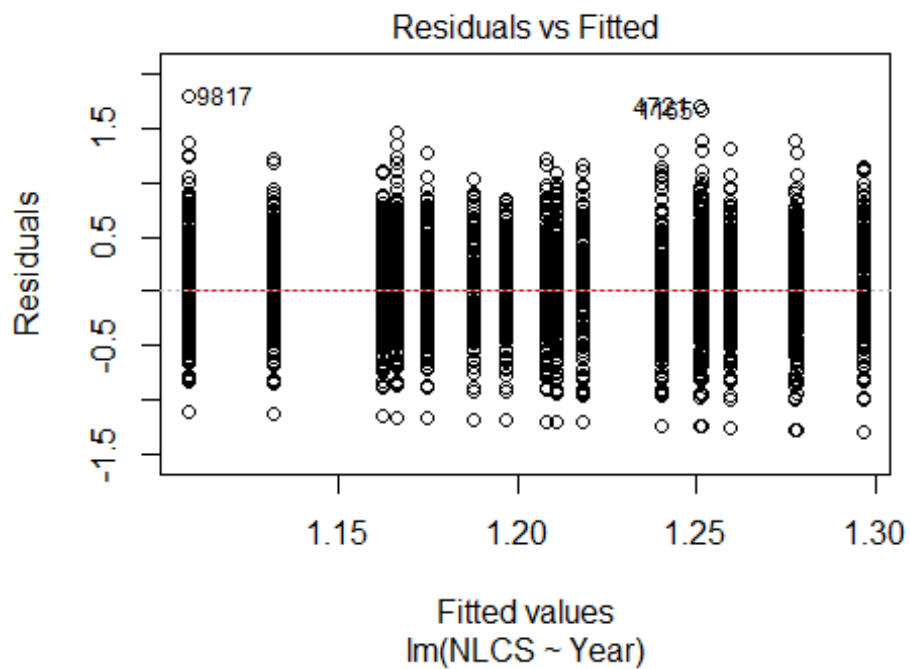
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3239 -0.2949 0.0179 0.3053 1.9239
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3085 0.0835 15.67 < 2e-16 ***
## LastAuthorFemale1 0.0154 0.0406 0.38 0.70396
## Year1997 -0.1859 0.1270 -1.46 0.14340
## Year1998 -0.0465 0.1022 -0.46 0.64909
## Year1999 0.1071 0.1030 1.04 0.29871
## Year2000 -0.1994 0.1055 -1.89 0.05890 .
## Year2001 -0.2687 0.1342 -2.00 0.04544 *
## Year2002 -0.3534 0.1056 -3.35 0.00084 ***
## Year2003 -0.1328 0.1097 -1.21 0.22616
## Year2004 -0.1286 0.1000 -1.29 0.19865
## Year2005 -0.1663 0.0931 -1.79 0.07429 .
## Year2006 -0.1208 0.0959 -1.26 0.20777
```

```

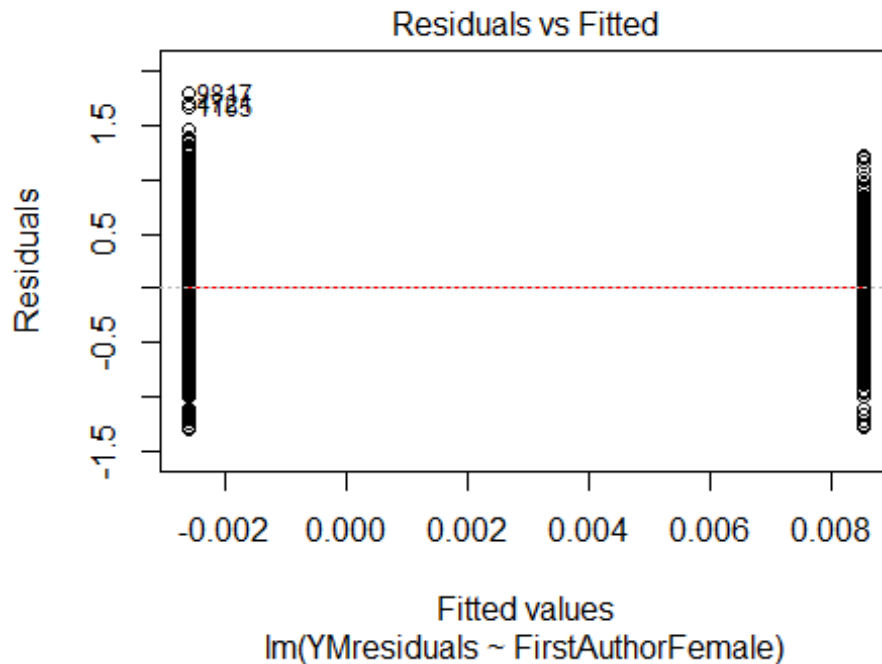
## Year2007          -0.1047      0.0915   -1.14   0.25285
## Year2008          -0.1567      0.0895   -1.75   0.08001 .
## Year2009          -0.1371      0.0899   -1.52   0.12767
## Year2010          -0.1831      0.0896   -2.04   0.04113 *
## Year2011          -0.1846      0.0877   -2.11   0.03538 *
## Year2012          -0.2364      0.0897   -2.63   0.00849 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.454
## Multiple R-squared:  0.0276, Adjusted R-squared:  0.0192
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 167 weights are ~ = 1. The remaining 1820 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.033  0.860  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      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: 1987"
## [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
##  490  512  521  438  489  405  431  432  451  520  560  524  542  537  582
## 2011 2012
##  621  627
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  184  168  213  145  160  125  189  194  194  202  234  234  251  250  287
## 2011 2012

```

```
## 325 334
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 138 140 158 123 116 99 157 157 159 166 167 179 194 198 215
## 2011 2012
## 258 256
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 40, df = 16, p-value = 9e-04
```

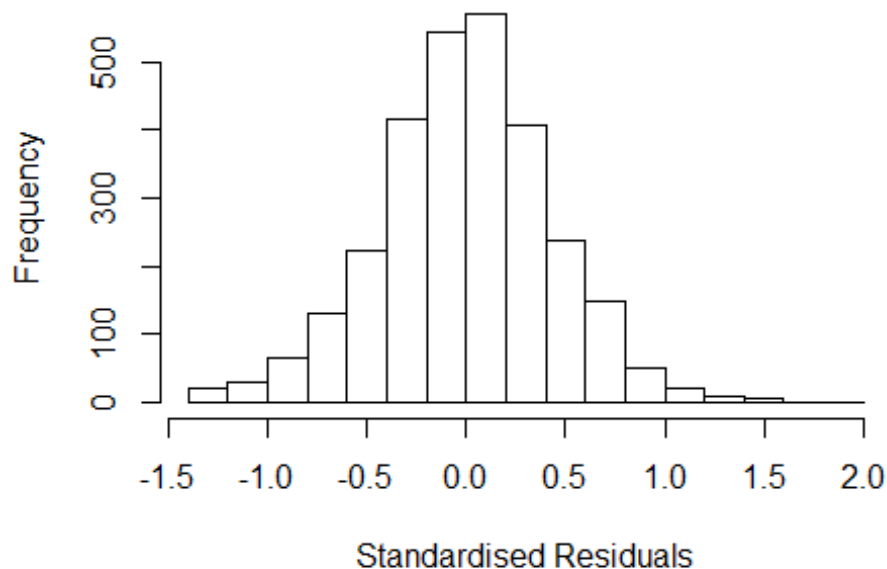


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8, df = 1, p-value = 0.005
```



```
## [1] "Female first author team size 2018 geometric mean: 3.26087829827548"
## [1] "Male first author team size 2018 geometric mean: 3.24725888581381"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 980, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.37788573568276"
## [1] "Male last author team size 2018 geometric mean: 3.22856664702059"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 650, 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 1.056 1          1.028
## LastAuthorFemale  1.032 1          1.016
## UniqueAuthors    1.217 4          1.025
## Year              1.251 16         1.007
```

Residuals from first and last author and team size



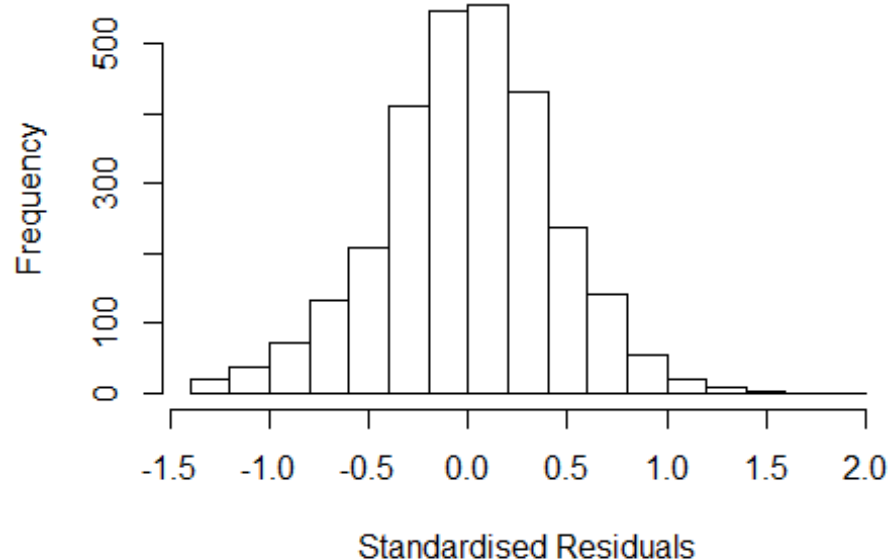
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37303 -0.26387  0.00424  0.27420  1.85577
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13862    0.06989   16.29 < 2e-16 ***
## FirstAuthorFemale1  0.00457    0.01851    0.25  0.8048
## LastAuthorFemale1  0.01750    0.02554    0.69  0.4933
## UniqueAuthors2     0.17395    0.05424    3.21  0.0014 **
## UniqueAuthors3     0.16500    0.05450    3.03  0.0025 **
## UniqueAuthors4     0.21691    0.05558    3.90 9.7e-05 ***
## UniqueAuthors5     0.26939    0.05627    4.79 1.8e-06 ***
## Year1997          -0.08783    0.06117   -1.44  0.1511
## Year1998          -0.07128    0.05909   -1.21  0.2278
## Year1999          -0.03349    0.06156   -0.54  0.5864
```

```

## Year2000      -0.12043    0.05909    -2.04    0.0417 *
## Year2001      -0.03038    0.06016    -0.50    0.6136
## Year2002      -0.02982    0.05391    -0.55    0.5802
## Year2003      -0.11098    0.05296    -2.10    0.0362 *
## Year2004      -0.06323    0.05524    -1.14    0.2525
## Year2005      -0.13010    0.05411    -2.40    0.0163 *
## Year2006      -0.07572    0.05455    -1.39    0.1652
## Year2007      -0.12011    0.05515    -2.18    0.0295 *
## Year2008      -0.13804    0.05502    -2.51    0.0122 *
## Year2009      -0.13537    0.05384    -2.51    0.0120 *
## Year2010      -0.21744    0.05434    -4.00    6.5e-05 ***
## Year2011      -0.25340    0.05397    -4.70    2.8e-06 ***
## Year2012      -0.19656    0.05278    -3.72    0.0002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.0415, Adjusted R-squared:  0.0341
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 248 weights are ~= 1. The remaining 2632 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.002  0.864  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      3.47e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.041 1      1.020
## LastAuthorFemale  1.023 1      1.012
## Year              1.058 16      1.002

```


Residuals from first and last author



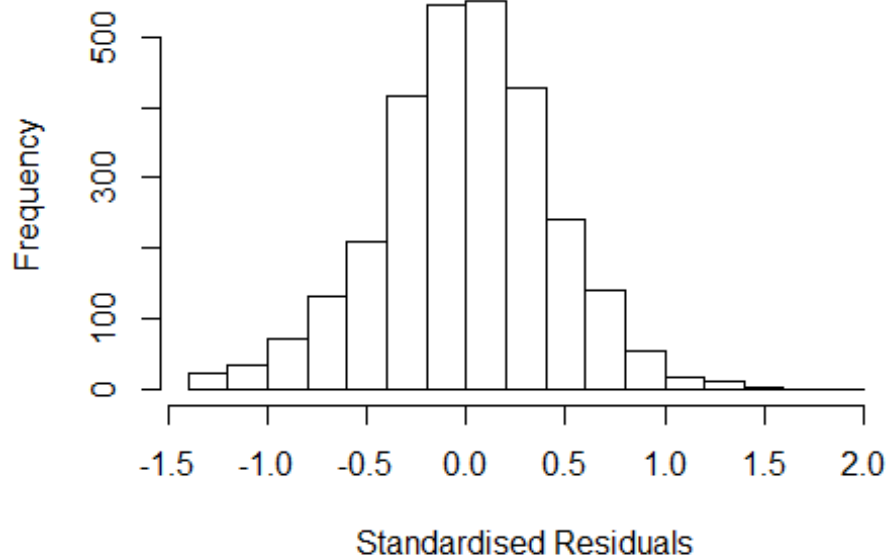
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34311 -0.26546 0.00349 0.27351 1.83263
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3163 0.0456 28.84 < 2e-16 ***
## FirstAuthorFemale1 0.0154 0.0185 0.84 0.40356
## LastAuthorFemale1 0.0268 0.0255 1.05 0.29251
## Year1997 -0.1000 0.0610 -1.64 0.10129
## Year1998 -0.0729 0.0595 -1.23 0.22052
## Year1999 -0.0483 0.0624 -0.77 0.43869
## Year2000 -0.1332 0.0601 -2.22 0.02675 *
## Year2001 -0.0248 0.0607 -0.41 0.68361
## Year2002 -0.0201 0.0542 -0.37 0.71086
## Year2003 -0.1163 0.0532 -2.19 0.02886 *
## Year2004 -0.0579 0.0561 -1.03 0.30186
## Year2005 -0.1233 0.0545 -2.26 0.02384 *
```

```

## Year2006          -0.0712      0.0548   -1.30   0.19432
## Year2007          -0.1146      0.0551   -2.08   0.03747 *
## Year2008          -0.1355      0.0555   -2.44   0.01477 *
## Year2009          -0.1232      0.0543   -2.27   0.02333 *
## Year2010          -0.2095      0.0548   -3.82   0.00013 ***
## Year2011          -0.2429      0.0548   -4.43   9.8e-06 ***
## Year2012          -0.1792      0.0531   -3.38   0.00074 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.408
## Multiple R-squared:  0.0246, Adjusted R-squared:  0.0185
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 260 weights are ~= 1. The remaining 2620 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0062 0.8620 0.9510 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.47e-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 1.038 1      1.019
## Year      1.038 16      1.001

```

Residuals from first author



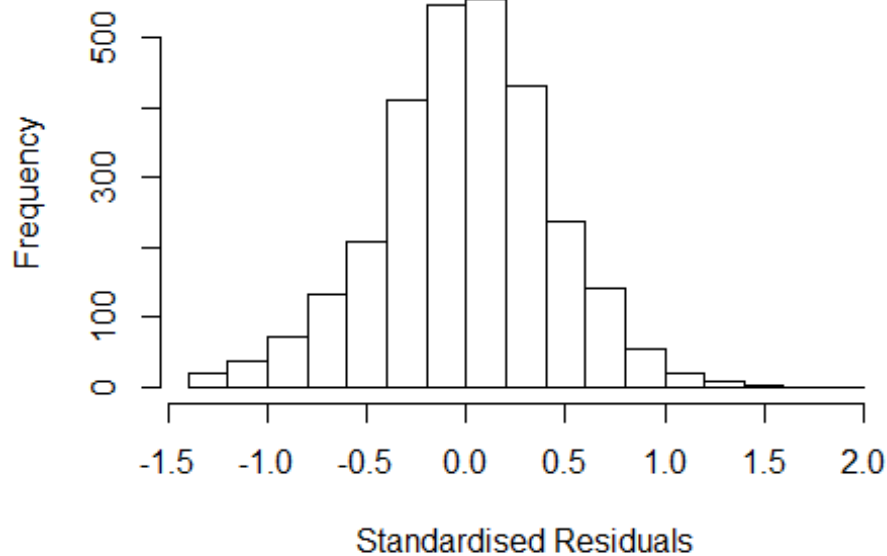
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3172 -0.2648 0.0039 0.2739 1.8290
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3172 0.0456 28.91 < 2e-16 ***
## FirstAuthorFemale1 0.0166 0.0185 0.90 0.36797
## Year1997 -0.0985 0.0609 -1.62 0.10595
## Year1998 -0.0722 0.0594 -1.22 0.22435
## Year1999 -0.0473 0.0623 -0.76 0.44763
## Year2000 -0.1317 0.0600 -2.20 0.02822 *
## Year2001 -0.0219 0.0606 -0.36 0.71792
## Year2002 -0.0196 0.0541 -0.36 0.71670
## Year2003 -0.1149 0.0531 -2.16 0.03058 *
## Year2004 -0.0571 0.0560 -1.02 0.30813
## Year2005 -0.1219 0.0545 -2.24 0.02531 *
## Year2006 -0.0691 0.0548 -1.26 0.20731
```

```

## Year2007          -0.1120      0.0549   -2.04  0.04142 *
## Year2008          -0.1324      0.0554   -2.39  0.01691 *
## Year2009          -0.1211      0.0542   -2.23  0.02553 *
## Year2010          -0.2072      0.0547   -3.79  0.00016 ***
## Year2011          -0.2402      0.0547   -4.39  1.2e-05 ***
## Year2012          -0.1770      0.0530   -3.34  0.00084 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.408
## Multiple R-squared:  0.0242, Adjusted R-squared:  0.0184
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 255 weights are ~= 1. The remaining 2625 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0073 0.8620 0.9520 0.8950 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.47e-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 1.02 1      1.010
## Year      1.02 16      1.001

```

Residuals from last author



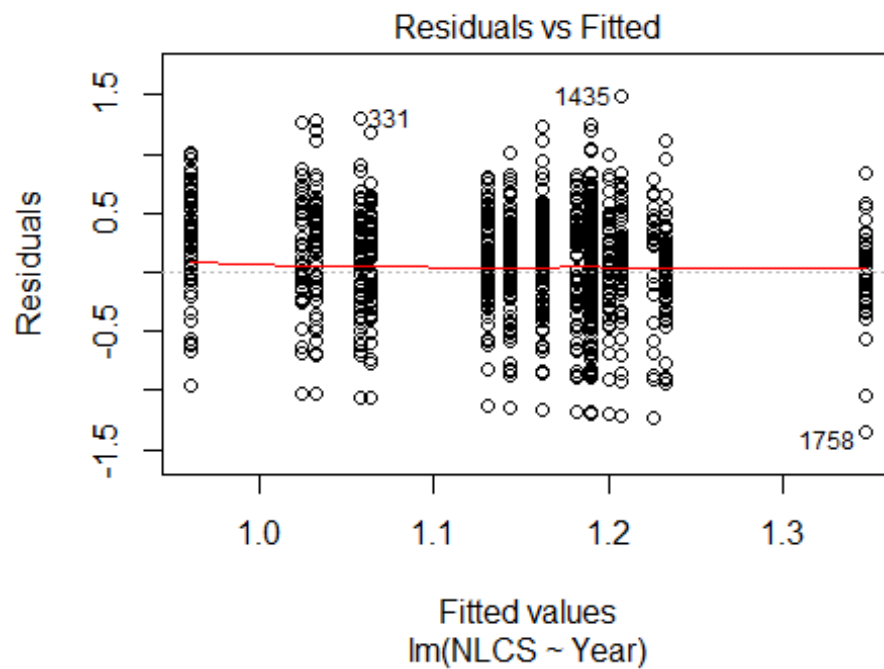
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34834 -0.26608 0.00542 0.27374 1.82834
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3202 0.0452 29.20 < 2e-16 ***
## LastAuthorFemale1 0.0281 0.0255 1.10 0.26981
## Year1997 -0.1013 0.0609 -1.66 0.09649 .
## Year1998 -0.0748 0.0594 -1.26 0.20798
## Year1999 -0.0492 0.0624 -0.79 0.43053
## Year2000 -0.1348 0.0600 -2.24 0.02490 *
## Year2001 -0.0260 0.0605 -0.43 0.66769
## Year2002 -0.0211 0.0541 -0.39 0.69636
## Year2003 -0.1166 0.0532 -2.19 0.02840 *
## Year2004 -0.0578 0.0561 -1.03 0.30279
## Year2005 -0.1230 0.0546 -2.25 0.02429 *
## Year2006 -0.0717 0.0548 -1.31 0.19067
```

```

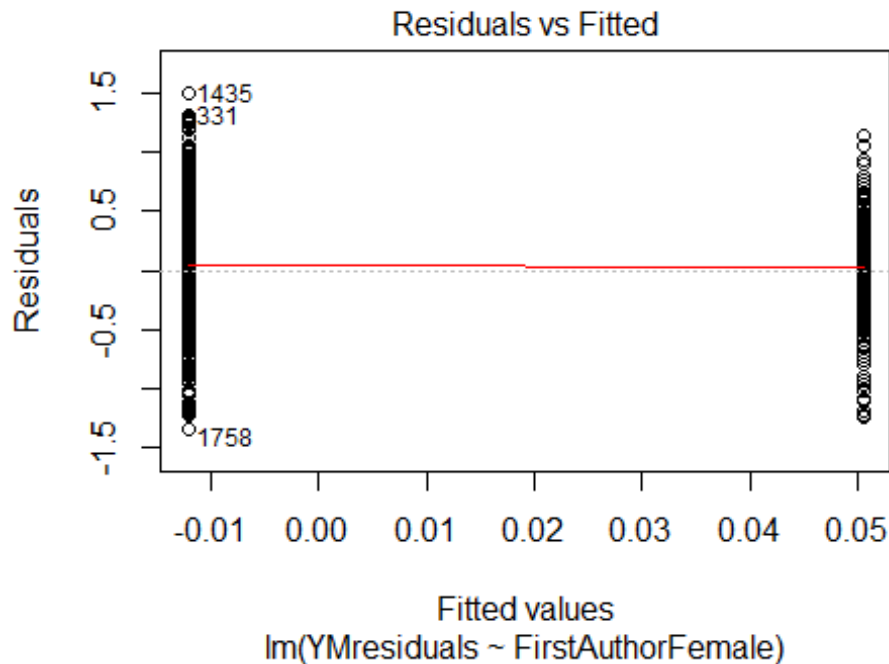
## Year2007          -0.1144      0.0550   -2.08   0.03753 *
## Year2008          -0.1350      0.0555   -2.43   0.01502 *
## Year2009          -0.1229      0.0543   -2.26   0.02375 *
## Year2010          -0.2096      0.0548   -3.83   0.00013 ***
## Year2011          -0.2426      0.0548   -4.43    1e-05 ***
## Year2012          -0.1791      0.0530   -3.38   0.00074 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.407
## Multiple R-squared:  0.0244, Adjusted R-squared:  0.0186
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 255 weights are ~= 1. The remaining 2625 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0068 0.8640 0.9510 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.47e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2880"
## [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
## 230 198 201 192 182 151 158 133 148 155 176 197 182 236 229
## 2011 2012
## 220 257
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 79 61 68 60 54 28 72 48 38 56 65 73 64 95 107
## 2011 2012

```

```
## 106 118
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 68 55 60 55 45 25 62 44 36 46 50 64 56 85 89
## 2011 2012
## 88 87
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 38, df = 16, p-value = 0.002
```

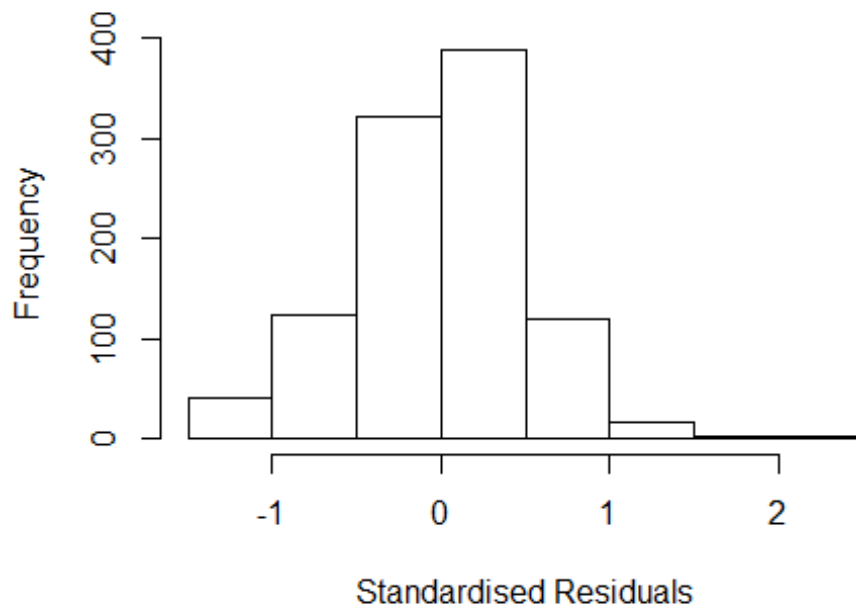


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.4, df = 1, p-value = 0.01
```



```
## [1] "Female first author team size 2018 geometric mean: 2.64769487910038"
## [1] "Male first author team size 2018 geometric mean: 2.83362858289364"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 550, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.60840937465693"
## [1] "Male last author team size 2018 geometric mean: 2.81316104552757"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 320, 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 1.102 1      1.050
## LastAuthorFemale  1.068 1      1.033
## UniqueAuthors    1.507 4      1.053
## Year             1.679 16      1.016
```


Residuals from first and last author and team size



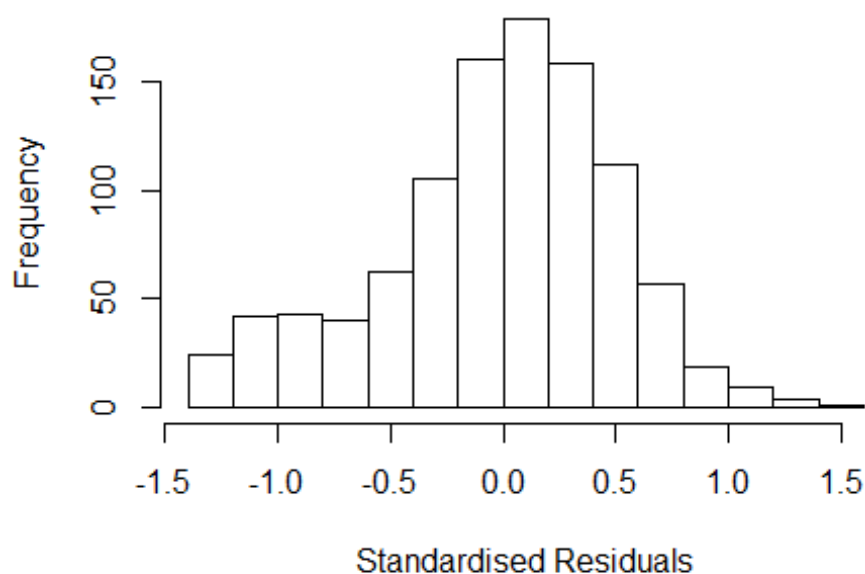
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3842 -0.3048 0.0157 0.3056 2.0189
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5593 0.0947 5.90 4.9e-09 ***
## FirstAuthorFemale1 0.0298 0.0352 0.85 0.398
## LastAuthorFemale1 -0.0769 0.0521 -1.47 0.141
## UniqueAuthors2 0.5151 0.0768 6.71 3.2e-11 ***
## UniqueAuthors3 0.5309 0.0789 6.73 3.0e-11 ***
## UniqueAuthors4 0.5654 0.0829 6.82 1.5e-11 ***
## UniqueAuthors5 0.6397 0.0796 8.03 2.7e-15 ***
## Year1997 0.0937 0.1143 0.82 0.413
## Year1998 0.0918 0.1166 0.79 0.431
## Year1999 0.2692 0.1101 2.45 0.015 *
```

```

## Year2000          0.1974      0.1170      1.69      0.092 .
## Year2001          0.2581      0.1331      1.94      0.053 .
## Year2002          0.1158      0.0928      1.25      0.212
## Year2003          0.0668      0.1240      0.54      0.590
## Year2004          0.2595      0.1079      2.41      0.016 *
## Year2005          0.1679      0.0978      1.72      0.087 .
## Year2006          0.1847      0.1071      1.72      0.085 .
## Year2007          0.2016      0.1011      1.99      0.046 *
## Year2008          0.0661      0.0974      0.68      0.498
## Year2009          0.1572      0.0982      1.60      0.110
## Year2010          0.0704      0.0936      0.75      0.452
## Year2011          0.0780      0.0958      0.81      0.416
## Year2012          0.0439      0.0960      0.46      0.648
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.445
## Multiple R-squared:  0.165, Adjusted R-squared:  0.146
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 83 weights are ~= 1. The remaining 932 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0038 0.8530 0.9510 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          9.85e-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 1.074 1          1.036
## LastAuthorFemale 1.048 1          1.024
## Year          1.125 16          1.004

```

Residuals from first and last author



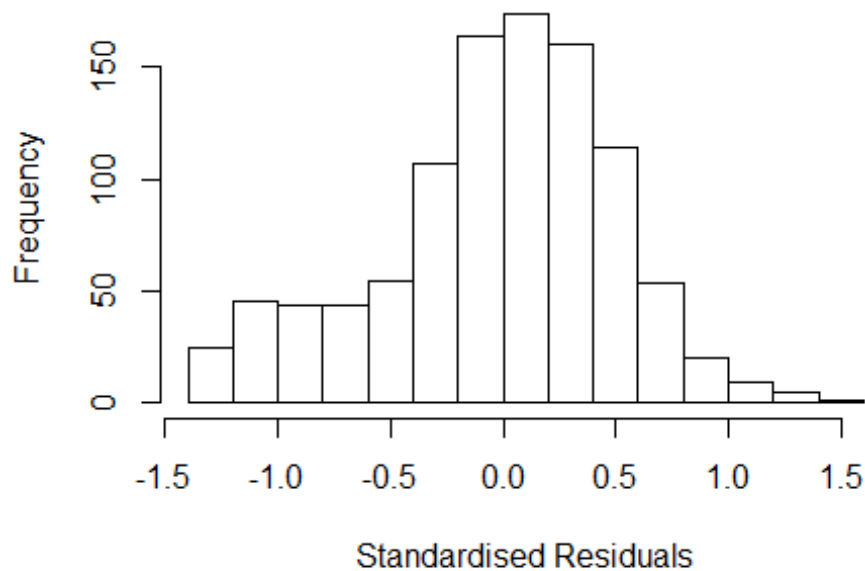
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3475 -0.3087  0.0273  0.3343  1.5142
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9612    0.1107   8.69  <2e-16 ***
## FirstAuthorFemale1  0.0657    0.0371   1.77   0.0769 .
## LastAuthorFemale1 -0.0860    0.0555  -1.55   0.1217
## Year1997          0.1102    0.1369   0.80   0.4213
## Year1998          0.0845    0.1468   0.58   0.5652
## Year1999          0.2787    0.1347   2.07   0.0388 *
## Year2000          0.2184    0.1390   1.57   0.1166
## Year2001          0.3453    0.1614   2.14   0.0326 *
## Year2002          0.2186    0.1216   1.80   0.0726 .
## Year2003          0.0674    0.1476   0.46   0.6478
## Year2004          0.3863    0.1266   3.05   0.0023 **
## Year2005          0.2463    0.1221   2.02   0.0440 *
```

```

## Year2006          0.2984      0.1301      2.29      0.0221 *
## Year2007          0.2844      0.1285      2.21      0.0270 *
## Year2008          0.1334      0.1282      1.04      0.2983
## Year2009          0.2808      0.1255      2.24      0.0254 *
## Year2010          0.1705      0.1206      1.41      0.1578
## Year2011          0.1917      0.1216      1.58      0.1153
## Year2012          0.1871      0.1214      1.54      0.1236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0366, Adjusted R-squared:  0.0192
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 106 weights are ~= 1. The remaining 909 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.283  0.849   0.947   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      9.85e-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 1.074 1      1.036
## Year              1.074 16      1.002

```

Residuals from first author



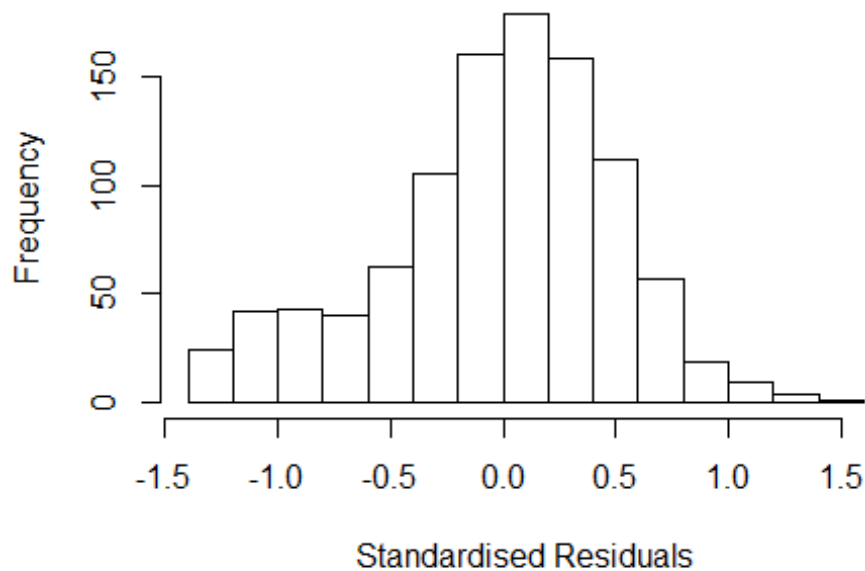
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3434 -0.3063 0.0274 0.3235 1.5181
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9591 0.1161 8.26 4.5e-16 ***
## FirstAuthorFemale1 0.0573 0.0377 1.52 0.1286
## Year1997 0.1111 0.1404 0.79 0.4288
## Year1998 0.0810 0.1500 0.54 0.5894
## Year1999 0.2812 0.1378 2.04 0.0416 *
## Year2000 0.2148 0.1414 1.52 0.1293
## Year2001 0.3372 0.1636 2.06 0.0396 *
## Year2002 0.2168 0.1263 1.72 0.0863 .
## Year2003 0.0642 0.1510 0.43 0.6709
## Year2004 0.3843 0.1306 2.94 0.0033 **
## Year2005 0.2500 0.1268 1.97 0.0490 *
## Year2006 0.2934 0.1340 2.19 0.0288 *
```

```

## Year2007          0.2785      0.1311      2.12      0.0339 *
## Year2008          0.1258      0.1324      0.95      0.3421
## Year2009          0.2747      0.1285      2.14      0.0327 *
## Year2010          0.1615      0.1243      1.30      0.1943
## Year2011          0.1870      0.1259      1.49      0.1376
## Year2012          0.1831      0.1253      1.46      0.1444
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.0344, Adjusted R-squared:  0.018
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 98 weights are ~= 1. The remaining 917 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.266  0.846   0.946   0.884   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      9.85e-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 1.054 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
## -1.3670 -0.3189  0.0228  0.3342  1.4963
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9773    0.1149   8.50  <2e-16 ***
## LastAuthorFemale1 -0.0747    0.0556  -1.34   0.179
## Year1997        0.1082    0.1401   0.77   0.440
## Year1998        0.0748    0.1489   0.50   0.616
## Year1999        0.2726    0.1376   1.98   0.048 *
## Year2000        0.2129    0.1424   1.50   0.135
## Year2001        0.3401    0.1636   2.08   0.038 *
## Year2002        0.2204    0.1259   1.75   0.080 .
## Year2003        0.0601    0.1516   0.40   0.692
## Year2004        0.3897    0.1309   2.98   0.003 **
## Year2005        0.2416    0.1262   1.91   0.056 .
## Year2006        0.2933    0.1333   2.20   0.028 *
```

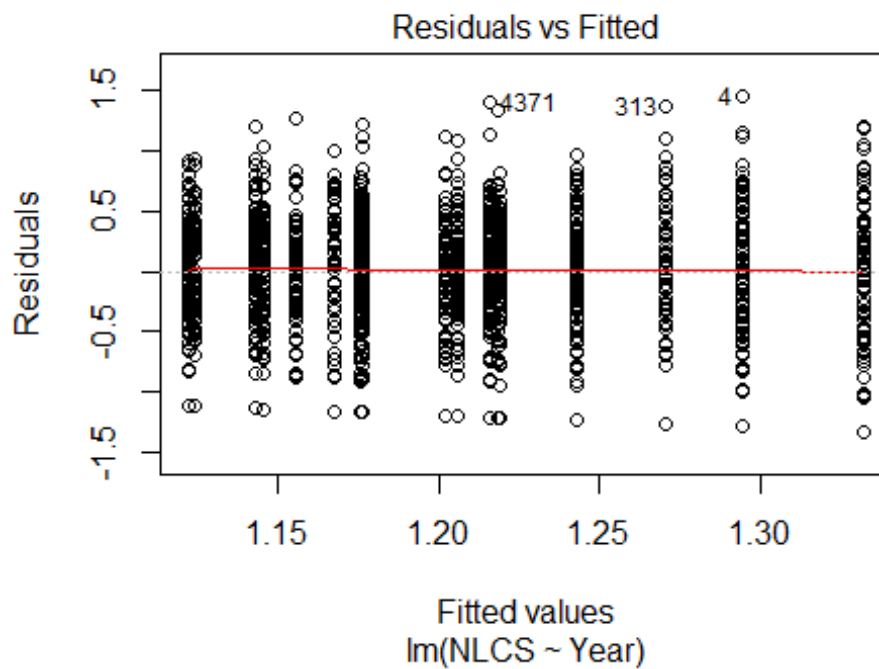
```

## Year2007          0.2848      0.1311      2.17      0.030 *
## Year2008          0.1359      0.1322      1.03      0.304
## Year2009          0.2879      0.1282      2.25      0.025 *
## Year2010          0.1690      0.1240      1.36      0.173
## Year2011          0.1861      0.1254      1.48      0.138
## Year2012          0.1880      0.1254      1.50      0.134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.0345, Adjusted R-squared:  0.018
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 93 weights are ~= 1. The remaining 922 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.280  0.846  0.947  0.884  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.85e-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: 1015"
## [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
## 260 233 241 185 218 167 170 212 206 232 271 244 240 187 247
## 2011 2012
## 245 279
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 91 70 78 58 70 54 75 101 86 85 102 104 120 88 130
## 2011 2012

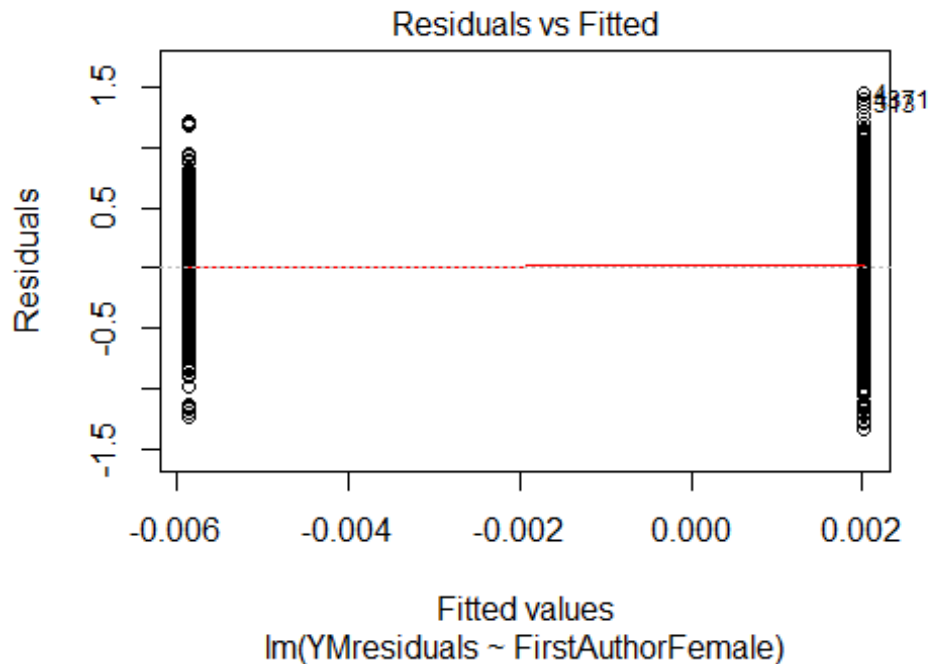
```



```
## 127 142
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 51 57 48 49 45 61 75 57 64 76 73 84 67 99
## 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 = 43, df = 16, p-value = 3e-04
```

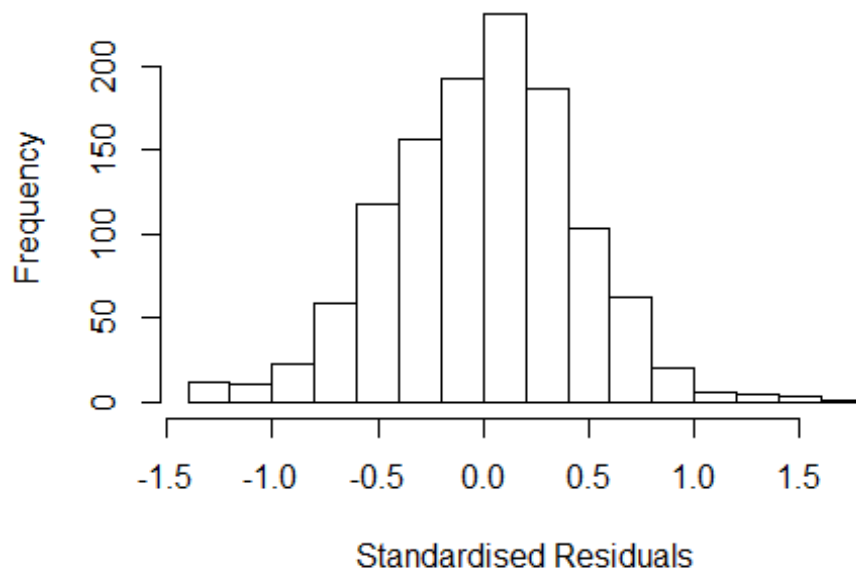


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.3, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 3.1657155591604"
## [1] "Male first author team size 2018 geometric mean: 3.28019742374317"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 760, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.0600904028133"
## [1] "Male last author team size 2018 geometric mean: 3.13959065890735"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 780, 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 1.129 1      1.062
## LastAuthorFemale  1.092 1      1.045
## UniqueAuthors    1.442 4      1.047
## Year              1.604 16     1.015
```

Residuals from first and last author and team size



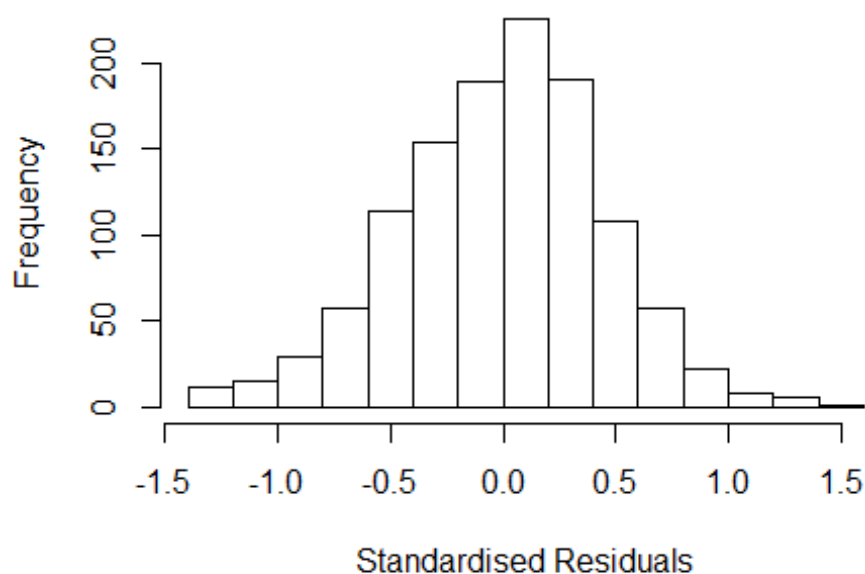
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3310 -0.2878 0.0155 0.2978 1.6771
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07088 0.09663 11.08 < 2e-16 ***
## FirstAuthorFemale1 -0.03106 0.03006 -1.03 0.30178
## LastAuthorFemale1 0.00145 0.04095 0.04 0.97180
## UniqueAuthors2 0.24199 0.07177 3.37 0.00077 ***
## UniqueAuthors3 0.26704 0.07159 3.73 0.00020 ***
## UniqueAuthors4 0.29679 0.07501 3.96 8.1e-05 ***
## UniqueAuthors5 0.36330 0.07618 4.77 2.1e-06 ***
## Year1997 -0.05303 0.10499 -0.51 0.61357
## Year1998 0.00636 0.10674 0.06 0.95252
## Year1999 -0.05711 0.11534 -0.50 0.62062
```

```

## Year2000      -0.16980      0.09399      -1.81      0.07109 .
## Year2001      -0.07748      0.10039      -0.77      0.44040
## Year2002      -0.03578      0.09202      -0.39      0.69746
## Year2003      -0.12652      0.08750      -1.45      0.14843
## Year2004      -0.07907      0.08990      -0.88      0.37930
## Year2005      -0.13525      0.08879      -1.52      0.12794
## Year2006      -0.07361      0.08652      -0.85      0.39507
## Year2007      -0.18290      0.08579      -2.13      0.03322 *
## Year2008      -0.09316      0.08599      -1.08      0.27885
## Year2009      -0.15490      0.08932      -1.73      0.08315 .
## Year2010      -0.16744      0.08643      -1.94      0.05295 .
## Year2011      -0.22198      0.08406      -2.64      0.00839 **
## Year2012      -0.13172      0.08187      -1.61      0.10790
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.0464, Adjusted R-squared:  0.0284
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 98 weights are ~= 1. The remaining 1092 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0851 0.8700 0.9470 0.9000 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.107 1      1.052
## LastAuthorFemale 1.074 1      1.036
## Year      1.161 16      1.005

```

Residuals from first and last author



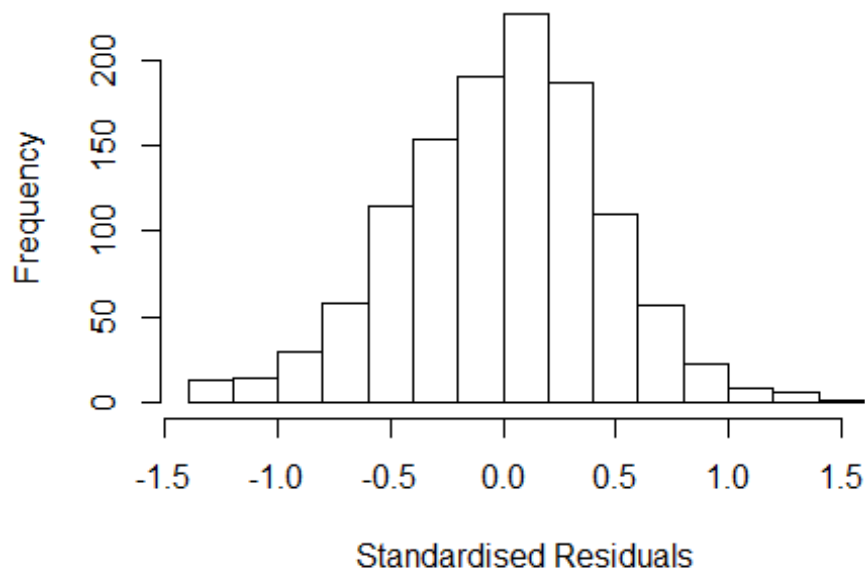
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3105 -0.2956 0.0214 0.2936 1.4375
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3105 0.0776 16.89 <2e-16 ***
## FirstAuthorFemale1 -0.0165 0.0301 -0.55 0.583
## LastAuthorFemale1 0.0177 0.0408 0.43 0.665
## Year1997 -0.0737 0.1119 -0.66 0.510
## Year1998 0.0159 0.1106 0.14 0.886
## Year1999 -0.0943 0.1210 -0.78 0.436
## Year2000 -0.1905 0.0990 -1.92 0.055 .
## Year2001 -0.0479 0.1010 -0.47 0.635
## Year2002 -0.0304 0.0942 -0.32 0.747
## Year2003 -0.1396 0.0904 -1.54 0.123
## Year2004 -0.0705 0.0952 -0.74 0.459
## Year2005 -0.1143 0.0909 -1.26 0.209
```

```

## Year2006          -0.0672      0.0900   -0.75    0.456
## Year2007          -0.1634      0.0889   -1.84    0.066 .
## Year2008          -0.0732      0.0901   -0.81    0.417
## Year2009          -0.1227      0.0915   -1.34    0.180
## Year2010          -0.1396      0.0896   -1.56    0.120
## Year2011          -0.1876      0.0876   -2.14    0.032 *
## Year2012          -0.0947      0.0851   -1.11    0.266
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.431
## Multiple R-squared:  0.0168, Adjusted R-squared:  0.00171
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1091 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.243  0.871  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      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 1.099 1      1.048
## Year              1.099 16      1.003

```

Residuals from first author



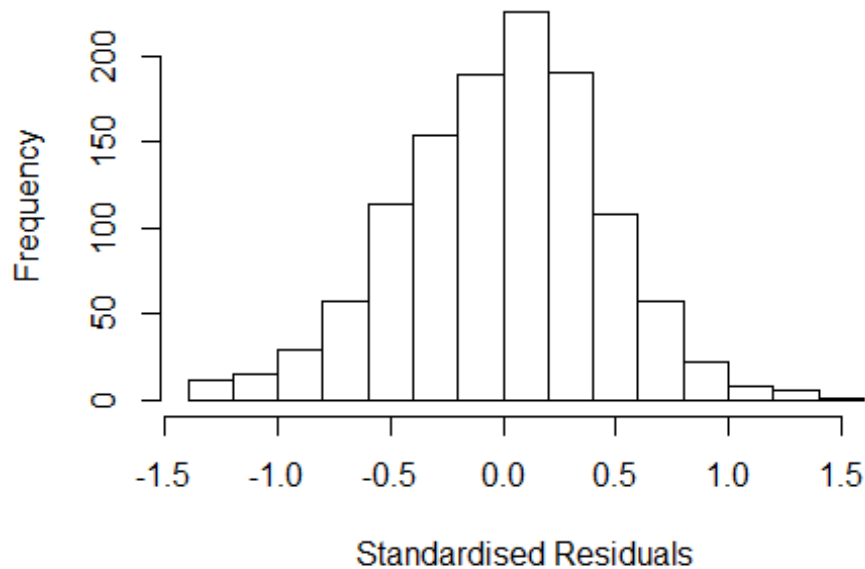
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3119 -0.2960 0.0228 0.2923 1.4361
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3119 0.0770 17.03 <2e-16 ***
## FirstAuthorFemale1 -0.0148 0.0300 -0.49 0.623
## Year1997 -0.0729 0.1123 -0.65 0.516
## Year1998 0.0152 0.1108 0.14 0.891
## Year1999 -0.0951 0.1197 -0.79 0.427
## Year2000 -0.1895 0.0989 -1.92 0.055 .
## Year2001 -0.0478 0.1003 -0.48 0.634
## Year2002 -0.0311 0.0932 -0.33 0.739
## Year2003 -0.1396 0.0899 -1.55 0.121
## Year2004 -0.0707 0.0948 -0.75 0.456
## Year2005 -0.1130 0.0906 -1.25 0.213
## Year2006 -0.0681 0.0895 -0.76 0.447
```

```

## Year2007          -0.1634      0.0888   -1.84    0.066 .
## Year2008          -0.0728      0.0895   -0.81    0.416
## Year2009          -0.1216      0.0913   -1.33    0.183
## Year2010          -0.1387      0.0889   -1.56    0.119
## Year2011          -0.1867      0.0872   -2.14    0.033 *
## Year2012          -0.0950      0.0846   -1.12    0.262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.0167, Adjusted R-squared:  0.00239
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 98 weights are ~= 1. The remaining 1092 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.246  0.871  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      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 1.063 1          1.031
## Year            1.063 16          1.002

```


Residuals from last author



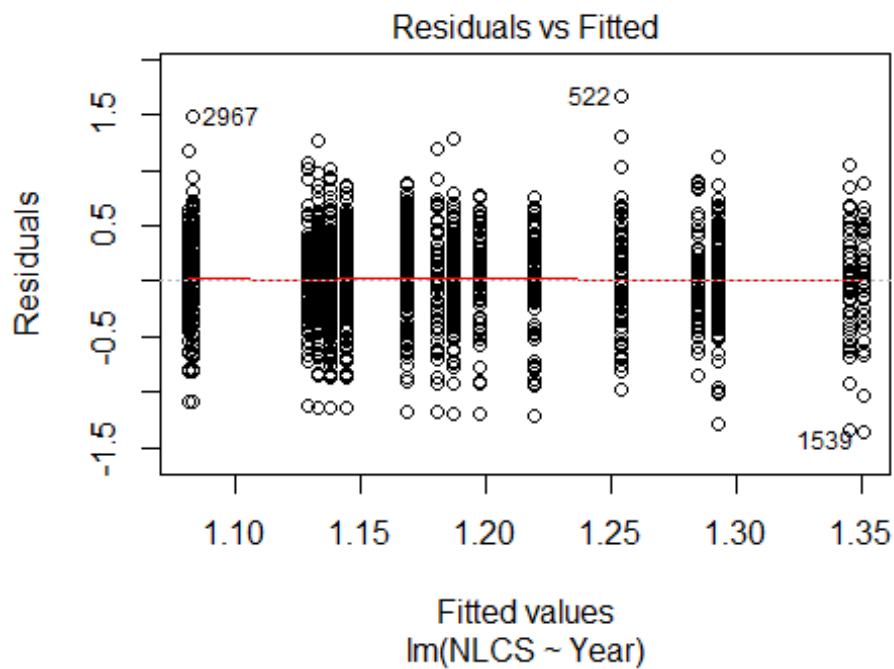
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.306 -0.296 0.020 0.291 1.442
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3064 0.0764 17.10 <2e-16 ***
## LastAuthorFemale1 0.0147 0.0406 0.36 0.718
## Year1997 -0.0709 0.1117 -0.63 0.526
## Year1998 0.0193 0.1102 0.18 0.861
## Year1999 -0.0929 0.1193 -0.78 0.436
## Year2000 -0.1887 0.0986 -1.91 0.056 .
## Year2001 -0.0482 0.1006 -0.48 0.632
## Year2002 -0.0273 0.0924 -0.30 0.768
## Year2003 -0.1399 0.0902 -1.55 0.121
## Year2004 -0.0721 0.0949 -0.76 0.447
## Year2005 -0.1146 0.0907 -1.26 0.206
## Year2006 -0.0663 0.0895 -0.74 0.459
```

```

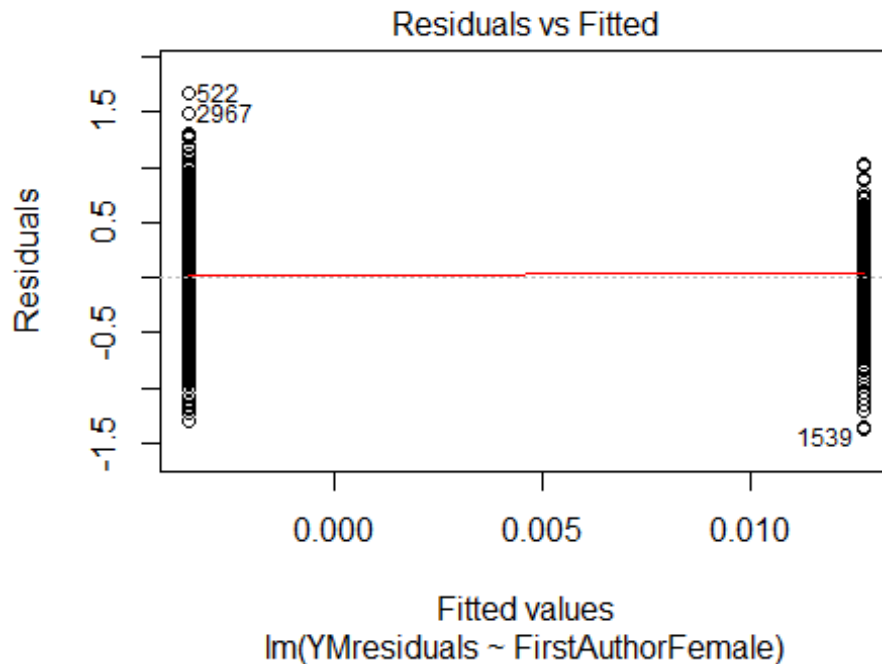
## Year2007          -0.1637      0.0890   -1.84    0.066 .
## Year2008          -0.0740      0.0896   -0.83    0.409
## Year2009          -0.1243      0.0913   -1.36    0.174
## Year2010          -0.1397      0.0889   -1.57    0.116
## Year2011          -0.1893      0.0873   -2.17    0.030 *
## Year2012          -0.0956      0.0848   -1.13    0.260
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.0166, Adjusted R-squared:  0.00231
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1093 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.243  0.870   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      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 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
## 223 238 244 193 252 174 248 216 214 239 330 315 334 370 383
## 2011 2012
## 388 378
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 65 73 81 52 64 41 90 60 75 99 117 134 138 172 190
## 2011 2012

```

```
## 194 195
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 53 56 53 40 44 34 73 52 65 80 80 104 106 129 137
## 2011 2012
## 154 137
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 38, df = 16, p-value = 0.001
```

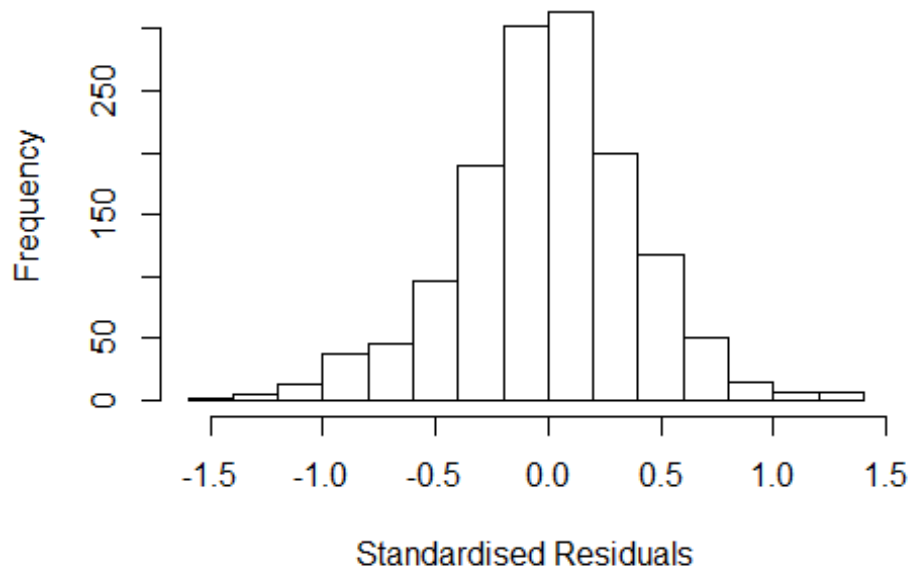


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.8, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 3.37840139473082"
## [1] "Male first author team size 2018 geometric mean: 3.19125612790539"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.56044643249711"
## [1] "Male last author team size 2018 geometric mean: 3.15474564825236"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, 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 1.060 1          1.029
## LastAuthorFemale  1.081 1          1.040
## UniqueAuthors    1.329 4          1.036
## Year              1.424 16         1.011
```

Residuals from first and last author and team size



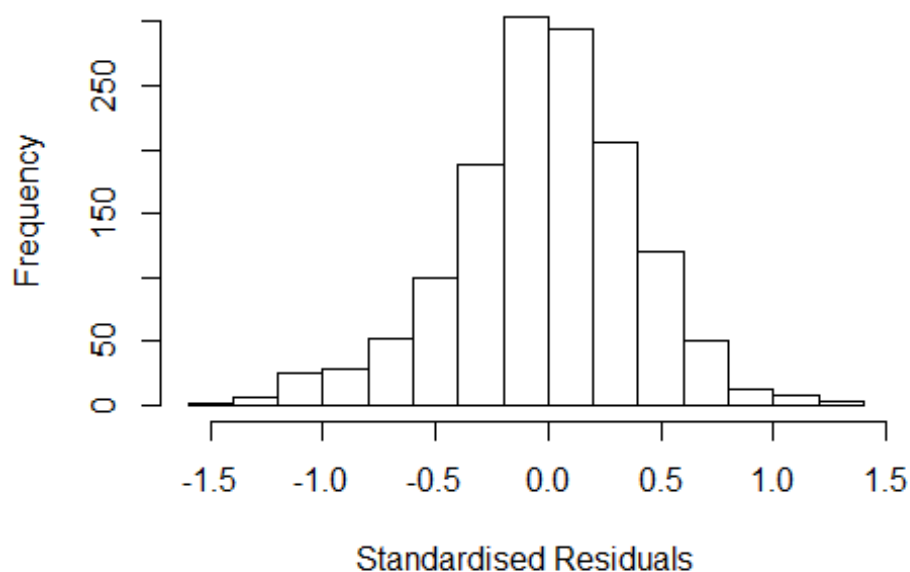
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.42870 -0.22599 0.00436 0.23012 1.31799
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05206 0.08720 12.06 < 2e-16 ***
## FirstAuthorFemale1 0.02119 0.02328 0.91 0.36282
## LastAuthorFemale1 0.00982 0.03248 0.30 0.76236
## UniqueAuthors2 0.21037 0.07366 2.86 0.00435 **
## UniqueAuthors3 0.22097 0.07400 2.99 0.00288 **
## UniqueAuthors4 0.29303 0.07561 3.88 0.00011 ***
## UniqueAuthors5 0.31718 0.07600 4.17 3.2e-05 ***
## Year1997 -0.04527 0.08194 -0.55 0.58072
## Year1998 -0.05050 0.08203 -0.62 0.53822
## Year1999 0.13447 0.07497 1.79 0.07309 .
```

```

## Year2000      -0.05015    0.08975   -0.56  0.57641
## Year2001      0.13641    0.10150    1.34  0.17918
## Year2002     -0.05758    0.06951   -0.83  0.40760
## Year2003      0.00633    0.06787    0.09  0.92568
## Year2004     -0.04128    0.07102   -0.58  0.56118
## Year2005     -0.14231    0.06894   -2.06  0.03916 *
## Year2006     -0.24818    0.07865   -3.16  0.00164 **
## Year2007     -0.05337    0.06844   -0.78  0.43566
## Year2008     -0.17503    0.06508   -2.69  0.00724 **
## Year2009     -0.10540    0.06817   -1.55  0.12232
## Year2010     -0.16455    0.06569   -2.51  0.01236 *
## Year2011     -0.17113    0.06420   -2.67  0.00778 **
## Year2012     -0.19024    0.06631   -2.87  0.00418 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.0798, Adjusted R-squared:  0.065
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 105 weights are ~= 1. The remaining 1292 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0828 0.8640 0.9550 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      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 1.043 1      1.021
## LastAuthorFemale  1.063 1      1.031
## Year              1.104 16      1.003

```

Residuals from first and last author



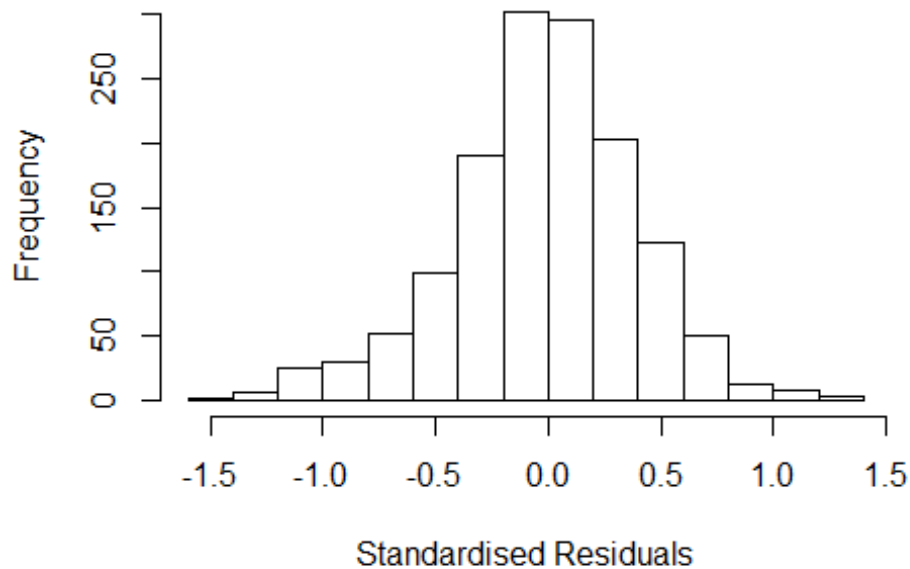
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.43287 -0.23493 -0.00274 0.23950 1.27283
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28065 0.06029 21.24 <2e-16 ***
## FirstAuthorFemale1 0.03134 0.02327 1.35 0.1784
## LastAuthorFemale1 0.01841 0.03252 0.57 0.5714
## Year1997 -0.06837 0.08287 -0.83 0.4095
## Year1998 -0.07316 0.08589 -0.85 0.3945
## Year1999 0.12088 0.08102 1.49 0.1359
## Year2000 -0.07126 0.09297 -0.77 0.4435
## Year2001 0.12007 0.09666 1.24 0.2144
## Year2002 -0.06313 0.07318 -0.86 0.3885
## Year2003 -0.00826 0.07035 -0.12 0.9066
## Year2004 -0.03280 0.07471 -0.44 0.6607
## Year2005 -0.14291 0.07203 -1.98 0.0475 *
```

```

## Year2006          -0.24844    0.08238   -3.02    0.0026 **
## Year2007          -0.05355    0.07053   -0.76    0.4478
## Year2008          -0.17106    0.06875   -2.49    0.0130 *
## Year2009          -0.09053    0.07068   -1.28    0.2005
## Year2010          -0.16272    0.06952   -2.34    0.0194 *
## Year2011          -0.15828    0.06743   -2.35    0.0191 *
## Year2012          -0.17090    0.06902   -2.48    0.0134 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.365
## Multiple R-squared:  0.0456, Adjusted R-squared:  0.0331
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 117 weights are ~= 1. The remaining 1280 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0893 0.8610 0.9530 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      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 1.043 1      1.021
## Year              1.043 16      1.001

```


Residuals from first author



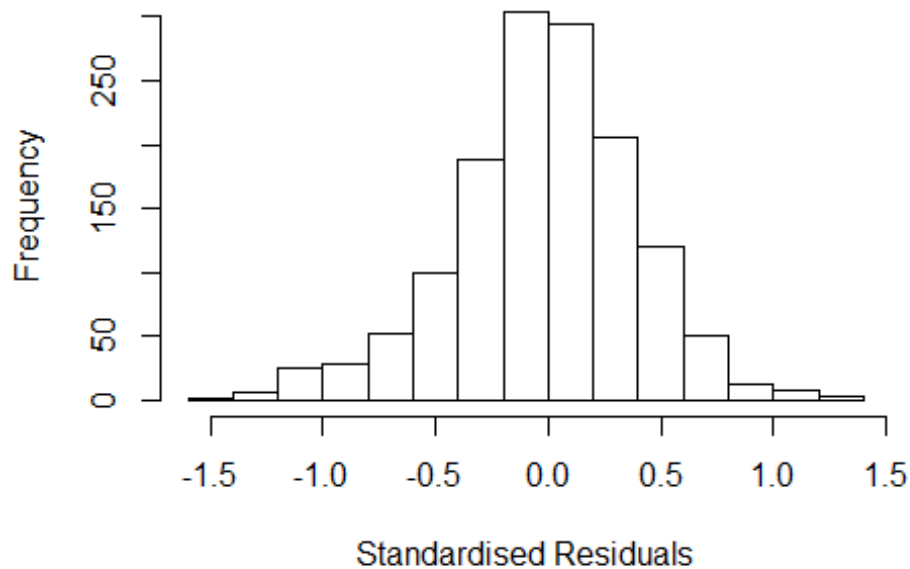
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.43420 -0.23676 -0.00393 0.24129 1.28967
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2813 0.0602 21.27 <2e-16 ***
## FirstAuthorFemale1 0.0322 0.0233 1.38 0.1672
## Year1997 -0.0666 0.0827 -0.81 0.4205
## Year1998 -0.0734 0.0859 -0.85 0.3933
## Year1999 0.1207 0.0810 1.49 0.1364
## Year2000 -0.0696 0.0928 -0.75 0.4534
## Year2001 0.1194 0.0966 1.24 0.2169
## Year2002 -0.0634 0.0731 -0.87 0.3860
## Year2003 -0.0077 0.0704 -0.11 0.9129
## Year2004 -0.0330 0.0747 -0.44 0.6589
## Year2005 -0.1409 0.0719 -1.96 0.0503 .
## Year2006 -0.2477 0.0824 -3.01 0.0027 **
```

```

## Year2007          -0.0517      0.0703   -0.74   0.4622
## Year2008          -0.1697      0.0687   -2.47   0.0136 *
## Year2009          -0.0885      0.0704   -1.26   0.2088
## Year2010          -0.1605      0.0692   -2.32   0.0206 *
## Year2011          -0.1567      0.0673   -2.33   0.0200 *
## Year2012          -0.1699      0.0690   -2.46   0.0139 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.365
## Multiple R-squared:  0.0455, Adjusted R-squared:  0.0337
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1279 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0877 0.8600 0.9520 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      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 1.065 1          1.032
## Year              1.065 16          1.002

```

Residuals from last author



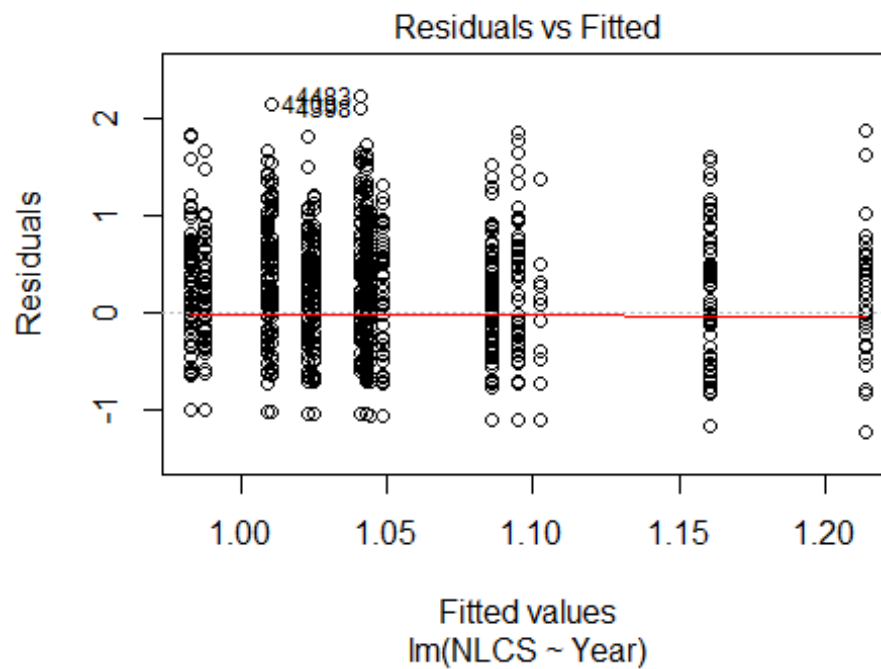
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4080 -0.2370 0.0011 0.2401 1.2622
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28773 0.05968 21.58 <2e-16 ***
## LastAuthorFemale1 0.02126 0.03269 0.65 0.5155
## Year1997 -0.07022 0.08260 -0.85 0.3954
## Year1998 -0.07729 0.08555 -0.90 0.3664
## Year1999 0.12025 0.08101 1.48 0.1380
## Year2000 -0.07015 0.09272 -0.76 0.4495
## Year2001 0.11631 0.09697 1.20 0.2306
## Year2002 -0.06383 0.07301 -0.87 0.3822
## Year2003 -0.00817 0.07037 -0.12 0.9076
## Year2004 -0.03064 0.07449 -0.41 0.6809
## Year2005 -0.14229 0.07181 -1.98 0.0477 *
## Year2006 -0.25095 0.08202 -3.06 0.0023 **
```

```

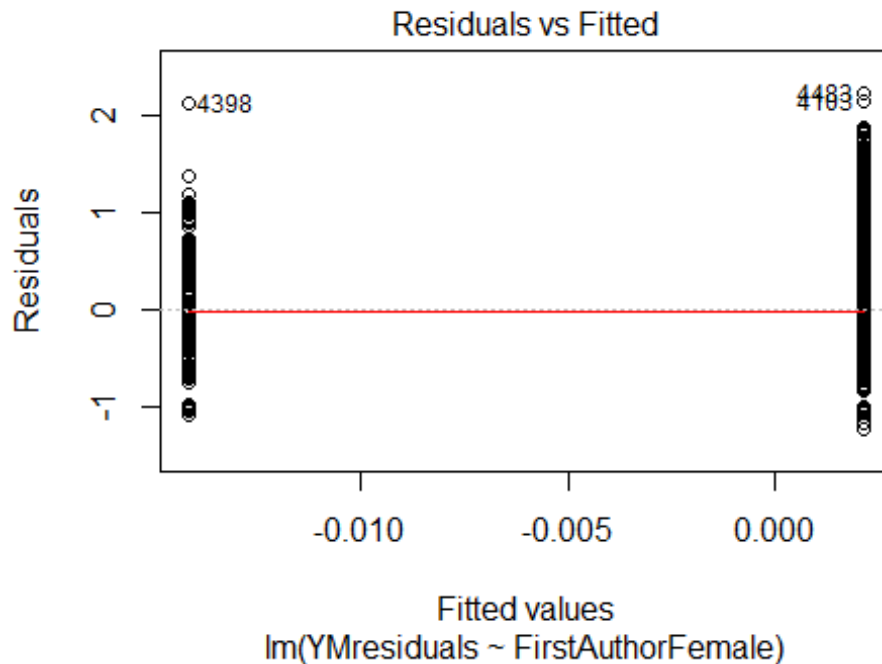
## Year2007          -0.05360      0.07017    -0.76    0.4451
## Year2008          -0.17289      0.06847    -2.53    0.0117 *
## Year2009          -0.09079      0.07056    -1.29    0.1984
## Year2010          -0.16057      0.06917    -2.32    0.0204 *
## Year2011          -0.15682      0.06720    -2.33    0.0198 *
## Year2012          -0.17017      0.06872    -2.48    0.0134 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.365
## Multiple R-squared:  0.0442, Adjusted R-squared:  0.0324
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 125 weights are ~= 1. The remaining 1272 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.105  0.860  0.953  0.890  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] "Sample size for the above analysis: 1397"
## [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
## 223 98 23 173 7 145 143 186 184 162 225 237 243 239 243
## 2011 2012
## 215 242
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 120 44 11 85 0 74 75 105 97 87 135 142 153 162 162
## 2011 2012

```

```
## 140 160
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 113 41 11 79 0 68 66 99 88 77 121 132 142 143 142
## 2011 2012
## 125 142
## [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 = 15, p-value = 0.002
```

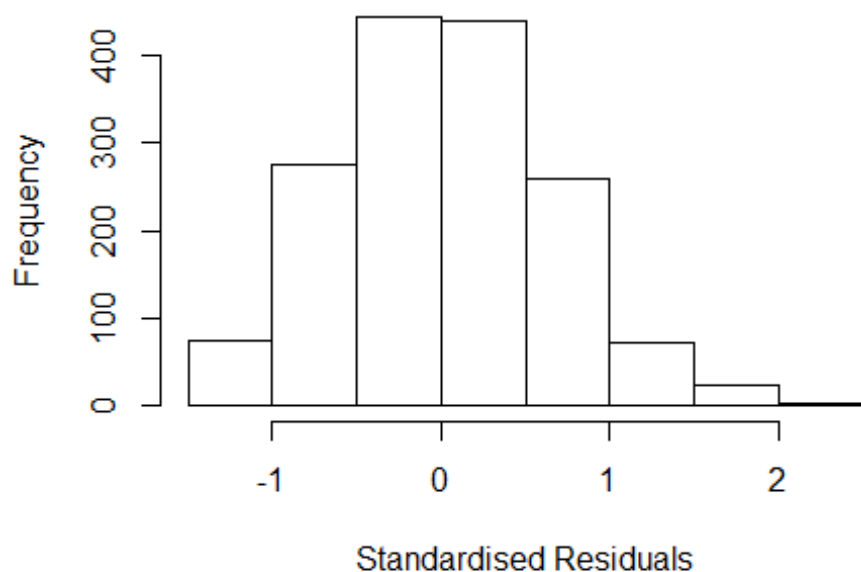


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.4, df = 1, p-value = 0.06
```



```
## [1] "Female first author team size 2018 geometric mean: 1.68223115744097"
## [1] "Male first author team size 2018 geometric mean: 1.39285889836071"
##
## 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: 1.55072839253126"
## [1] "Male last author team size 2018 geometric mean: 1.42349781425291"
##
## 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 1.675 1      1.294
## LastAuthorFemale  1.670 1      1.292
## UniqueAuthors    1.265 4      1.030
## Year              1.333 15     1.010
```

Residuals from first and last author and team size



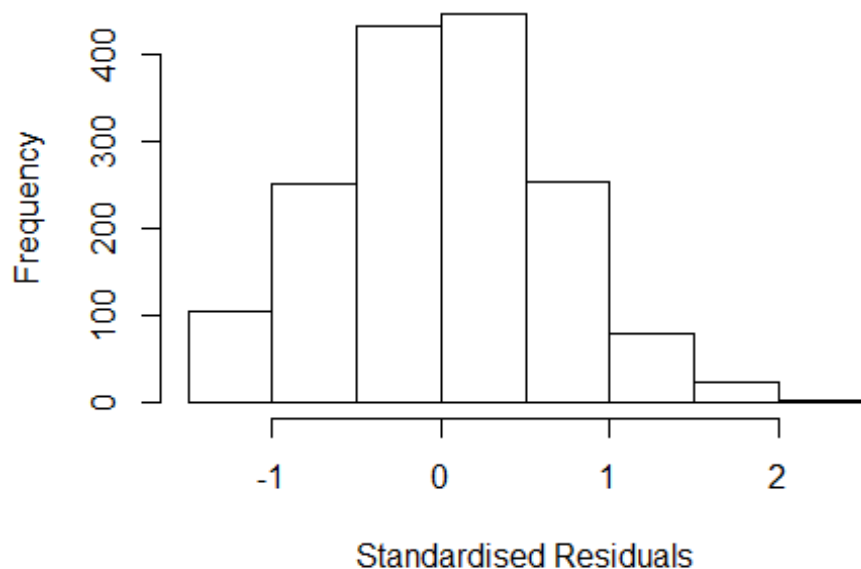
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24829 -0.42451 0.00099 0.42811 2.29317
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.92654 0.06270 14.78 < 2e-16 ***
## FirstAuthorFemale1 -0.04032 0.05703 -0.71 0.47963
## LastAuthorFemale1 -0.01494 0.05933 -0.25 0.80115
## UniqueAuthors2 0.14041 0.04113 3.41 0.00066 ***
## UniqueAuthors3 0.27822 0.06104 4.56 5.6e-06 ***
## UniqueAuthors4 0.44315 0.13329 3.32 0.00091 ***
## UniqueAuthors5 0.53728 0.14809 3.63 0.00029 ***
## Year1997 0.27489 0.11841 2.32 0.02039 *
## Year1998 0.14370 0.20177 0.71 0.47644
## Year1999 0.02724 0.09087 0.30 0.76440
```

```

## Year2001      0.06374      0.10164      0.63  0.53068
## Year2002      0.08507      0.11344      0.75  0.45340
## Year2003      0.06287      0.08484      0.74  0.45875
## Year2004      0.08874      0.08973      0.99  0.32287
## Year2005      0.15858      0.10509      1.51  0.13150
## Year2006      0.11794      0.08201      1.44  0.15061
## Year2007      0.02948      0.08007      0.37  0.71284
## Year2008      0.05932      0.08199      0.72  0.46944
## Year2009      0.00242      0.08189      0.03  0.97648
## Year2010      0.04354      0.08786      0.50  0.62030
## Year2011     -0.03666      0.08788     -0.42  0.67665
## Year2012     -0.01944      0.09564     -0.20  0.83896
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.65
## Multiple R-squared:  0.0303, Adjusted R-squared:  0.0173
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 133 weights are ~= 1. The remaining 1456 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.187  0.866  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      6.29e-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 1.597 1 1.264
## LastAuthorFemale 1.578 1 1.256
## Year 1.112 15 1.004

```


Residuals from first and last author



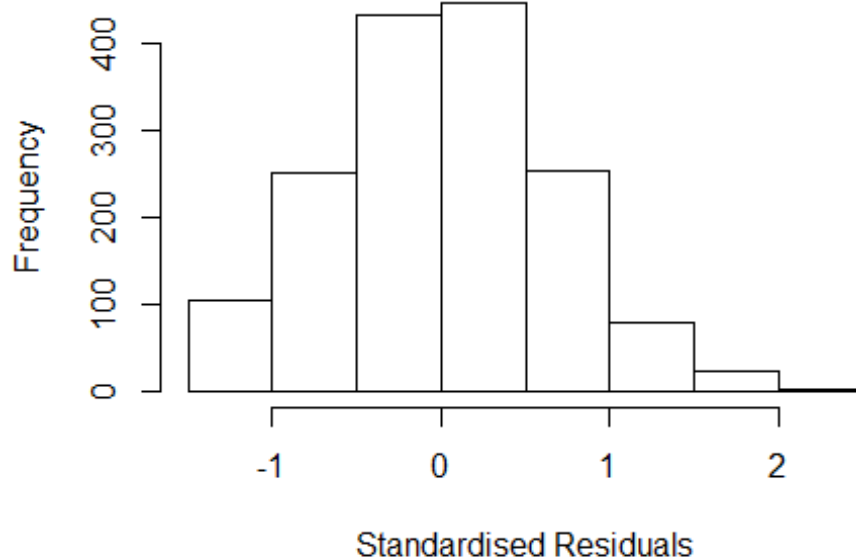
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23534 -0.42459 0.00599 0.44964 2.28999
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95603 0.06067 15.76 <2e-16 ***
## FirstAuthorFemale1 -0.02721 0.05655 -0.48 0.631
## LastAuthorFemale1 0.00302 0.05879 0.05 0.959
## Year1997 0.27931 0.12223 2.29 0.022 *
## Year1998 0.12867 0.20140 0.64 0.523
## Year1999 0.02390 0.08988 0.27 0.790
## Year2001 0.07685 0.10062 0.76 0.445
## Year2002 0.10475 0.11527 0.91 0.364
## Year2003 0.06900 0.08426 0.82 0.413
## Year2004 0.09252 0.08967 1.03 0.302
## Year2005 0.18334 0.10452 1.75 0.080 .
## Year2006 0.14547 0.08122 1.79 0.073 .
```

```

## Year2007          0.06066    0.07892    0.77    0.442
## Year2008          0.09081    0.08194    1.11    0.268
## Year2009          0.04151    0.07982    0.52    0.603
## Year2010          0.07709    0.08724    0.88    0.377
## Year2011          0.00360    0.08643    0.04    0.967
## Year2012          0.02598    0.09417    0.28    0.783
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.00803,    Adjusted R-squared:  -0.00271
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1469 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.204  0.871  0.952  0.913  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.29e-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 1.05 1      1.025
## Year              1.05 15      1.002

```

Residuals from first author



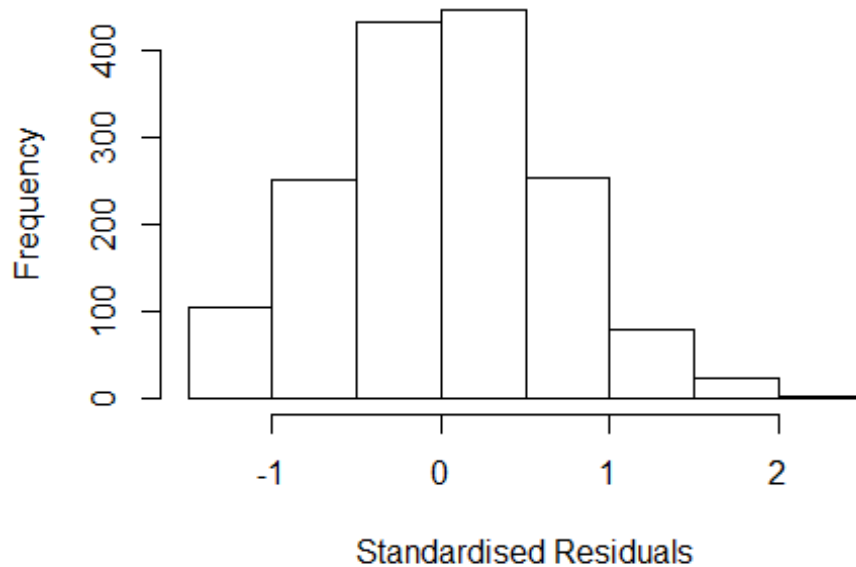
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23532 -0.42408 0.00603 0.44972 2.29003
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95608 0.06068 15.76 <2e-16 ***
## FirstAuthorFemale1 -0.02517 0.04638 -0.54 0.587
## Year1997 0.27924 0.12219 2.29 0.022 *
## Year1998 0.12867 0.20136 0.64 0.523
## Year1999 0.02394 0.08988 0.27 0.790
## Year2001 0.07688 0.10063 0.76 0.445
## Year2002 0.10476 0.11530 0.91 0.364
## Year2003 0.06928 0.08369 0.83 0.408
## Year2004 0.09268 0.08959 1.03 0.301
## Year2005 0.18327 0.10452 1.75 0.080 .
## Year2006 0.14574 0.08095 1.80 0.072 .
## Year2007 0.06068 0.07894 0.77 0.442
```

```

## Year2008          0.09085    0.08192    1.11    0.268
## Year2009          0.04161    0.07976    0.52    0.602
## Year2010          0.07720    0.08724    0.88    0.376
## Year2011          0.00357    0.08644    0.04    0.967
## Year2012          0.02590    0.09417    0.27    0.783
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.659
## Multiple R-squared:  0.00803,    Adjusted R-squared:  -0.00207
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1469 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.203  0.871  0.952  0.913  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.29e-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 1.038 1          1.019
## Year            1.038 15          1.001

```

Residuals from last author



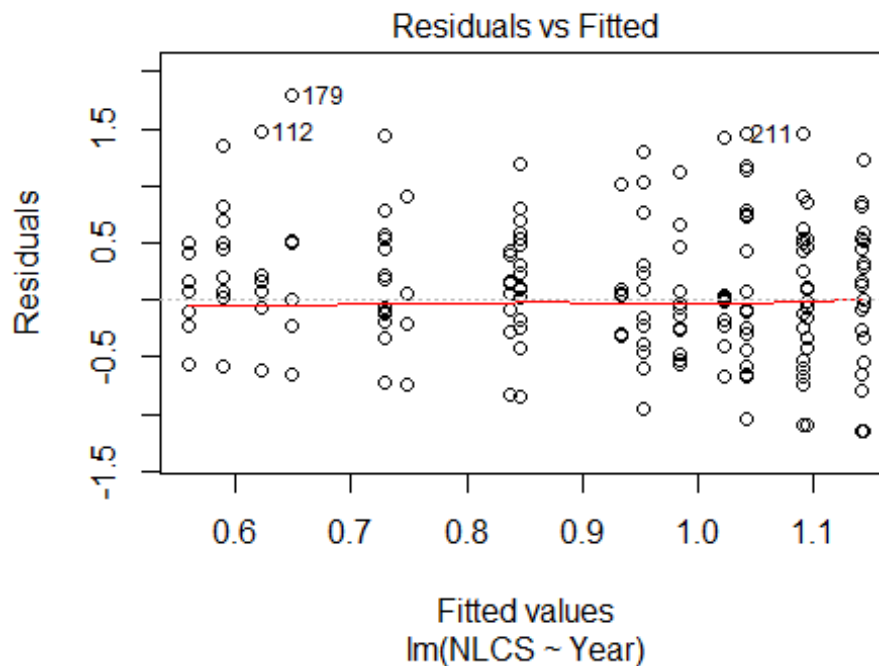
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23328 -0.42640 0.00765 0.44699 2.29165
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95540 0.06061 15.76 <2e-16 ***
## LastAuthorFemale1 -0.01525 0.04787 -0.32 0.750
## Year1997 0.27787 0.12186 2.28 0.023 *
## Year1998 0.12855 0.20182 0.64 0.524
## Year1999 0.02361 0.08987 0.26 0.793
## Year2001 0.07661 0.10061 0.76 0.446
## Year2002 0.10214 0.11477 0.89 0.374
## Year2003 0.06983 0.08414 0.83 0.407
## Year2004 0.09260 0.08968 1.03 0.302
## Year2005 0.18247 0.10443 1.75 0.081 .
## Year2006 0.14523 0.08122 1.79 0.074 .
## Year2007 0.06000 0.07902 0.76 0.448
```

```

## Year2008          0.08952    0.08184    1.09    0.274
## Year2009          0.04071    0.07985    0.51    0.610
## Year2010          0.07750    0.08717    0.89    0.374
## Year2011          0.00239    0.08636    0.03    0.978
## Year2012          0.02495    0.09399    0.27    0.791
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.00791,    Adjusted R-squared:  -0.00219
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1470 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.203  0.871  0.953  0.913  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.29e-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: 1589"
## [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
##   29   21    6   16   15   13   15   23   23   23   22   21   25   18   26
## 2011 2012
##   27   20
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18   13    4   10    8    8    8    9   17   11   11   11   19   12   22
## 2011 2012
##   15   12

```

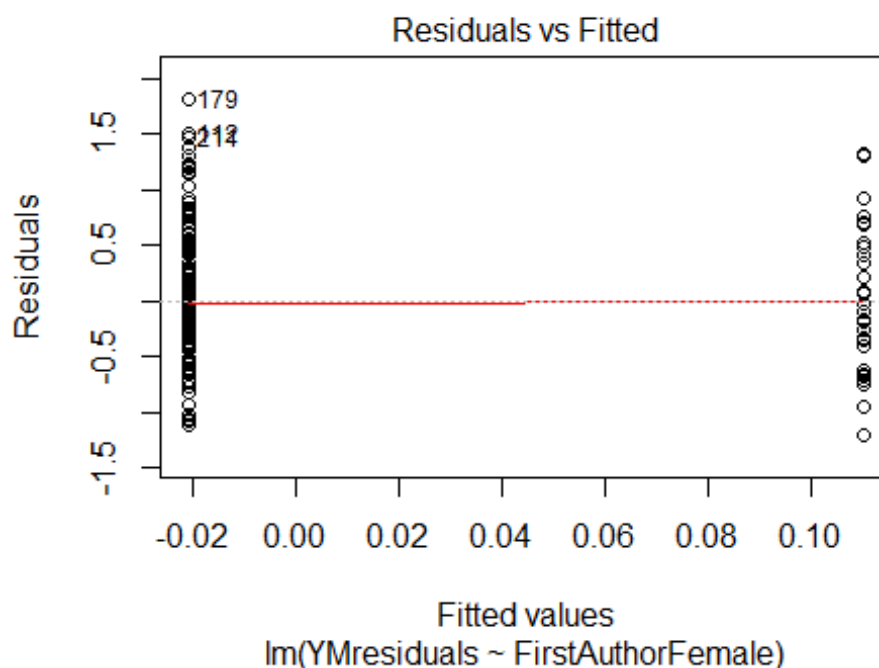
```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 17 12 4 8 8 8 7 8 15 11 11 11 16 12 21
## 2011 2012
## 13 11
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 12, df = 16, p-value = 0.7
```



```
##
## 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: 1.41421356237309"
## [1] "Male first author team size 2018 geometric mean: 1.08005973889231"
## Warning in wilcox.test.default(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 = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.25992104989487"
## [1] "Male last author team size 2018 geometric mean: 1.14869835499704"

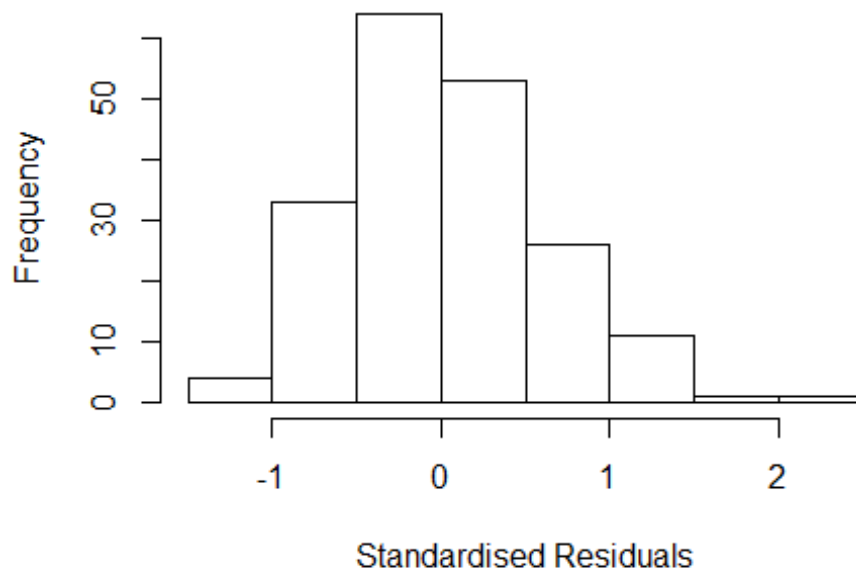
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17, p-value = 0.7
## 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	4.272	1	2.067
LastAuthorFemale	4.046	1	2.011
UniqueAuthors	2.111	3	1.133
Year	3.759	16	1.042

Residuals from first and last author and team size



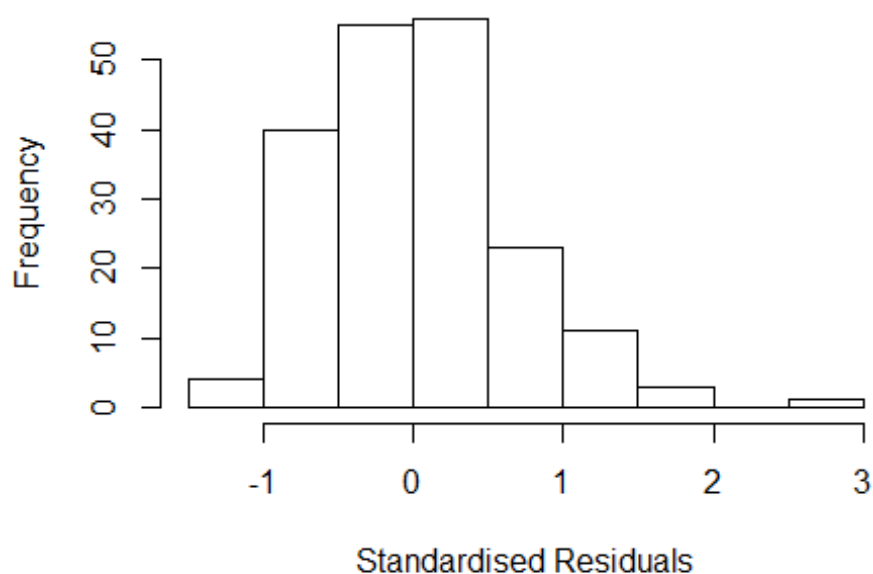
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0898 -0.4198 -0.0217 0.3469 2.4328
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6709 0.1441 4.66 6.4e-06 ***
## FirstAuthorFemale1 0.3549 0.2075 1.71 0.0891 .
## LastAuthorFemale1 -0.4022 0.2290 -1.76 0.0808 .
## UniqueAuthors2 0.1913 0.1078 1.77 0.0778 .
## UniqueAuthors3 0.6189 0.1924 3.22 0.0015 **
## UniqueAuthors4 0.1397 0.4016 0.35 0.7284
## Year1997 0.0788 0.2929 0.27 0.7884
## Year1998 -0.1216 0.2398 -0.51 0.6128
## Year1999 0.1631 0.1711 0.95 0.3417
## Year2000 0.1841 0.1982 0.93 0.3542
```

```

## Year2001      -0.1914      0.2379      -0.80      0.4223
## Year2002      0.1905      0.2077      0.92      0.3603
## Year2003     -0.4518      0.2283     -1.98      0.0495 *
## Year2004      0.1419      0.2349      0.60      0.5467
## Year2005      0.3989      0.2542      1.57      0.1184
## Year2006     -0.1471      0.1774     -0.83      0.4082
## Year2007      0.2276      0.2229      1.02      0.3086
## Year2008      0.0425      0.2301      0.18      0.8535
## Year2009      0.3368      0.2757      1.22      0.2234
## Year2010      0.2256      0.2200      1.03      0.3066
## Year2011     -0.1298      0.2620     -0.50      0.6209
## Year2012      0.1826      0.2208      0.83      0.4094
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.57
## Multiple R-squared:  0.193, Adjusted R-squared:  0.0937
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 174 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0294 0.8630 0.9440 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      5.18e-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 5.839 1 2.416
## LastAuthorFemale 4.963 1 2.228
## Year 2.321 16 1.027

```

Residuals from first and last author



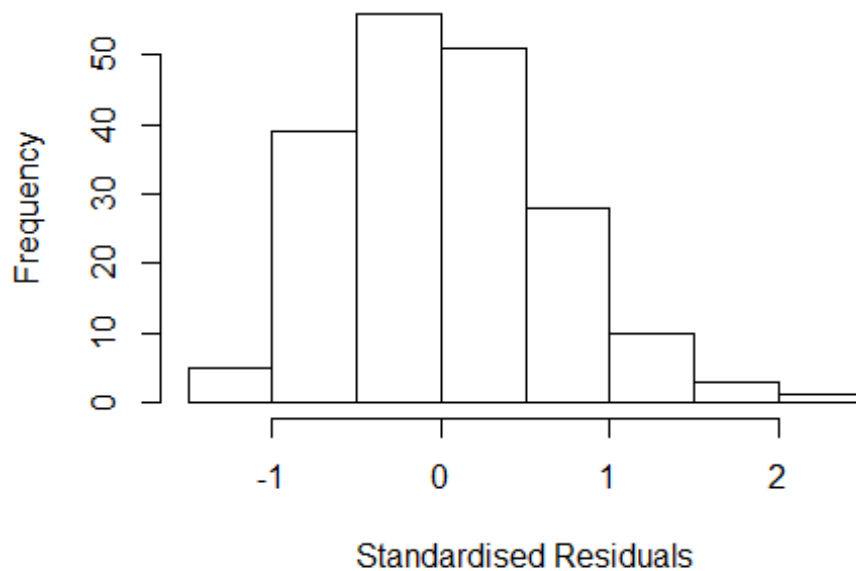
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 179 0037345050 2.441 2003      2601      2      2.698
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1319 -0.3958 -0.0223  0.4115  2.6980
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.700      0.141    4.95 1.7e-06 ***
## FirstAuthorFemale1  0.490      0.237    2.06  0.040 *
## LastAuthorFemale1 -0.554      0.274   -2.03  0.044 *
## Year1997          0.193      0.283    0.68  0.495
## Year1998          0.023      0.361    0.06  0.949
## Year1999          0.164      0.171    0.96  0.340
## Year2000          0.155      0.196    0.79  0.431
## Year2001         -0.162      0.253   -0.64  0.522
## Year2002          0.198      0.211    0.94  0.349
## Year2003         -0.402      0.216   -1.86  0.065 .
## Year2004          0.212      0.244    0.87  0.386
## Year2005          0.432      0.249    1.73  0.085 .
```

```

## Year2006          -0.140      0.181   -0.78    0.439
## Year2007          0.332      0.197    1.69    0.093 .
## Year2008          0.048      0.227    0.21    0.832
## Year2009          0.415      0.277    1.50    0.135
## Year2010          0.292      0.250    1.17    0.244
## Year2011         -0.110      0.258   -0.42    0.672
## Year2012          0.321      0.207    1.55    0.123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.582
## Multiple R-squared:  0.154, Adjusted R-squared:  0.0662
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 70 is an outlier with |weight| <= 0.00042 ( < 0.00052);
## 16 weights are ~= 1. The remaining 176 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.409  0.873   0.949   0.906   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.18e-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 1.398 1          1.183
## Year              1.398 16          1.011

```

Residuals from first author



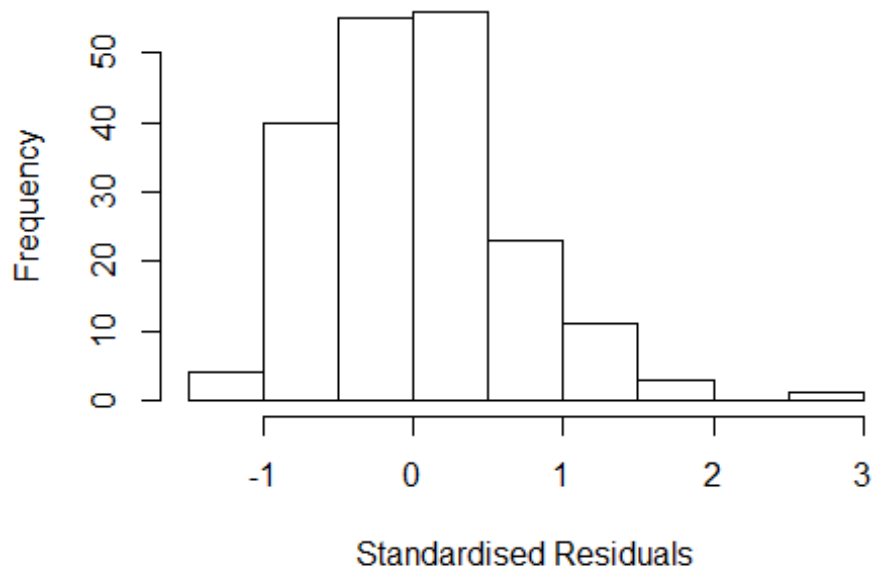
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 179 0037345050 2.441 2003    2601      2    2.698
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1467 -0.3948 -0.0421  0.4230  2.0739
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7032    0.1486   4.73 4.6e-06 ***
## FirstAuthorFemale1 0.1072    0.1434   0.75  0.455
## Year1997        0.1887    0.2761   0.68  0.495
## Year1998        0.0210    0.3618   0.06  0.954
## Year1999        0.1156    0.1800   0.64  0.521
## Year2000        0.1503    0.2016   0.75  0.457
## Year2001       -0.1904    0.2582  -0.74  0.462
## Year2002        0.1697    0.2221   0.76  0.446
## Year2003       -0.3361    0.2589  -1.30  0.196
## Year2004        0.2385    0.2747   0.87  0.387
## Year2005        0.4435    0.2634   1.68  0.094 .
## Year2006       -0.1438    0.1862  -0.77  0.441
```

```

## Year2007          0.3777      0.2138      1.77      0.079 .
## Year2008          0.0182      0.2365      0.08      0.939
## Year2009          0.4356      0.2670      1.63      0.105
## Year2010          0.2720      0.2383      1.14      0.255
## Year2011         -0.1289      0.2580     -0.50      0.618
## Year2012          0.2827      0.2167      1.30      0.194
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.597
## Multiple R-squared:  0.117, Adjusted R-squared:  0.0312
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 175 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.202  0.876  0.942  0.902  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.18e-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 1.495 1      1.223
## Year              1.495 16      1.013

```

Residuals from last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 179 0037345050 2.441 2003      2601      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
## -1.1944 -0.4015 -0.0481  0.4228  2.2054
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7021     0.1493   4.70 5.2e-06 ***
## LastAuthorFemale1 -0.1174     0.1446  -0.81  0.418
## Year1997          0.1964     0.2853   0.69  0.492
## Year1998          0.0225     0.3615   0.06  0.951
## Year1999          0.1743     0.1791   0.97  0.332
## Year2000          0.1511     0.2022   0.75  0.456
## Year2001         -0.1494     0.2607  -0.57  0.567
## Year2002          0.2069     0.2155   0.96  0.338
## Year2003         -0.3492     0.2567  -1.36  0.175
## Year2004          0.3027     0.2613   1.16  0.248
## Year2005          0.4797     0.2678   1.79  0.075 .
## Year2006         -0.1428     0.1867  -0.76  0.445
```

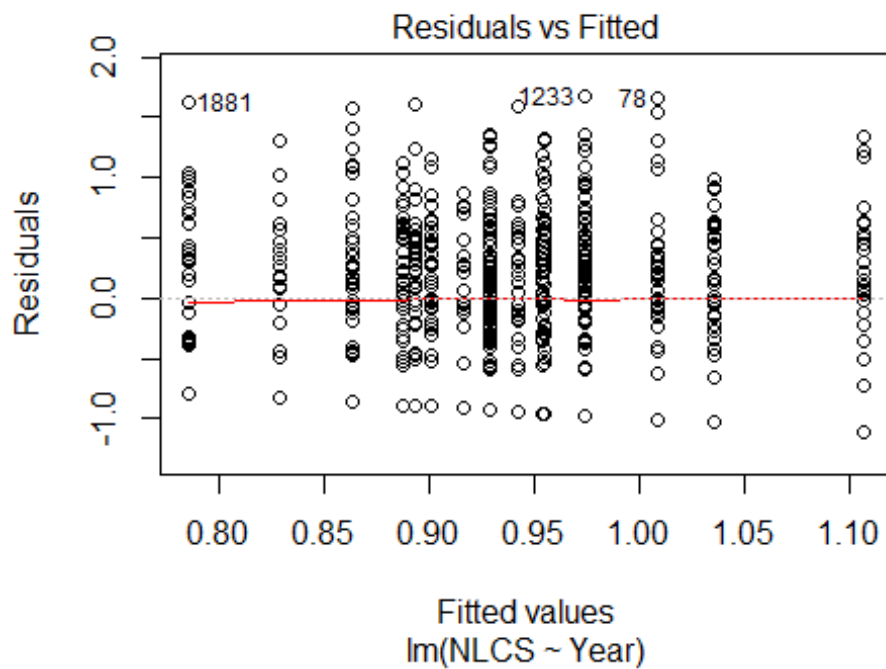
```

## Year2007          0.4192      0.2129      1.97      0.051 .
## Year2008          0.0558      0.2302      0.24      0.809
## Year2009          0.4923      0.2603      1.89      0.060 .
## Year2010          0.3162      0.2345      1.35      0.179
## Year2011         -0.1034      0.2595     -0.40      0.691
## Year2012          0.3315      0.2111      1.57      0.118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.602
## Multiple R-squared:  0.119, Adjusted R-squared:  0.0336
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 179 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.151  0.878  0.945  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      5.18e-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: 193"
## [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
##   81   90   69   94   83   78   67   65   91   96   83  121   99  111   99
## 2011 2012
##  101  114
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   49   43   35   44   41   42   41   37   47   58   52   84   66   65   57
## 2011 2012

```



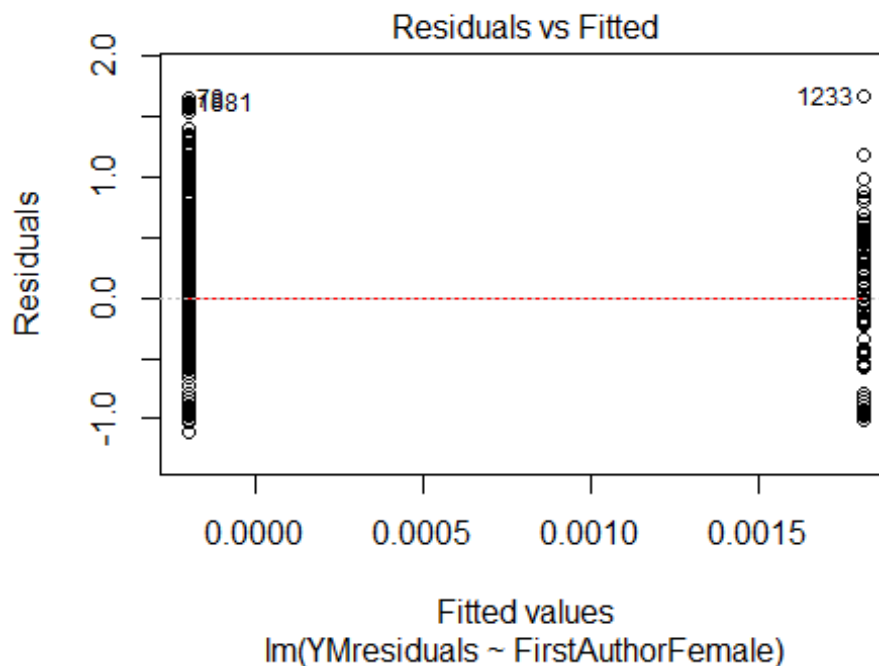
```
## 66 82
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 42 39 29 39 31 39 36 35 46 52 51 76 59 61 51
## 2011 2012
## 60 69
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 20, df = 16, p-value = 0.2
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.028, df = 1, p-value = 0.9
## [1] "Female first author team size 2018 geometric mean: 1.10408951367381"
## [1] "Male first author team size 2018 geometric mean: 1.30382731611844"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

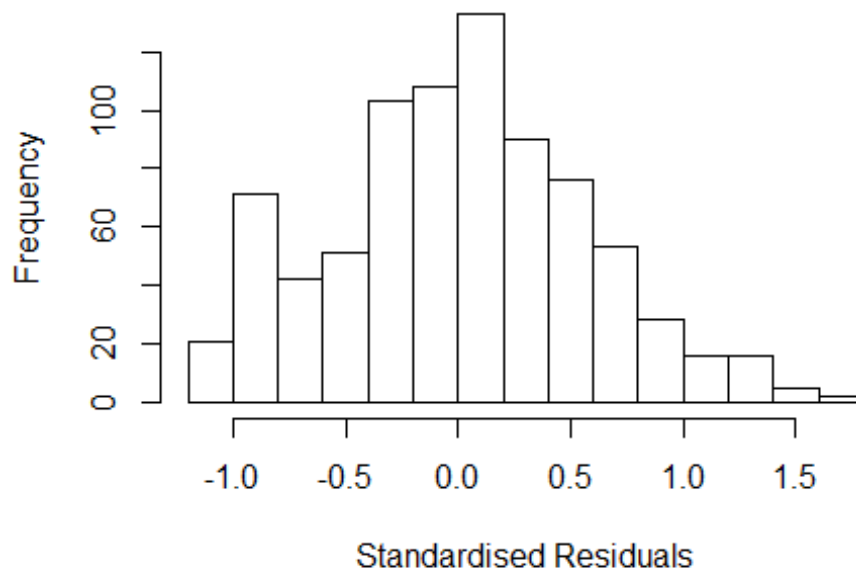
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.16993081275869"
## [1] "Male last author team size 2018 geometric mean: 1.29308292343605"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.4
## 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 1.798 1      1.341
## LastAuthorFemale  1.600 1      1.265
## UniqueAuthors    18.369 4      1.439
## Year              19.109 16     1.097
```

Residuals from first and last author and team size



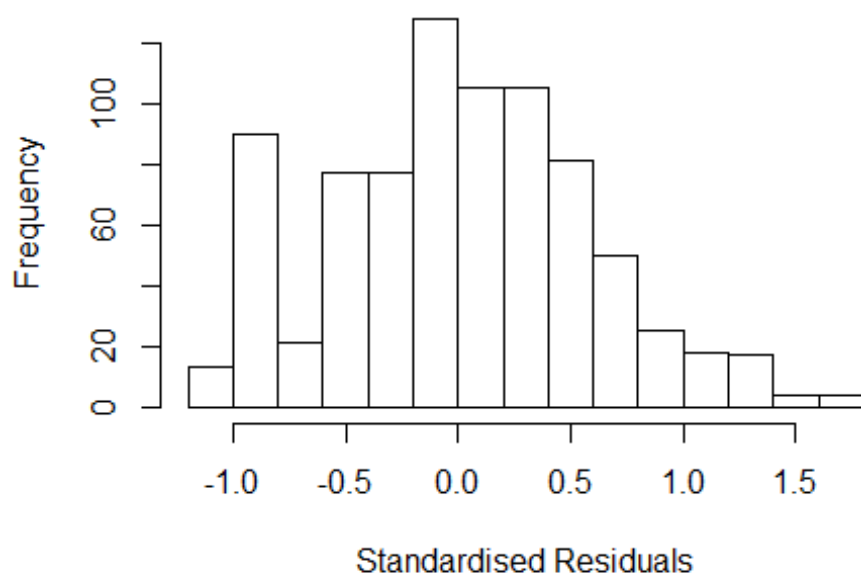
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1414 -0.3725 0.0289 0.3790 1.6418
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0222 0.0995 10.27 < 2e-16 ***
## FirstAuthorFemale1 0.0279 0.0807 0.35 0.72928
## LastAuthorFemale1 -0.1044 0.0948 -1.10 0.27085
## UniqueAuthors2 0.1975 0.0541 3.65 0.00028 ***
## UniqueAuthors3 0.1224 0.0925 1.32 0.18611
## UniqueAuthors4 0.2028 0.1851 1.10 0.27349
## UniqueAuthors5 0.8982 0.0759 11.84 < 2e-16 ***
## Year1997 -0.0861 0.1313 -0.66 0.51235
## Year1998 -0.1701 0.1501 -1.13 0.25728
## Year1999 0.0350 0.1460 0.24 0.81050
```

```

## Year2000          -0.0444      0.1441   -0.31  0.75826
## Year2001          -0.1164      0.1407   -0.83  0.40815
## Year2002          -0.1332      0.1377   -0.97  0.33393
## Year2003          -0.2128      0.1462   -1.46  0.14598
## Year2004          -0.1085      0.1202   -0.90  0.36712
## Year2005          -0.1746      0.1266   -1.38  0.16802
## Year2006          -0.0782      0.1385   -0.56  0.57267
## Year2007          -0.1265      0.1214   -1.04  0.29764
## Year2008          -0.2079      0.1157   -1.80  0.07277 .
## Year2009          -0.2114      0.1237   -1.71  0.08795 .
## Year2010          -0.3102      0.1343   -2.31  0.02120 *
## Year2011          -0.2940      0.1311   -2.24  0.02520 *
## Year2012          -0.1200      0.1266   -0.95  0.34345
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.579
## Multiple R-squared:  0.0434, Adjusted R-squared:  0.0169
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 738 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.402  0.861  0.951  0.909  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          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 1.577 1          1.256
## LastAuthorFemale  1.501 1          1.225
## Year              1.222 16          1.006

```

Residuals from first and last author



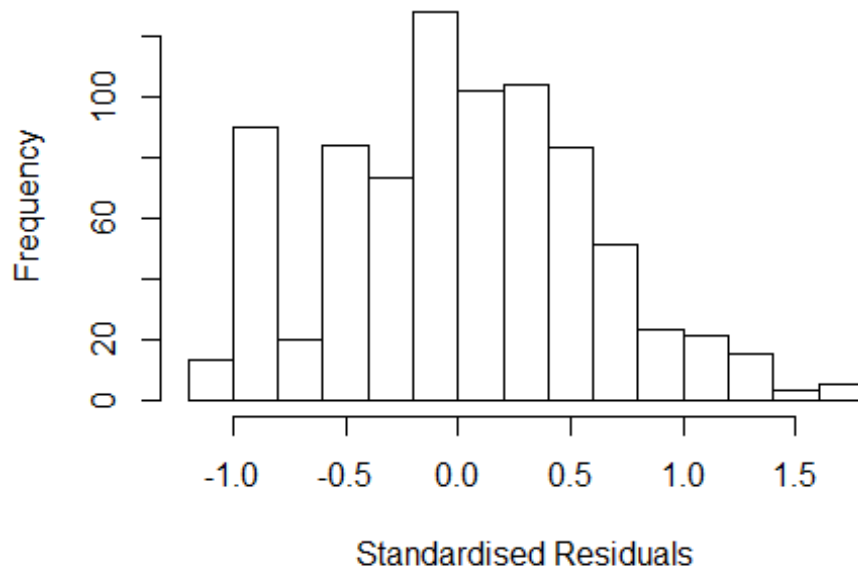
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.10530 -0.38888  0.00242  0.39172  1.66807
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0603     0.1009   10.51  <2e-16 ***
## FirstAuthorFemale1  0.0791     0.0819    0.97   0.334
## LastAuthorFemale1 -0.1352     0.0968   -1.40   0.163
## Year1997          -0.1010     0.1313   -0.77   0.442
## Year1998          -0.1566     0.1463   -1.07   0.285
## Year1999           0.0450     0.1482    0.30   0.762
## Year2000          -0.0431     0.1495   -0.29   0.773
## Year2001          -0.1239     0.1424   -0.87   0.384
## Year2002          -0.1148     0.1398   -0.82   0.412
## Year2003          -0.2199     0.1483   -1.48   0.139
## Year2004          -0.1069     0.1238   -0.86   0.388
## Year2005          -0.1700     0.1280   -1.33   0.184
```

```

## Year2006          -0.0780      0.1426   -0.55    0.585
## Year2007          -0.1157      0.1238   -0.94    0.350
## Year2008          -0.1880      0.1166   -1.61    0.107
## Year2009          -0.1748      0.1266   -1.38    0.168
## Year2010          -0.3224      0.1348   -2.39    0.017 *
## Year2011          -0.2480      0.1350   -1.84    0.066 .
## Year2012          -0.0941      0.1287   -0.73    0.465
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0222, Adjusted R-squared:  0.000129
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 738 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.399  0.861  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      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 1.129 1         1.063
## Year              1.129 16         1.004

```

Residuals from first author



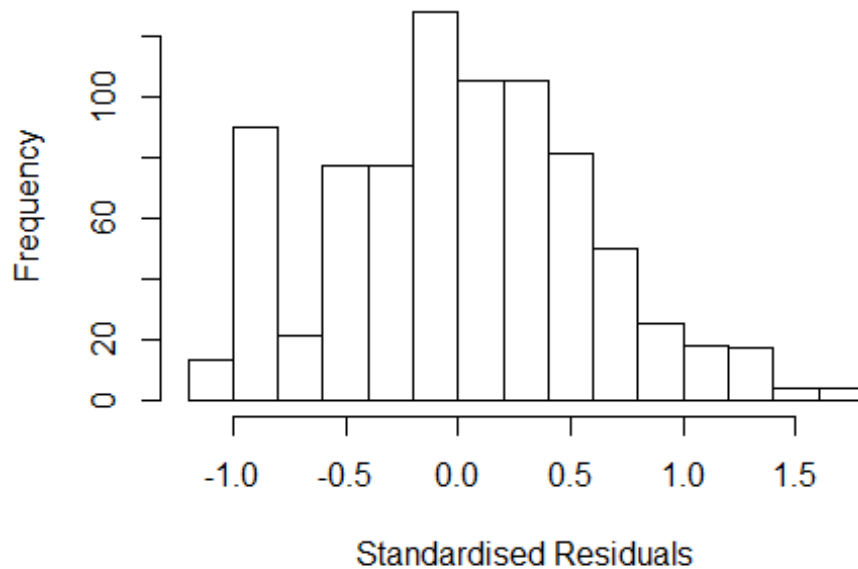
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.10893 -0.41348 -0.00314 0.39394 1.66710
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05896 0.10230 10.35 <2e-16 ***
## FirstAuthorFemale1 0.00737 0.07138 0.10 0.918
## Year1997 -0.10267 0.13214 -0.78 0.437
## Year1998 -0.15987 0.14805 -1.08 0.281
## Year1999 0.04997 0.14823 0.34 0.736
## Year2000 -0.04125 0.15115 -0.27 0.785
## Year2001 -0.13269 0.14252 -0.93 0.352
## Year2002 -0.11968 0.13995 -0.86 0.393
## Year2003 -0.23365 0.14831 -1.58 0.116
## Year2004 -0.11003 0.12514 -0.88 0.380
## Year2005 -0.17073 0.12921 -1.32 0.187
## Year2006 -0.08843 0.14352 -0.62 0.538
```

```

## Year2007      -0.11971    0.12512   -0.96    0.339
## Year2008      -0.18605    0.11786   -1.58    0.115
## Year2009      -0.17418    0.12797   -1.36    0.174
## Year2010      -0.31482    0.13613   -2.31    0.021 *
## Year2011      -0.24847    0.13508   -1.84    0.066 .
## Year2012      -0.09896    0.13001   -0.76    0.447
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.02,   Adjusted R-squared:  -0.000928
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 731 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.399  0.857   0.948   0.908   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 1.068 1          1.033
## Year              1.068 16          1.002

```


Residuals from last author



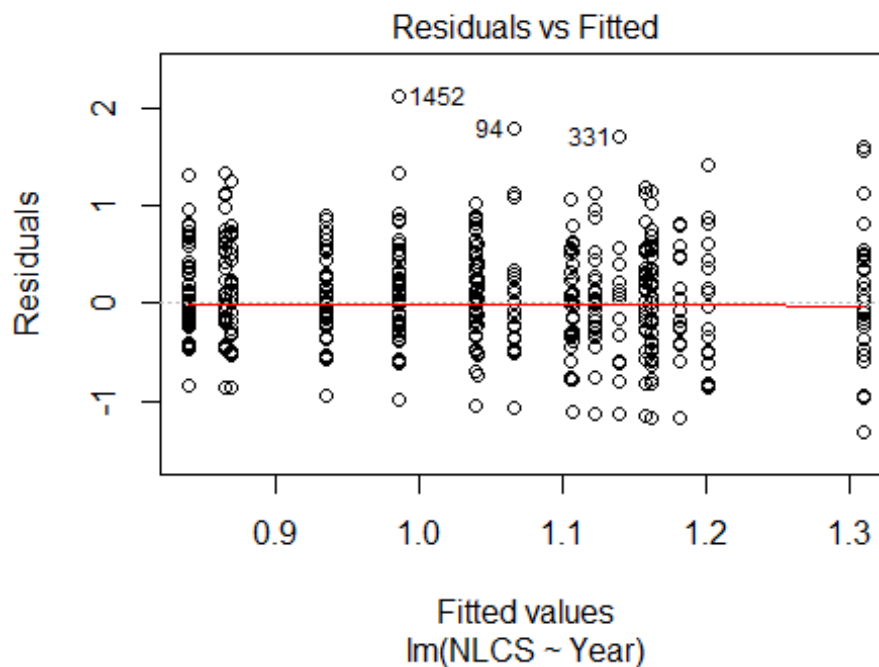
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11548 -0.38233  0.00244  0.38889  1.66378
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0646     0.1005  10.60  <2e-16 ***
## LastAuthorFemale1 -0.0822     0.0807  -1.02   0.309
## Year1997         -0.1041     0.1310  -0.79   0.427
## Year1998         -0.1583     0.1464  -1.08   0.280
## Year1999          0.0509     0.1471   0.35   0.729
## Year2000         -0.0393     0.1490  -0.26   0.792
## Year2001         -0.1276     0.1424  -0.90   0.371
## Year2002         -0.1193     0.1392  -0.86   0.392
## Year2003         -0.2168     0.1472  -1.47   0.141
## Year2004         -0.1093     0.1234  -0.89   0.376
## Year2005         -0.1719     0.1280  -1.34   0.179
## Year2006         -0.0833     0.1418  -0.59   0.557
```

```

## Year2007          -0.1176      0.1237   -0.95      0.342
## Year2008          -0.1850      0.1162   -1.59      0.112
## Year2009          -0.1714      0.1264   -1.36      0.176
## Year2010          -0.3184      0.1349   -2.36      0.018 *
## Year2011          -0.2468      0.1342   -1.84      0.066 .
## Year2012          -0.0977      0.1282   -0.76      0.446
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0212, Adjusted R-squared:  0.000363
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 737 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.400  0.858  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      1.23e-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 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
##   62   60   47   42   48   56   45   28   55   68   77   80   87  115  104
## 2011 2012
##   89   85
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   30   25   18   16   17   17   19   12   28   39   41   34   51   59   62
## 2011 2012

```

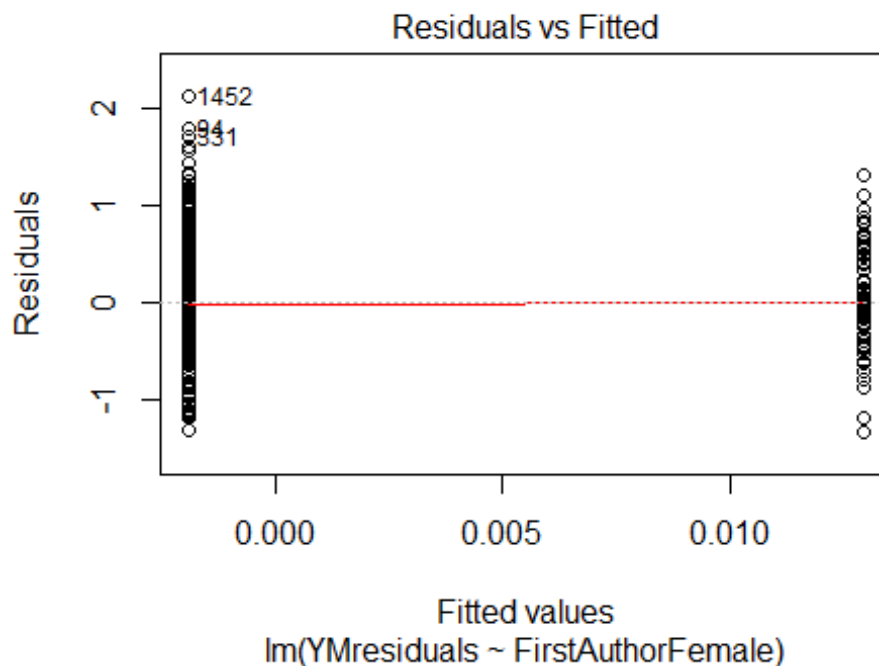
```
## 50 52
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 29 20 17 15 14 14 13 10 24 34 37 32 41 50 52
## 2011 2012
## 46 41
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 11, df = 16, p-value = 0.8
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.19, df = 1, p-value = 0.7
## [1] "Female first author team size 2018 geometric mean: 1.36426160182137"
## [1] "Male first author team size 2018 geometric mean: 1.42938420364513"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.12246204830937"
## [1] "Male last author team size 2018 geometric mean: 1.4684031608208"

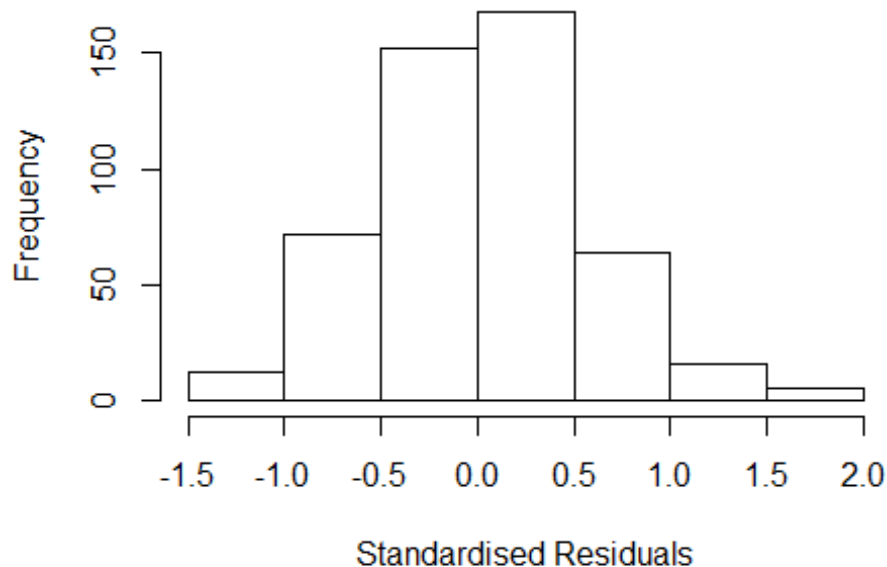
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 76, p-value = 0.1
## 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	2.063	1	1.436
LastAuthorFemale	2.002	1	1.415
UniqueAuthors	1.412	3	1.059
Year	1.760	16	1.018

Residuals from first and last author and team size



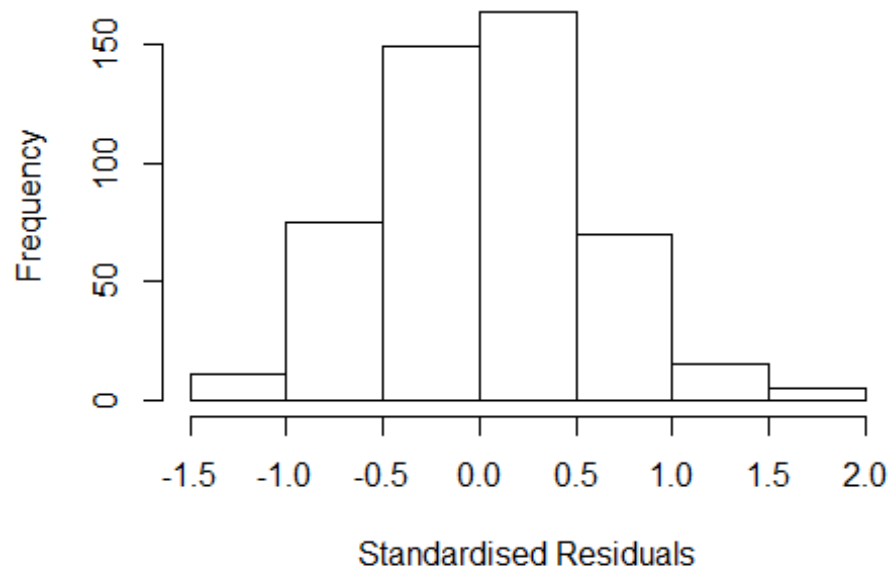
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2685 -0.3599 0.0227 0.3580 1.8469
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2685 0.1354 9.37 <2e-16 ***
## FirstAuthorFemale1 0.1380 0.1032 1.34 0.1819
## LastAuthorFemale1 -0.1551 0.0989 -1.57 0.1175
## UniqueAuthors2 0.0565 0.0587 0.96 0.3363
## UniqueAuthors3 0.1231 0.0960 1.28 0.2002
## UniqueAuthors4 0.0865 0.2427 0.36 0.7216
## Year1997 -0.3249 0.1694 -1.92 0.0557 .
## Year1998 -0.2329 0.1744 -1.34 0.1825
## Year1999 -0.1494 0.1863 -0.80 0.4230
## Year2000 -0.1843 0.1974 -0.93 0.3509
```

```

## Year2001          -0.3088      0.2287   -1.35    0.1777
## Year2002          -0.2626      0.1895   -1.39    0.1664
## Year2003          -0.2034      0.1803   -1.13    0.2599
## Year2004          -0.2521      0.1785   -1.41    0.1586
## Year2005          -0.1595      0.1679   -0.95    0.3426
## Year2006          -0.1708      0.1708   -1.00    0.3177
## Year2007          -0.4476      0.1757   -2.55    0.0112 *
## Year2008          -0.3297      0.1612   -2.04    0.0414 *
## Year2009          -0.4010      0.1575   -2.55    0.0112 *
## Year2010          -0.3514      0.1524   -2.31    0.0215 *
## Year2011          -0.5006      0.1587   -3.15    0.0017 **
## Year2012          -0.5058      0.1719   -2.94    0.0034 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.0698, Adjusted R-squared:  0.0279
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 457 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.219  0.868  0.956  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          2.04e-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 2.066 1 1.437
## LastAuthorFemale 2.045 1 1.430
## Year 1.304 16 1.008

```

Residuals from first and last author



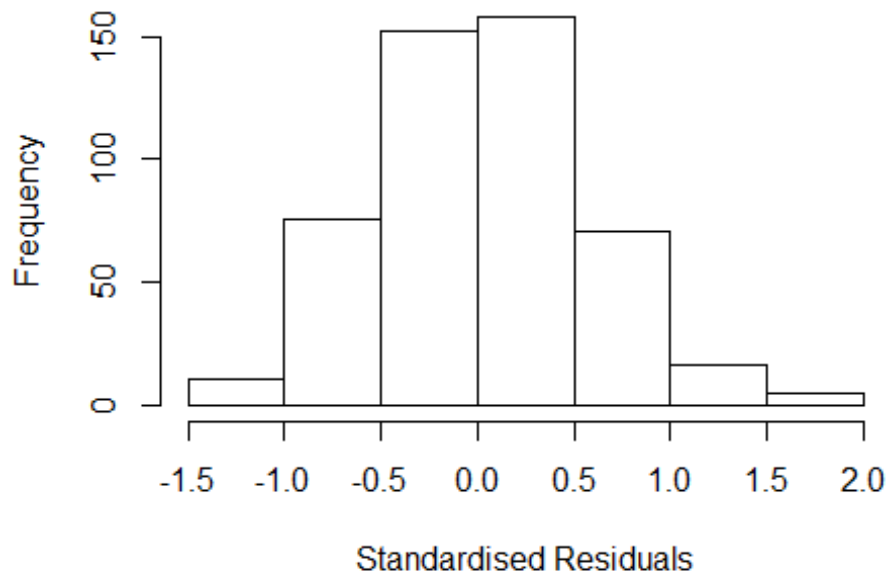
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2906 -0.3719  0.0385  0.3370  1.8796
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.291      0.135   9.59  <2e-16 ***
## FirstAuthorFemale1  0.143      0.103   1.39  0.1656
## LastAuthorFemale1 -0.160      0.101  -1.58  0.1140
## Year1997          -0.323      0.170  -1.90  0.0574 .
## Year1998          -0.248      0.174  -1.43  0.1539
## Year1999          -0.138      0.189  -0.73  0.4640
## Year2000          -0.201      0.196  -1.02  0.3060
## Year2001          -0.295      0.226  -1.31  0.1916
## Year2002          -0.266      0.190  -1.40  0.1609
## Year2003          -0.181      0.184  -0.98  0.3252
## Year2004          -0.250      0.176  -1.43  0.1545
## Year2005          -0.152      0.164  -0.93  0.3552
```

```

## Year2006          -0.164      0.171   -0.96   0.3357
## Year2007          -0.450      0.174   -2.58   0.0102 *
## Year2008          -0.319      0.159   -2.01   0.0449 *
## Year2009          -0.396      0.154   -2.57   0.0106 *
## Year2010          -0.345      0.152   -2.27   0.0234 *
## Year2011          -0.501      0.158   -3.17   0.0016 **
## Year2012          -0.486      0.171   -2.84   0.0047 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.0648, Adjusted R-squared:  0.029
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 450 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.862  0.954  0.904  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.04e-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 1.185 1      1.089
## Year              1.185 16      1.005

```


Residuals from first author



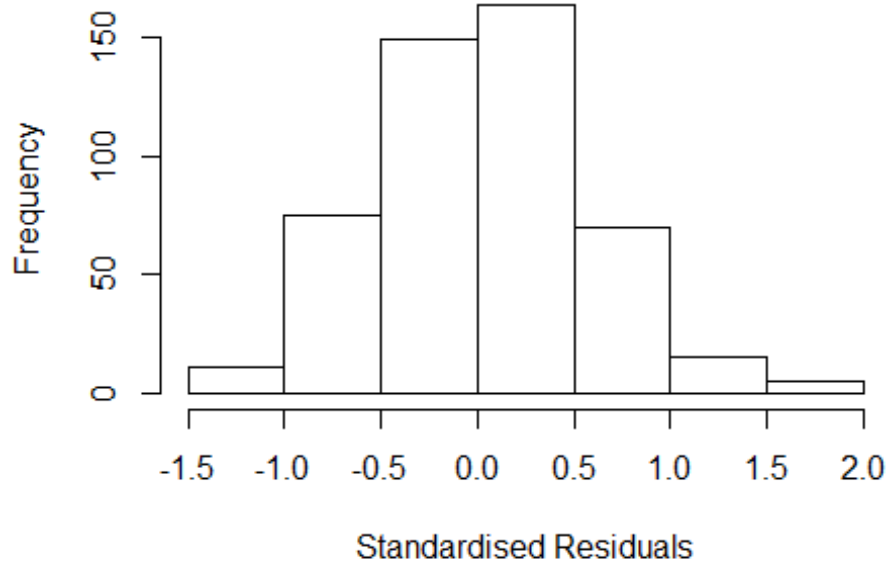
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3212 -0.3648 0.0337 0.3495 1.8912
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2871 0.1352 9.52 <2e-16 ***
## FirstAuthorFemale1 0.0341 0.0790 0.43 0.6665
## Year1997 -0.3313 0.1704 -1.94 0.0525 .
## Year1998 -0.2492 0.1757 -1.42 0.1568
## Year1999 -0.1479 0.1873 -0.79 0.4301
## Year2000 -0.2069 0.1956 -1.06 0.2905
## Year2001 -0.2904 0.2262 -1.28 0.1998
## Year2002 -0.2553 0.1916 -1.33 0.1833
## Year2003 -0.1960 0.1828 -1.07 0.2841
## Year2004 -0.2412 0.1750 -1.38 0.1689
## Year2005 -0.1526 0.1665 -0.92 0.3598
## Year2006 -0.1609 0.1718 -0.94 0.3493
```

```

## Year2007          -0.4473      0.1767    -2.53    0.0117 *
## Year2008          -0.3190      0.1590    -2.01    0.0454 *
## Year2009          -0.3950      0.1547    -2.55    0.0110 *
## Year2010          -0.3418      0.1521    -2.25    0.0251 *
## Year2011          -0.5143      0.1573    -3.27    0.0012 **
## Year2012          -0.4901      0.1708    -2.87    0.0043 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.541
## Multiple R-squared:  0.0613, Adjusted R-squared:  0.0274
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 452 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.868  0.956  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      2.04e-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 1.173 1          1.083
## Year            1.173 16          1.005

```

Residuals from last author



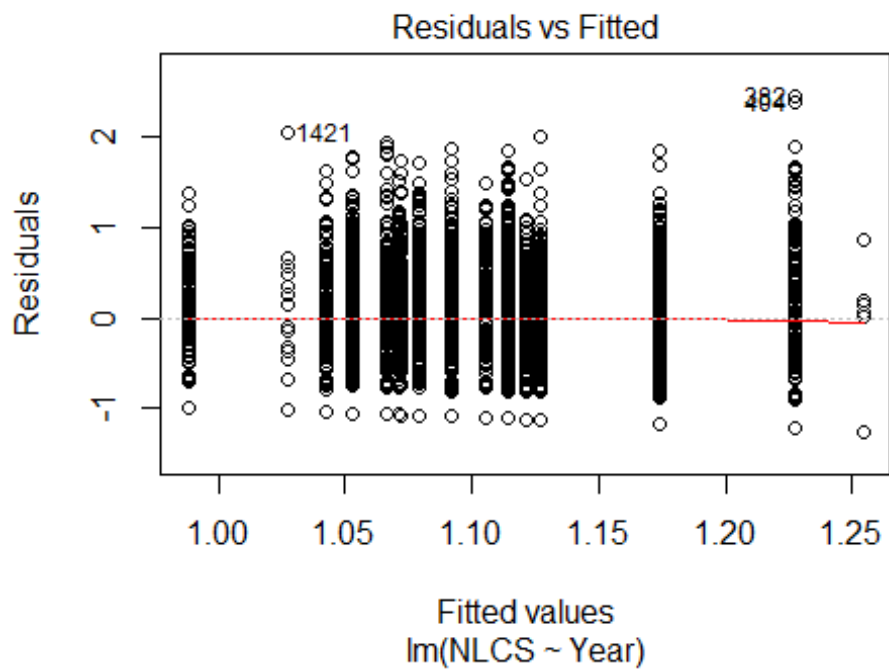
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2944 -0.3776 0.0295 0.3393 1.8690
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2944 0.1341 9.66 <2e-16 ***
## LastAuthorFemale1 -0.0551 0.0769 -0.72 0.4739
## Year1997 -0.3164 0.1690 -1.87 0.0618 .
## Year1998 -0.2366 0.1748 -1.35 0.1766
## Year1999 -0.1507 0.1869 -0.81 0.4206
## Year2000 -0.1972 0.1973 -1.00 0.3181
## Year2001 -0.2976 0.2254 -1.32 0.1875
## Year2002 -0.2545 0.1885 -1.35 0.1775
## Year2003 -0.1970 0.1823 -1.08 0.2804
## Year2004 -0.2354 0.1734 -1.36 0.1752
## Year2005 -0.1498 0.1661 -0.90 0.3678
## Year2006 -0.1613 0.1708 -0.94 0.3454
```

```

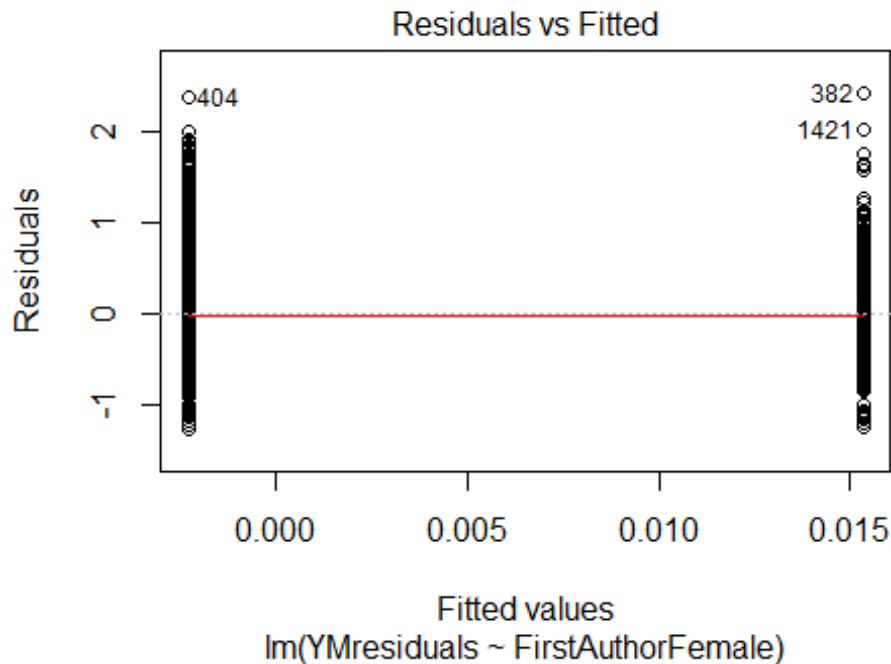
## Year2007          -0.4458      0.1737   -2.57   0.0106 *
## Year2008          -0.3117      0.1575   -1.98   0.0484 *
## Year2009          -0.3939      0.1539   -2.56   0.0108 *
## Year2010          -0.3449      0.1508   -2.29   0.0226 *
## Year2011          -0.5032      0.1574   -3.20   0.0015 **
## Year2012          -0.4846      0.1715   -2.83   0.0049 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.0616, Adjusted R-squared:  0.0277
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 453 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.210  0.874  0.955  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      2.04e-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: 489"
## [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
## 385 325 332 24 36 403 370 291 325 362 460 504 555 658 575
## 2011 2012
## 562 583
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 176 131 122 6 16 148 195 139 139 186 224 252 305 368 332
## 2011 2012

```

```
## 313 323
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 160 109 110 5 14 118 159 117 117 155 177 202 247 288 259
## 2011 2012
## 246 268
## [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
```

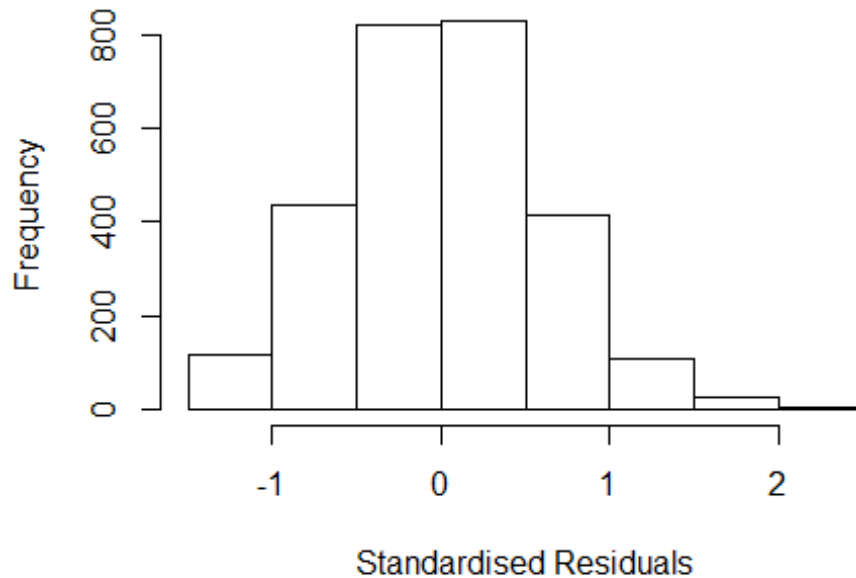


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.1, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 1.86349733927431"
## [1] "Male first author team size 2018 geometric mean: 1.93673341052427"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3600, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.8988837854649"
## [1] "Male last author team size 2018 geometric mean: 1.92873057968116"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3200, 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 1.152 1      1.073
## LastAuthorFemale  1.133 1      1.064
## UniqueAuthors     1.160 4      1.019
## Year              1.180 16     1.005
```

Residuals from first and last author and team size



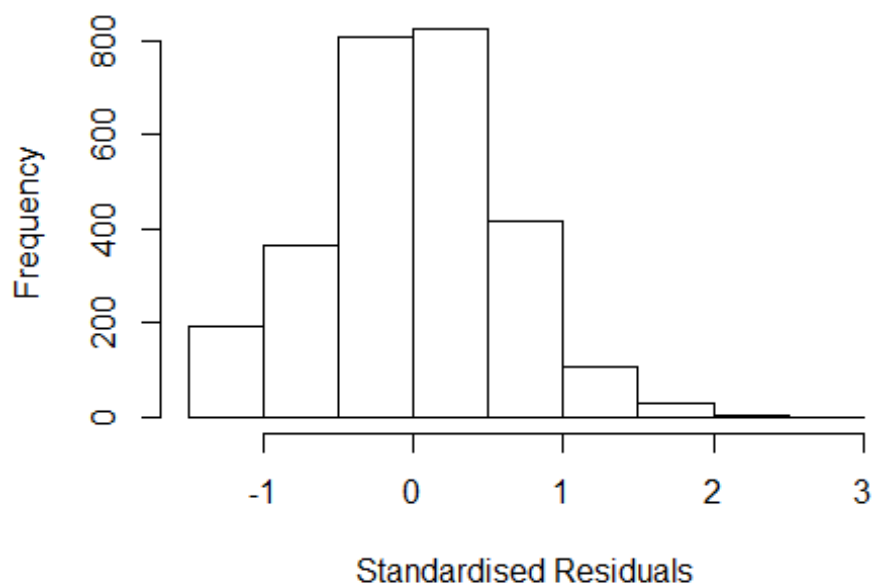
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.43169 -0.39002 0.00152 0.40170 2.44056
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0897 0.0605 18.00 < 2e-16 ***
## FirstAuthorFemale1 -0.0240 0.0341 -0.70 0.4819
## LastAuthorFemale1 0.0319 0.0393 0.81 0.4178
## UniqueAuthors2 0.1667 0.0274 6.08 1.4e-09 ***
## UniqueAuthors3 0.2352 0.0341 6.89 6.8e-12 ***
## UniqueAuthors4 0.1744 0.0642 2.72 0.0067 **
## UniqueAuthors5 0.2977 0.0707 4.21 2.6e-05 ***
## Year1997 -0.0513 0.0850 -0.60 0.5467
## Year1998 -0.1399 0.0822 -1.70 0.0891 .
## Year1999 0.1067 0.3089 0.35 0.7297
```

```

## Year2000      -0.2558      0.1829      -1.40      0.1620
## Year2001      -0.0953      0.0858      -1.11      0.2668
## Year2002      -0.2130      0.0832      -2.56      0.0106 *
## Year2003      -0.1515      0.0842      -1.80      0.0721 .
## Year2004      -0.2032      0.0771      -2.63      0.0085 **
## Year2005      -0.1488      0.0790      -1.88      0.0599 .
## Year2006      -0.0853      0.0724      -1.18      0.2385
## Year2007      -0.1052      0.0714      -1.47      0.1405
## Year2008      -0.1265      0.0722      -1.75      0.0796 .
## Year2009      -0.1615      0.0712      -2.27      0.0234 *
## Year2010      -0.0699      0.0695      -1.01      0.3143
## Year2011      -0.1658      0.0721      -2.30      0.0216 *
## Year2012      -0.2336      0.0710      -3.29      0.0010 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.589
## Multiple R-squared:  0.0338, Adjusted R-squared:  0.026
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 232 weights are ~= 1. The remaining 2519 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0471 0.8680 0.9500 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          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 1.124 1 1.060
## LastAuthorFemale 1.126 1 1.061
## Year 1.043 16 1.001

```


Residuals from first and last author



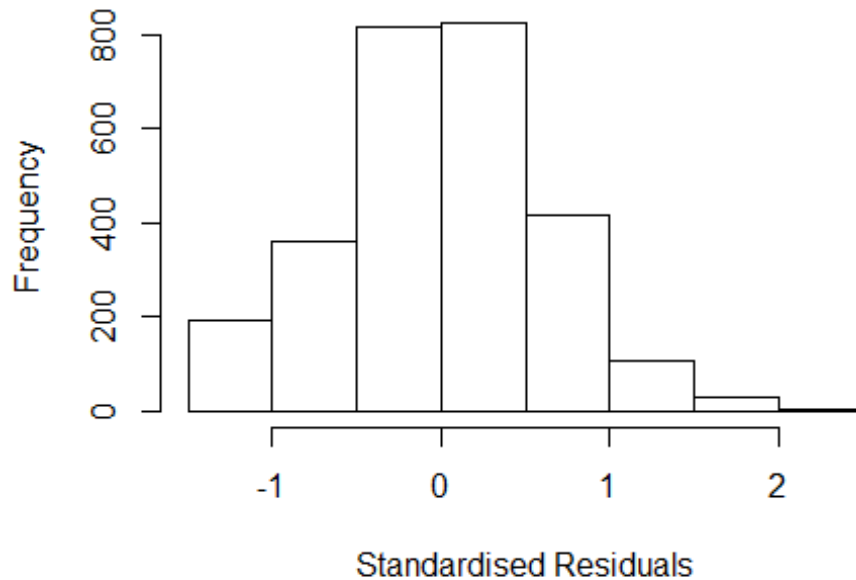
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 382 0030151578 3.673 1996      1711      3      2.51
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32163 -0.38831  0.00577  0.40644  2.51017
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1594     0.0598   19.40  <2e-16 ***
## FirstAuthorFemale1  0.0034     0.0341    0.10   0.921
## LastAuthorFemale1  0.0339     0.0406    0.84   0.403
## Year1997         -0.0505     0.0839   -0.60   0.547
## Year1998         -0.1221     0.0830   -1.47   0.142
## Year1999          0.1622     0.3138    0.52   0.605
## Year2000         -0.2394     0.1900   -1.26   0.208
## Year2001         -0.0513     0.0847   -0.61   0.545
## Year2002         -0.1573     0.0825   -1.91   0.057 .
## Year2003         -0.1250     0.0854   -1.46   0.143
## Year2004         -0.1701     0.0771   -2.20   0.028 *
## Year2005         -0.1071     0.0782   -1.37   0.171
```

```

## Year2006          -0.0447      0.0721    -0.62      0.536
## Year2007          -0.0595      0.0715    -0.83      0.405
## Year2008          -0.0776      0.0720    -1.08      0.281
## Year2009          -0.1032      0.0712    -1.45      0.148
## Year2010          -0.0236      0.0692    -0.34      0.733
## Year2011          -0.1064      0.0718    -1.48      0.138
## Year2012          -0.1697      0.0704    -2.41      0.016 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.00787,    Adjusted R-squared:  0.00134
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 239 weights are ~= 1. The remaining 2512 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0344 0.8650 0.9480 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      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 1.021 1      1.011
## Year              1.021 16      1.001

```

Residuals from first author



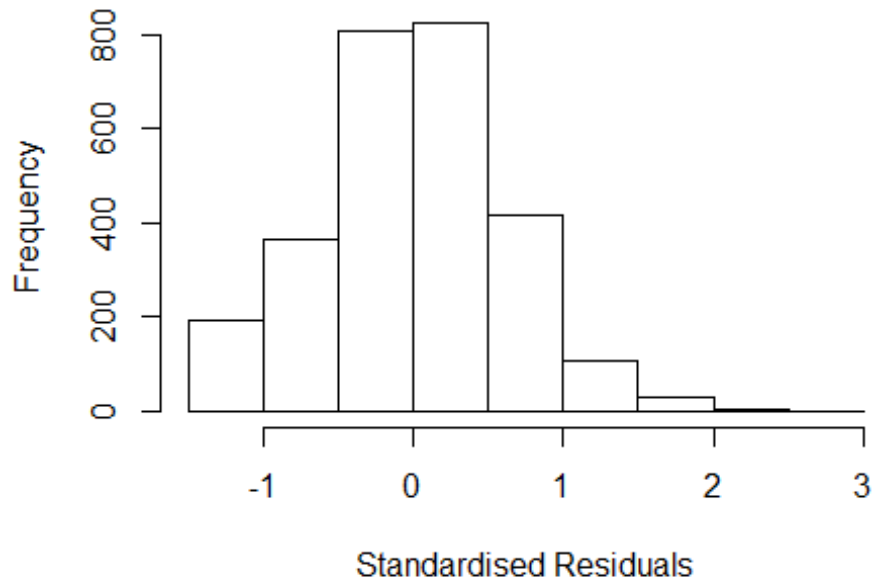
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 382 0030151578 3.673 1996      1711      3      2.51
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.31843 -0.38810  0.00334  0.40505  2.49798
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1609     0.0597   19.44  <2e-16 ***
## FirstAuthorFemale1  0.0141     0.0325    0.43   0.665
## Year1997        -0.0482     0.0837   -0.58   0.565
## Year1998        -0.1215     0.0830   -1.46   0.143
## Year1999         0.1575     0.3139    0.50   0.616
## Year2000        -0.2412     0.1905   -1.27   0.206
## Year2001        -0.0523     0.0846   -0.62   0.537
## Year2002        -0.1580     0.0826   -1.91   0.056 .
## Year2003        -0.1218     0.0851   -1.43   0.152
## Year2004        -0.1689     0.0772   -2.19   0.029 *
## Year2005        -0.1073     0.0782   -1.37   0.170
## Year2006        -0.0421     0.0721   -0.58   0.559
```

```

## Year2007          -0.0595      0.0714   -0.83    0.405
## Year2008          -0.0777      0.0719   -1.08    0.280
## Year2009          -0.1022      0.0712   -1.44    0.151
## Year2010          -0.0230      0.0691   -0.33    0.740
## Year2011          -0.1063      0.0717   -1.48    0.139
## Year2012          -0.1693      0.0704   -2.40    0.016 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.00762,    Adjusted R-squared:  0.00145
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 247 weights are ~= 1. The remaining 2504 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0377 0.8660 0.9480 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      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 1.023 1          1.011
## Year              1.023 16          1.001

```

Residuals from last author



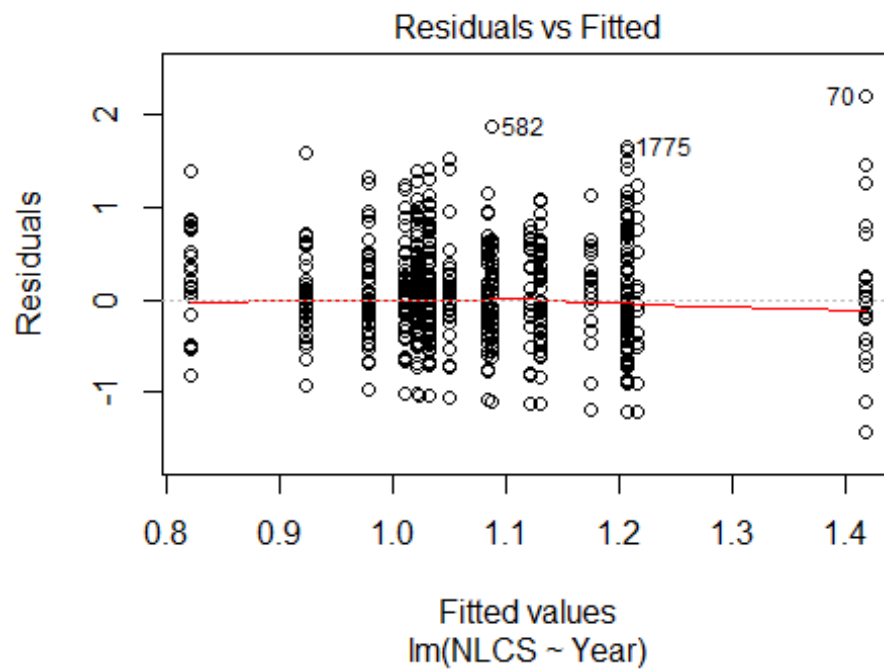
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 382 0030151578 3.673 1996      1711      3      2.51
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32269 -0.38875  0.00532  0.40609  2.51341
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1596     0.0597   19.41  <2e-16 ***
## LastAuthorFemale1  0.0353     0.0387    0.91   0.362
## Year1997        -0.0504     0.0839   -0.60   0.548
## Year1998        -0.1219     0.0830   -1.47   0.142
## Year1999         0.1631     0.3133    0.52   0.603
## Year2000        -0.2396     0.1900   -1.26   0.207
## Year2001        -0.0512     0.0847   -0.60   0.546
## Year2002        -0.1571     0.0824   -1.91   0.057 .
## Year2003        -0.1250     0.0854   -1.46   0.143
## Year2004        -0.1701     0.0771   -2.20   0.028 *
## Year2005        -0.1068     0.0781   -1.37   0.172
## Year2006        -0.0446     0.0721   -0.62   0.536
```

```

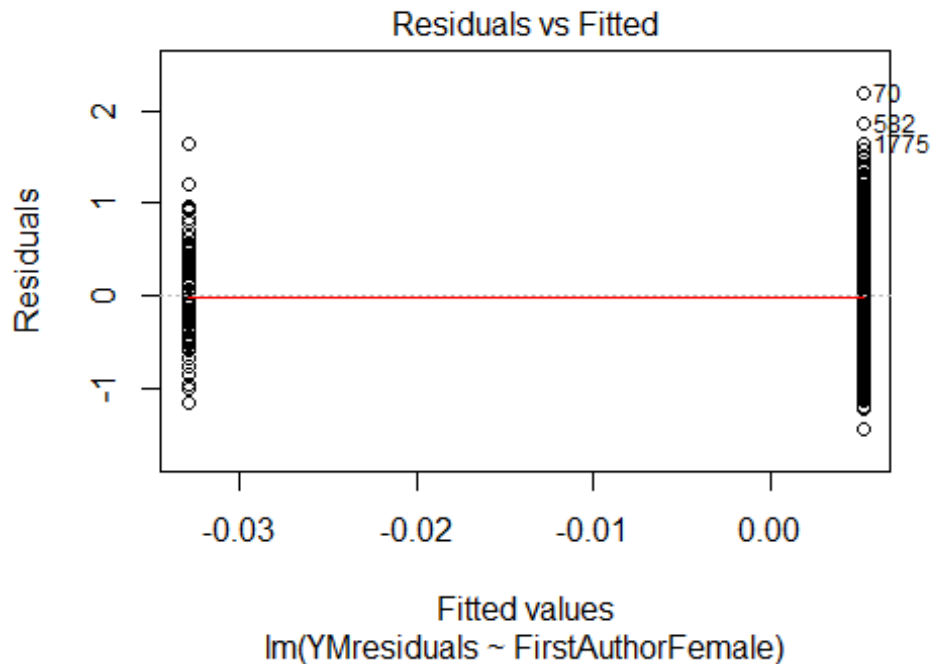
## Year2007          -0.0594      0.0715   -0.83    0.406
## Year2008          -0.0774      0.0720   -1.08    0.282
## Year2009          -0.1029      0.0712   -1.45    0.148
## Year2010          -0.0235      0.0692   -0.34    0.734
## Year2011          -0.1062      0.0717   -1.48    0.139
## Year2012          -0.1695      0.0704   -2.41    0.016 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.593
## Multiple R-squared:  0.00787,    Adjusted R-squared:  0.0017
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 240 weights are ~= 1. The remaining 2511 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0335 0.8650 0.9480 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      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: 2751"
## [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
##   72   71   65   55   64   64   69   65   93   73  105  113  112  152  132
## 2011 2012
##  145  106
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   29   29   16   23   22   34   25   33   39   50   53   52   87   75
## 2011 2012

```

```
##      87      63
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    22    27    27    15    21    20    28    23    30    37    41    41    40    69    69
## 2011 2012
##    70    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 = 20, df = 16, p-value = 0.2
```

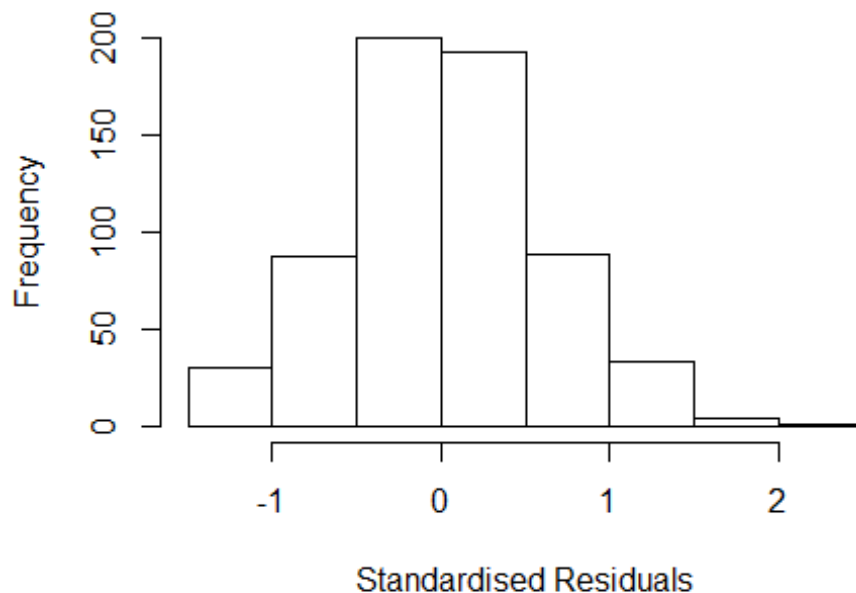


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.7, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 3.12031850602576"
## [1] "Male first author team size 2018 geometric mean: 2.21579602489623"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 280, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.44948974278318"
## [1] "Male last author team size 2018 geometric mean: 2.29192515034791"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 72, 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 1.504 1      1.226
## LastAuthorFemale  1.460 1      1.208
## UniqueAuthors    1.828 4      1.078
## Year              2.134 16     1.024
```


Residuals from first and last author and team size



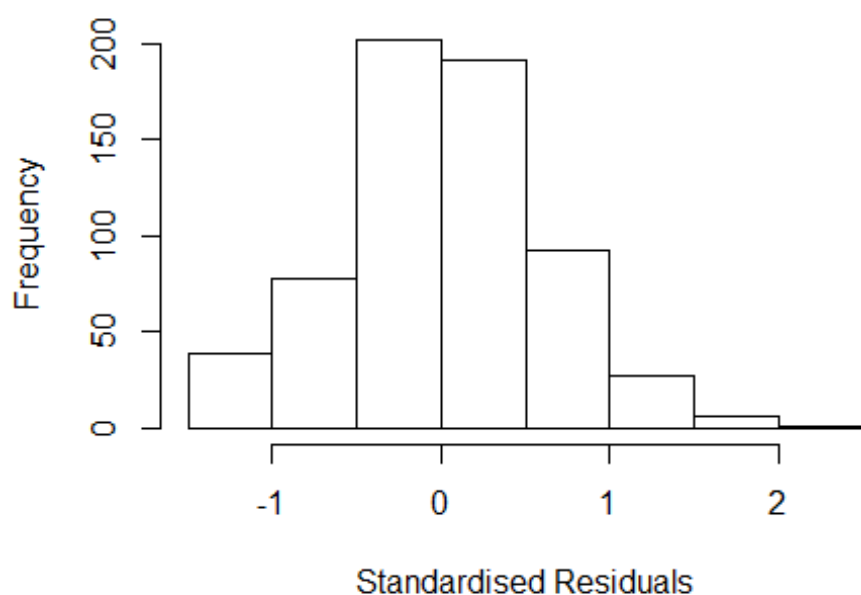
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.48005 -0.33877  0.00327  0.37057  2.22975
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.31011    0.16182   8.10 3.1e-15 ***
## FirstAuthorFemale1 -0.00928    0.06813  -0.14  0.8917
## LastAuthorFemale1 -0.09021    0.09006  -1.00  0.3169
## UniqueAuthors2     0.06714    0.05862   1.15  0.2525
## UniqueAuthors3     0.13443    0.07221   1.86  0.0631 .
## UniqueAuthors4     0.17983    0.10257   1.75  0.0800 .
## UniqueAuthors5     0.37677    0.13950   2.70  0.0071 **
## Year1997         -0.46437    0.21270  -2.18  0.0294 *
## Year1998         -0.12942    0.20640  -0.63  0.5309
## Year1999         -0.06507    0.29637  -0.22  0.8263
```

```

## Year2000      -0.26415      0.19117      -1.38      0.1675
## Year2001      -0.21865      0.20156      -1.08      0.2784
## Year2002      -0.31562      0.18350      -1.72      0.0859 .
## Year2003      -0.23341      0.20343      -1.15      0.2517
## Year2004      -0.46739      0.18860      -2.48      0.0135 *
## Year2005      -0.38238      0.17935      -2.13      0.0334 *
## Year2006      -0.30641      0.19113      -1.60      0.1094
## Year2007      -0.26396      0.18500      -1.43      0.1541
## Year2008      -0.41082      0.18869      -2.18      0.0298 *
## Year2009      -0.41078      0.17589      -2.34      0.0198 *
## Year2010      -0.20683      0.17868      -1.16      0.2475
## Year2011      -0.33795      0.17606      -1.92      0.0554 .
## Year2012      -0.36818      0.17108      -2.15      0.0318 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.052, Adjusted R-squared:  0.0179
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 568 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0646 0.8380 0.9500 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          1.57e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.363 1 1.168
## LastAuthorFemale 1.389 1 1.179
## Year 1.267 16 1.007

```

Residuals from first and last author



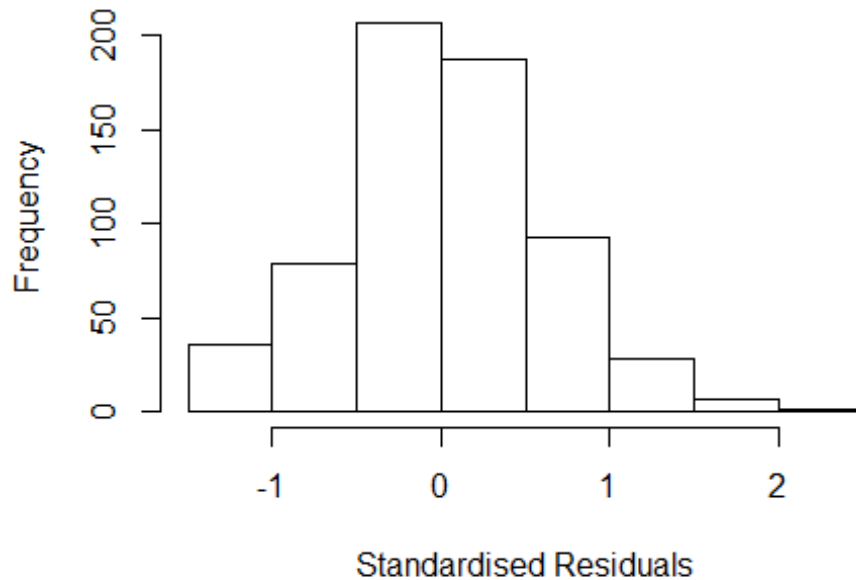
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.36654 -0.34313 -0.00574 0.37873 2.24046
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3665 0.1555 8.79 <2e-16 ***
## FirstAuthorFemale1 0.0211 0.0665 0.32 0.751
## LastAuthorFemale1 -0.1102 0.0905 -1.22 0.224
## Year1997 -0.4767 0.2105 -2.26 0.024 *
## Year1998 -0.1572 0.2016 -0.78 0.436
## Year1999 -0.0833 0.2872 -0.29 0.772
## Year2000 -0.2852 0.1893 -1.51 0.132
## Year2001 -0.2124 0.1946 -1.09 0.275
## Year2002 -0.3128 0.1821 -1.72 0.086 .
## Year2003 -0.2440 0.2006 -1.22 0.224
## Year2004 -0.4546 0.1822 -2.49 0.013 *
## Year2005 -0.3350 0.1743 -1.92 0.055 .
```

```

## Year2006          -0.2770      0.1867   -1.48    0.139
## Year2007          -0.2702      0.1798   -1.50    0.133
## Year2008          -0.4050      0.1847   -2.19    0.029 *
## Year2009          -0.3613      0.1723   -2.10    0.036 *
## Year2010          -0.1956      0.1773   -1.10    0.270
## Year2011          -0.3141      0.1723   -1.82    0.069 .
## Year2012          -0.3313      0.1689   -1.96    0.050 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.032, Adjusted R-squared:  0.00374
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 573 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0661 0.8520 0.9520 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      1.57e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.138 1          1.067
## Year              1.138 16          1.004

```

Residuals from first author



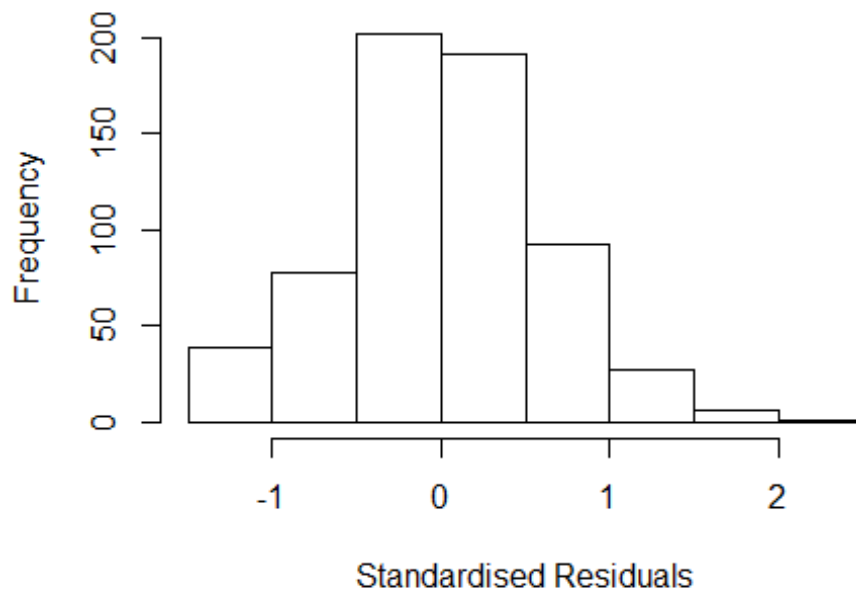
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3631 -0.3474 -0.0108 0.3769 2.2439
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3631 0.1573 8.66 <2e-16 ***
## FirstAuthorFemale1 -0.0134 0.0616 -0.22 0.828
## Year1997 -0.4785 0.2142 -2.23 0.026 *
## Year1998 -0.1541 0.2034 -0.76 0.449
## Year1999 -0.0889 0.2872 -0.31 0.757
## Year2000 -0.2797 0.1921 -1.46 0.146
## Year2001 -0.2176 0.1981 -1.10 0.272
## Year2002 -0.3093 0.1846 -1.68 0.094 .
## Year2003 -0.2432 0.2023 -1.20 0.230
## Year2004 -0.4475 0.1833 -2.44 0.015 *
## Year2005 -0.3495 0.1757 -1.99 0.047 *
## Year2006 -0.2749 0.1879 -1.46 0.144
```

```

## Year2007          -0.2767      0.1815    -1.52     0.128
## Year2008          -0.4016      0.1853    -2.17     0.031 *
## Year2009          -0.3702      0.1738    -2.13     0.034 *
## Year2010          -0.1941      0.1794    -1.08     0.280
## Year2011          -0.3176      0.1738    -1.83     0.068 .
## Year2012          -0.3309      0.1712    -1.93     0.054 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.557
## Multiple R-squared:  0.029, Adjusted R-squared:  0.00234
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 572 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0675 0.8520 0.9470 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      1.57e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.142 1          1.069
## Year            1.142 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
## -1.3673 -0.3465 -0.0026 0.3759 2.2397
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3673 0.1552 8.81 <2e-16 ***
## LastAuthorFemale1 -0.1003 0.0825 -1.21 0.225
## Year1997 -0.4780 0.2106 -2.27 0.024 *
## Year1998 -0.1552 0.2011 -0.77 0.441
## Year1999 -0.0828 0.2870 -0.29 0.773
## Year2000 -0.2833 0.1892 -1.50 0.135
## Year2001 -0.2086 0.1942 -1.07 0.283
## Year2002 -0.3113 0.1820 -1.71 0.088 .
## Year2003 -0.2406 0.2004 -1.20 0.230
## Year2004 -0.4522 0.1814 -2.49 0.013 *
## Year2005 -0.3355 0.1742 -1.93 0.055 .
## Year2006 -0.2746 0.1860 -1.48 0.140
```

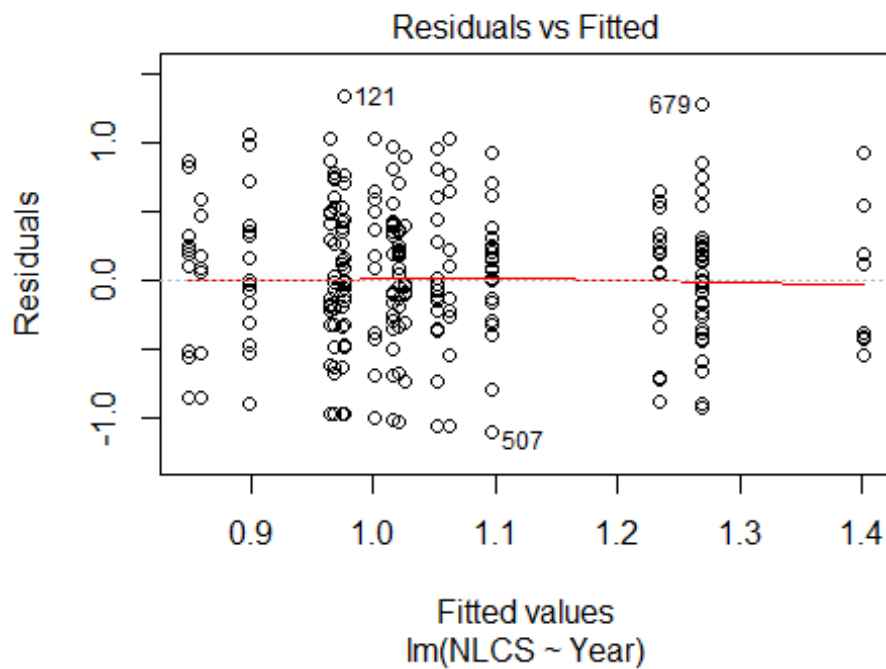
```

## Year2007          -0.2690      0.1795   -1.50    0.134
## Year2008          -0.4028      0.1840   -2.19    0.029 *
## Year2009          -0.3606      0.1722   -2.09    0.037 *
## Year2010          -0.1925      0.1763   -1.09    0.275
## Year2011          -0.3126      0.1719   -1.82    0.069 .
## Year2012          -0.3306      0.1689   -1.96    0.051 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.0318, Adjusted R-squared:  0.0052
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 573 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0668 0.8530 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      1.57e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 636"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
##   24   23   33   27   30   34   33   23   17   27   32   47   41   40   48
## 2011 2012
##   39   33
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   14    7   17   11   13   11   15    8    8   12   16   24   21   17   34
## 2011 2012

```



```
## 20 18
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 6 14 10 11 11 15 8 7 12 13 21 17 16 29
## 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 = 6.7, df = 16, p-value = 1
```



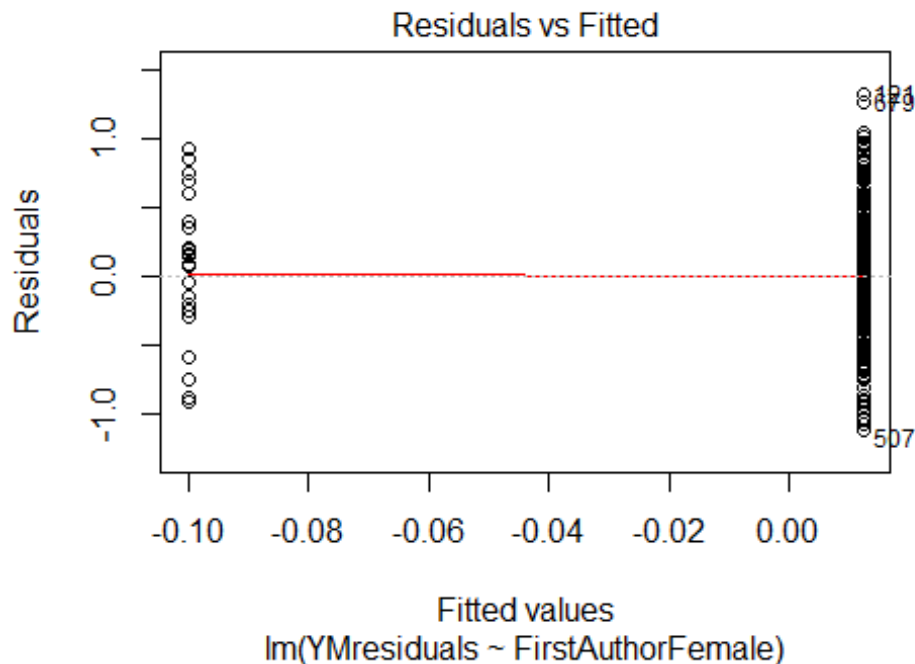
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.00056, df = 1, p-value = 1

## [1] "Female first author team size 2018 geometric mean: 1.81712059283214"
## [1] "Male first author team size 2018 geometric mean: 2.12845045174094"

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

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 62, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.22090615485233"
## [1] "Male last author team size 2018 geometric mean: 2.09495998679878"

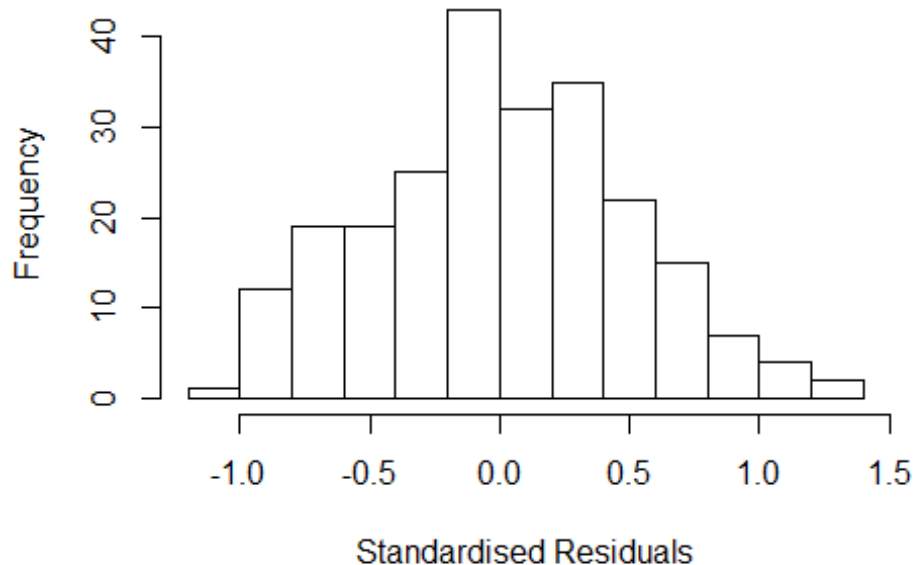
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.647	1	1.283
LastAuthorFemale	1.634	1	1.278
UniqueAuthors	53.676	4	1.645
Year	68.136	16	1.141

Residuals from first and last author and team size



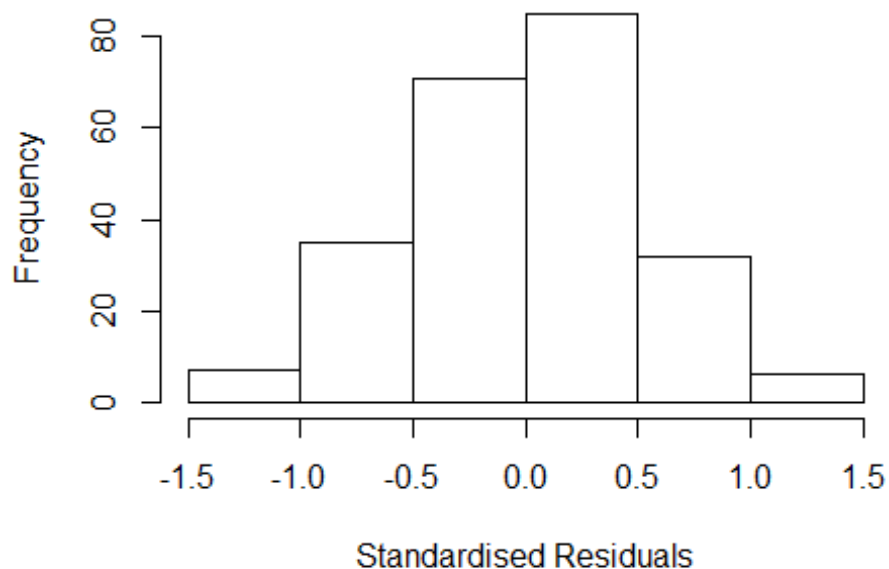
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18113 -0.32293 -0.00366 0.33599 1.28784
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.81222 0.16852 4.82 2.7e-06 ***
## FirstAuthorFemale1 -0.30769 0.11323 -2.72 0.00712 **
## LastAuthorFemale1 0.22011 0.11100 1.98 0.04865 *
## UniqueAuthors2 0.23610 0.09535 2.48 0.01407 *
## UniqueAuthors3 0.32993 0.09791 3.37 0.00089 ***
## UniqueAuthors4 0.37387 0.17049 2.19 0.02939 *
## UniqueAuthors5 0.42572 0.14914 2.85 0.00474 **
## Year1997 0.05023 0.22135 0.23 0.82068
## Year1998 -0.01571 0.22181 -0.07 0.94359
## Year1999 0.10463 0.29845 0.35 0.72624
```

```

## Year2000          0.12614      0.26137      0.48  0.62986
## Year2001         -0.03443      0.23294     -0.15  0.88263
## Year2002          0.15649      0.20318      0.77  0.44205
## Year2003          0.00539      0.23342      0.02  0.98160
## Year2004          0.39947      0.22263      1.79  0.07419 .
## Year2005          0.03873      0.21209      0.18  0.85526
## Year2006         -0.08750      0.20911     -0.42  0.67604
## Year2007          0.17795      0.19198      0.93  0.35501
## Year2008          0.03898      0.20295      0.19  0.84786
## Year2009          0.19117      0.19656      0.97  0.33188
## Year2010          0.21485      0.18000      1.19  0.23397
## Year2011         -0.04493      0.20187     -0.22  0.82407
## Year2012         -0.12480      0.22020     -0.57  0.57147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.506
## Multiple R-squared:  0.151, Adjusted R-squared:  0.0629
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 212 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.496  0.857  0.953  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 1.309 1      1.144
## LastAuthorFemale  1.434 1      1.198
## Year              1.772 16      1.018

```

Residuals from first and last author



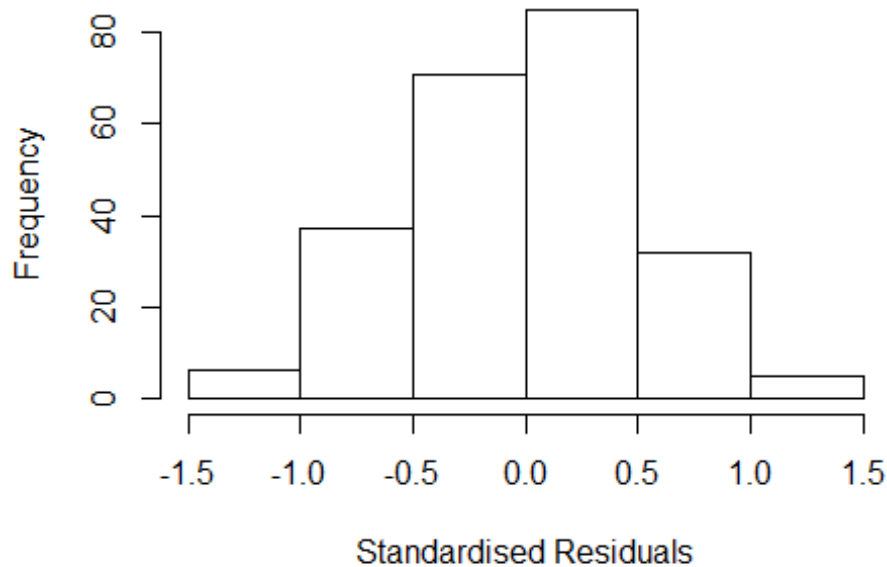
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1385 -0.3514 0.0274 0.3052 1.4454
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94765 0.17416 5.44 1.4e-07 ***
## FirstAuthorFemale1 -0.27610 0.11304 -2.44 0.0154 *
## LastAuthorFemale1 0.30052 0.11110 2.70 0.0074 **
## Year1997 0.00944 0.21701 0.04 0.9654
## Year1998 -0.08106 0.22977 -0.35 0.7246
## Year1999 0.09390 0.31083 0.30 0.7629
## Year2000 0.12656 0.30605 0.41 0.6796
## Year2001 -0.09256 0.26813 -0.35 0.7303
## Year2002 0.10641 0.21883 0.49 0.6273
## Year2003 -0.03499 0.24858 -0.14 0.8882
## Year2004 0.45304 0.26600 1.70 0.0900 .
## Year2005 0.06217 0.20985 0.30 0.7673
```

```

## Year2006      -0.12181    0.22617   -0.54    0.5907
## Year2007      0.19082    0.19892    0.96    0.3385
## Year2008      0.12209    0.20853    0.59    0.5588
## Year2009      0.25575    0.20466    1.25    0.2128
## Year2010      0.22790    0.19118    1.19    0.2345
## Year2011      0.06278    0.21668    0.29    0.7723
## Year2012     -0.00164    0.22892   -0.01    0.9943
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.505
## Multiple R-squared:  0.0925, Adjusted R-squared:  0.0172
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 214 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.393  0.849  0.955  0.904  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.189 1      1.090
## Year      1.189 16      1.005

```

Residuals from first author



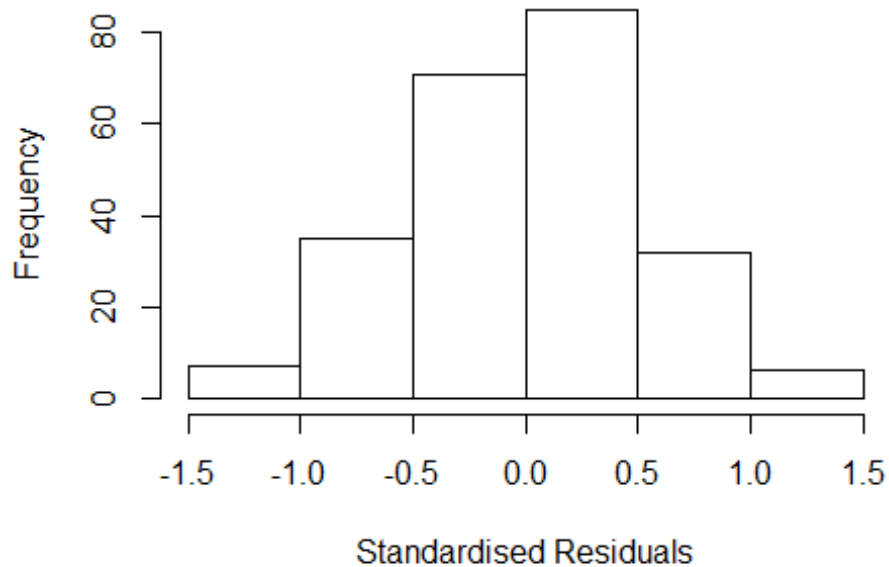
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1468 -0.3429 0.0217 0.3189 1.4377
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9793 0.1899 5.16 5.6e-07 ***
## FirstAuthorFemale1 -0.1545 0.1160 -1.33 0.185
## Year1997 0.0574 0.2513 0.23 0.820
## Year1998 -0.1050 0.2437 -0.43 0.667
## Year1999 0.1024 0.3032 0.34 0.736
## Year2000 0.0940 0.2926 0.32 0.748
## Year2001 -0.0863 0.2822 -0.31 0.760
## Year2002 0.0851 0.2369 0.36 0.720
## Year2003 0.0532 0.2478 0.21 0.830
## Year2004 0.4781 0.2786 1.72 0.088 .
## Year2005 0.0425 0.2282 0.19 0.853
## Year2006 -0.1381 0.2456 -0.56 0.574
```

```

## Year2007          0.1675      0.2118      0.79      0.430
## Year2008          0.0745      0.2236      0.33      0.739
## Year2009          0.2774      0.2263      1.23      0.222
## Year2010          0.2168      0.2055      1.05      0.293
## Year2011          0.0450      0.2302      0.20      0.845
## Year2012         -0.0528      0.2398     -0.22      0.826
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.0722, Adjusted R-squared:  -0.000196
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 214 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.417  0.850  0.956  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      4.24e-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 1.651 1      1.285
## Year      1.651 16      1.016

```


Residuals from last author



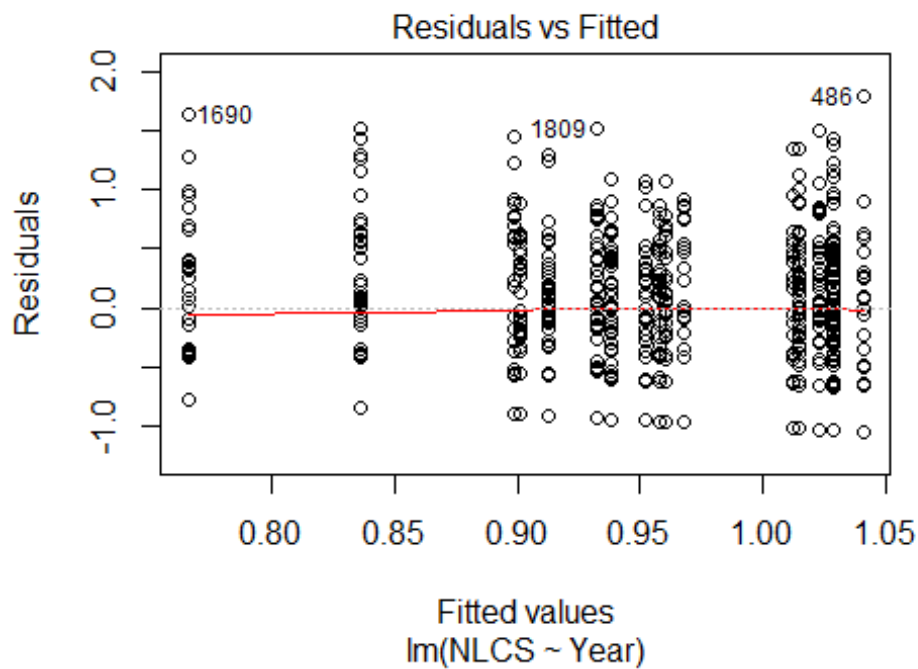
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2246 -0.3234 0.0275 0.3250 1.4889
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9376 0.1753 5.35 2.2e-07 ***
## LastAuthorFemale1 0.1915 0.1191 1.61 0.11
## Year1997 0.0133 0.2120 0.06 0.95
## Year1998 -0.1145 0.2236 -0.51 0.61
## Year1999 0.1172 0.3048 0.38 0.70
## Year2000 0.0954 0.3373 0.28 0.78
## Year2001 -0.1188 0.2678 -0.44 0.66
## Year2002 0.1004 0.2219 0.45 0.65
## Year2003 -0.0188 0.2460 -0.08 0.94
## Year2004 0.4829 0.2657 1.82 0.07 .
## Year2005 0.0474 0.2187 0.22 0.83
## Year2006 -0.1053 0.2301 -0.46 0.65
```

```

## Year2007          0.1940      0.2023      0.96      0.34
## Year2008          0.0985      0.2158      0.46      0.65
## Year2009          0.2402      0.2073      1.16      0.25
## Year2010          0.2347      0.1917      1.22      0.22
## Year2011          0.0598      0.2162      0.28      0.78
## Year2012         -0.0367      0.2298     -0.16      0.87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.503
## Multiple R-squared:  0.0766, Adjusted R-squared:  0.00461
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 218 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.361  0.857  0.957  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      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 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
##   82   65   75   75   85   57   63   54   56   88   83   85   93  106   82
## 2011 2012
##   62   84
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   45   37   35   27   24   28   34   30   33   57   47   49   52   69   50
## 2011 2012

```

```
## 45 59
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 41 33 34 25 22 22 31 28 29 50 42 45 48 62 45
## 2011 2012
## 40 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 = 14, df = 16, p-value = 0.6
```



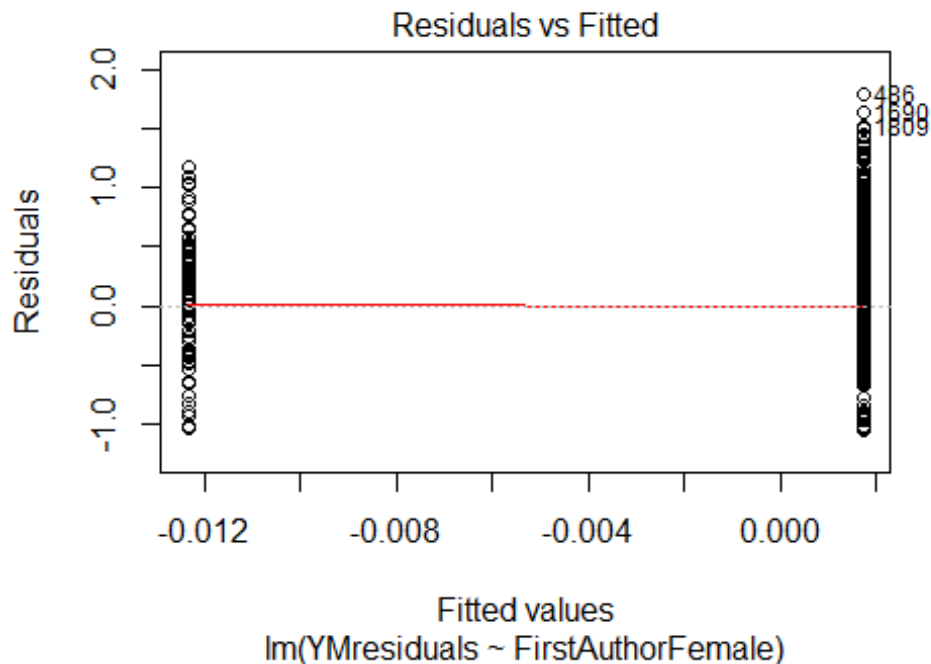
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.43, df = 1, p-value = 0.5

## [1] "Female first author team size 2018 geometric mean: 1.25103340485907"
## [1] "Male first author team size 2018 geometric mean: 1.69468595369776"

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

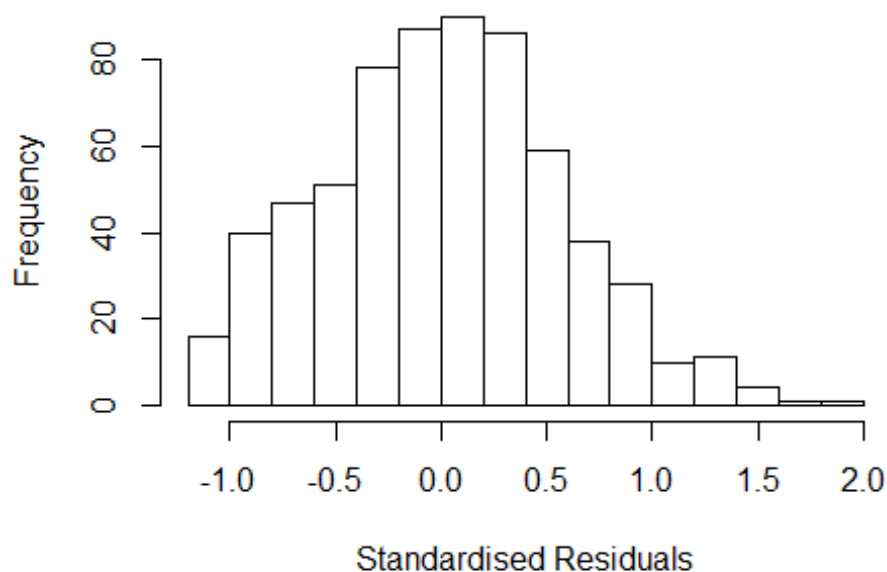
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 110, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.65814935213846"
## [1] "Male last author team size 2018 geometric mean: 1.59295758932319"

## Warning in wilcox.test.default(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 = 0.8
## 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 1.351 1      1.162
## UniqueAuthors    1.514 4      1.053
## Year              1.773 16     1.018
```

Residuals from first and last author and team size



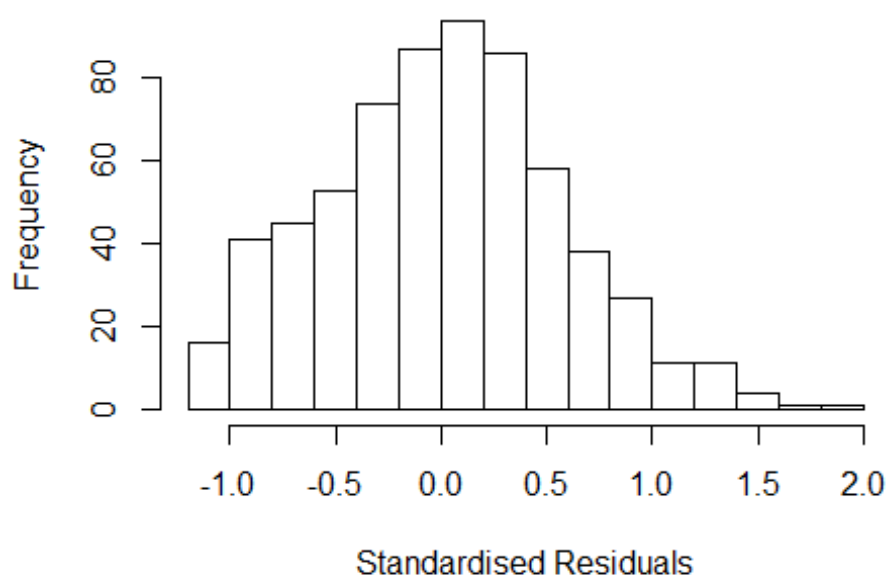
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0479 -0.3753  0.0238  0.3834  1.8738
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03948    0.09600   10.83  <2e-16 ***
## FirstAuthorFemale1  0.07871    0.07035    1.12   0.264
## LastAuthorFemale1 -0.12454    0.07663   -1.63   0.105
## UniqueAuthors2    -0.01727    0.05496   -0.31   0.753
## UniqueAuthors3     0.04613    0.07487    0.62   0.538
## UniqueAuthors4     0.03710    0.11503    0.32   0.747
## UniqueAuthors5    -0.21851    0.18983   -1.15   0.250
## Year1997          -0.04418    0.12761   -0.35   0.729
## Year1998          -0.15103    0.12830   -1.18   0.240
## Year1999          -0.09110    0.16529   -0.55   0.582
```

```

## Year2000      -0.06900    0.16831   -0.41    0.682
## Year2001      -0.01108    0.15576   -0.07    0.943
## Year2002      -0.11958    0.13689   -0.87    0.383
## Year2003      -0.16951    0.17902   -0.95    0.344
## Year2004      -0.13104    0.14482   -0.90    0.366
## Year2005      -0.00046    0.12842    0.00    0.997
## Year2006      -0.03427    0.13154   -0.26    0.795
## Year2007      -0.06810    0.12472   -0.55    0.585
## Year2008      -0.12517    0.11538   -1.08    0.278
## Year2009       0.00845    0.11992    0.07    0.944
## Year2010      -0.29827    0.12505   -2.39    0.017 *
## Year2011      -0.09428    0.13836   -0.68    0.496
## Year2012      -0.25198    0.13339   -1.89    0.059 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.584
## Multiple R-squared:  0.0313, Adjusted R-squared:  -0.0028
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 604 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.283  0.877  0.957  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      1.55e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.184 1      1.088
## LastAuthorFemale  1.316 1      1.147
## Year              1.222 16      1.006

```

Residuals from first and last author



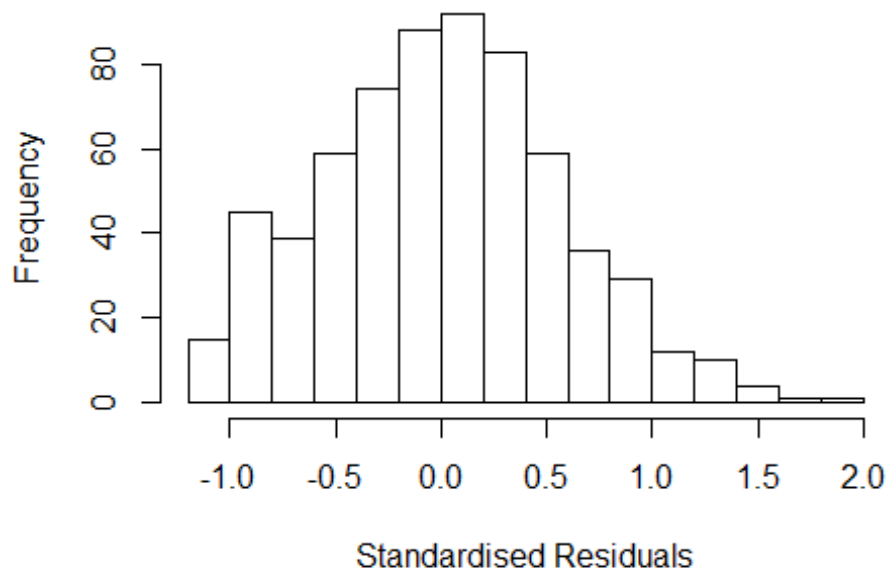
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0491 -0.3820 0.0208 0.3772 1.8555
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03581 0.09488 10.92 <2e-16 ***
## FirstAuthorFemale1 0.07513 0.06960 1.08 0.281
## LastAuthorFemale1 -0.11997 0.07663 -1.57 0.118
## Year1997 -0.04030 0.12737 -0.32 0.752
## Year1998 -0.15097 0.12904 -1.17 0.242
## Year1999 -0.08930 0.16566 -0.54 0.590
## Year2000 -0.06429 0.16800 -0.38 0.702
## Year2001 -0.00913 0.15661 -0.06 0.954
## Year2002 -0.11765 0.13657 -0.86 0.389
## Year2003 -0.16568 0.17894 -0.93 0.355
## Year2004 -0.13512 0.14312 -0.94 0.345
## Year2005 0.00291 0.12842 0.02 0.982
```

```

## Year2006      -0.03536    0.13141   -0.27    0.788
## Year2007      -0.06728    0.12319   -0.55    0.585
## Year2008      -0.13116    0.11568   -1.13    0.257
## Year2009       0.01327    0.11921    0.11    0.911
## Year2010      -0.29536    0.12552   -2.35    0.019 *
## Year2011      -0.08905    0.13685   -0.65    0.515
## Year2012      -0.24777    0.13114   -1.89    0.059 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.029, Adjusted R-squared:  0.00121
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 605 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.295  0.880  0.958  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      1.55e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.046 1      1.023
## Year      1.046 16      1.001

```


Residuals from first author



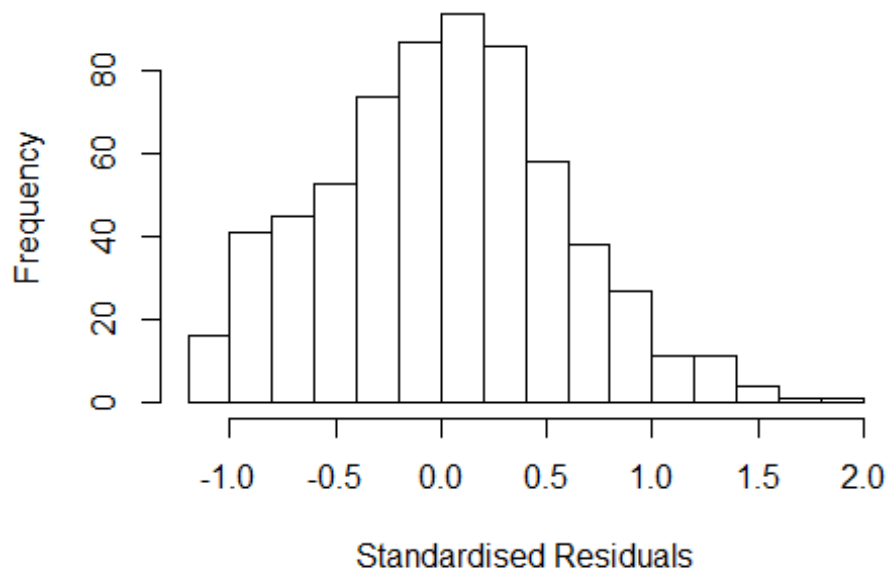
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0399 -0.3881 0.0116 0.3809 1.8617
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0230 0.0948 10.79 <2e-16 ***
## FirstAuthorFemale1 0.0309 0.0652 0.47 0.636
## Year1997 -0.0330 0.1273 -0.26 0.796
## Year1998 -0.1364 0.1295 -1.05 0.293
## Year1999 -0.0873 0.1681 -0.52 0.604
## Year2000 -0.0577 0.1716 -0.34 0.737
## Year2001 -0.0143 0.1584 -0.09 0.928
## Year2002 -0.1067 0.1373 -0.78 0.437
## Year2003 -0.1540 0.1784 -0.86 0.388
## Year2004 -0.1189 0.1420 -0.84 0.403
## Year2005 0.0130 0.1280 0.10 0.919
## Year2006 -0.0401 0.1308 -0.31 0.759
```

```

## Year2007          -0.0615      0.1237   -0.50    0.619
## Year2008          -0.1242      0.1163   -1.07    0.286
## Year2009           0.0168      0.1191    0.14    0.888
## Year2010          -0.2939      0.1260   -2.33    0.020 *
## Year2011          -0.1056      0.1347   -0.78    0.433
## Year2012          -0.2536      0.1319   -1.92    0.055 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0253, Adjusted R-squared:  -0.00107
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 594 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.293  0.874  0.954  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      1.55e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.156 1      1.075
## Year              1.156 16      1.005

```

Residuals from last author



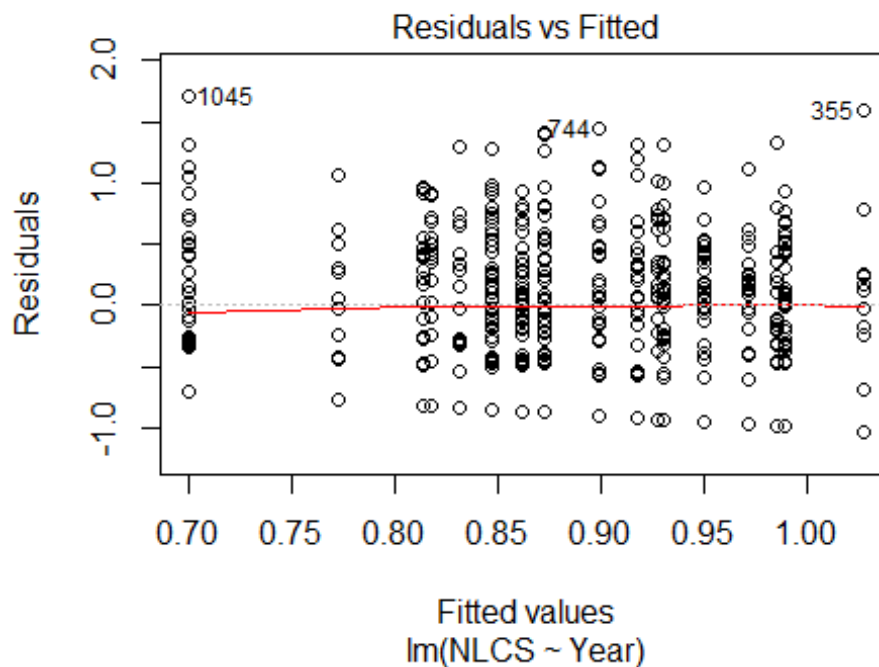
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0577 -0.3711 0.0152 0.3783 1.8477
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04153 0.09446 11.03 <2e-16 ***
## LastAuthorFemale1 -0.09058 0.07262 -1.25 0.213
## Year1997 -0.04287 0.12750 -0.34 0.737
## Year1998 -0.14582 0.12842 -1.14 0.257
## Year1999 -0.09104 0.16568 -0.55 0.583
## Year2000 -0.06225 0.16668 -0.37 0.709
## Year2001 -0.00864 0.15705 -0.06 0.956
## Year2002 -0.11632 0.13707 -0.85 0.396
## Year2003 -0.16981 0.17784 -0.95 0.340
## Year2004 -0.13359 0.14329 -0.93 0.352
## Year2005 0.00832 0.12892 0.06 0.949
## Year2006 -0.03897 0.13107 -0.30 0.766
```

```

## Year2007          -0.06508      0.12363    -0.53      0.599
## Year2008          -0.13236      0.11567    -1.14      0.253
## Year2009           0.01616      0.11887     0.14      0.892
## Year2010          -0.29562      0.12579    -2.35      0.019 *
## Year2011          -0.09369      0.13676    -0.69      0.494
## Year2012          -0.24727      0.13090    -1.89      0.059 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0276, Adjusted R-squared:  0.00131
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 600 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.302  0.877   0.958   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      1.55e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 647"
## [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
##   52   51   43   40   40   28   28   30   50   59   63   56   65   75   66
## 2011 2012
##   70   71
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   25   23   18   23   23   12   14   18   29   34   39   33   41   53   48
## 2011 2012

```

```
## 49 50
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 24 20 18 20 20 9 14 18 26 32 34 29 37 48 43
## 2011 2012
## 44 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 = 8.1, df = 16, p-value = 0.9
```



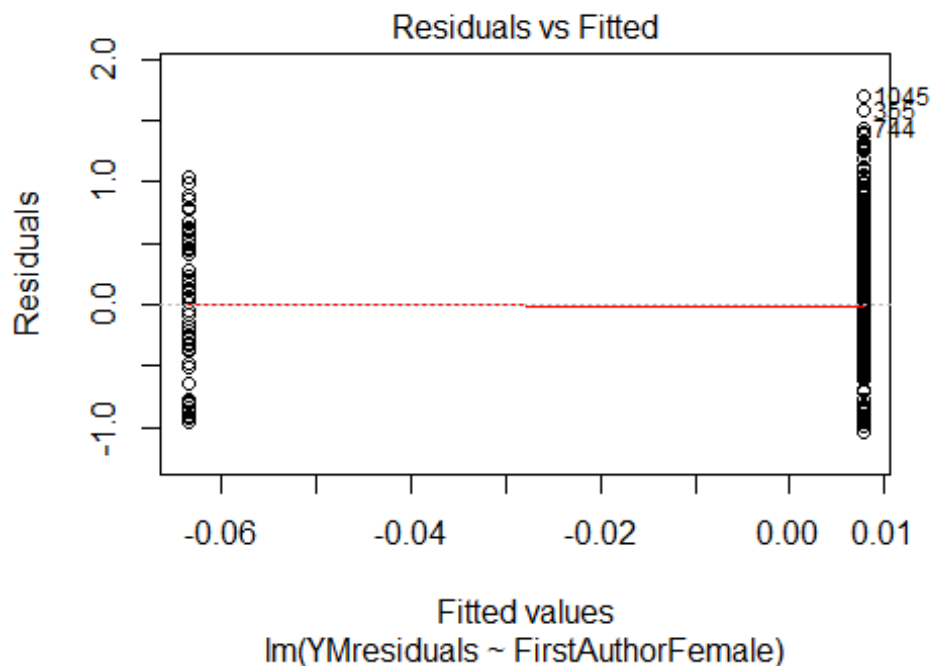
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0047, df = 1, p-value = 0.9

## [1] "Female first author team size 2018 geometric mean: 1.29170834209075"
## [1] "Male first author team size 2018 geometric mean: 1.49565774559143"

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

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 90, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.29683955465101"
## [1] "Male last author team size 2018 geometric mean: 1.50121084011208"

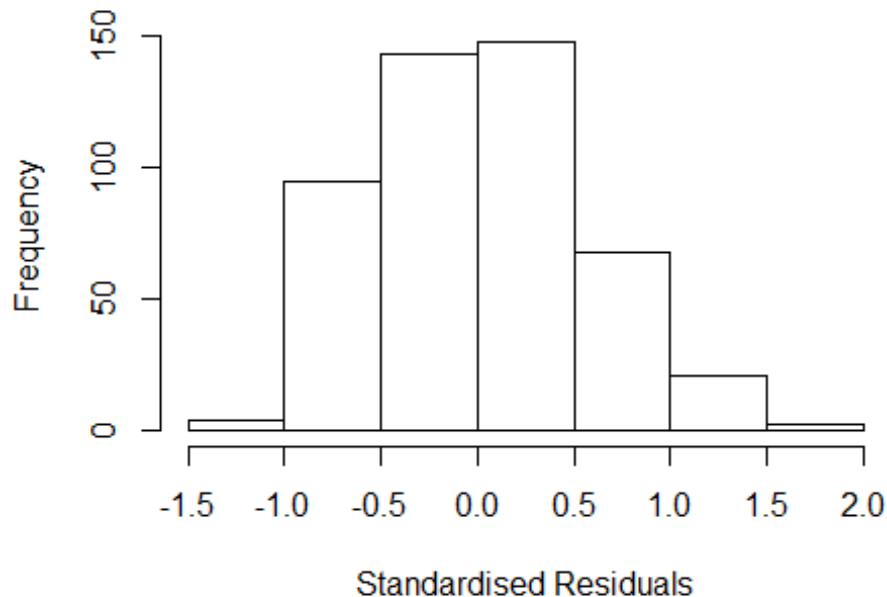
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## 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	1.809	1	1.345
LastAuthorFemale	1.726	1	1.314
UniqueAuthors	18.033	4	1.436
Year	19.002	16	1.096

Residuals from first and last author and team size



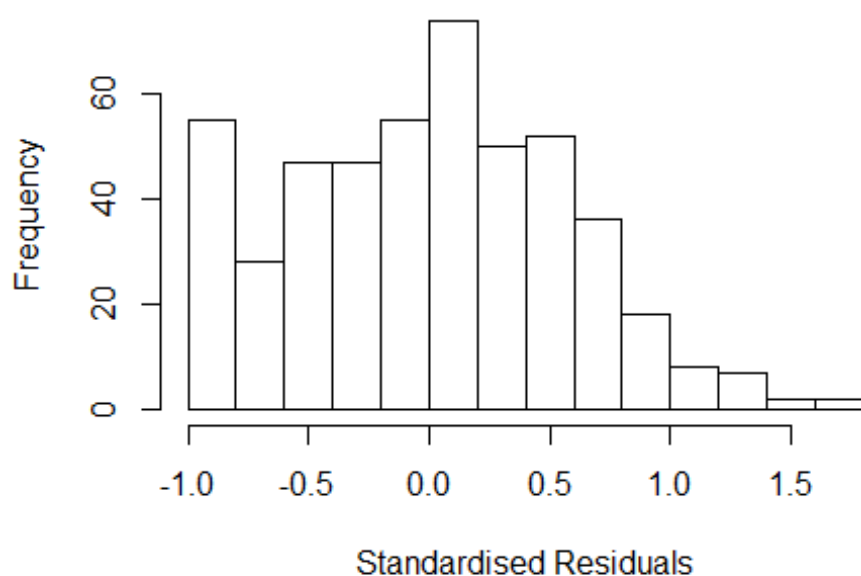
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17e+00 -3.87e-01 -2.06e-05 4.06e-01 1.85e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97384 0.10046 9.69 <2e-16 ***
## FirstAuthorFemale1 -0.16340 0.09661 -1.69 0.091 .
## LastAuthorFemale1 0.04779 0.10555 0.45 0.651
## UniqueAuthors2 0.20054 0.06941 2.89 0.004 **
## UniqueAuthors3 0.24746 0.10235 2.42 0.016 *
## UniqueAuthors4 0.27588 0.17034 1.62 0.106
## UniqueAuthors5 0.21671 0.08969 2.42 0.016 *
## Year1997 -0.13886 0.19162 -0.72 0.469
## Year1998 -0.05121 0.14972 -0.34 0.732
## Year1999 -0.25256 0.18156 -1.39 0.165
```

```

## Year2000      -0.17565    0.17160   -1.02    0.307
## Year2001      -0.20335    0.23526   -0.86    0.388
## Year2002      -0.24055    0.20539   -1.17    0.242
## Year2003      -0.19312    0.17600   -1.10    0.273
## Year2004      -0.21617    0.15418   -1.40    0.162
## Year2005      -0.04942    0.14447   -0.34    0.732
## Year2006      -0.16090    0.13932   -1.15    0.249
## Year2007      -0.13083    0.16776   -0.78    0.436
## Year2008      -0.00853    0.13741   -0.06    0.951
## Year2009      -0.22955    0.13237   -1.73    0.084 .
## Year2010      -0.37938    0.13115   -2.89    0.004 **
## Year2011      -0.18996    0.13152   -1.44    0.149
## Year2012      -0.25082    0.12869   -1.95    0.052 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.595
## Multiple R-squared:  0.0542, Adjusted R-squared:  0.00877
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 448 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.313  0.871  0.955  0.919  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      2.08e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.462 1      1.209
## LastAuthorFemale  1.600 1      1.265
## Year              1.285 16      1.008

```


Residuals from first and last author



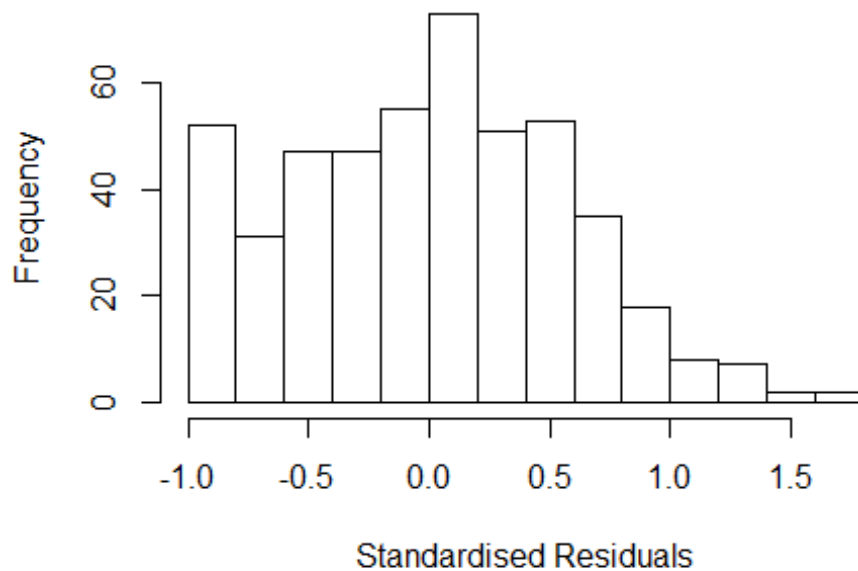
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.9869 -0.4194 0.0228 0.4122 1.7836
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0151 0.0972 10.44 <2e-16 ***
## FirstAuthorFemale1 -0.1208 0.0982 -1.23 0.2191
## LastAuthorFemale1 0.0417 0.1107 0.38 0.7067
## Year1997 -0.1384 0.1878 -0.74 0.4615
## Year1998 -0.0418 0.1498 -0.28 0.7805
## Year1999 -0.2514 0.1884 -1.33 0.1828
## Year2000 -0.1543 0.1714 -0.90 0.3686
## Year2001 -0.1767 0.2576 -0.69 0.4931
## Year2002 -0.2490 0.1951 -1.28 0.2027
## Year2003 -0.2046 0.1799 -1.14 0.2559
## Year2004 -0.1953 0.1546 -1.26 0.2072
## Year2005 -0.0366 0.1432 -0.26 0.7982
```

```

## Year2006          -0.1451      0.1419   -1.02   0.3068
## Year2007          -0.1329      0.1700   -0.78   0.4348
## Year2008          -0.0281      0.1350   -0.21   0.8350
## Year2009          -0.1776      0.1315   -1.35   0.1774
## Year2010          -0.3728      0.1332   -2.80   0.0053 **
## Year2011          -0.1351      0.1291   -1.05   0.2960
## Year2012          -0.1907      0.1284   -1.48   0.1384
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.6
## Multiple R-squared:  0.0284, Adjusted R-squared:  -0.00943
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 448 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.357  0.859   0.951   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      2.08e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



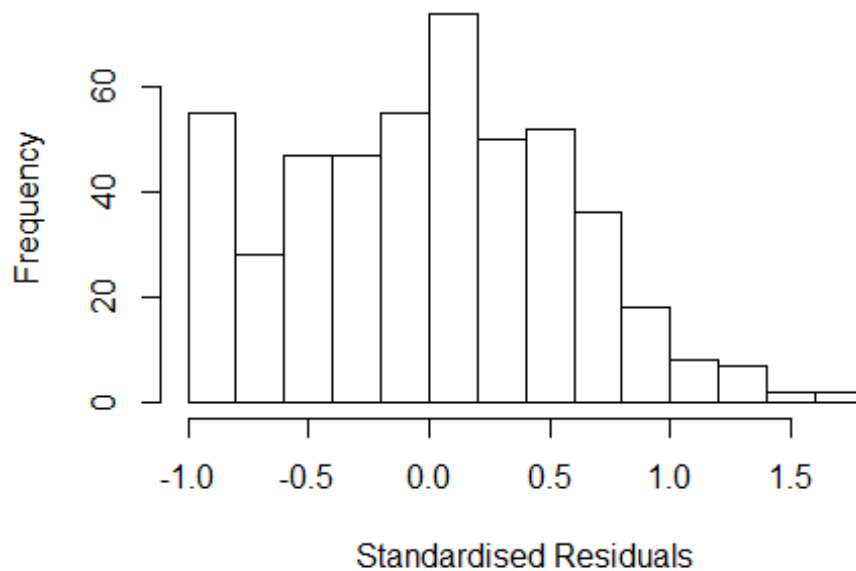
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9892 -0.4212  0.0238  0.4115  1.7788
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0142     0.0964   10.52  <2e-16 ***
## FirstAuthorFemale1 -0.0971     0.0848   -1.15   0.2526
## Year1997         -0.1383     0.1875   -0.74   0.4610
## Year1998         -0.0438     0.1493   -0.29   0.7695
## Year1999         -0.2509     0.1880   -1.33   0.1827
## Year2000         -0.1538     0.1708   -0.90   0.3684
## Year2001         -0.1710     0.2547   -0.67   0.5023
## Year2002         -0.2481     0.1948   -1.27   0.2034
## Year2003         -0.2039     0.1802   -1.13   0.2584
## Year2004         -0.1936     0.1544   -1.25   0.2103
## Year2005         -0.0360     0.1428   -0.25   0.8010
## Year2006         -0.1415     0.1406   -1.01   0.3146
```

```

## Year2007          -0.1277      0.1668   -0.77   0.4442
## Year2008          -0.0250      0.1338   -0.19   0.8518
## Year2009          -0.1725      0.1277   -1.35   0.1775
## Year2010          -0.3702      0.1322   -2.80   0.0053 **
## Year2011          -0.1289      0.1265   -1.02   0.3084
## Year2012          -0.1889      0.1273   -1.48   0.1385
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.6
## Multiple R-squared:  0.0281, Adjusted R-squared:  -0.00754
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 449 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.359  0.860  0.951  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      2.08e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.182 1          1.087
## Year            1.182 16          1.005

```

Residuals from last author



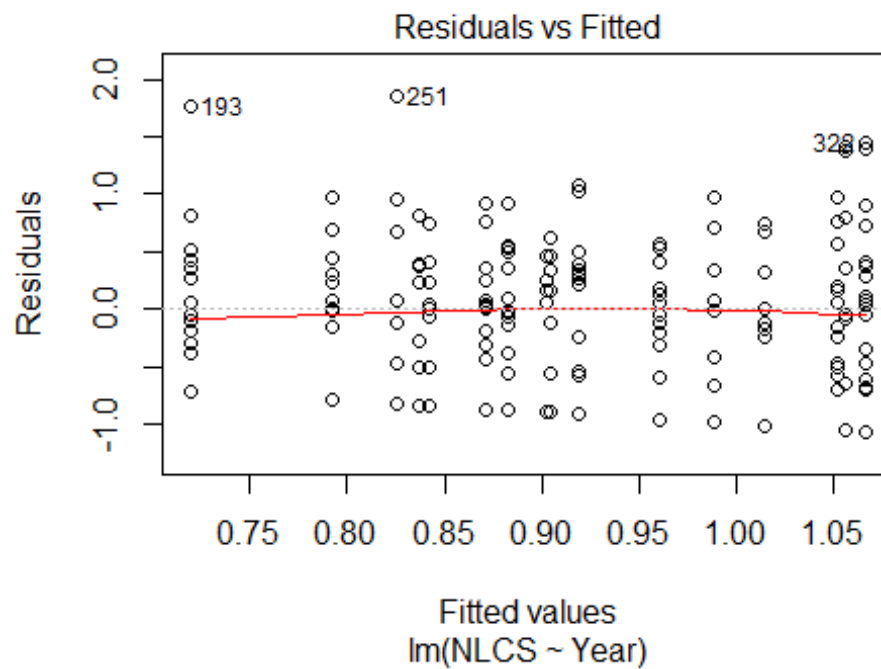
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.9812 -0.4191 0.0159 0.4171 1.7959
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0000 0.0961 10.41 <2e-16 ***
## LastAuthorFemale1 -0.0337 0.0974 -0.35 0.7292
## Year1997 -0.1311 0.1844 -0.71 0.4776
## Year1998 -0.0415 0.1472 -0.28 0.7783
## Year1999 -0.2457 0.1856 -1.32 0.1860
## Year2000 -0.1500 0.1717 -0.87 0.3827
## Year2001 -0.1739 0.2602 -0.67 0.5042
## Year2002 -0.2340 0.1946 -1.20 0.2298
## Year2003 -0.1900 0.1811 -1.05 0.2946
## Year2004 -0.1825 0.1554 -1.17 0.2410
## Year2005 -0.0281 0.1429 -0.20 0.8440
## Year2006 -0.1325 0.1423 -0.93 0.3523
```

```

## Year2007          -0.1315      0.1688   -0.78    0.4366
## Year2008          -0.0189      0.1343   -0.14    0.8884
## Year2009          -0.1617      0.1300   -1.24    0.2144
## Year2010          -0.3633      0.1340   -2.71    0.0069 **
## Year2011          -0.1198      0.1282   -0.93    0.3506
## Year2012          -0.1850      0.1281   -1.44    0.1495
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.026, Adjusted R-squared:  -0.00973
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 34 weights are ~= 1. The remaining 447 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.349  0.861  0.950  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      2.08e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 481"
## [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
##   21   14   18   16   24   14    8   11   10    9   13   13   17   24   26
## 2011 2012
##   12   16
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   14   11   14   14   14    9    5    8    7    7   10    8   10   13   20
## 2011 2012

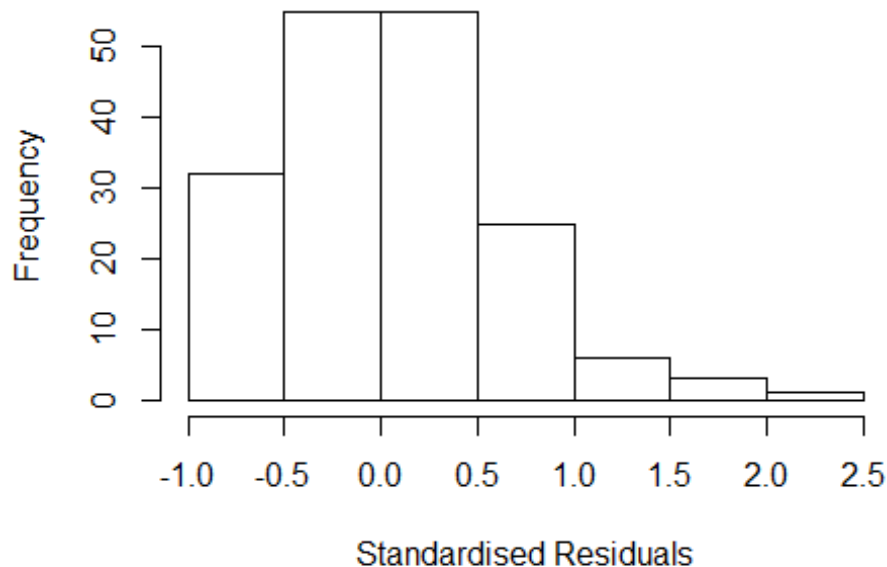
```

```
## 10 14
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 11 11 14 14 9 5 8 7 7 9 7 9 12 18
## 2011 2012
## 10 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 = 18, df = 16, p-value = 0.3
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.7, df = 1, p-value = 0.2
```


Residuals from first and last author and team size



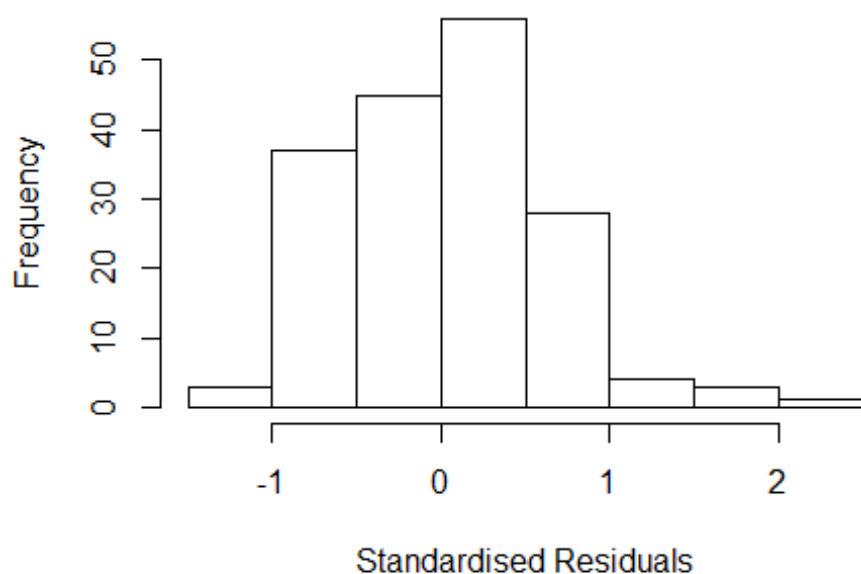
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.9851 -0.3814 0.0113 0.3266 2.2845
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.77037 0.18007 4.28 3.3e-05 ***
## FirstAuthorFemale1 0.25003 0.16985 1.47 0.1430
## LastAuthorFemale1 -0.31658 0.17135 -1.85 0.0666 .
## UniqueAuthors2 0.33677 0.10572 3.19 0.0017 **
## UniqueAuthors3 0.43055 0.26174 1.64 0.1020
## UniqueAuthors4 2.17372 0.19025 11.43 < 2e-16 ***
## Year1997 -0.08567 0.23702 -0.36 0.7183
## Year1998 -0.11174 0.27965 -0.40 0.6900
## Year1999 0.12834 0.20515 0.63 0.5325
## Year2000 0.17170 0.20677 0.83 0.4076
```

```

## Year2001      0.21478    0.25841    0.83    0.4072
## Year2002      0.09126    0.27888    0.33    0.7439
## Year2003     -0.05019    0.23072   -0.22    0.8281
## Year2004     -0.04694    0.22919   -0.20    0.8380
## Year2005     -0.46408    0.25440   -1.82    0.0700 .
## Year2006     -0.12596    0.20803   -0.61    0.5458
## Year2007      0.00687    0.36925    0.02    0.9852
## Year2008     -0.37782    0.30757   -1.23    0.2212
## Year2009     -0.04529    0.21054   -0.22    0.8300
## Year2010      0.16901    0.23697    0.71    0.4768
## Year2011      0.01699    0.55377    0.03    0.9756
## Year2012      0.05180    0.22139    0.23    0.8153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.184, Adjusted R-squared:  0.0731
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 163 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0584 0.8750 0.9560 0.9060 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          5.65e-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 5.207 1 2.282
## LastAuthorFemale 4.946 1 2.224
## Year 2.083 16 1.023

```

Residuals from first and last author



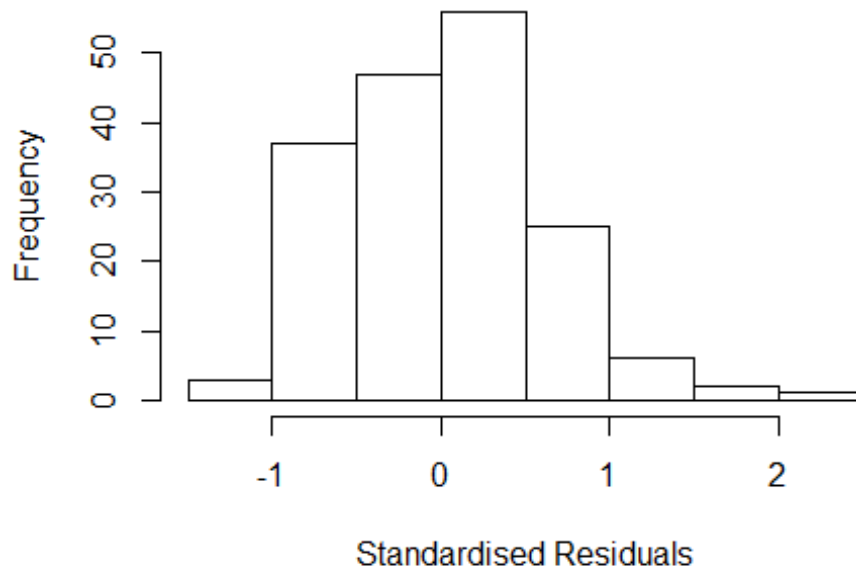
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0244 -0.3818 0.0122 0.3998 2.1715
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.86200 0.16730 5.15 7.6e-07 ***
## FirstAuthorFemale1 0.43952 0.21138 2.08 0.039 *
## LastAuthorFemale1 -0.49616 0.22339 -2.22 0.028 *
## Year1997 -0.14218 0.22344 -0.64 0.525
## Year1998 -0.17766 0.27311 -0.65 0.516
## Year1999 0.11424 0.19852 0.58 0.566
## Year2000 0.18060 0.21359 0.85 0.399
## Year2001 0.16243 0.23807 0.68 0.496
## Year2002 0.07478 0.26715 0.28 0.780
## Year2003 0.12735 0.27082 0.47 0.639
## Year2004 0.06435 0.27076 0.24 0.812
## Year2005 -0.34451 0.28225 -1.22 0.224
```

```

## Year2006      -0.11360    0.22373   -0.51    0.612
## Year2007      0.11266    0.34703    0.32    0.746
## Year2008     -0.35653    0.30592   -1.17    0.246
## Year2009     -0.13160    0.20298   -0.65    0.518
## Year2010      0.13970    0.24398    0.57    0.568
## Year2011      0.10053    0.46837    0.21    0.830
## Year2012     -0.00966    0.21239   -0.05    0.964
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.602
## Multiple R-squared:  0.0793, Adjusted R-squared:  -0.0256
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 156 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.166  0.875   0.944   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      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 1.533 1      1.238
## Year              1.533 16      1.013

```

Residuals from first author



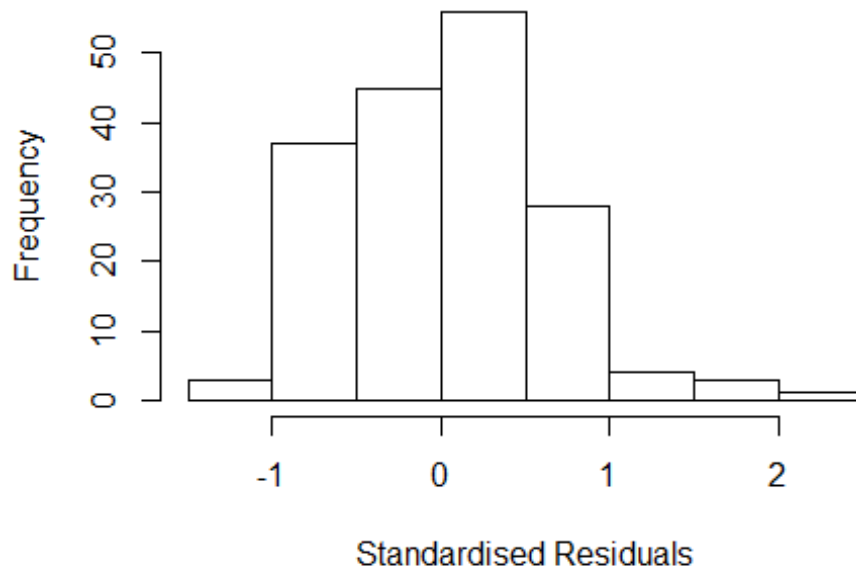
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.02405 -0.44757 0.00955 0.40642 2.16183
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.833445 0.168366 4.95 1.9e-06 ***
## FirstAuthorFemale1 0.046126 0.138972 0.33 0.74
## Year1997 -0.113667 0.224043 -0.51 0.61
## Year1998 -0.110243 0.266774 -0.41 0.68
## Year1999 0.136135 0.202047 0.67 0.50
## Year2000 0.201130 0.213210 0.94 0.35
## Year2001 0.190609 0.238681 0.80 0.43
## Year2002 0.101894 0.267976 0.38 0.70
## Year2003 0.017243 0.239448 0.07 0.94
## Year2004 0.091938 0.270911 0.34 0.73
## Year2005 -0.308190 0.286818 -1.07 0.28
## Year2006 -0.049398 0.242346 -0.20 0.84
```

```

## Year2007          0.141369    0.345491    0.41    0.68
## Year2008          -0.318276    0.310103   -1.03    0.31
## Year2009          -0.118647    0.216421   -0.55    0.58
## Year2010          0.187335    0.244002    0.77    0.44
## Year2011          0.134936    0.460683    0.29    0.77
## Year2012          0.000998    0.216764    0.00    1.00
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.0609, Adjusted R-squared:  -0.0395
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 152 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.189  0.874  0.943  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      5.65e-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 1.624 1          1.274
## Year            1.624 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
## -1.05219 -0.40444 0.00615 0.38556 2.15753
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.88088 0.16477 5.35 3.1e-07 ***
## LastAuthorFemale1 -0.12585 0.11670 -1.08 0.28
## Year1997 -0.16112 0.22122 -0.73 0.47
## Year1998 -0.13544 0.27791 -0.49 0.63
## Year1999 0.09920 0.19469 0.51 0.61
## Year2000 0.16819 0.21310 0.79 0.43
## Year2001 0.14301 0.23622 0.61 0.55
## Year2002 0.05383 0.26559 0.20 0.84
## Year2003 0.00569 0.23256 0.02 0.98
## Year2004 0.04409 0.26819 0.16 0.87
## Year2005 -0.35219 0.28599 -1.23 0.22
## Year2006 -0.09356 0.24128 -0.39 0.70
```

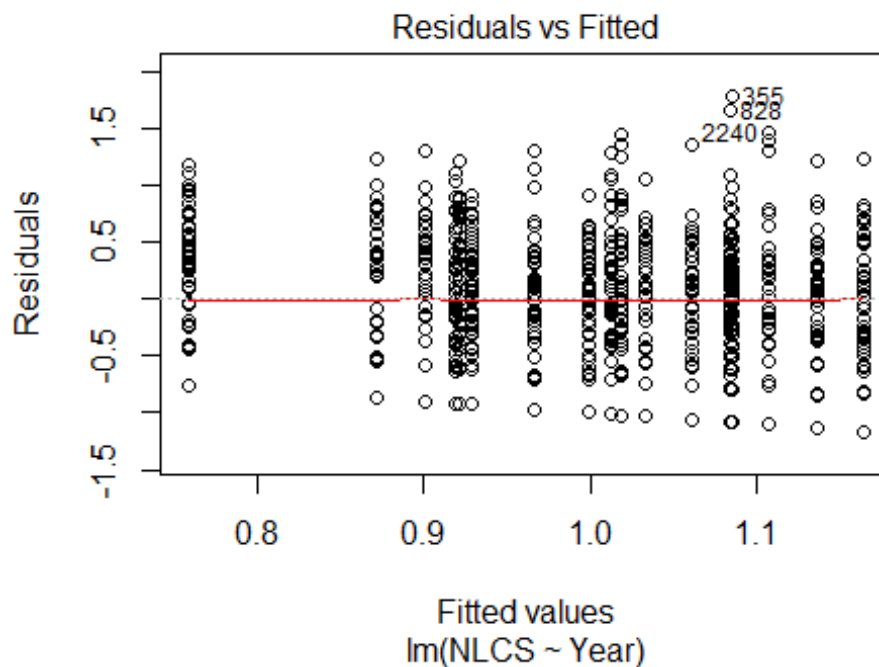
```

## Year2007      0.09400      0.34312      0.27      0.78
## Year2008     -0.36141      0.30909     -1.17      0.24
## Year2009     -0.10148      0.21388     -0.47      0.64
## Year2010      0.17131      0.24167      0.71      0.48
## Year2011      0.08997      0.44844      0.20      0.84
## Year2012     -0.01624      0.20949     -0.08      0.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.619
## Multiple R-squared:  0.0633, Adjusted R-squared:  -0.0369
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 156 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.199  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      5.65e-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 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
##   95   95   93   96  102  119  121  149  110  104  100  102  112   93   83
## 2011 2012
##   76   80
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   44   40   49   42   47   57   73   47   55   49   52   70   47   46
## 2011 2012

```



```
## 41 58
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 31 35 34 40 38 37 48 66 43 47 45 49 62 41 43
## 2011 2012
## 35 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 = 19, df = 16, p-value = 0.3
```



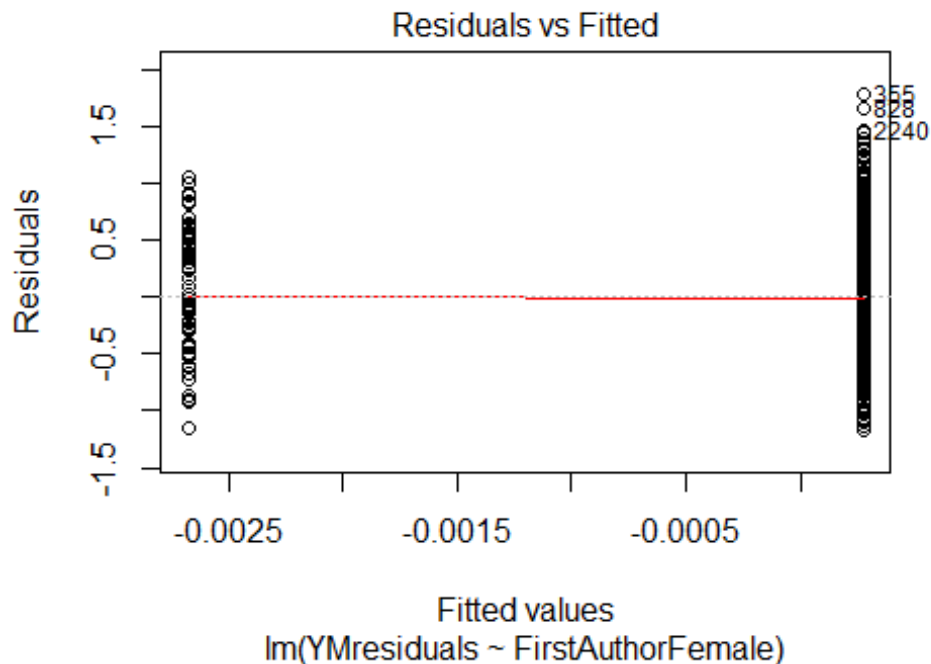
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0034, df = 1, p-value = 1

## [1] "Female first author team size 2018 geometric mean: 1.41421356237309"
## [1] "Male first author team size 2018 geometric mean: 1.55713438680895"

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

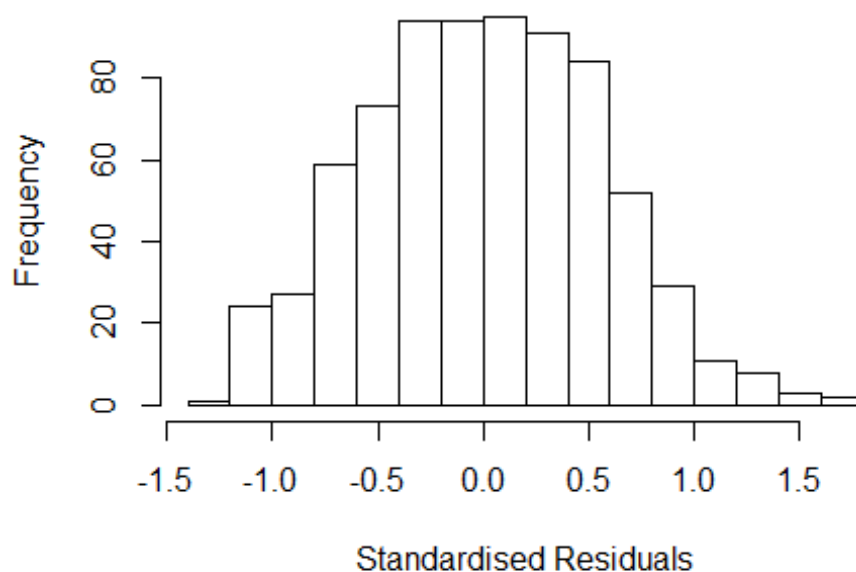
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 36, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1"
## [1] "Male last author team size 2018 geometric mean: 1.56742483791451"

## Warning in wilcox.test.default(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 = 0.4
## 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 1.261  1      1.123
## LastAuthorFemale  1.293  1      1.137
## UniqueAuthors    1.487  4      1.051
## Year              1.646 16      1.016
```

Residuals from first and last author and team size



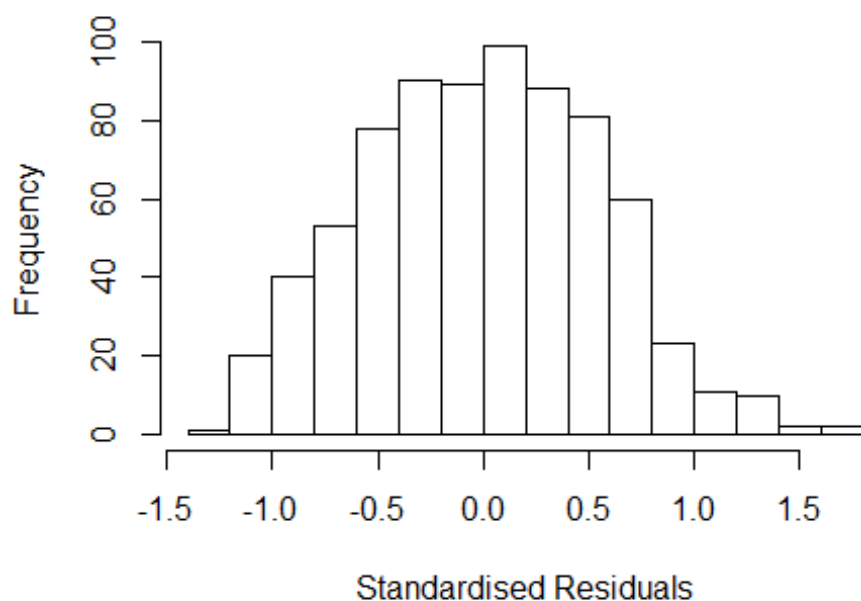
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.227430 -0.389849 0.000222 0.405193 1.716380
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01478 0.10250 9.90 <2e-16 ***
## FirstAuthorFemale1 0.03598 0.08123 0.44 0.658
## LastAuthorFemale1 -0.10153 0.08128 -1.25 0.212
## UniqueAuthors2 0.10681 0.04939 2.16 0.031 *
## UniqueAuthors3 0.08834 0.06501 1.36 0.175
## UniqueAuthors4 0.26916 0.12140 2.22 0.027 *
## UniqueAuthors5 0.21063 0.16550 1.27 0.204
## Year1997 0.00395 0.13176 0.03 0.976
## Year1998 0.03403 0.13193 0.26 0.797
## Year1999 -0.14501 0.12655 -1.15 0.252
```

```

## Year2000      0.06396    0.12556    0.51    0.611
## Year2001     -0.09903    0.15229   -0.65    0.516
## Year2002      0.03255    0.12775    0.25    0.799
## Year2003     -0.33078    0.13540   -2.44    0.015 *
## Year2004     -0.21804    0.14886   -1.46    0.143
## Year2005     -0.08500    0.12430   -0.68    0.494
## Year2006     -0.03511    0.12286   -0.29    0.775
## Year2007     -0.13797    0.13579   -1.02    0.310
## Year2008     -0.14215    0.11602   -1.23    0.221
## Year2009     -0.13710    0.13851   -0.99    0.323
## Year2010      0.06986    0.13720    0.51    0.611
## Year2011      0.02728    0.14494    0.19    0.851
## Year2012     -0.08437    0.13211   -0.64    0.523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.584
## Multiple R-squared:  0.0515, Adjusted R-squared:  0.0227
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 689 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.368  0.882  0.950  0.917  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.34e-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 1.253 1      1.120
## LastAuthorFemale  1.283 1      1.133
## Year              1.171 16      1.005

```

Residuals from first and last author

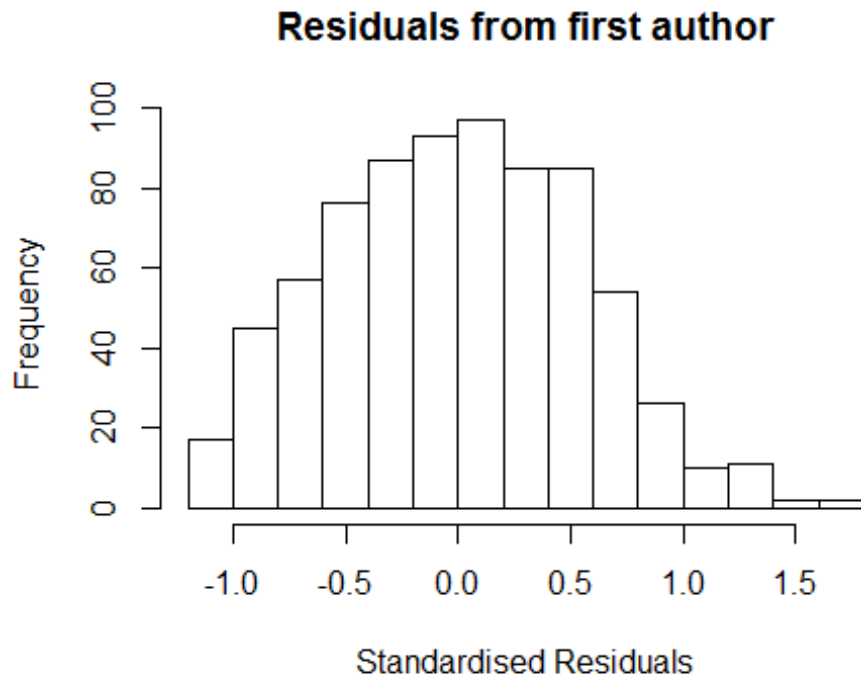


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2082 -0.4135  0.0036  0.4065  1.7652
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0795     0.1018  10.60  <2e-16 ***
## FirstAuthorFemale1  0.0509     0.0832   0.61   0.541
## LastAuthorFemale1 -0.0983     0.0840  -1.17   0.242
## Year1997          0.0209     0.1361   0.15   0.878
## Year1998          0.0273     0.1336   0.20   0.838
## Year1999         -0.1609     0.1298  -1.24   0.216
## Year2000          0.0299     0.1272   0.24   0.814
## Year2001         -0.0995     0.1543  -0.64   0.519
## Year2002          0.0335     0.1304   0.26   0.797
## Year2003         -0.3338     0.1378  -2.42   0.016 *
## Year2004         -0.2142     0.1470  -1.46   0.145
## Year2005         -0.1051     0.1271  -0.83   0.408
```

```

## Year2006          -0.0637      0.1265   -0.50    0.615
## Year2007          -0.1561      0.1392   -1.12    0.263
## Year2008          -0.1678      0.1179   -1.42    0.155
## Year2009          -0.1652      0.1384   -1.19    0.233
## Year2010           0.0778      0.1409    0.55    0.581
## Year2011           0.0442      0.1462    0.30    0.763
## Year2012          -0.0771      0.1329   -0.58    0.562
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.574
## Multiple R-squared:  0.0413, Adjusted R-squared:  0.0176
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 57 weights are ~= 1. The remaining 690 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.325  0.878   0.948   0.914   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      1.34e-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 1.066 1      1.033
## Year              1.066 16      1.002

```

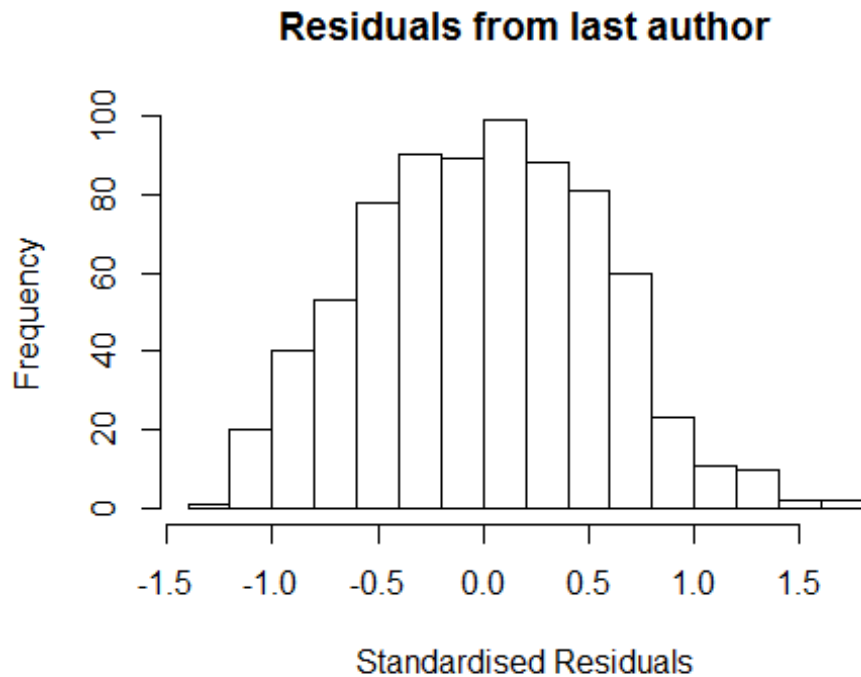


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.17338 -0.41007 -0.00305  0.41003  1.77095
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0812    0.1031   10.49  <2e-16 ***
## FirstAuthorFemale1  0.0151    0.0784    0.19   0.847
## Year1997          0.0205    0.1372    0.15   0.882
## Year1998          0.0198    0.1340    0.15   0.882
## Year1999         -0.1721    0.1294   -1.33   0.184
## Year2000          0.0263    0.1280    0.21   0.837
## Year2001         -0.1149    0.1533   -0.75   0.454
## Year2002          0.0270    0.1317    0.21   0.838
## Year2003         -0.3370    0.1387   -2.43   0.015 *
## Year2004         -0.2201    0.1480   -1.49   0.137
## Year2005         -0.1116    0.1274   -0.88   0.381
## Year2006         -0.0714    0.1272   -0.56   0.575
```

```

## Year2007          -0.1622      0.1404   -1.16    0.248
## Year2008          -0.1731      0.1189   -1.46    0.146
## Year2009          -0.1700      0.1388   -1.23    0.221
## Year2010           0.0771      0.1423    0.54    0.588
## Year2011           0.0381      0.1463    0.26    0.795
## Year2012          -0.0824      0.1337   -0.62    0.538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.573
## Multiple R-squared:  0.04,   Adjusted R-squared:  0.0176
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 691 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.320  0.879   0.949   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      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 1.081 1      1.040
## Year              1.081 16      1.002

```

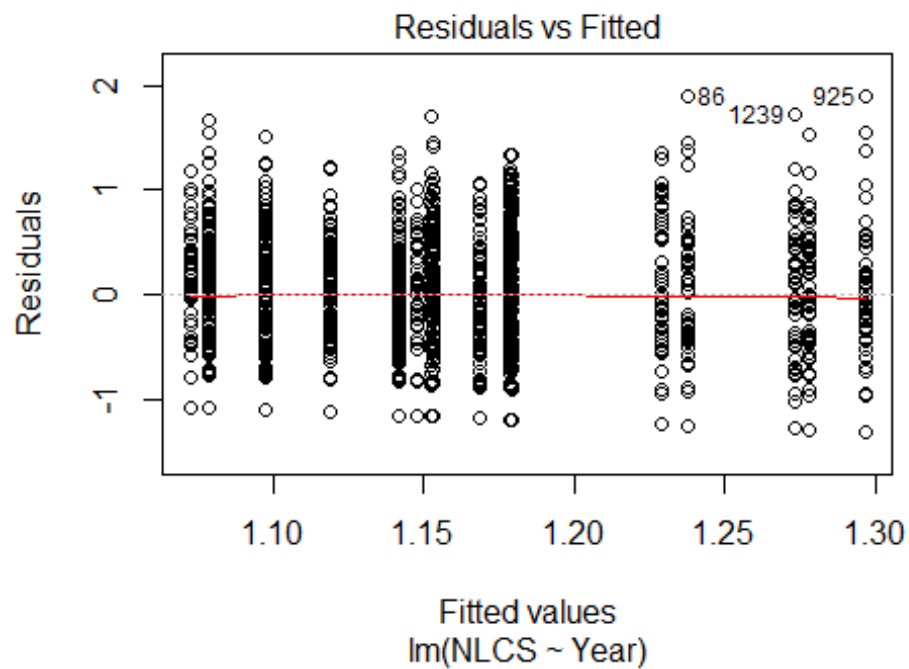
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.16379 -0.41606 0.00266 0.40312 1.76009
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0822 0.1023 10.58 <2e-16 ***
## LastAuthorFemale1 -0.0761 0.0763 -1.00 0.319
## Year1997 0.0198 0.1367 0.14 0.885
## Year1998 0.0297 0.1346 0.22 0.825
## Year1999 -0.1620 0.1302 -1.24 0.214
## Year2000 0.0281 0.1277 0.22 0.826
## Year2001 -0.1041 0.1539 -0.68 0.499
## Year2002 0.0345 0.1311 0.26 0.792
## Year2003 -0.3341 0.1372 -2.44 0.015 *
## Year2004 -0.2136 0.1475 -1.45 0.148
## Year2005 -0.1065 0.1276 -0.83 0.404
## Year2006 -0.0639 0.1267 -0.50 0.614
```

```

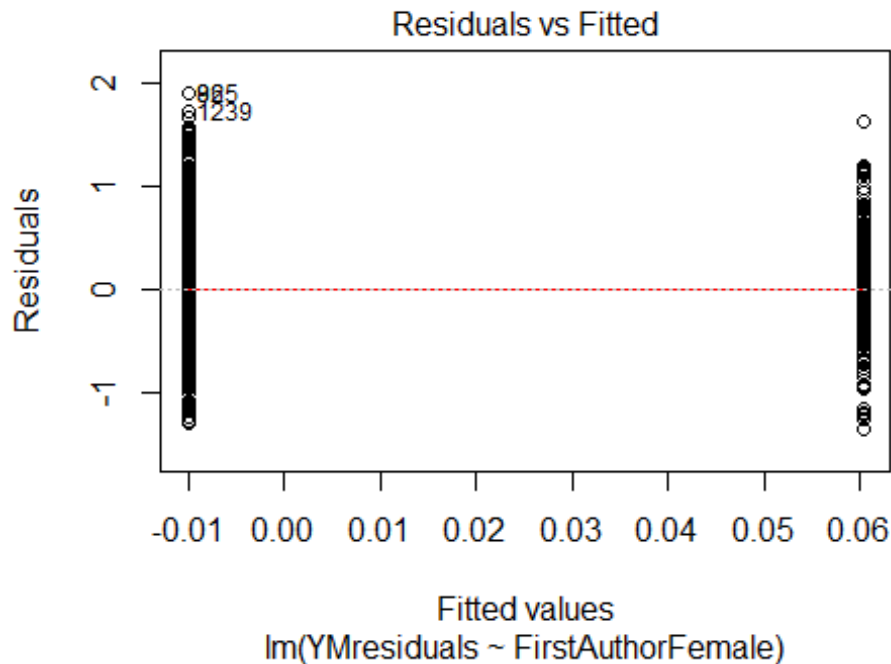
## Year2007          -0.1567      0.1390    -1.13      0.260
## Year2008          -0.1651      0.1185    -1.39      0.164
## Year2009          -0.1646      0.1387    -1.19      0.236
## Year2010           0.0816      0.1418      0.58      0.565
## Year2011           0.0487      0.1465      0.33      0.740
## Year2012          -0.0762      0.1336    -0.57      0.569
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.582
## Multiple R-squared:  0.0408, Adjusted R-squared:  0.0184
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 691 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.340  0.881  0.950  0.916  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.34e-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: 747"
## [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
##  142  150  144  136  137  137  135  112  124  142  202  259  253  328  270
## 2011 2012
##  304  283
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   50   53   53   45   51   57   52   53   58   70  107  128  130  157  141
## 2011 2012

```

```
## 171 168
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 45 41 49 40 41 44 46 44 46 61 85 100 105 127 114
## 2011 2012
## 132 140
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 42, df = 16, p-value = 3e-04
```

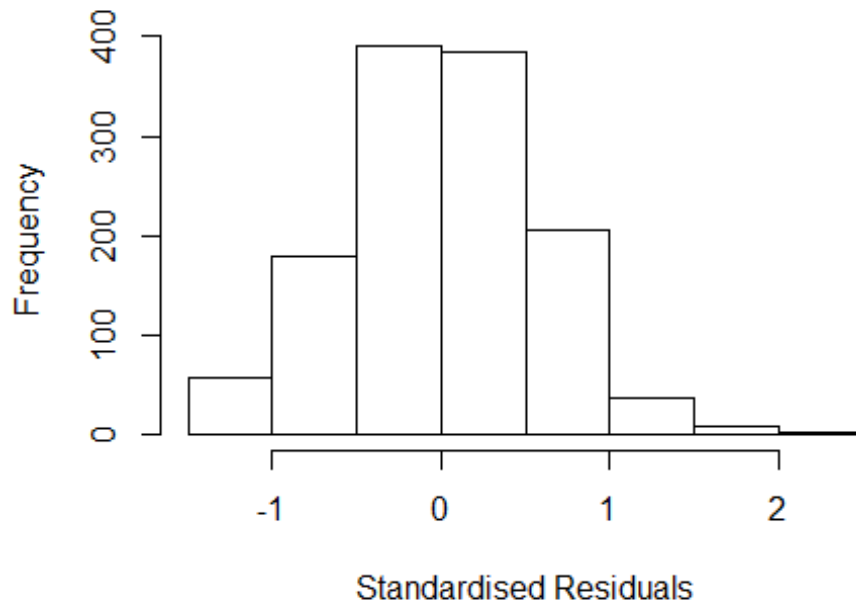


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.87, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 2.54349955445047"
## [1] "Male first author team size 2018 geometric mean: 2.29414008758552"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.7834411251135"
## [1] "Male last author team size 2018 geometric mean: 2.27079124992831"
##
## 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 1.118 1      1.057
## LastAuthorFemale  1.101 1      1.049
## UniqueAuthors     1.246 4      1.028
## Year              1.341 16     1.009
```

Residuals from first and last author and team size



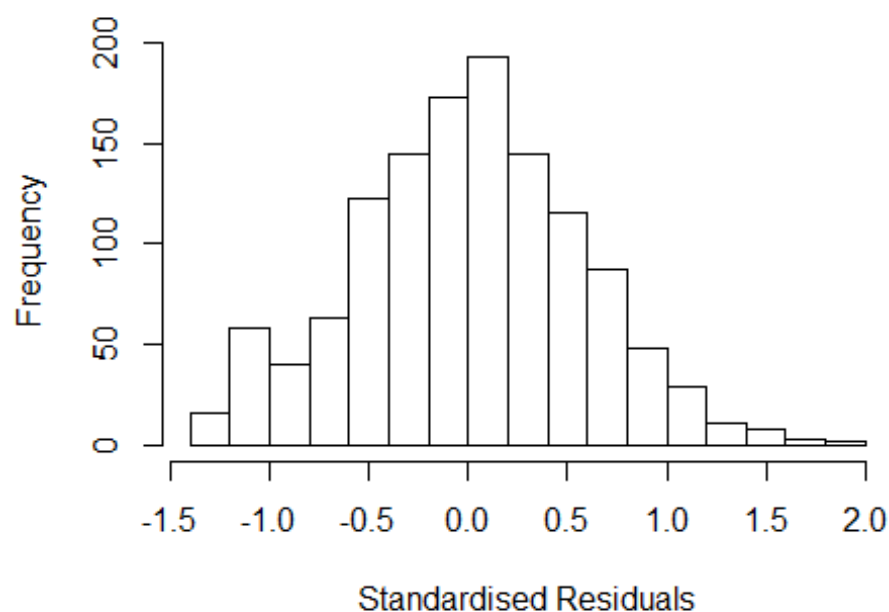
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.33551 -0.38130 0.00282 0.37099 2.12758
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0996 0.1173 9.37 < 2e-16 ***
## FirstAuthorFemale1 0.1120 0.0471 2.38 0.01748 *
## LastAuthorFemale1 -0.0752 0.0598 -1.26 0.20904
## UniqueAuthors2 0.1549 0.0458 3.38 0.00074 ***
## UniqueAuthors3 0.2672 0.0490 5.46 5.9e-08 ***
## UniqueAuthors4 0.2190 0.0679 3.22 0.00130 **
## UniqueAuthors5 0.2500 0.0913 2.74 0.00628 **
## Year1997 0.0492 0.1703 0.29 0.77269
## Year1998 0.0235 0.1430 0.16 0.86925
## Year1999 -0.1054 0.1531 -0.69 0.49141
```

```

## Year2000          -0.1057      0.1700   -0.62  0.53427
## Year2001          -0.0372      0.1404   -0.27  0.79100
## Year2002           0.0322      0.1481    0.22  0.82795
## Year2003          -0.1493      0.1387   -1.08  0.28208
## Year2004          -0.0310      0.1481   -0.21  0.83401
## Year2005          -0.0914      0.1307   -0.70  0.48454
## Year2006          -0.1033      0.1314   -0.79  0.43183
## Year2007          -0.0681      0.1234   -0.55  0.58098
## Year2008          -0.1210      0.1204   -1.01  0.31499
## Year2009          -0.1865      0.1231   -1.52  0.12989
## Year2010          -0.0746      0.1196   -0.62  0.53254
## Year2011          -0.1538      0.1179   -1.30  0.19232
## Year2012          -0.2111      0.1205   -1.75  0.08003 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.558
## Multiple R-squared:  0.0431, Adjusted R-squared:  0.0261
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 111 weights are ~= 1. The remaining 1149 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.114  0.870  0.950  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          7.94e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.088 1          1.043
## LastAuthorFemale  1.095 1          1.047
## Year              1.112 16          1.003

```

Residuals from first and last author



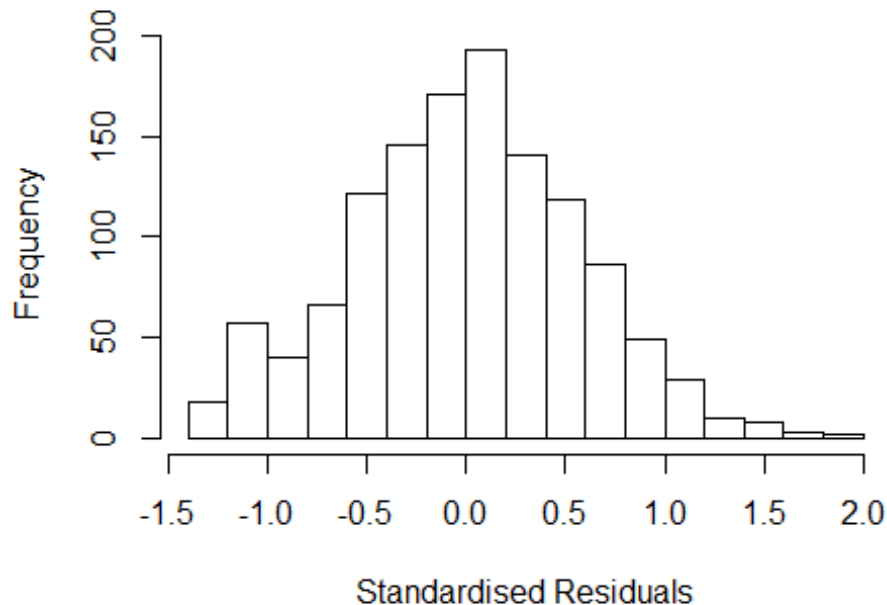
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33437 -0.37168  0.00915  0.38234  1.98490
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.21109    0.11040   10.97  <2e-16 ***
## FirstAuthorFemale1 0.13737    0.04626    2.97   0.003 **
## LastAuthorFemale1 -0.06887    0.06021   -1.14   0.253
## Year1997          0.02881    0.17001    0.17   0.865
## Year1998          0.03185    0.14668    0.22   0.828
## Year1999         -0.11039    0.15081   -0.73   0.464
## Year2000         -0.06385    0.17024   -0.38   0.708
## Year2001         -0.00599    0.14347   -0.04   0.967
## Year2002          0.04145    0.15145    0.27   0.784
## Year2003         -0.12258    0.14217   -0.86   0.389
## Year2004         -0.01409    0.14919   -0.09   0.925
## Year2005         -0.03657    0.13060   -0.28   0.779
```

```

## Year2006          -0.05386    0.13016   -0.41    0.679
## Year2007          -0.03137    0.12294   -0.26    0.799
## Year2008          -0.09189    0.12049   -0.76    0.446
## Year2009          -0.13961    0.12346   -1.13    0.258
## Year2010          -0.03017    0.11969   -0.25    0.801
## Year2011          -0.09672    0.11782   -0.82    0.412
## Year2012          -0.14769    0.12036   -1.23    0.220
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.568
## Multiple R-squared:  0.0158, Adjusted R-squared:  0.00155
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 110 weights are ~= 1. The remaining 1150 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.871  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      7.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000      0
##      psi            subsampling            cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.043 1      1.022
## Year              1.043 16      1.001

```


Residuals from first author



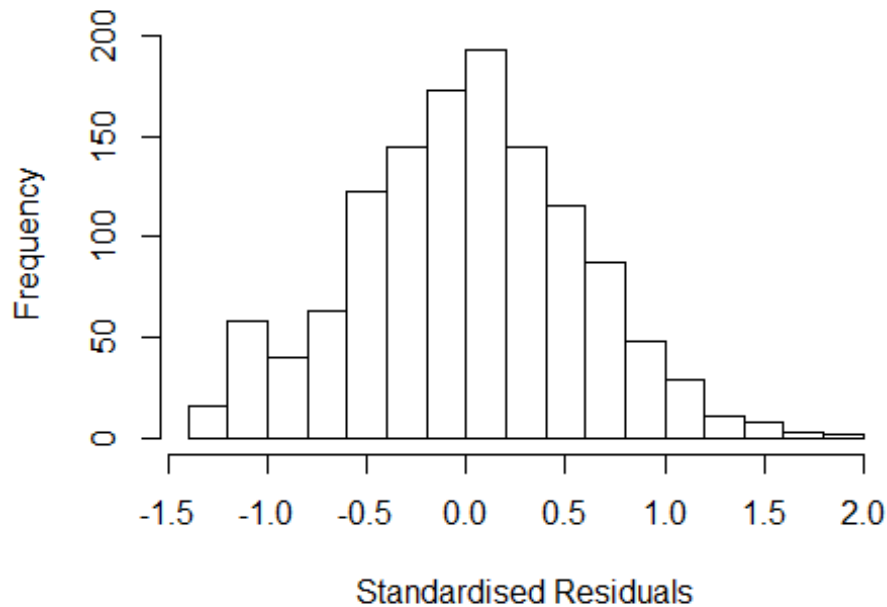
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3656 -0.3823 0.0142 0.3800 1.9841
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21e+00 1.11e-01 10.87 <2e-16 ***
## FirstAuthorFemale1 1.24e-01 4.56e-02 2.73 0.0065 **
## Year1997 2.71e-02 1.72e-01 0.16 0.8744
## Year1998 3.52e-02 1.47e-01 0.24 0.8107
## Year1999 -1.08e-01 1.51e-01 -0.71 0.4766
## Year2000 -6.76e-02 1.71e-01 -0.40 0.6924
## Year2001 -8.24e-05 1.44e-01 0.00 0.9995
## Year2002 4.56e-02 1.51e-01 0.30 0.7635
## Year2003 -1.22e-01 1.43e-01 -0.85 0.3948
## Year2004 -1.32e-02 1.50e-01 -0.09 0.9297
## Year2005 -3.49e-02 1.32e-01 -0.27 0.7909
## Year2006 -6.18e-02 1.31e-01 -0.47 0.6368
```

```

## Year2007          -2.79e-02   1.24e-01   -0.23   0.8215
## Year2008          -9.37e-02   1.21e-01   -0.77   0.4403
## Year2009          -1.40e-01   1.24e-01   -1.13   0.2598
## Year2010          -2.95e-02   1.20e-01   -0.25   0.8060
## Year2011          -9.95e-02   1.19e-01   -0.84   0.4018
## Year2012          -1.45e-01   1.21e-01   -1.20   0.2302
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.568
## Multiple R-squared:  0.0147, Adjusted R-squared:  0.00126
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 1153 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.869  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      7.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.045 1      1.022
## Year              1.045 16      1.001

```

Residuals from last author



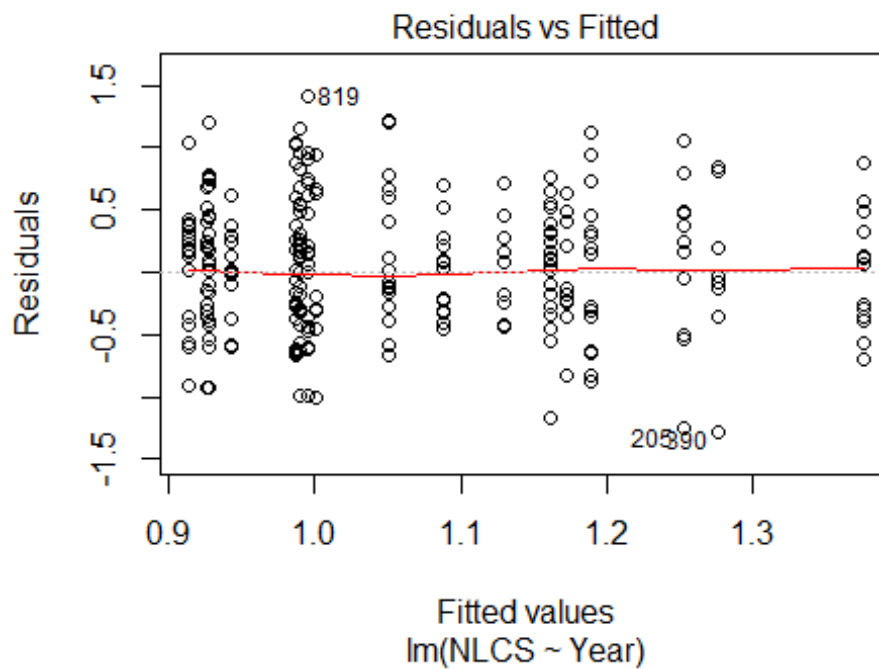
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26672 -0.38297 0.00562 0.38546 1.97580
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21939 0.10819 11.27 <2e-16 ***
## LastAuthorFemale1 -0.03266 0.05882 -0.56 0.58
## Year1997 0.02934 0.16798 0.17 0.86
## Year1998 0.03245 0.14426 0.22 0.82
## Year1999 -0.09754 0.14824 -0.66 0.51
## Year2000 -0.05652 0.16832 -0.34 0.74
## Year2001 -0.00519 0.14191 -0.04 0.97
## Year2002 0.04733 0.14958 0.32 0.75
## Year2003 -0.11468 0.14166 -0.81 0.42
## Year2004 -0.01789 0.14625 -0.12 0.90
## Year2005 -0.02187 0.12982 -0.17 0.87
## Year2006 -0.05385 0.12906 -0.42 0.68
```

```

## Year2007          -0.01581      0.12110      -0.13      0.90
## Year2008          -0.08753      0.11869      -0.74      0.46
## Year2009          -0.13069      0.12137      -1.08      0.28
## Year2010          -0.02106      0.11787      -0.18      0.86
## Year2011          -0.08831      0.11585      -0.76      0.45
## Year2012          -0.13069      0.11823      -1.11      0.27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.57
## Multiple R-squared:  0.00938,    Adjusted R-squared:  -0.00418
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1138 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.204  0.862  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      7.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1260"
## [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
##   49   46   45   40   26   32   32   20   31   40   41   50   54   41   41
## 2011 2012
##   49   45
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   18   13   12    8   10   11    8   12   16   13   20   21   23   22
## 2011 2012

```

```
## 31 22
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 16 11 9 6 10 9 7 11 13 13 13 13 17 15
## 2011 2012
## 29 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 = 19, df = 16, p-value = 0.3
```



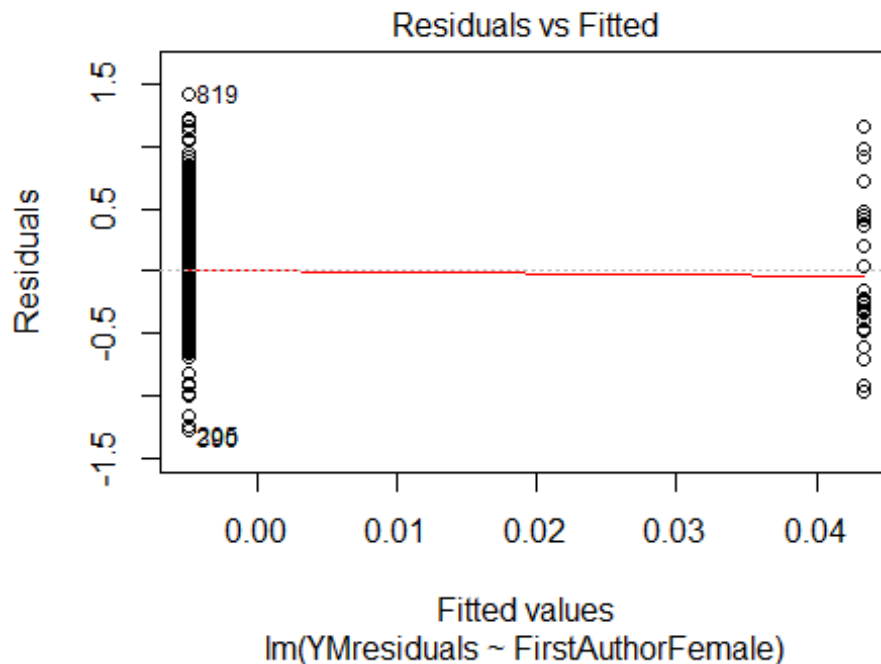
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.38, df = 1, p-value = 0.5

## [1] "Female first author team size 2018 geometric mean: 1.41421356237309"
## [1] "Male first author team size 2018 geometric mean: 1.68206416228297"

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

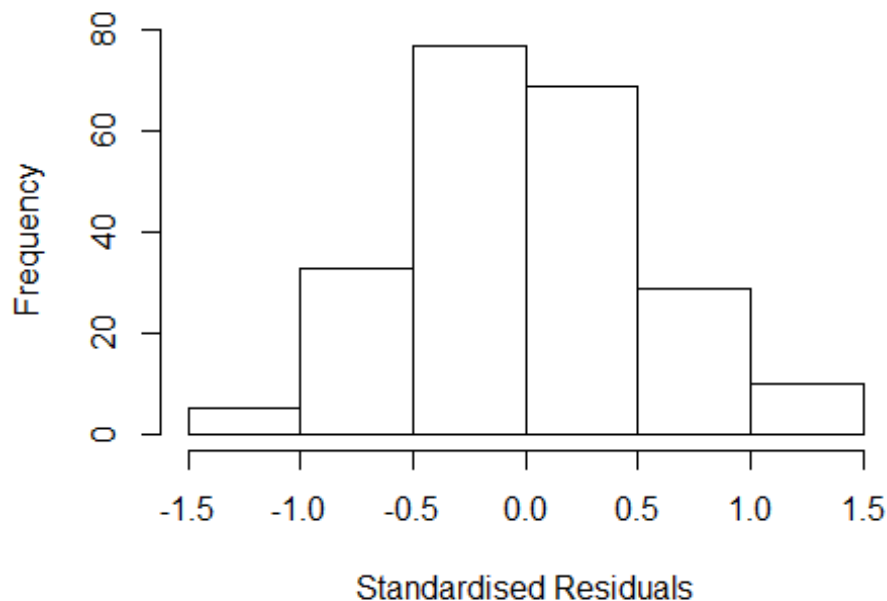
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 32, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.25992104989487"
## [1] "Male last author team size 2018 geometric mean: 1.69535311563252"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.096  1      1.760
## LastAuthorFemale  2.958  1      1.720
## UniqueAuthors    2.596  3      1.172
## Year              4.693 16      1.050
```

Residuals from first and last author and team size



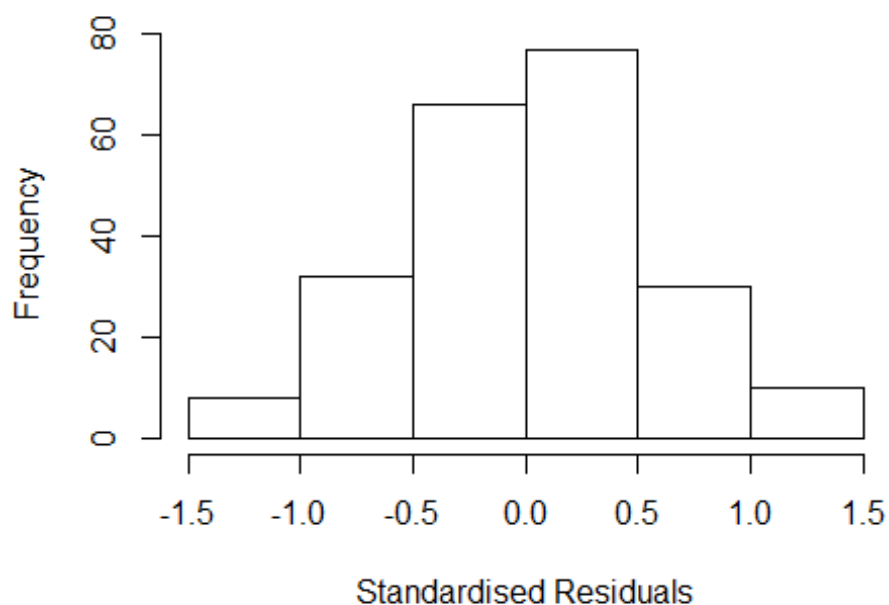
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.332 -0.354 -0.026 0.344 1.357
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94850 0.11816 8.03 8.2e-14 ***
## FirstAuthorFemale1 -0.18015 0.19693 -0.91 0.361
## LastAuthorFemale1 0.21147 0.18014 1.17 0.242
## UniqueAuthors2 0.20309 0.08796 2.31 0.022 *
## UniqueAuthors3 0.30586 0.14370 2.13 0.035 *
## UniqueAuthors4 -0.16776 0.23757 -0.71 0.481
## Year1997 -0.10130 0.16782 -0.60 0.547
## Year1998 0.07024 0.15366 0.46 0.648
## Year1999 0.18074 0.28861 0.63 0.532
## Year2000 -0.10649 0.38487 -0.28 0.782
```

```

## Year2001      0.12789    0.16801    0.76    0.447
## Year2002      0.14169    0.19353    0.73    0.465
## Year2003      0.29299    0.27545    1.06    0.289
## Year2004     -0.09494    0.17867   -0.53    0.596
## Year2005      0.18636    0.26022    0.72    0.475
## Year2006      0.27964    0.19251    1.45    0.148
## Year2007     -0.10866    0.19289   -0.56    0.574
## Year2008      0.29389    0.17680    1.66    0.098 .
## Year2009     -0.38023    0.21719   -1.75    0.082 .
## Year2010     -0.00802    0.28566   -0.03    0.978
## Year2011     -0.08286    0.16777   -0.49    0.622
## Year2012     -0.08016    0.16224   -0.49    0.622
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.14,   Adjusted R-squared:  0.0501
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 207 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.498  0.877  0.956  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      4.48e-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 2.508 1      1.584
## LastAuthorFemale  2.280 1      1.510
## Year              1.981 16      1.022

```


Residuals from first and last author



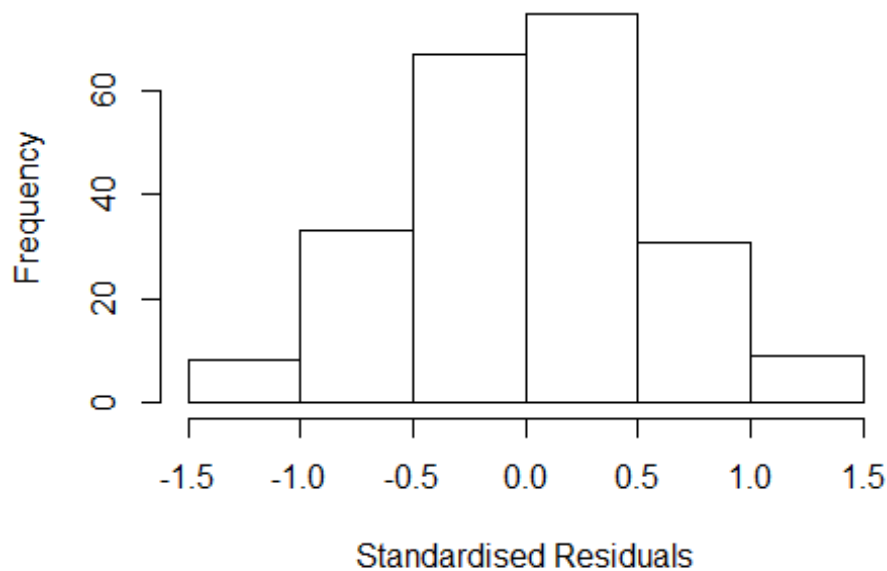
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3759 -0.3663 0.0211 0.3372 1.4125
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9972 0.1031 9.67 <2e-16 ***
## FirstAuthorFemale1 -0.1356 0.1910 -0.71 0.479
## LastAuthorFemale1 0.1799 0.1801 1.00 0.319
## Year1997 -0.0850 0.1543 -0.55 0.582
## Year1998 0.1127 0.1458 0.77 0.440
## Year1999 0.2513 0.2878 0.87 0.384
## Year2000 -0.0644 0.3577 -0.18 0.857
## Year2001 0.1089 0.1776 0.61 0.540
## Year2002 0.1835 0.1866 0.98 0.327
## Year2003 0.3787 0.2490 1.52 0.130
## Year2004 -0.0735 0.1703 -0.43 0.667
## Year2005 0.2176 0.2442 0.89 0.374
```

```

## Year2006          0.3661      0.1717      2.13      0.034 *
## Year2007          -0.0838      0.1948     -0.43      0.667
## Year2008           0.3117      0.1692      1.84      0.067 .
## Year2009          -0.2818      0.1969     -1.43      0.154
## Year2010           0.0627      0.2890      0.22      0.829
## Year2011          -0.0319      0.1574     -0.20      0.839
## Year2012          -0.0509      0.1544     -0.33      0.742
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0217
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 210 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.480  0.874  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      4.48e-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 1.499 1          1.224
## Year              1.499 16          1.013

```

Residuals from first author



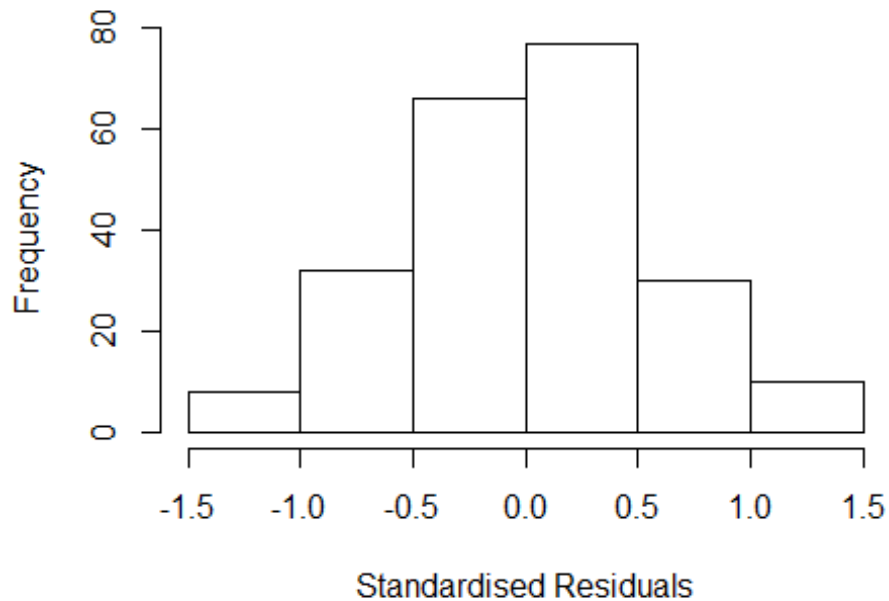
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3978 -0.3471  0.0224  0.3337  1.4063
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9898    0.1009    9.81  <2e-16 ***
## FirstAuthorFemale1 -0.0458    0.1469   -0.31    0.755
## Year1997       -0.0784    0.1528   -0.51    0.609
## Year1998        0.1114    0.1452    0.77    0.444
## Year1999        0.2744    0.2987    0.92    0.359
## Year2000       -0.0591    0.3467   -0.17    0.865
## Year2001        0.1468    0.1685    0.87    0.385
## Year2002        0.1907    0.1851    1.03    0.304
## Year2003        0.4080    0.2651    1.54    0.125
## Year2004       -0.0679    0.1737   -0.39    0.696
## Year2005        0.2206    0.2445    0.90    0.368
## Year2006        0.3812    0.1675    2.28    0.024 *
```

```

## Year2007          -0.0587      0.1878   -0.31    0.755
## Year2008           0.3177      0.1714    1.85    0.065 .
## Year2009          -0.2682      0.1917   -1.40    0.163
## Year2010           0.0672      0.2807    0.24    0.811
## Year2011          -0.0140      0.1537   -0.09    0.927
## Year2012          -0.0395      0.1523   -0.26    0.796
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.0945, Adjusted R-squared:  0.0194
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 209 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.494  0.869  0.958  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      4.48e-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 1.329 1          1.153
## Year            1.329 16          1.009

```

Residuals from last author



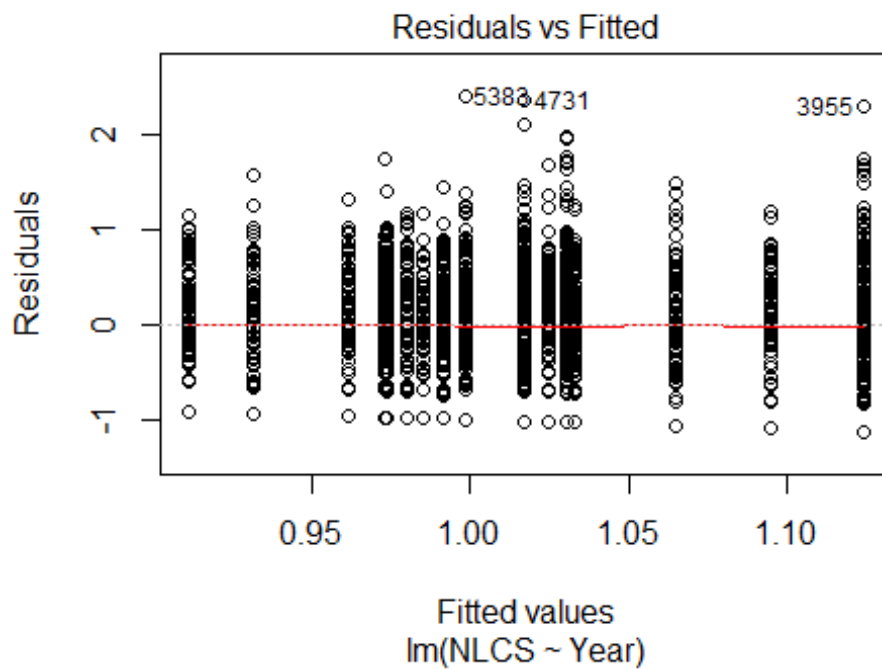
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3796 -0.3631 0.0315 0.3437 1.4163
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9855 0.0988 9.98 <2e-16 ***
## LastAuthorFemale1 0.0992 0.1463 0.68 0.499
## Year1997 -0.0762 0.1516 -0.50 0.616
## Year1998 0.1118 0.1451 0.77 0.442
## Year1999 0.2594 0.2889 0.90 0.370
## Year2000 -0.0956 0.3439 -0.28 0.781
## Year2001 0.1035 0.1770 0.58 0.559
## Year2002 0.1945 0.1835 1.06 0.290
## Year2003 0.3941 0.2549 1.55 0.124
## Year2004 -0.0811 0.1692 -0.48 0.632
## Year2005 0.2239 0.2436 0.92 0.359
## Year2006 0.3744 0.1700 2.20 0.029 *
```

```

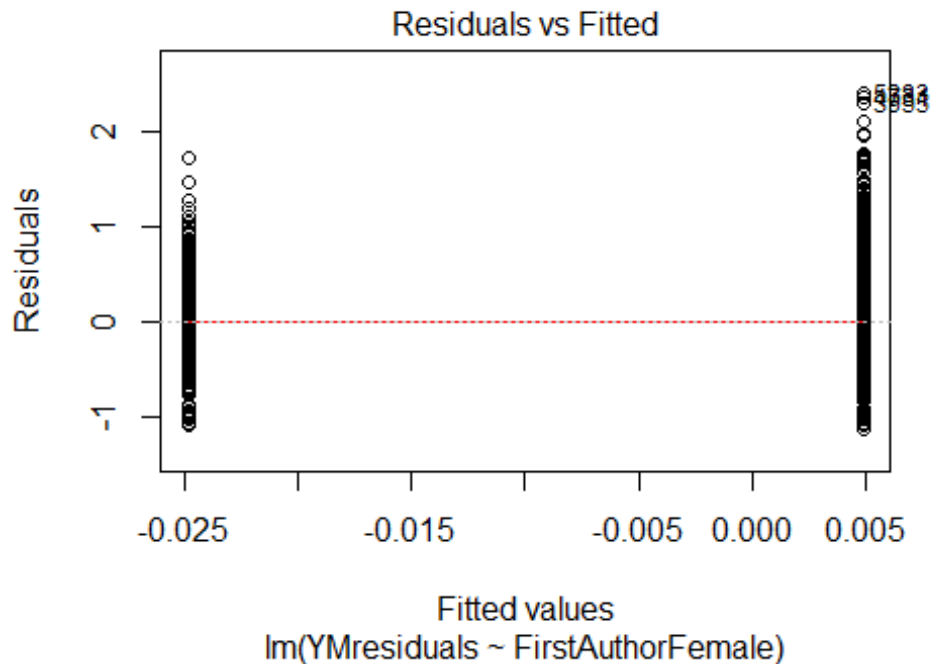
## Year2007          -0.0607      0.1884   -0.32    0.748
## Year2008           0.3027      0.1702    1.78    0.077 .
## Year2009          -0.2738      0.1893   -1.45    0.150
## Year2010           0.0683      0.2713    0.25    0.801
## Year2011          -0.0279      0.1554   -0.18    0.858
## Year2012          -0.0510      0.1524   -0.33    0.738
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.57
## Multiple R-squared:  0.0931, Adjusted R-squared:  0.0179
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 208 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.517  0.881  0.959  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      4.48e-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: 223"
## [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
## 175 194 152 130 154 198 196 145 165 198 161 266 295 353 355
## 2011 2012
## 354 348
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 97 97 77 55 88 94 116 85 93 110 90 153 165 214 218
## 2011 2012

```

```
## 220 210
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 89 90 68 50 75 86 96 68 80 97 72 134 140 172 177
## 2011 2012
## 178 173
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 20, df = 16, p-value = 0.2
```

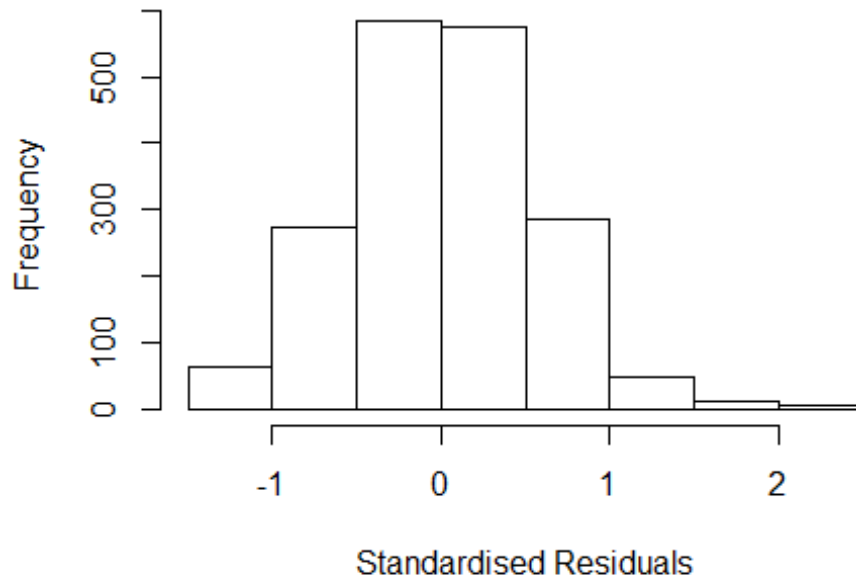


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.9, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 2.19594664982872"
## [1] "Male first author team size 2018 geometric mean: 1.88019877394175"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.81463635666791"
## [1] "Male last author team size 2018 geometric mean: 1.94219102335883"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 900, 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 1.226 1      1.107
## LastAuthorFemale  1.234 1      1.111
## UniqueAuthors     1.229 4      1.026
## Year              1.278 16      1.008
```


Residuals from first and last author and team size



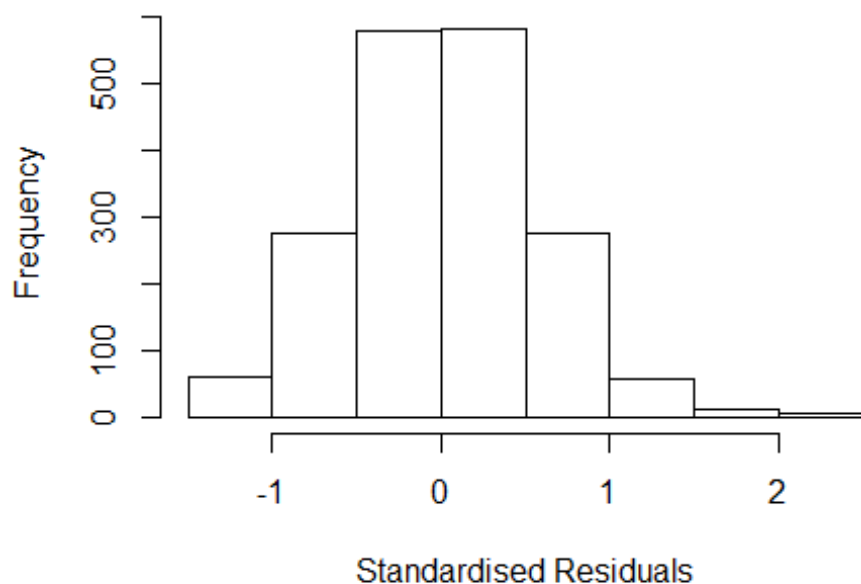
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21224 -0.35883 0.00102 0.36727 2.49823
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96039 0.06470 14.84 < 2e-16 ***
## FirstAuthorFemale1 -0.03084 0.03911 -0.79 0.43
## LastAuthorFemale1 -0.00182 0.04035 -0.05 0.96
## UniqueAuthors2 0.13434 0.03148 4.27 2.1e-05 ***
## UniqueAuthors3 0.18246 0.03849 4.74 2.3e-06 ***
## UniqueAuthors4 0.25440 0.05984 4.25 2.2e-05 ***
## UniqueAuthors5 0.37532 0.07977 4.71 2.7e-06 ***
## Year1997 -0.10436 0.08978 -1.16 0.25
## Year1998 -0.05416 0.10050 -0.54 0.59
## Year1999 -0.04435 0.09933 -0.45 0.66
```

```

## Year2000      -0.13970    0.09453   -1.48    0.14
## Year2001      -0.11128    0.09145   -1.22    0.22
## Year2002      -0.06907    0.08744   -0.79    0.43
## Year2003      -0.07013    0.09274   -0.76    0.45
## Year2004      -0.07521    0.09415   -0.80    0.42
## Year2005       0.00244    0.08452    0.03    0.98
## Year2006       0.11729    0.09155    1.28    0.20
## Year2007      -0.05300    0.07860   -0.67    0.50
## Year2008      -0.00257    0.07725   -0.03    0.97
## Year2009      -0.07341    0.07795   -0.94    0.35
## Year2010       0.06939    0.07917    0.88    0.38
## Year2011      -0.06962    0.07936   -0.88    0.38
## Year2012      -0.08112    0.07708   -1.05    0.29
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.041, Adjusted R-squared:  0.0294
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 180 weights are ~= 1. The remaining 1665 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0042 0.8610 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      5.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.208 1      1.099
## LastAuthorFemale 1.233 1      1.110
## Year              1.072 16      1.002

```

Residuals from first and last author



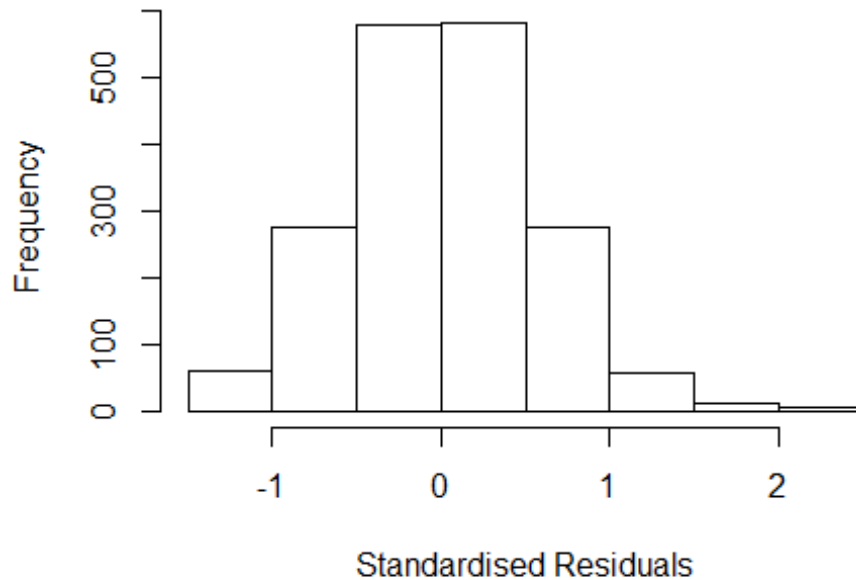
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17740 -0.37040 0.00613 0.37652 2.40656
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00901 0.06356 15.87 <2e-16 ***
## FirstAuthorFemale1 -0.01020 0.03925 -0.26 0.795
## LastAuthorFemale1 -0.00255 0.04112 -0.06 0.951
## Year1997 -0.08451 0.08974 -0.94 0.346
## Year1998 -0.02417 0.10299 -0.23 0.814
## Year1999 -0.03023 0.09878 -0.31 0.760
## Year2000 -0.13281 0.09501 -1.40 0.162
## Year2001 -0.06899 0.09029 -0.76 0.445
## Year2002 -0.03561 0.08594 -0.41 0.679
## Year2003 -0.01896 0.09308 -0.20 0.839
## Year2004 -0.02814 0.09386 -0.30 0.764
## Year2005 0.06919 0.08386 0.83 0.409
```

```

## Year2006          0.16839    0.09300    1.81    0.070 .
## Year2007          -0.01795    0.07785   -0.23    0.818
## Year2008          0.03766    0.07659    0.49    0.623
## Year2009          -0.00660    0.07757   -0.09    0.932
## Year2010          0.12970    0.07846    1.65    0.098 .
## Year2011          0.00422    0.07782    0.05    0.957
## Year2012          -0.01157    0.07619   -0.15    0.879
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.0142, Adjusted R-squared:  0.00446
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 152 weights are ~= 1. The remaining 1693 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0218 0.8650 0.9510 0.9040 0.9880 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.021 1      1.011
## Year              1.021 16      1.001

```

Residuals from first author



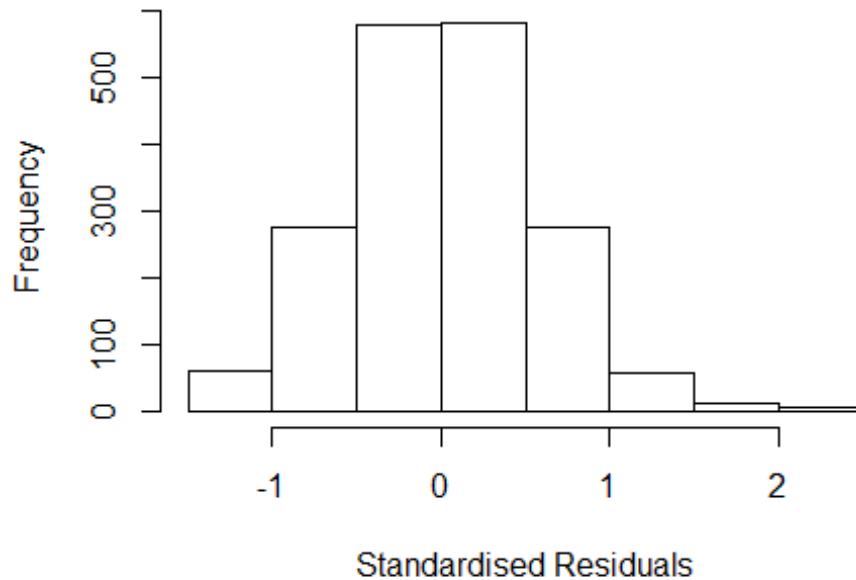
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17705 -0.37023 0.00582 0.37592 2.40658
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0089 0.0635 15.90 <2e-16 ***
## FirstAuthorFemale1 -0.0112 0.0361 -0.31 0.757
## Year1997 -0.0847 0.0896 -0.94 0.345
## Year1998 -0.0242 0.1030 -0.24 0.814
## Year1999 -0.0304 0.0987 -0.31 0.758
## Year2000 -0.1330 0.0948 -1.40 0.161
## Year2001 -0.0690 0.0903 -0.76 0.445
## Year2002 -0.0356 0.0859 -0.41 0.679
## Year2003 -0.0189 0.0931 -0.20 0.839
## Year2004 -0.0281 0.0938 -0.30 0.764
## Year2005 0.0691 0.0838 0.82 0.410
## Year2006 0.1682 0.0930 1.81 0.071 .
```

```

## Year2007          -0.0180      0.0778   -0.23    0.817
## Year2008           0.0376      0.0766    0.49    0.624
## Year2009          -0.0067      0.0776   -0.09    0.931
## Year2010           0.1297      0.0785    1.65    0.098 .
## Year2011           0.0042      0.0778    0.05    0.957
## Year2012          -0.0115      0.0761   -0.15    0.880
## ---
## 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:  0.00501
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 151 weights are ~= 1. The remaining 1694 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0221 0.8650 0.9510 0.9050 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      5.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.042 1          1.021
## Year              1.042 16          1.001

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17608 -0.36987 0.00713 0.37706 2.40813
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00841 0.06348 15.89 <2e-16 ***
## LastAuthorFemale1 -0.00709 0.03789 -0.19 0.852
## Year1997 -0.08394 0.08965 -0.94 0.349
## Year1998 -0.02441 0.10302 -0.24 0.813
## Year1999 -0.03055 0.09871 -0.31 0.757
## Year2000 -0.13360 0.09497 -1.41 0.160
## Year2001 -0.06973 0.09026 -0.77 0.440
## Year2002 -0.03576 0.08596 -0.42 0.677
## Year2003 -0.01954 0.09303 -0.21 0.834
## Year2004 -0.02881 0.09393 -0.31 0.759
## Year2005 0.06911 0.08387 0.82 0.410
## Year2006 0.16767 0.09302 1.80 0.072 .
```

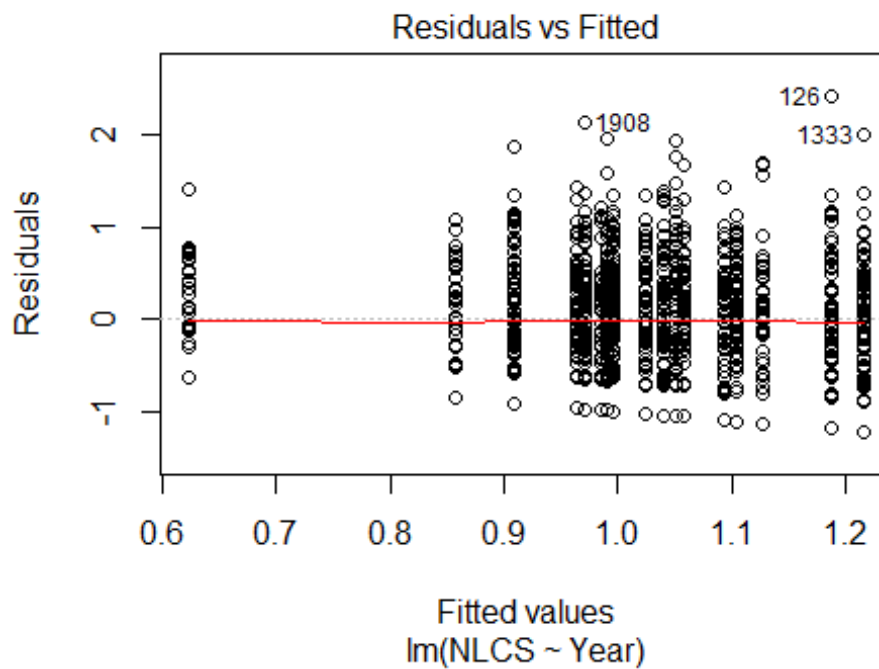
```

## Year2007          -0.01887      0.07789    -0.24      0.809
## Year2008           0.03730      0.07662      0.49      0.626
## Year2009          -0.00682      0.07758    -0.09      0.930
## Year2010           0.12902      0.07850      1.64      0.100
## Year2011           0.00353      0.07783      0.05      0.964
## Year2012          -0.01255      0.07614    -0.16      0.869
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.0141, Adjusted R-squared:  0.00497
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 148 weights are ~= 1. The remaining 1697 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0213 0.8660 0.9510 0.9050 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      5.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1845"
## [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
## 125  81  93  75  108  107  121  100  83  126  156  127  155  203  142
## 2011 2012
## 135  162
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 62  38  51  41  50  60  74  68  57  80  101  80  102  136  92
## 2011 2012

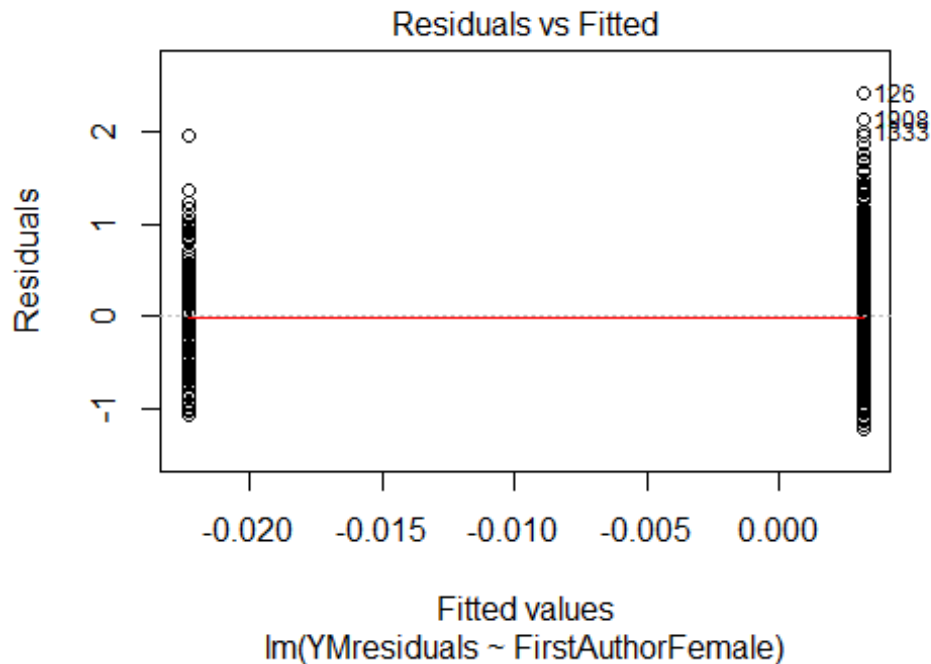
```



```
## 95 110
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 55 32 46 37 41 55 60 62 51 65 81 67 91 116 77
## 2011 2012
## 81 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 = 13, df = 16, p-value = 0.6
```

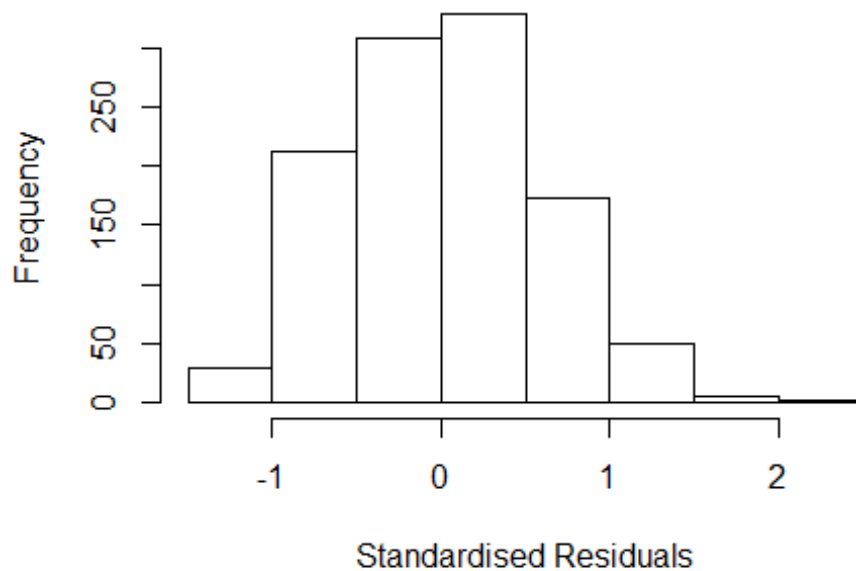


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.3, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 1.93991037446139"
## [1] "Male first author team size 2018 geometric mean: 1.68840737295724"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 500, p-value = 0.2
## 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: 1.71475393524561"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 380, 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 1.194 1      1.093
## LastAuthorFemale  1.153 1      1.074
## UniqueAuthors    1.308 4      1.034
## Year              1.344 16     1.009
```

Residuals from first and last author and team size



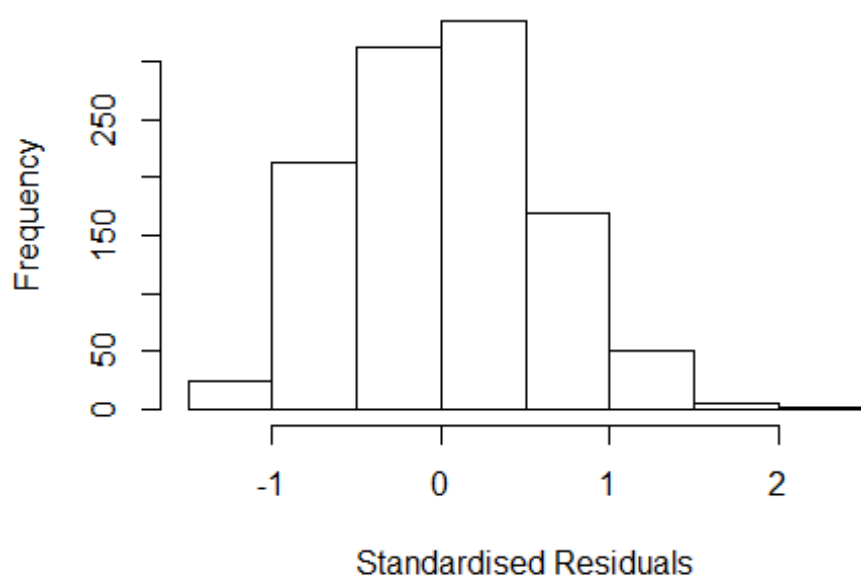
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23710 -0.41485 0.00718 0.40778 2.37325
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1982 0.0999 12.00 < 2e-16 ***
## FirstAuthorFemale1 -0.0112 0.0578 -0.19 0.8463
## LastAuthorFemale1 -0.0684 0.0611 -1.12 0.2631
## UniqueAuthors2 0.0355 0.0435 0.82 0.4142
## UniqueAuthors3 0.0895 0.0532 1.68 0.0925 .
## UniqueAuthors4 0.1393 0.0898 1.55 0.1212
## UniqueAuthors5 0.1564 0.1538 1.02 0.3094
## Year1997 -0.4186 0.1401 -2.99 0.0029 **
## Year1998 -0.2711 0.1269 -2.14 0.0330 *
## Year1999 -0.5781 0.1377 -4.20 2.9e-05 ***
```

```

## Year2000          -0.2254      0.1362    -1.66    0.0982 .
## Year2001          -0.2063      0.1302    -1.58    0.1134
## Year2002          -0.2613      0.1347    -1.94    0.0526 .
## Year2003          -0.2290      0.1226    -1.87    0.0622 .
## Year2004          -0.1759      0.1288    -1.37    0.1724
## Year2005          -0.0506      0.1227    -0.41    0.6802
## Year2006          -0.1495      0.1192    -1.25    0.2102
## Year2007          -0.2882      0.1268    -2.27    0.0232 *
## Year2008          -0.2880      0.1172    -2.46    0.0141 *
## Year2009          -0.2370      0.1136    -2.09    0.0372 *
## Year2010          -0.3517      0.1242    -2.83    0.0047 **
## Year2011          -0.2132      0.1216    -1.75    0.0798 .
## Year2012          -0.2859      0.1191    -2.40    0.0166 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.616
## Multiple R-squared:  0.0387, Adjusted R-squared:  0.0193
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 101 weights are ~= 1. The remaining 1009 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.105  0.883  0.948  0.914  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.01e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.155 1          1.075
## LastAuthorFemale  1.151 1          1.073
## Year              1.077 16          1.002

```

Residuals from first and last author



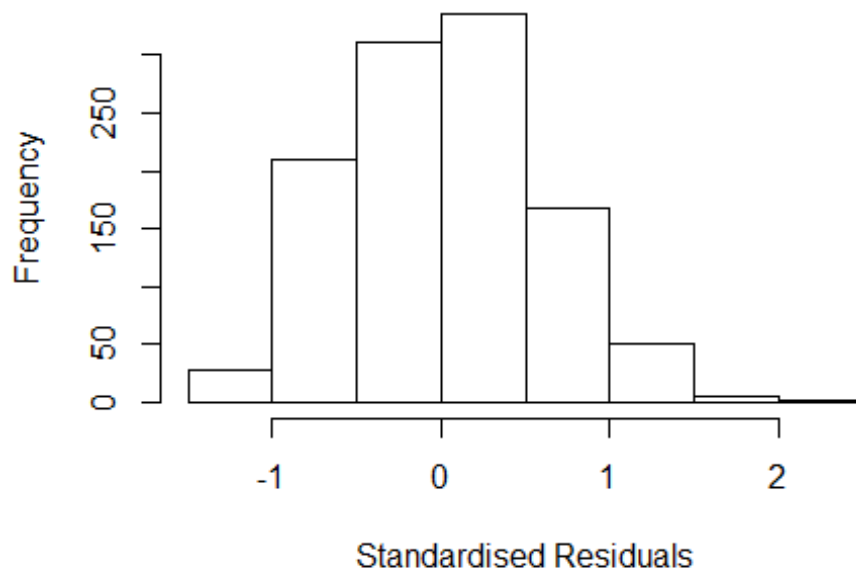
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21682 -0.43553 0.00854 0.39935 2.39018
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21682 0.09990 12.18 <2e-16 ***
## FirstAuthorFemale1 -0.00278 0.05723 -0.05 0.9613
## LastAuthorFemale1 -0.06468 0.06181 -1.05 0.2956
## Year1997 -0.40693 0.14077 -2.89 0.0039 **
## Year1998 -0.27232 0.12736 -2.14 0.0327 *
## Year1999 -0.57806 0.13800 -4.19 3e-05 ***
## Year2000 -0.20549 0.13596 -1.51 0.1310
## Year2001 -0.19120 0.13075 -1.46 0.1439
## Year2002 -0.24771 0.13552 -1.83 0.0678 .
## Year2003 -0.21995 0.12275 -1.79 0.0734 .
## Year2004 -0.16834 0.12909 -1.30 0.1925
## Year2005 -0.03011 0.12205 -0.25 0.8052
```

```

## Year2006      -0.13233    0.12029   -1.10    0.2715
## Year2007      -0.27843    0.12718   -2.19    0.0288 *
## Year2008      -0.27161    0.11767   -2.31    0.0212 *
## Year2009      -0.22539    0.11401   -1.98    0.0483 *
## Year2010      -0.33884    0.12442   -2.72    0.0066 **
## Year2011      -0.18844    0.12169   -1.55    0.1218
## Year2012      -0.26670    0.11939   -2.23    0.0257 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.0346, Adjusted R-squared:  0.0186
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 94 weights are ~= 1. The remaining 1016 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0971 0.8830 0.9480 0.9140 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.043 1      1.021
## Year      1.043 16      1.001

```

Residuals from first author



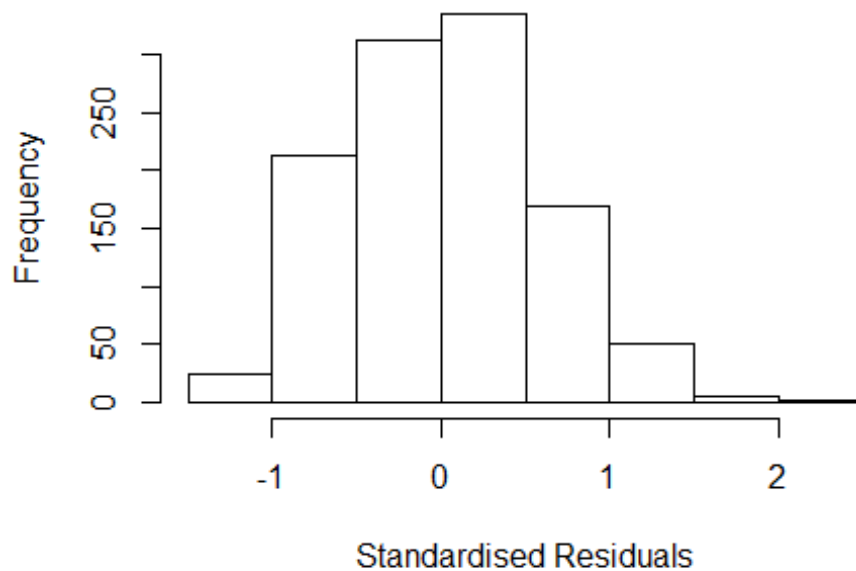
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20910 -0.43209 0.00936 0.40680 2.39790
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2091 0.0998 12.12 < 2e-16 ***
## FirstAuthorFemale1 -0.0249 0.0547 -0.45 0.6495
## Year1997 -0.4014 0.1410 -2.85 0.0045 **
## Year1998 -0.2656 0.1278 -2.08 0.0379 *
## Year1999 -0.5772 0.1384 -4.17 3.3e-05 ***
## Year2000 -0.2035 0.1367 -1.49 0.1370
## Year2001 -0.1890 0.1310 -1.44 0.1495
## Year2002 -0.2453 0.1355 -1.81 0.0705 .
## Year2003 -0.2140 0.1228 -1.74 0.0817 .
## Year2004 -0.1652 0.1289 -1.28 0.2001
## Year2005 -0.0244 0.1219 -0.20 0.8414
## Year2006 -0.1306 0.1206 -1.08 0.2791
```

```

## Year2007          -0.2763      0.1276   -2.16   0.0306 *
## Year2008          -0.2664      0.1176   -2.27   0.0237 *
## Year2009          -0.2237      0.1143   -1.96   0.0506 .
## Year2010          -0.3401      0.1247   -2.73   0.0065 **
## Year2011          -0.1857      0.1219   -1.52   0.1278
## Year2012          -0.2648      0.1196   -2.21   0.0270 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.0335, Adjusted R-squared:  0.0184
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1013 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0935 0.8840 0.9480 0.9140 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.037 1          1.018
## Year          1.037 16          1.001

```


Residuals from last author



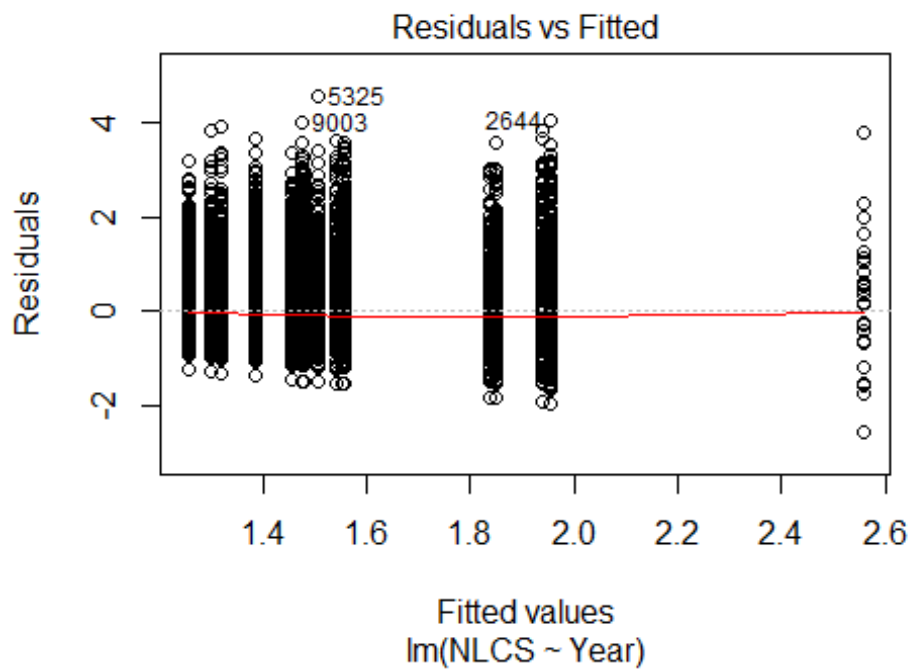
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21660 -0.43527 0.00882 0.39673 2.39040
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2166 0.1001 12.15 < 2e-16 ***
## LastAuthorFemale1 -0.0657 0.0587 -1.12 0.2636
## Year1997 -0.4069 0.1407 -2.89 0.0039 **
## Year1998 -0.2725 0.1269 -2.15 0.0319 *
## Year1999 -0.5780 0.1381 -4.19 3.1e-05 ***
## Year2000 -0.2054 0.1361 -1.51 0.1316
## Year2001 -0.1912 0.1307 -1.46 0.1437
## Year2002 -0.2476 0.1356 -1.83 0.0680 .
## Year2003 -0.2200 0.1226 -1.80 0.0729 .
## Year2004 -0.1682 0.1291 -1.30 0.1929
## Year2005 -0.0302 0.1216 -0.25 0.8037
## Year2006 -0.1323 0.1202 -1.10 0.2712
```

```

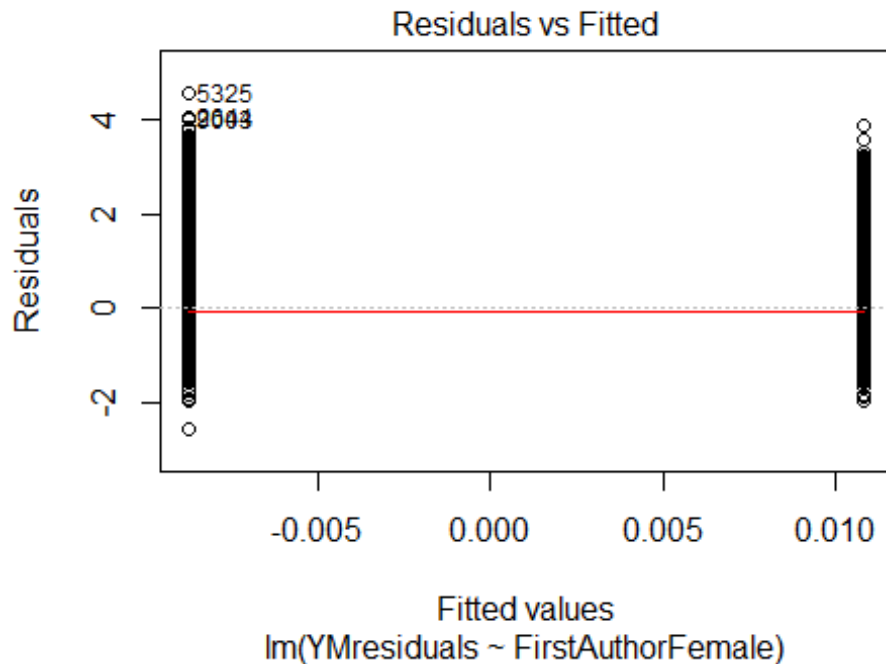
## Year2007          -0.2785      0.1270    -2.19    0.0286 *
## Year2008          -0.2717      0.1174    -2.31    0.0208 *
## Year2009          -0.2255      0.1135    -1.99    0.0473 *
## Year2010          -0.3389      0.1243    -2.73    0.0065 **
## Year2011          -0.1886      0.1215    -1.55    0.1209
## Year2012          -0.2666      0.1195    -2.23    0.0259 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.0346, Adjusted R-squared:  0.0196
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 95 weights are ~= 1. The remaining 1015 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.096  0.882  0.948  0.914  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1110"
## [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
##  609  623  621  625   76  686  633  580  588  687  969 1137 1271 1319 1625
## 2011 2012
## 1236 1366
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  289  320  317  310   31  209  471  438  455  535  752  880 1022 1040 1255
## 2011 2012

```

```
## 1012 1087
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 260 295 286 281 30 194 417 403 398 483 665 784 924 933 1107
## 2011 2012
## 908 978
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 500, df = 16, p-value <2e-16
```

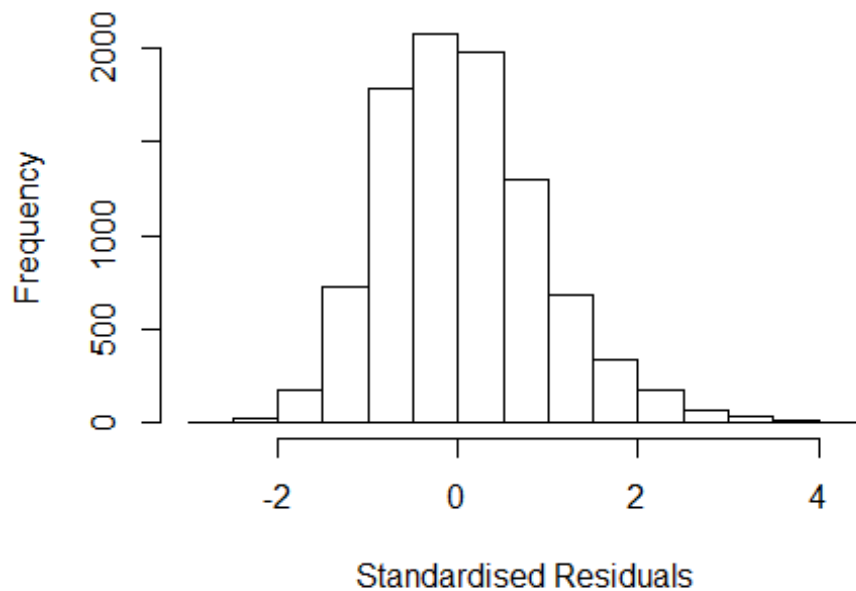


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 25, df = 1, p-value = 5e-07
```



```
## [1] "Female first author team size 2018 geometric mean: 3.91439961029845"
## [1] "Male first author team size 2018 geometric mean: 2.96046050631809"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24000, p-value = 4e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.57135526344093"
## [1] "Male last author team size 2018 geometric mean: 3.28271134872981"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 21000, 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 1.205 1          1.098
## LastAuthorFemale  1.207 1          1.099
## UniqueAuthors     1.099 4          1.012
## Year               1.100 16         1.003
```

Residuals from first and last author and team size



```
## [1] "List of 103 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 214      0030603333 5.098 1996      2700      1      2.656
## 274      0029799559 5.048 1996      2700      1      2.606
## 316      0029793301 3.898 1996      2700      1      2.632
## 356      0030056515 4.523 1996      2700      1      2.746
## 372      0029814354 4.211 1996      2700      1      2.971
## 450      0029862973 5.794 1996      2700      1      3.352
## 486      0006590545 5.593 1996      2700      1      3.151
## 560      0001259453 5.013 1996      2700      1      3.773
## 715      0031588519 5.136 1997      2700      1      2.680
## 856      0031583760 4.794 1997      2700      1      3.540
## 904      0030808429 5.491 1997      2700      1      3.035
## 963      0030800795 5.042 1997      2700      1      2.577
## 1000     0030811896 5.020 1997      2700      1      3.052
## 1138     0031029373 5.109 1997      2700      1      3.000
## 1151     0030981335 5.288 1997      2700      1      2.823
## 1183     0031024865 5.035 1997      2700      1      3.252
## 1184     0031017667 5.126 1997      2700      1      3.185
## 1781     0032555023 4.631 1998      2700      1      2.697
## 1783     2642599026 5.421 1998      2700      1      4.193
## 1820     0032554695 3.912 1998      2700      1      2.657
## 2426     0033595136 5.107 1999      2700      1      2.980
## 2644     0343360868 5.997 1999      2700      1      3.504
## 2654     0033528281 4.383 1999      2700      1      3.111
## 2777     0033193134 4.710 1999      2700      1      2.575
## 2876     0034430489 0.000 2000      2700      1     -2.733
```

##	2942	0034688194	6.362	2000	2700	1	4.316
##	3254	0035975409	4.758	2001	2700	1	3.008
##	3417	0035906286	4.818	2001	2700	1	2.892
##	3519	0035843250	4.421	2001	2700	1	2.671
##	3763	0036440553	4.664	2002	2700	1	3.312
##	4001	0037167013	4.710	2002	2700	1	3.192
##	4188	0037045882	3.961	2002	2700	1	2.601
##	4243	0037176525	3.951	2002	2700	1	3.128
##	4351	0036834665	4.324	2002	2700	1	3.474
##	4513	0344420330	4.990	2003	2700	1	2.945
##	4514	0642364501	4.434	2003	2700	1	2.904
##	4741	0037976821	4.681	2003	2700	1	2.636
##	4786	0013165966	4.910	2003	2700	1	2.838
##	4869	0037443934	4.690	2003	2700	1	2.985
##	4912	0037456354	5.153	2003	2700	1	3.100
##	4960	0037413466	5.051	2003	2700	1	3.006
##	5318	4544230001	4.656	2004	2700	1	2.637
##	5391	3442886513	4.914	2004	2700	1	2.903
##	5447	3042738109	4.386	2004	2700	1	2.864
##	5589	1442339082	3.761	2004	2700	1	2.952
##	5959	27844535729	4.362	2005	2700	1	2.678
##	6017	27544511378	4.326	2005	2700	1	2.809
##	6136	24044485648	5.177	2005	2700	1	3.145
##	6301	19744368662	4.104	2005	2700	1	2.745
##	6312	21844449936	4.142	2005	2700	1	2.757
##	6314	21844456036	3.452	2005	2700	1	2.622
##	6315	21844458935	4.430	2005	2700	1	3.574
##	6372	15744389846	4.744	2005	2700	1	2.694
##	6382	20044383911	4.809	2005	2700	1	3.953
##	6747	33847712261	4.255	2006	2700	1	2.942
##	6789	34748880596	3.824	2006	2700	1	3.058
##	7139	34548410860	3.330	2006	2700	1	2.537
##	7238	33846236864	3.442	2006	2700	1	2.676
##	7239	34247562149	4.092	2006	2700	1	3.299
##	7293	33745881036	4.378	2006	2700	1	3.612
##	7305	33745400514	4.384	2006	2700	1	2.755
##	7474	33646130849	3.945	2006	2700	1	2.650
##	7493	33645500931	3.462	2006	2700	1	2.669
##	7495	33645497961	4.742	2006	2700	1	3.976
##	7680	313444440606	4.440	2006	2700	1	3.145
##	7928	39549119621	3.335	2007	2700	1	2.522
##	8148	35348967810	5.076	2007	2700	1	3.079
##	8149	35348983928	3.291	2007	2700	1	2.505
##	8252	34548772315	4.777	2007	2700	1	2.789
##	8266	34548387504	3.335	2007	2700	1	2.549
##	8415	38449084393	3.345	2007	2700	1	2.532
##	8464	34547666899	4.192	2007	2700	1	3.406
##	8703	34547850842	3.795	2007	2700	1	3.009
##	8873	33847303686	4.719	2007	2700	1	3.377
##	8941	33847172327	4.158	2007	1000	2	2.843

```

## 8950 33847759152 4.625 2007 2700 1 2.637
## 9003 33846407655 5.477 2007 2700 1 3.489
## 9558 53249119070 3.698 2008 2700 1 2.914
## 10380 51649111466 3.394 2008 2700 1 2.610
## 10556 38949213691 4.140 2008 2700 1 2.827
## 10826 77955587813 3.204 2009 2700 1 2.532
## 11217 70349155477 4.446 2009 2700 1 2.572
## 11293 77953805846 3.268 2009 2700 1 2.596
## 11513 67649407404 4.381 2009 2700 1 2.507
## 11871 62649164844 4.746 2009 2700 1 2.872
## 11873 65349171524 4.140 2009 2700 1 2.596
## 11915 61849155730 5.054 2009 2700 1 3.180
## 12228 70449653474 5.051 2009 2700 1 3.177
## 12543 78649906763 4.032 2010 2700 2 3.452
## 13205 77956635947 3.646 2010 2700 1 2.537
## 13895 77950273246 5.118 2010 2700 1 3.336
## 14190 76649099557 4.472 2010 2700 1 2.671
## 14205 75849147039 4.355 2010 2700 1 2.564
## 14923 80455173556 4.657 2011 2700 1 2.827
## 15020 80054938852 3.423 2011 2700 1 2.777
## 15192 80052868407 3.552 2011 2700 1 2.906
## 15618 84858996807 4.302 2011 2700 1 2.802
## 15747 79956110208 3.923 2011 2700 1 2.616
## 15877 79955132796 4.367 2011 2700 1 2.545
## 15895 79953775836 5.237 2011 2700 1 3.389
## 16095 79951995822 4.559 2011 2700 1 2.729
## 17892 84857815393 4.076 2012 2700 1 2.850
## 18065 84856529575 4.439 2012 2700 1 2.708
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -2.7330 -0.5839 -0.0238 0.5803 4.3160
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23962 0.08628 14.37 < 2e-16 ***
## FirstAuthorFemale1 0.00812 0.02008 0.40 0.686
## LastAuthorFemale1 0.01832 0.02077 0.88 0.378
## UniqueAuthors2 0.52887 0.02713 19.49 < 2e-16 ***
## UniqueAuthors3 0.68700 0.02755 24.94 < 2e-16 ***
## UniqueAuthors4 0.85417 0.03049 28.01 < 2e-16 ***
## UniqueAuthors5 1.20214 0.02842 42.29 < 2e-16 ***
## Year1997 0.01472 0.11168 0.13 0.895

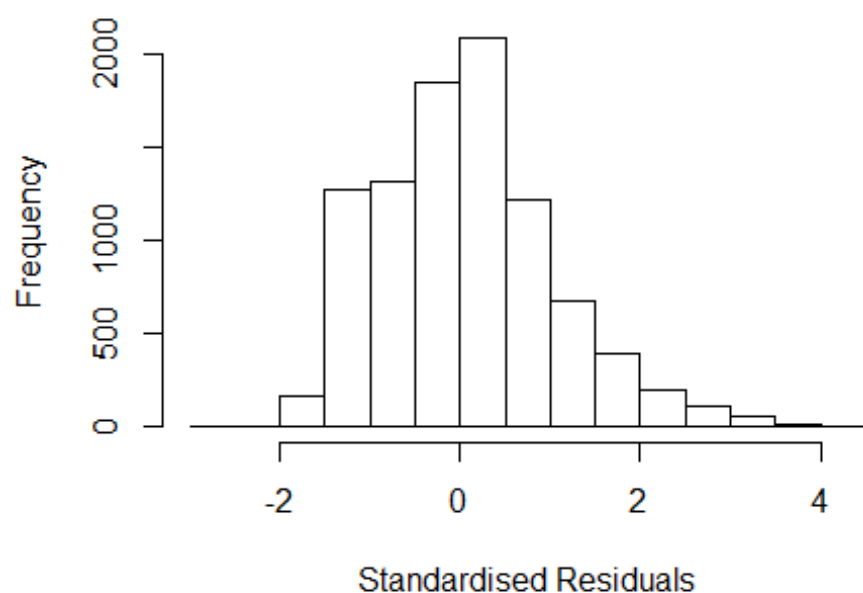
```

```

## Year1998      -0.01121    0.10852   -0.10    0.918
## Year1999      0.03282    0.10495    0.31    0.754
## Year2000      0.80640    0.32698    2.47    0.014 *
## Year2001     -0.01854    0.11428   -0.16    0.871
## Year2002     -0.41631    0.09735   -4.28   1.9e-05 ***
## Year2003     -0.39655    0.09952   -3.98   6.8e-05 ***
## Year2004     -0.43094    0.09713   -4.44   9.2e-06 ***
## Year2005     -0.40958    0.09471   -4.32   1.5e-05 ***
## Year2006     -0.47333    0.09177   -5.16   2.6e-07 ***
## Year2007     -0.45337    0.09063   -5.00   5.8e-07 ***
## Year2008     -0.45588    0.08986   -5.07   4.0e-07 ***
## Year2009     -0.56768    0.08905   -6.37   1.9e-10 ***
## Year2010     -0.65934    0.08722   -7.56   4.4e-14 ***
## Year2011     -0.62005    0.08839   -7.01   2.5e-12 ***
## Year2012     -0.71886    0.08740   -8.22   2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.817
## Multiple R-squared:  0.233, Adjusted R-squared:  0.231
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 4 observations c(718,1142,2843,3510)
## are outliers with |weight| = 0 ( < 1.1e-05);
## 741 weights are ~= 1. The remaining 8601 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0009 0.8640 0.9460 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      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 1.138 1          1.067
## LastAuthorFemale 1.127 1          1.062
## Year 1.022 16          1.001

```


Residuals from first and last author



```
## [1] "List of 180 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 10      0029806876 4.368 1996    2700      1    2.587
## 183     0029907525 4.716 1996    2700      1    2.935
## 214     0030603333 5.098 1996    2700      1    3.317
## 274     0029799559 5.048 1996    2700      1    3.267
## 325     9344249534 4.358 1996    2700      1    2.577
## 356     0030056515 4.523 1996    2700      1    2.666
## 373     0029938206 4.446 1996    2700      1    2.665
## 426     0029880393 4.724 1996    2700      1    3.011
## 427     0029990981 4.429 1996    2700      1    2.648
## 450     0029862973 5.794 1996    2700      1    4.013
## 486     0006590545 5.593 1996    2700      1    3.812
## 496     0030093546 4.950 1996    2700      1    3.161
## 553     0030047924 4.654 1996    2700      1    2.873
## 560     0001259453 5.013 1996    2700      1    3.232
## 715     0031588519 5.136 1997    2700      1    3.317
## 856     0031583760 4.794 1997    2700      1    2.975
## 904     0030808429 5.491 1997    2700      1    3.672
## 962     18544412975 4.558 1997    2700      1    2.662
## 963     0030800795 5.042 1997    2700      1    3.146
## 976     1842338658 4.759 1997    2700      1    2.940
## 1000    0030811896 5.020 1997    2700      1    3.192
## 1002    0031578297 4.624 1997    2700      1    2.805
## 1025    0030908220 4.463 1997    2700      1    2.567
## 1030    0030943882 4.662 1997    2700      1    2.910
## 1138    0031029373 5.109 1997    2700      1    3.290
```

##	1151	0030981335	5.288	1997	2700	1	3.392
##	1177	0031048342	4.524	1997	2700	1	2.705
##	1183	0031024865	5.035	1997	2700	1	3.216
##	1184	0031017667	5.126	1997	2700	1	3.307
##	1212	0031022182	4.739	1997	2700	1	2.843
##	1269	0031035439	4.806	1997	2700	1	2.910
##	1420	0032539180	4.542	1998	2700	1	2.754
##	1687	0032511413	4.449	1998	2700	1	2.585
##	1736	0032562946	4.354	1998	2700	1	2.566
##	1781	0032555023	4.631	1998	2700	1	2.911
##	1783	2642599026	5.421	1998	2700	1	3.633
##	1866	6844236394	4.749	1998	2700	1	2.953
##	1888	0032033812	4.367	1998	2700	1	2.571
##	1913	0032507027	4.854	1998	2700	1	3.066
##	1916	0032484918	4.822	1998	2700	1	3.034
##	1955	0032495544	4.571	1998	2700	1	2.783
##	2281	0032717944	4.535	1999	2700	1	2.654
##	2426	0033595136	5.107	1999	2700	1	3.226
##	2606	0033580398	4.314	1999	2700	1	2.501
##	2644	0343360868	5.997	1999	2700	1	4.184
##	2654	0033528281	4.383	1999	2700	1	2.502
##	2777	0033193134	4.710	1999	2700	1	2.752
##	2876	0034430489	0.000	2000	2700	1	-2.613
##	2880	53949090522	0.000	2000	2700	2	-2.613
##	2937	85008539534	0.000	2000	2700	2	-2.613
##	2940	85007753361	0.000	2000	2700	1	-2.613
##	2942	0034688194	6.362	2000	2700	1	3.749
##	3106	0035965633	4.678	2001	2700	1	2.933
##	3169	0035904362	4.784	2001	2700	1	3.039
##	3254	0035975409	4.758	2001	2700	1	3.013
##	3284	0035833493	4.848	2001	2700	1	3.095
##	3383	0035810589	4.879	2001	2700	1	3.134
##	3417	0035906286	4.818	2001	2700	1	3.141
##	3512	0035941514	4.680	2001	2700	1	2.927
##	3519	0035843250	4.421	2001	2700	1	2.676
##	3763	0036440553	4.664	2002	2700	1	3.239
##	3907	0037151906	3.930	2002	2700	1	2.505
##	3999	0037178575	3.986	2002	2700	1	2.561
##	4001	0037167013	4.710	2002	2700	1	3.209
##	4179	0037076029	4.344	2002	2700	1	2.843
##	4196	0037035122	4.319	2002	2700	1	2.894
##	4243	0037176525	3.951	2002	2700	1	2.526
##	4246	0037160563	4.315	2002	2700	1	2.890
##	4351	0036834665	4.324	2002	2700	1	2.891
##	4454	0346555267	4.287	2003	2700	1	2.886
##	4513	0344420330	4.990	2003	2700	1	3.589
##	4514	0642364501	4.434	2003	2700	1	3.033
##	4595	0141763668	4.176	2003	2700	1	2.699
##	4726	0037832413	3.991	2003	2700	1	2.590
##	4741	0037976821	4.681	2003	2700	1	3.280

##	4786	0013165966	4.910	2003	2700	1	3.501
##	4826	0242578548	4.063	2003	2700	1	2.662
##	4834	0037414162	4.528	2003	2700	1	3.127
##	4869	0037443934	4.690	2003	2700	1	3.213
##	4875	0037420266	4.266	2003	2700	1	2.789
##	4912	0037456354	5.153	2003	2700	1	3.676
##	4960	0037413466	5.051	2003	2700	1	3.650
##	4979	0037380645	4.167	2003	2700	1	2.690
##	5076	1542494137	4.153	2003	2700	1	2.744
##	5217	9244237570	4.055	2004	2700	1	2.655
##	5318	4544230001	4.656	2004	2700	1	3.180
##	5377	4143096183	4.405	2004	2700	1	3.005
##	5391	3442886513	4.914	2004	2700	1	3.514
##	5420	3142705643	4.188	2004	2700	1	2.712
##	5447	3042738109	4.386	2004	2700	1	2.977
##	5498	2542570187	4.122	2004	2700	1	2.790
##	5560	1842430968	3.938	2004	2700	1	2.606
##	5637	1442309702	4.035	2004	2700	1	2.635
##	5678	1642455902	3.970	2004	2700	1	2.570
##	5959	27844535729	4.362	2005	2700	1	2.892
##	5960	28044462641	4.500	2005	2700	1	3.022
##	6017	27544511378	4.326	2005	2700	1	2.856
##	6021	26844508456	4.078	2005	2700	1	2.608
##	6136	240444485648	5.177	2005	2700	1	3.707
##	6301	19744368662	4.104	2005	2700	1	2.634
##	6312	218444449936	4.142	2005	2700	1	2.664
##	6315	21844458935	4.430	2005	2700	1	2.952
##	6372	15744389846	4.744	2005	2700	1	3.342
##	6382	20044383911	4.809	2005	2700	1	3.331
##	6747	33847712261	4.255	2006	2700	1	2.941
##	6935	33750958903	3.908	2006	2700	1	2.526
##	7206	33747108061	3.894	2006	2700	1	2.512
##	7239	34247562149	4.092	2006	2700	1	2.701
##	7293	33745881036	4.378	2006	2700	1	2.996
##	7305	33745400514	4.384	2006	2700	1	2.926
##	7431	33646681444	4.429	2006	2700	1	2.971
##	7474	33646130849	3.945	2006	2700	1	2.563
##	7495	33645497961	4.742	2006	2700	1	3.360
##	7540	33645399023	4.082	2006	2700	1	2.624
##	7637	31844434994	4.007	2006	2700	1	2.625
##	7680	313444440606	4.440	2006	2700	1	3.058
##	7837	37149034178	4.371	2007	2700	1	2.902
##	8048	36749016190	3.920	2007	2700	1	2.595
##	8061	35948976968	4.234	2007	2700	1	2.765
##	8148	35348967810	5.076	2007	2700	1	3.607
##	8163	34948904387	4.484	2007	2700	1	3.091
##	8252	34548772315	4.777	2007	2700	1	3.384
##	8263	34548603259	4.369	2007	2700	1	2.976
##	8278	34548445010	4.263	2007	2700	1	2.794
##	8464	34547666899	4.192	2007	2700	1	2.799

##	8473	34547167458	4.103	2007	2700	1	2.634
##	8750	34247468559	4.385	2007	2700	1	3.060
##	8873	33847303686	4.719	2007	2700	1	3.317
##	8941	33847172327	4.158	2007	1000	2	2.765
##	8950	33847759152	4.625	2007	2700	1	3.232
##	8957	33846806078	4.110	2007	2700	1	2.717
##	9003	33846407655	5.477	2007	2700	1	4.084
##	9196	58149145568	4.036	2008	2700	1	2.628
##	9491	56649088178	4.078	2008	2700	1	2.670
##	9669	51649088482	4.382	2008	2700	1	2.897
##	9788	49749141653	4.171	2008	2700	1	2.686
##	9859	48349118873	4.078	2008	2700	1	2.661
##	9898	47649130481	3.937	2008	2700	1	2.597
##	10525	38349073481	4.179	2008	2700	1	2.771
##	10556	38949213691	4.140	2008	2700	1	2.732
##	11091	70349766844	4.197	2009	2700	1	2.954
##	11217	70349155477	4.446	2009	2700	1	3.135
##	11340	69249208815	4.005	2009	2700	1	2.762
##	11513	67649407404	4.381	2009	2700	1	3.070
##	11654	65649088288	4.306	2009	2700	1	3.063
##	11661	67650035326	4.048	2009	2700	1	2.737
##	11715	69049117564	4.182	2009	2700	1	2.939
##	11871	62649164844	4.746	2009	2700	1	3.435
##	11873	65349171524	4.140	2009	2700	1	2.897
##	11915	61849155730	5.054	2009	2700	1	3.743
##	11922	65649085603	4.033	2009	2700	1	2.722
##	12028	61449244692	3.743	2009	2700	1	2.500
##	12228	70449653474	5.051	2009	2700	1	3.740
##	12543	78649906763	4.032	2010	2700	2	2.783
##	13239	77956056661	4.268	2010	2700	1	3.019
##	13895	77950273246	5.118	2010	2700	1	3.869
##	13904	77951240308	3.906	2010	2700	1	2.657
##	14190	76649099557	4.472	2010	2700	1	3.291
##	14205	75849147039	4.355	2010	2700	1	3.030
##	14889	81255143853	3.841	2011	2700	1	2.592
##	14923	80455173556	4.657	2011	2700	1	3.341
##	14959	82455205788	3.890	2011	2700	1	2.641
##	15326	80052073777	3.748	2011	2700	1	2.508
##	15421	84859006414	3.927	2011	2700	1	2.687
##	15469	79959748876	3.820	2011	2700	1	2.571
##	15618	84858996807	4.302	2011	2700	1	3.053
##	15747	79956110208	3.923	2011	2700	1	2.683
##	15851	79957546192	3.765	2011	2700	1	2.516
##	15855	79955474494	3.788	2011	2700	1	2.548
##	15877	79955132796	4.367	2011	2700	1	3.127
##	15895	79953775836	5.237	2011	2700	1	3.988
##	16023	79953726311	3.918	2011	2700	1	2.602
##	16095	79951995822	4.559	2011	2700	1	3.243
##	16108	79951649740	3.820	2011	2700	1	2.580
##	16915	84867161086	3.880	2012	2700	1	2.662

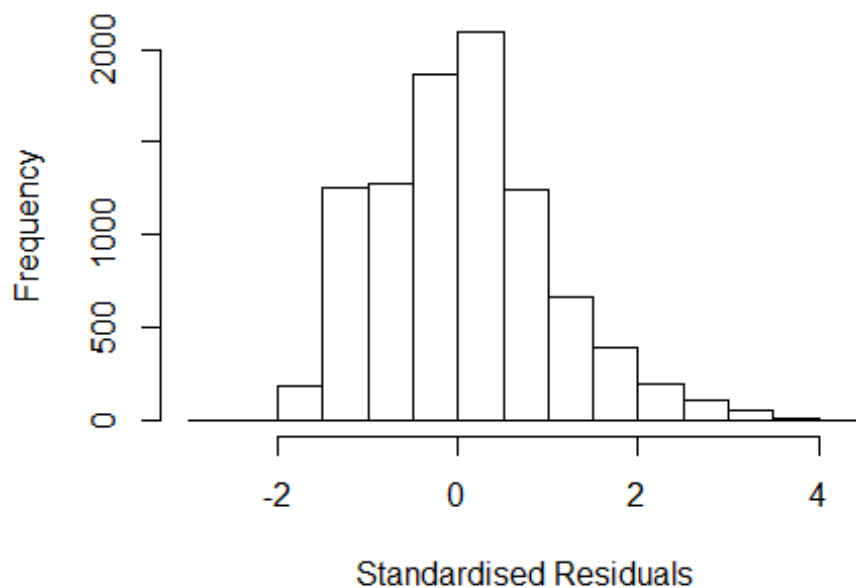
```

## 17359 84863895902 3.806 2012      2700      1      2.580
## 17892 84857815393 4.076 2012      2700      1      2.926
## 18065 84856529575 4.439 2012      2700      1      3.145
## 18182 84860538806 4.001 2012      2700      1      2.783
## 18235 84866517922 3.859 2012      2700      1      2.641
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6127 -0.6238  0.0174  0.6045  4.1835
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.78125    0.09388   18.97 < 2e-16 ***
## FirstAuthorFemale1  0.07614    0.02175    3.50 0.00047 ***
## LastAuthorFemale1 -0.06791    0.02273   -2.99 0.00282 **
## Year1997          0.03820    0.12671    0.30 0.76309
## Year1998          0.00701    0.11880    0.06 0.95294
## Year1999          0.10015    0.11994    0.84 0.40373
## Year2000          0.83142    0.33490    2.48 0.01306 *
## Year2001         -0.03622    0.13114   -0.28 0.78243
## Year2002         -0.35672    0.10995   -3.24 0.00118 **
## Year2003         -0.38031    0.11122   -3.42 0.00063 ***
## Year2004         -0.38091    0.10902   -3.49 0.00048 ***
## Year2005         -0.31112    0.10420   -2.99 0.00284 **
## Year2006         -0.39890    0.10129   -3.94 8.3e-05 ***
## Year2007         -0.38796    0.10014   -3.87 0.00011 ***
## Year2008         -0.37288    0.09877   -3.78 0.00016 ***
## Year2009         -0.47064    0.09858   -4.77 1.8e-06 ***
## Year2010         -0.53205    0.09627   -5.53 3.3e-08 ***
## Year2011         -0.54092    0.09814   -5.51 3.7e-08 ***
## Year2012         -0.56303    0.09653   -5.83 5.6e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.939
## Multiple R-squared:  0.039, Adjusted R-squared:  0.0372
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 819 weights are ~ 1. The remaining 8527 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.0092  0.8450  0.9500  0.8980  0.9860  0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00        5.00e-01        4.69e+00        1.00e-07

```

```
##          rel.tol          solve.tol          eps.outlier          eps.x
##          1.00e-07          1.00e-07          1.07e-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 1.016 1          1.008
## Year              1.016 16          1.000
```

Residuals from first author



```
## [1] "List of 180 outliers with residuals above 2.5"
##           ScopusId NLCS Year OneField Fields residuals
## 10      0029806876 4.368 1996      2700      1      2.587
## 183     0029907525 4.716 1996      2700      1      2.935
## 214     0030603333 5.098 1996      2700      1      3.317
## 274     0029799559 5.048 1996      2700      1      3.267
## 325     9344249534 4.358 1996      2700      1      2.577
## 356     0030056515 4.523 1996      2700      1      2.666
## 373     0029938206 4.446 1996      2700      1      2.665
```

## 426	0029880393	4.724	1996	2700	1	3.011
## 427	0029990981	4.429	1996	2700	1	2.648
## 450	0029862973	5.794	1996	2700	1	4.013
## 486	0006590545	5.593	1996	2700	1	3.812
## 496	0030093546	4.950	1996	2700	1	3.161
## 553	0030047924	4.654	1996	2700	1	2.873
## 560	0001259453	5.013	1996	2700	1	3.232
## 715	0031588519	5.136	1997	2700	1	3.317
## 856	0031583760	4.794	1997	2700	1	2.975
## 904	0030808429	5.491	1997	2700	1	3.672
## 962	18544412975	4.558	1997	2700	1	2.662
## 963	0030800795	5.042	1997	2700	1	3.146
## 976	1842338658	4.759	1997	2700	1	2.940
## 1000	0030811896	5.020	1997	2700	1	3.192
## 1002	0031578297	4.624	1997	2700	1	2.805
## 1025	0030908220	4.463	1997	2700	1	2.567
## 1030	0030943882	4.662	1997	2700	1	2.910
## 1138	0031029373	5.109	1997	2700	1	3.290
## 1151	0030981335	5.288	1997	2700	1	3.392
## 1177	0031048342	4.524	1997	2700	1	2.705
## 1183	0031024865	5.035	1997	2700	1	3.216
## 1184	0031017667	5.126	1997	2700	1	3.307
## 1212	0031022182	4.739	1997	2700	1	2.843
## 1269	0031035439	4.806	1997	2700	1	2.910
## 1420	0032539180	4.542	1998	2700	1	2.754
## 1687	0032511413	4.449	1998	2700	1	2.585
## 1736	0032562946	4.354	1998	2700	1	2.566
## 1781	0032555023	4.631	1998	2700	1	2.911
## 1783	2642599026	5.421	1998	2700	1	3.633
## 1866	6844236394	4.749	1998	2700	1	2.953
## 1888	0032033812	4.367	1998	2700	1	2.571
## 1913	0032507027	4.854	1998	2700	1	3.066
## 1916	0032484918	4.822	1998	2700	1	3.034
## 1955	0032495544	4.571	1998	2700	1	2.783
## 2281	0032717944	4.535	1999	2700	1	2.654
## 2426	0033595136	5.107	1999	2700	1	3.226
## 2606	0033580398	4.314	1999	2700	1	2.501
## 2644	0343360868	5.997	1999	2700	1	4.184
## 2654	0033528281	4.383	1999	2700	1	2.502
## 2777	0033193134	4.710	1999	2700	1	2.752
## 2876	0034430489	0.000	2000	2700	1	-2.613
## 2880	53949090522	0.000	2000	2700	2	-2.613
## 2937	85008539534	0.000	2000	2700	2	-2.613
## 2940	85007753361	0.000	2000	2700	1	-2.613
## 2942	0034688194	6.362	2000	2700	1	3.749
## 3106	0035965633	4.678	2001	2700	1	2.933
## 3169	0035904362	4.784	2001	2700	1	3.039
## 3254	0035975409	4.758	2001	2700	1	3.013
## 3284	0035833493	4.848	2001	2700	1	3.095
## 3383	0035810589	4.879	2001	2700	1	3.134

##	3417	0035906286	4.818	2001	2700	1	3.141
##	3512	0035941514	4.680	2001	2700	1	2.927
##	3519	0035843250	4.421	2001	2700	1	2.676
##	3763	0036440553	4.664	2002	2700	1	3.239
##	3907	0037151906	3.930	2002	2700	1	2.505
##	3999	0037178575	3.986	2002	2700	1	2.561
##	4001	0037167013	4.710	2002	2700	1	3.209
##	4179	0037076029	4.344	2002	2700	1	2.843
##	4196	0037035122	4.319	2002	2700	1	2.894
##	4243	0037176525	3.951	2002	2700	1	2.526
##	4246	0037160563	4.315	2002	2700	1	2.890
##	4351	0036834665	4.324	2002	2700	1	2.891
##	4454	0346555267	4.287	2003	2700	1	2.886
##	4513	0344420330	4.990	2003	2700	1	3.589
##	4514	0642364501	4.434	2003	2700	1	3.033
##	4595	0141763668	4.176	2003	2700	1	2.699
##	4726	0037832413	3.991	2003	2700	1	2.590
##	4741	0037976821	4.681	2003	2700	1	3.280
##	4786	0013165966	4.910	2003	2700	1	3.501
##	4826	0242578548	4.063	2003	2700	1	2.662
##	4834	0037414162	4.528	2003	2700	1	3.127
##	4869	0037443934	4.690	2003	2700	1	3.213
##	4875	0037420266	4.266	2003	2700	1	2.789
##	4912	0037456354	5.153	2003	2700	1	3.676
##	4960	0037413466	5.051	2003	2700	1	3.650
##	4979	0037380645	4.167	2003	2700	1	2.690
##	5076	1542494137	4.153	2003	2700	1	2.744
##	5217	9244237570	4.055	2004	2700	1	2.655
##	5318	4544230001	4.656	2004	2700	1	3.180
##	5377	4143096183	4.405	2004	2700	1	3.005
##	5391	3442886513	4.914	2004	2700	1	3.514
##	5420	3142705643	4.188	2004	2700	1	2.712
##	5447	3042738109	4.386	2004	2700	1	2.977
##	5498	2542570187	4.122	2004	2700	1	2.790
##	5560	1842430968	3.938	2004	2700	1	2.606
##	5637	1442309702	4.035	2004	2700	1	2.635
##	5678	1642455902	3.970	2004	2700	1	2.570
##	5959	27844535729	4.362	2005	2700	1	2.892
##	5960	28044462641	4.500	2005	2700	1	3.022
##	6017	27544511378	4.326	2005	2700	1	2.856
##	6021	26844508456	4.078	2005	2700	1	2.608
##	6136	24044485648	5.177	2005	2700	1	3.707
##	6301	19744368662	4.104	2005	2700	1	2.634
##	6312	21844449936	4.142	2005	2700	1	2.664
##	6315	21844458935	4.430	2005	2700	1	2.952
##	6372	15744389846	4.744	2005	2700	1	3.342
##	6382	20044383911	4.809	2005	2700	1	3.331
##	6747	33847712261	4.255	2006	2700	1	2.941
##	6935	33750958903	3.908	2006	2700	1	2.526
##	7206	33747108061	3.894	2006	2700	1	2.512

##	7239	34247562149	4.092	2006	2700	1	2.701
##	7293	33745881036	4.378	2006	2700	1	2.996
##	7305	33745400514	4.384	2006	2700	1	2.926
##	7431	33646681444	4.429	2006	2700	1	2.971
##	7474	33646130849	3.945	2006	2700	1	2.563
##	7495	33645497961	4.742	2006	2700	1	3.360
##	7540	33645399023	4.082	2006	2700	1	2.624
##	7637	31844434994	4.007	2006	2700	1	2.625
##	7680	31344440606	4.440	2006	2700	1	3.058
##	7837	37149034178	4.371	2007	2700	1	2.902
##	8048	36749016190	3.920	2007	2700	1	2.595
##	8061	35948976968	4.234	2007	2700	1	2.765
##	8148	35348967810	5.076	2007	2700	1	3.607
##	8163	34948904387	4.484	2007	2700	1	3.091
##	8252	34548772315	4.777	2007	2700	1	3.384
##	8263	34548603259	4.369	2007	2700	1	2.976
##	8278	34548445010	4.263	2007	2700	1	2.794
##	8464	34547666899	4.192	2007	2700	1	2.799
##	8473	34547167458	4.103	2007	2700	1	2.634
##	8750	34247468559	4.385	2007	2700	1	3.060
##	8873	33847303686	4.719	2007	2700	1	3.317
##	8941	33847172327	4.158	2007	1000	2	2.765
##	8950	33847759152	4.625	2007	2700	1	3.232
##	8957	33846806078	4.110	2007	2700	1	2.717
##	9003	33846407655	5.477	2007	2700	1	4.084
##	9196	58149145568	4.036	2008	2700	1	2.628
##	9491	56649088178	4.078	2008	2700	1	2.670
##	9669	51649088482	4.382	2008	2700	1	2.897
##	9788	49749141653	4.171	2008	2700	1	2.686
##	9859	48349118873	4.078	2008	2700	1	2.661
##	9898	47649130481	3.937	2008	2700	1	2.597
##	10525	38349073481	4.179	2008	2700	1	2.771
##	10556	38949213691	4.140	2008	2700	1	2.732
##	11091	70349766844	4.197	2009	2700	1	2.954
##	11217	70349155477	4.446	2009	2700	1	3.135
##	11340	69249208815	4.005	2009	2700	1	2.762
##	11513	67649407404	4.381	2009	2700	1	3.070
##	11654	65649088288	4.306	2009	2700	1	3.063
##	11661	67650035326	4.048	2009	2700	1	2.737
##	11715	69049117564	4.182	2009	2700	1	2.939
##	11871	62649164844	4.746	2009	2700	1	3.435
##	11873	65349171524	4.140	2009	2700	1	2.897
##	11915	61849155730	5.054	2009	2700	1	3.743
##	11922	65649085603	4.033	2009	2700	1	2.722
##	12028	61449244692	3.743	2009	2700	1	2.500
##	12228	70449653474	5.051	2009	2700	1	3.740
##	12543	78649906763	4.032	2010	2700	2	2.783
##	13239	77956056661	4.268	2010	2700	1	3.019
##	13895	77950273246	5.118	2010	2700	1	3.869
##	13904	77951240308	3.906	2010	2700	1	2.657

```

## 14190 76649099557 4.472 2010      2700      1      3.291
## 14205 75849147039 4.355 2010      2700      1      3.030
## 14889 81255143853 3.841 2011      2700      1      2.592
## 14923 80455173556 4.657 2011      2700      1      3.341
## 14959 82455205788 3.890 2011      2700      1      2.641
## 15326 80052073777 3.748 2011      2700      1      2.508
## 15421 84859006414 3.927 2011      2700      1      2.687
## 15469 79959748876 3.820 2011      2700      1      2.571
## 15618 84858996807 4.302 2011      2700      1      3.053
## 15747 79956110208 3.923 2011      2700      1      2.683
## 15851 79957546192 3.765 2011      2700      1      2.516
## 15855 79955474494 3.788 2011      2700      1      2.548
## 15877 79955132796 4.367 2011      2700      1      3.127
## 15895 79953775836 5.237 2011      2700      1      3.988
## 16023 79953726311 3.918 2011      2700      1      2.602
## 16095 79951995822 4.559 2011      2700      1      3.243
## 16108 79951649740 3.820 2011      2700      1      2.580
## 16915 84867161086 3.880 2012      2700      1      2.662
## 17359 84863895902 3.806 2012      2700      1      2.580
## 17892 84857815393 4.076 2012      2700      1      2.926
## 18065 84856529575 4.439 2012      2700      1      3.145
## 18182 84860538806 4.001 2012      2700      1      2.783
## 18235 84866517922 3.859 2012      2700      1      2.641
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6035 -0.6296  0.0184  0.6106  4.1237
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.77099    0.09344   18.95 < 2e-16 ***
## FirstAuthorFemale1 0.05093    0.02062    2.47  0.01355 *
## Year1997        0.03610    0.12661    0.29  0.77555
## Year1998        0.00265    0.11841    0.02  0.98217
## Year1999        0.10233    0.11963    0.86  0.39237
## Year2000        0.83251    0.33509    2.48  0.01299 *
## Year2001       -0.03671    0.13085   -0.28  0.77906
## Year2002       -0.35894    0.10976   -3.27  0.00108 **
## Year2003       -0.38507    0.11095   -3.47  0.00052 ***
## Year2004       -0.38474    0.10885   -3.53  0.00041 ***
## Year2005       -0.31549    0.10397   -3.03  0.00242 **
## Year2006       -0.40332    0.10106   -3.99  6.6e-05 ***
## Year2007       -0.39182    0.09984   -3.92  8.8e-05 ***
## Year2008       -0.37884    0.09850   -3.85  0.00012 ***
## Year2009       -0.47184    0.09832   -4.80  1.6e-06 ***

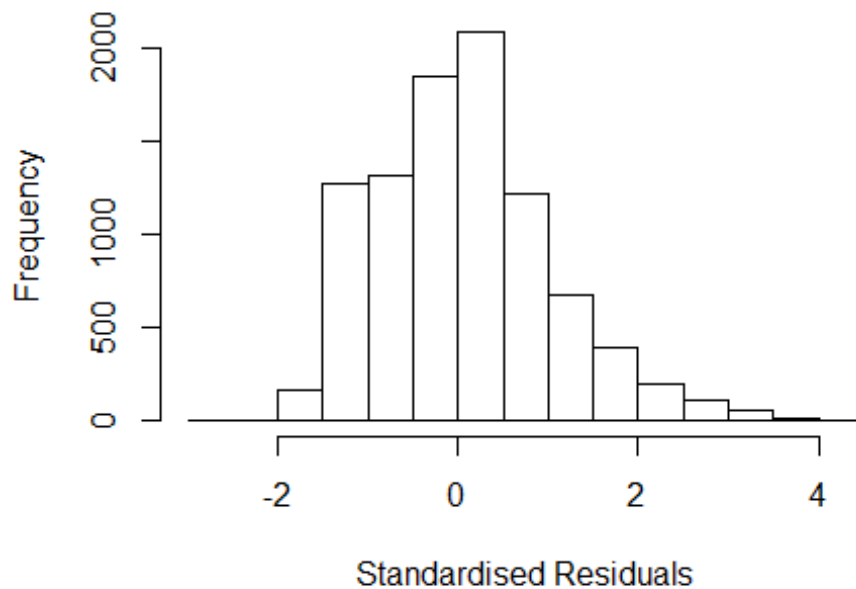
```

```

## Year2010          -0.53257    0.09597   -5.55  2.9e-08 ***
## Year2011          -0.54292    0.09787   -5.55  3.0e-08 ***
## Year2012          -0.56426    0.09626   -5.86  4.7e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.939
## Multiple R-squared:  0.0381, Adjusted R-squared:  0.0364
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 843 weights are ~= 1. The remaining 8503 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0147 0.8430 0.9500 0.8980 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.07e-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 1.006 1          1.003
## Year            1.006 16          1.000

```

Residuals from last author



```
## [1] "List of 180 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 10      0029806876 4.368 1996      2700      1      2.587
## 183     0029907525 4.716 1996      2700      1      2.935
## 214     0030603333 5.098 1996      2700      1      3.317
## 274     0029799559 5.048 1996      2700      1      3.267
## 325     9344249534 4.358 1996      2700      1      2.577
## 356     0030056515 4.523 1996      2700      1      2.666
## 373     0029938206 4.446 1996      2700      1      2.665
## 426     0029880393 4.724 1996      2700      1      3.011
## 427     0029990981 4.429 1996      2700      1      2.648
## 450     0029862973 5.794 1996      2700      1      4.013
## 486     0006590545 5.593 1996      2700      1      3.812
## 496     0030093546 4.950 1996      2700      1      3.161
## 553     0030047924 4.654 1996      2700      1      2.873
## 560     0001259453 5.013 1996      2700      1      3.232
## 715     0031588519 5.136 1997      2700      1      3.317
## 856     0031583760 4.794 1997      2700      1      2.975
## 904     0030808429 5.491 1997      2700      1      3.672
## 962     18544412975 4.558 1997      2700      1      2.662
## 963     0030800795 5.042 1997      2700      1      3.146
## 976     1842338658 4.759 1997      2700      1      2.940
## 1000     0030811896 5.020 1997      2700      1      3.192
## 1002     0031578297 4.624 1997      2700      1      2.805
## 1025     0030908220 4.463 1997      2700      1      2.567
## 1030     0030943882 4.662 1997      2700      1      2.910
## 1138     0031029373 5.109 1997      2700      1      3.290
```

## 1151	0030981335	5.288	1997	2700	1	3.392
## 1177	0031048342	4.524	1997	2700	1	2.705
## 1183	0031024865	5.035	1997	2700	1	3.216
## 1184	0031017667	5.126	1997	2700	1	3.307
## 1212	0031022182	4.739	1997	2700	1	2.843
## 1269	0031035439	4.806	1997	2700	1	2.910
## 1420	0032539180	4.542	1998	2700	1	2.754
## 1687	0032511413	4.449	1998	2700	1	2.585
## 1736	0032562946	4.354	1998	2700	1	2.566
## 1781	0032555023	4.631	1998	2700	1	2.911
## 1783	2642599026	5.421	1998	2700	1	3.633
## 1866	6844236394	4.749	1998	2700	1	2.953
## 1888	0032033812	4.367	1998	2700	1	2.571
## 1913	0032507027	4.854	1998	2700	1	3.066
## 1916	0032484918	4.822	1998	2700	1	3.034
## 1955	0032495544	4.571	1998	2700	1	2.783
## 2281	0032717944	4.535	1999	2700	1	2.654
## 2426	0033595136	5.107	1999	2700	1	3.226
## 2606	0033580398	4.314	1999	2700	1	2.501
## 2644	0343360868	5.997	1999	2700	1	4.184
## 2654	0033528281	4.383	1999	2700	1	2.502
## 2777	0033193134	4.710	1999	2700	1	2.752
## 2876	0034430489	0.000	2000	2700	1	-2.613
## 2880	53949090522	0.000	2000	2700	2	-2.613
## 2937	85008539534	0.000	2000	2700	2	-2.613
## 2940	85007753361	0.000	2000	2700	1	-2.613
## 2942	0034688194	6.362	2000	2700	1	3.749
## 3106	0035965633	4.678	2001	2700	1	2.933
## 3169	0035904362	4.784	2001	2700	1	3.039
## 3254	0035975409	4.758	2001	2700	1	3.013
## 3284	0035833493	4.848	2001	2700	1	3.095
## 3383	0035810589	4.879	2001	2700	1	3.134
## 3417	0035906286	4.818	2001	2700	1	3.141
## 3512	0035941514	4.680	2001	2700	1	2.927
## 3519	0035843250	4.421	2001	2700	1	2.676
## 3763	0036440553	4.664	2002	2700	1	3.239
## 3907	0037151906	3.930	2002	2700	1	2.505
## 3999	0037178575	3.986	2002	2700	1	2.561
## 4001	0037167013	4.710	2002	2700	1	3.209
## 4179	0037076029	4.344	2002	2700	1	2.843
## 4196	0037035122	4.319	2002	2700	1	2.894
## 4243	0037176525	3.951	2002	2700	1	2.526
## 4246	0037160563	4.315	2002	2700	1	2.890
## 4351	0036834665	4.324	2002	2700	1	2.891
## 4454	0346555267	4.287	2003	2700	1	2.886
## 4513	0344420330	4.990	2003	2700	1	3.589
## 4514	0642364501	4.434	2003	2700	1	3.033
## 4595	0141763668	4.176	2003	2700	1	2.699
## 4726	0037832413	3.991	2003	2700	1	2.590
## 4741	0037976821	4.681	2003	2700	1	3.280

##	4786	0013165966	4.910	2003	2700	1	3.501
##	4826	0242578548	4.063	2003	2700	1	2.662
##	4834	0037414162	4.528	2003	2700	1	3.127
##	4869	0037443934	4.690	2003	2700	1	3.213
##	4875	0037420266	4.266	2003	2700	1	2.789
##	4912	0037456354	5.153	2003	2700	1	3.676
##	4960	0037413466	5.051	2003	2700	1	3.650
##	4979	0037380645	4.167	2003	2700	1	2.690
##	5076	1542494137	4.153	2003	2700	1	2.744
##	5217	9244237570	4.055	2004	2700	1	2.655
##	5318	4544230001	4.656	2004	2700	1	3.180
##	5377	4143096183	4.405	2004	2700	1	3.005
##	5391	3442886513	4.914	2004	2700	1	3.514
##	5420	3142705643	4.188	2004	2700	1	2.712
##	5447	3042738109	4.386	2004	2700	1	2.977
##	5498	2542570187	4.122	2004	2700	1	2.790
##	5560	1842430968	3.938	2004	2700	1	2.606
##	5637	1442309702	4.035	2004	2700	1	2.635
##	5678	1642455902	3.970	2004	2700	1	2.570
##	5959	27844535729	4.362	2005	2700	1	2.892
##	5960	28044462641	4.500	2005	2700	1	3.022
##	6017	27544511378	4.326	2005	2700	1	2.856
##	6021	26844508456	4.078	2005	2700	1	2.608
##	6136	240444485648	5.177	2005	2700	1	3.707
##	6301	19744368662	4.104	2005	2700	1	2.634
##	6312	218444449936	4.142	2005	2700	1	2.664
##	6315	21844458935	4.430	2005	2700	1	2.952
##	6372	15744389846	4.744	2005	2700	1	3.342
##	6382	20044383911	4.809	2005	2700	1	3.331
##	6747	33847712261	4.255	2006	2700	1	2.941
##	6935	33750958903	3.908	2006	2700	1	2.526
##	7206	33747108061	3.894	2006	2700	1	2.512
##	7239	34247562149	4.092	2006	2700	1	2.701
##	7293	33745881036	4.378	2006	2700	1	2.996
##	7305	33745400514	4.384	2006	2700	1	2.926
##	7431	33646681444	4.429	2006	2700	1	2.971
##	7474	33646130849	3.945	2006	2700	1	2.563
##	7495	33645497961	4.742	2006	2700	1	3.360
##	7540	33645399023	4.082	2006	2700	1	2.624
##	7637	31844434994	4.007	2006	2700	1	2.625
##	7680	313444440606	4.440	2006	2700	1	3.058
##	7837	37149034178	4.371	2007	2700	1	2.902
##	8048	36749016190	3.920	2007	2700	1	2.595
##	8061	35948976968	4.234	2007	2700	1	2.765
##	8148	35348967810	5.076	2007	2700	1	3.607
##	8163	34948904387	4.484	2007	2700	1	3.091
##	8252	34548772315	4.777	2007	2700	1	3.384
##	8263	34548603259	4.369	2007	2700	1	2.976
##	8278	34548445010	4.263	2007	2700	1	2.794
##	8464	34547666899	4.192	2007	2700	1	2.799

##	8473	34547167458	4.103	2007	2700	1	2.634
##	8750	34247468559	4.385	2007	2700	1	3.060
##	8873	33847303686	4.719	2007	2700	1	3.317
##	8941	33847172327	4.158	2007	1000	2	2.765
##	8950	33847759152	4.625	2007	2700	1	3.232
##	8957	33846806078	4.110	2007	2700	1	2.717
##	9003	33846407655	5.477	2007	2700	1	4.084
##	9196	58149145568	4.036	2008	2700	1	2.628
##	9491	56649088178	4.078	2008	2700	1	2.670
##	9669	51649088482	4.382	2008	2700	1	2.897
##	9788	49749141653	4.171	2008	2700	1	2.686
##	9859	48349118873	4.078	2008	2700	1	2.661
##	9898	47649130481	3.937	2008	2700	1	2.597
##	10525	38349073481	4.179	2008	2700	1	2.771
##	10556	38949213691	4.140	2008	2700	1	2.732
##	11091	70349766844	4.197	2009	2700	1	2.954
##	11217	70349155477	4.446	2009	2700	1	3.135
##	11340	69249208815	4.005	2009	2700	1	2.762
##	11513	67649407404	4.381	2009	2700	1	3.070
##	11654	65649088288	4.306	2009	2700	1	3.063
##	11661	67650035326	4.048	2009	2700	1	2.737
##	11715	69049117564	4.182	2009	2700	1	2.939
##	11871	62649164844	4.746	2009	2700	1	3.435
##	11873	65349171524	4.140	2009	2700	1	2.897
##	11915	61849155730	5.054	2009	2700	1	3.743
##	11922	65649085603	4.033	2009	2700	1	2.722
##	12028	61449244692	3.743	2009	2700	1	2.500
##	12228	70449653474	5.051	2009	2700	1	3.740
##	12543	78649906763	4.032	2010	2700	2	2.783
##	13239	77956056661	4.268	2010	2700	1	3.019
##	13895	77950273246	5.118	2010	2700	1	3.869
##	13904	77951240308	3.906	2010	2700	1	2.657
##	14190	76649099557	4.472	2010	2700	1	3.291
##	14205	75849147039	4.355	2010	2700	1	3.030
##	14889	81255143853	3.841	2011	2700	1	2.592
##	14923	80455173556	4.657	2011	2700	1	3.341
##	14959	82455205788	3.890	2011	2700	1	2.641
##	15326	80052073777	3.748	2011	2700	1	2.508
##	15421	84859006414	3.927	2011	2700	1	2.687
##	15469	79959748876	3.820	2011	2700	1	2.571
##	15618	84858996807	4.302	2011	2700	1	3.053
##	15747	79956110208	3.923	2011	2700	1	2.683
##	15851	79957546192	3.765	2011	2700	1	2.516
##	15855	79955474494	3.788	2011	2700	1	2.548
##	15877	79955132796	4.367	2011	2700	1	3.127
##	15895	79953775836	5.237	2011	2700	1	3.988
##	16023	79953726311	3.918	2011	2700	1	2.602
##	16095	79951995822	4.559	2011	2700	1	3.243
##	16108	79951649740	3.820	2011	2700	1	2.580
##	16915	84867161086	3.880	2012	2700	1	2.662

```

## 17359 84863895902 3.806 2012      2700      1      2.580
## 17892 84857815393 4.076 2012      2700      1      2.926
## 18065 84856529575 4.439 2012      2700      1      3.145
## 18182 84860538806 4.001 2012      2700      1      2.783
## 18235 84866517922 3.859 2012      2700      1      2.641
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.641 -0.625  0.021  0.609  4.134
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.79848    0.09379   19.18 < 2e-16 ***
## LastAuthorFemale1 -0.03755    0.02144   -1.75  0.07996 .
## Year1997         0.03929    0.12712    0.31  0.75726
## Year1998         0.00721    0.11902    0.06  0.95172
## Year1999         0.10246    0.12035    0.85  0.39461
## Year2000         0.84250    0.33601    2.51  0.01218 *
## Year2001        -0.02809    0.13166   -0.21  0.83104
## Year2002        -0.35295    0.10997   -3.21  0.00133 **
## Year2003        -0.37699    0.11129   -3.39  0.00071 ***
## Year2004        -0.37792    0.10903   -3.47  0.00053 ***
## Year2005        -0.30326    0.10427   -2.91  0.00364 **
## Year2006        -0.39011    0.10133   -3.85  0.00012 ***
## Year2007        -0.38152    0.10018   -3.81  0.00014 ***
## Year2008        -0.36393    0.09880   -3.68  0.00023 ***
## Year2009        -0.46177    0.09861   -4.68  2.9e-06 ***
## Year2010        -0.52204    0.09628   -5.42  6.0e-08 ***
## Year2011        -0.53498    0.09821   -5.45  5.3e-08 ***
## Year2012        -0.55220    0.09652   -5.72  1.1e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.94
## Multiple R-squared:  0.0379, Adjusted R-squared:  0.0361
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 812 weights are ~1. The remaining 8534 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.0143  0.8440  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

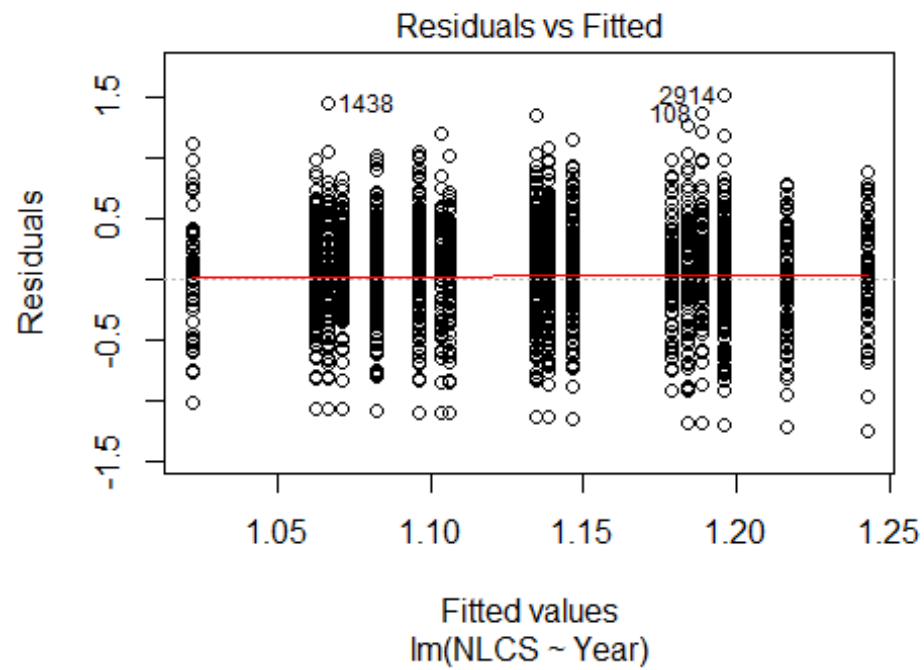
```



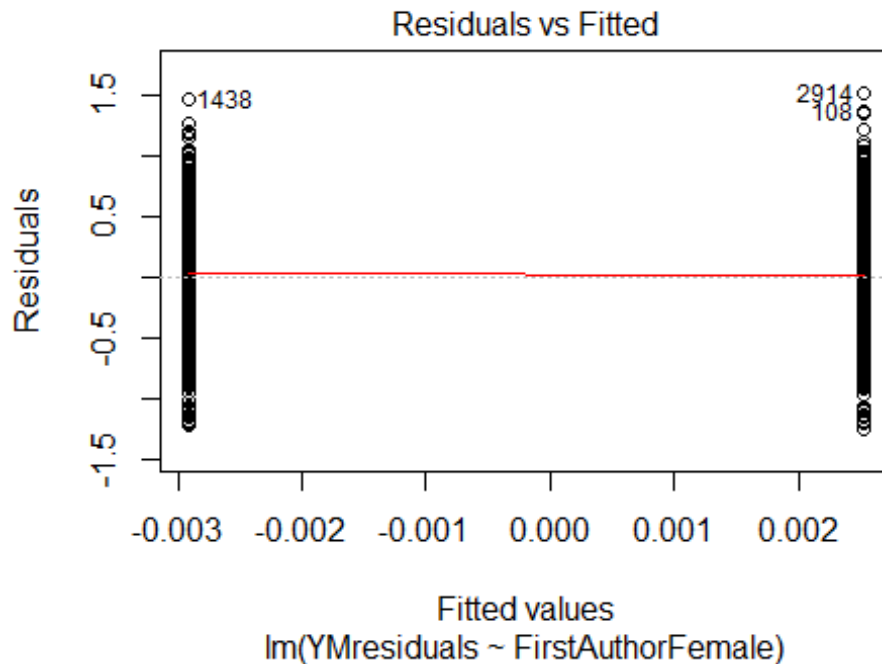
```

##          1.00e-07          1.00e-07          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: 9346"
## [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
## 144 136 114 133 139 153 114 146 140 164 195 240 250 288 288
## 2011 2012
## 303 323
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 92 91 71 93 82 57 81 109 115 119 142 178 192 209 213
## 2011 2012
## 206 236
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 78 79 62 80 72 50 72 95 101 105 128 155 174 185 191
## 2011 2012
## 184 204
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 27, df = 16, p-value = 0.04

```

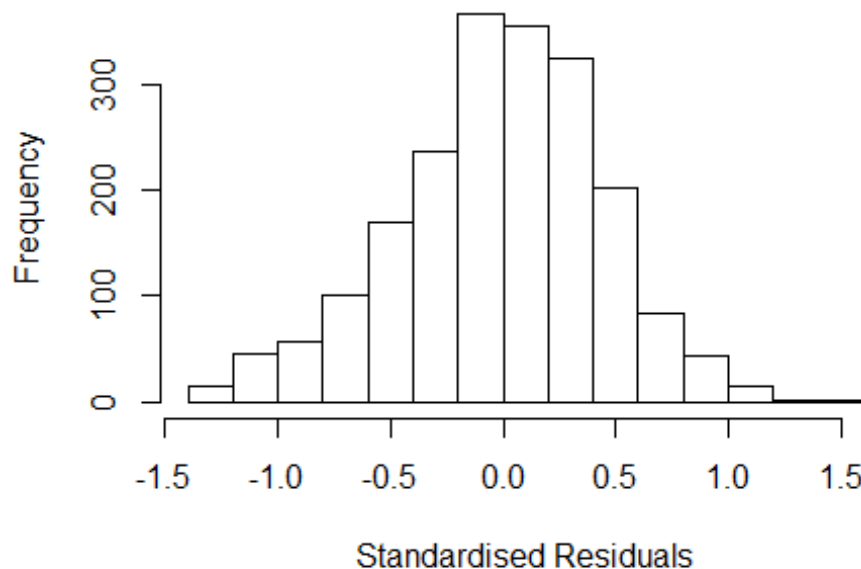


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.8, df = 1, p-value = 0.003
```



```
## [1] "Female first author team size 2018 geometric mean: 4.71081745053592"
## [1] "Male first author team size 2018 geometric mean: 4.64404601835313"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8100, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.70333672280837"
## [1] "Male last author team size 2018 geometric mean: 4.66859272235484"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8100, 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  1.113 1      1.055
## UniqueAuthors     1.291 4      1.032
## Year              1.354 16     1.010
```

Residuals from first and last author and team size



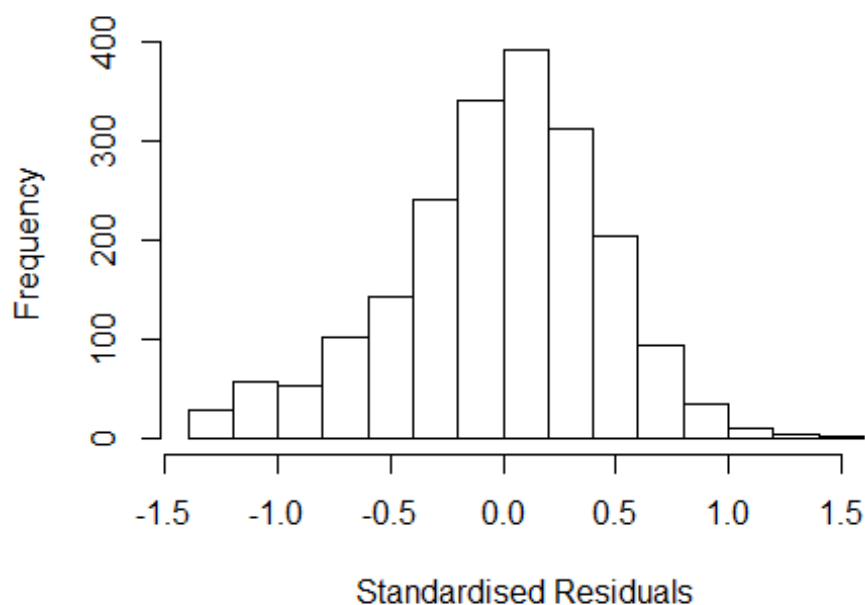
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33397 -0.28756  0.00749  0.29720  1.56244
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1097    0.0708   15.67 < 2e-16 ***
## FirstAuthorFemale1 -0.0116    0.0220   -0.53  0.59769
## LastAuthorFemale1 -0.0478    0.0216   -2.21  0.02723 *
## UniqueAuthors2     0.1666    0.0491    3.39  0.00071 ***
## UniqueAuthors3     0.1584    0.0467    3.39  0.00070 ***
## UniqueAuthors4     0.1993    0.0481    4.15  3.5e-05 ***
## UniqueAuthors5     0.2812    0.0449    6.26  4.8e-10 ***
## Year1997          -0.0178    0.0893   -0.20  0.84187
## Year1998           0.0692    0.0841    0.82  0.41092
## Year1999          -0.0810    0.0775   -1.05  0.29561
```

```

## Year2000          -0.1598      0.0857    -1.86   0.06234 .
## Year2001          -0.2007      0.1054    -1.90   0.05705 .
## Year2002          -0.0458      0.0791    -0.58   0.56254
## Year2003          -0.1605      0.0727    -2.21   0.02733 *
## Year2004          -0.2327      0.0727    -3.20   0.00140 **
## Year2005          -0.1525      0.0712    -2.14   0.03231 *
## Year2006          -0.2215      0.0709    -3.12   0.00180 **
## Year2007          -0.2031      0.0690    -2.94   0.00330 **
## Year2008          -0.1067      0.0689    -1.55   0.12159
## Year2009          -0.0634      0.0666    -0.95   0.34148
## Year2010          -0.1121      0.0699    -1.60   0.10910
## Year2011          -0.2047      0.0674    -3.04   0.00242 **
## Year2012          -0.1154      0.0693    -1.67   0.09597 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.0573, Adjusted R-squared:  0.0469
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 171 weights are ~= 1. The remaining 1844 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.163  0.862  0.950  0.897  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.96e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.113 1          1.055
## LastAuthorFemale  1.100 1          1.049
## Year              1.081 16          1.002

```

Residuals from first and last author



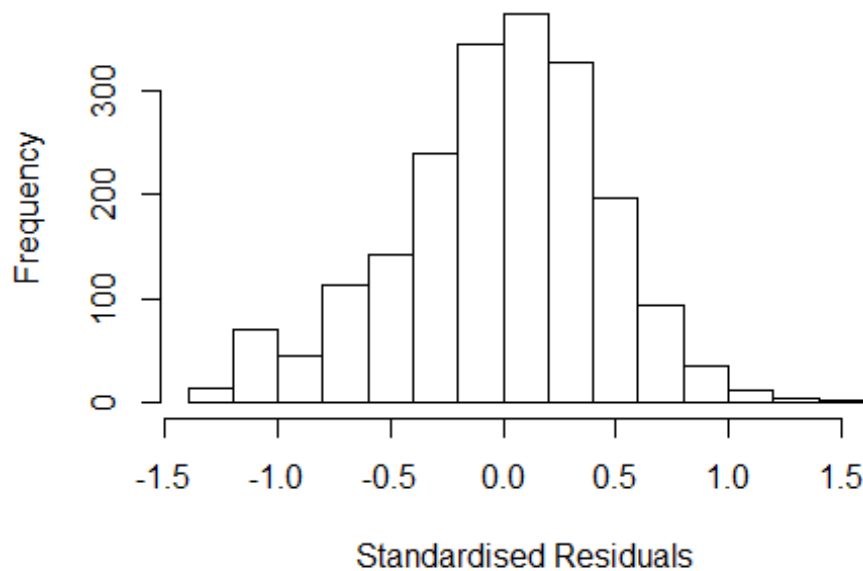
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3257 -0.2941  0.0188  0.2987  1.5355
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.263176   0.059954   21.07  <2e-16 ***
## FirstAuthorFemale1 -0.000999   0.022197   -0.05   0.9641
## LastAuthorFemale1 -0.058104   0.021799   -2.67   0.0078 **
## Year1997        -0.031202   0.089026   -0.35   0.7260
## Year1998         0.063525   0.085439    0.74   0.4573
## Year1999        -0.074872   0.075590   -0.99   0.3220
## Year2000        -0.130376   0.085420   -1.53   0.1271
## Year2001        -0.196692   0.112338   -1.75   0.0801 .
## Year2002        -0.016783   0.077955   -0.22   0.8296
## Year2003        -0.135017   0.072560   -1.86   0.0629 .
## Year2004        -0.200530   0.073080   -2.74   0.0061 **
## Year2005        -0.104819   0.070453   -1.49   0.1370
```

```

## Year2006          -0.183389    0.071135    -2.58    0.0100 *
## Year2007          -0.172882    0.068793    -2.51    0.0120 *
## Year2008          -0.068547    0.068540    -1.00    0.3174
## Year2009          -0.023557    0.066782    -0.35    0.7243
## Year2010          -0.057105    0.069262    -0.82    0.4098
## Year2011          -0.147016    0.066466    -2.21    0.0271 *
## Year2012          -0.059515    0.068122    -0.87    0.3824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0253, Adjusted R-squared:  0.0165
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 184 weights are ~= 1. The remaining 1831 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.184  0.862  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      4.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.048 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
## -1.2920 -0.2974 0.0201 0.2973 1.5089
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.24420 0.05912 21.05 <2e-16 ***
## FirstAuthorFemale1 -0.01802 0.02160 -0.83 0.4042
## Year1997 -0.02765 0.08934 -0.31 0.7570
## Year1998 0.06582 0.08498 0.77 0.4387
## Year1999 -0.06641 0.07517 -0.88 0.3771
## Year2000 -0.12510 0.08494 -1.47 0.1410
## Year2001 -0.19695 0.11096 -1.78 0.0760 .
## Year2002 -0.00917 0.07766 -0.12 0.9060
## Year2003 -0.12931 0.07235 -1.79 0.0741 .
## Year2004 -0.19767 0.07280 -2.72 0.0067 **
## Year2005 -0.10729 0.07016 -1.53 0.1264
## Year2006 -0.17422 0.07076 -2.46 0.0139 *
```

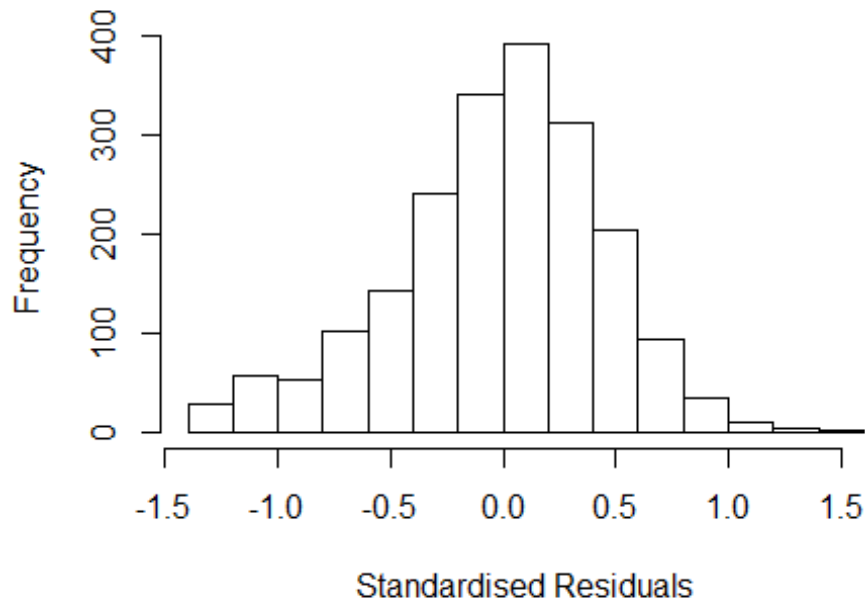


```

## Year2007          -0.16937    0.06858   -2.47   0.0136 *
## Year2008          -0.06492    0.06844   -0.95   0.3429
## Year2009          -0.01911    0.06644   -0.29   0.7737
## Year2010          -0.05480    0.06913   -0.79   0.4280
## Year2011          -0.14683    0.06620   -2.22   0.0267 *
## Year2012          -0.05645    0.06780   -0.83   0.4052
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.435
## Multiple R-squared:  0.0216, Adjusted R-squared:  0.0133
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 180 weights are ~= 1. The remaining 1835 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.204  0.864  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      4.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.037 1          1.018
## Year            1.037 16          1.001

```

Residuals from last author



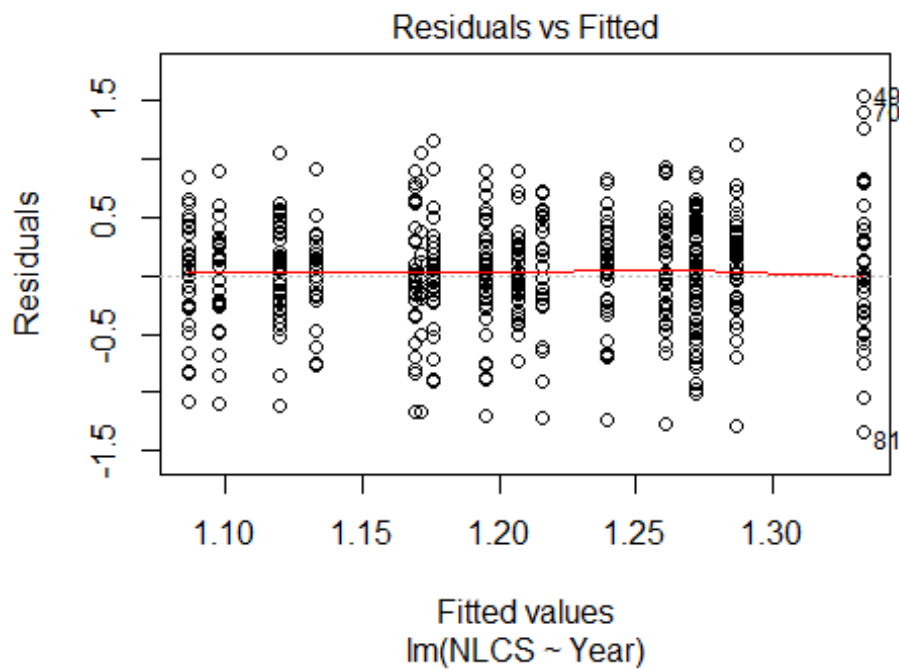
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3262 -0.2943  0.0188  0.2987  1.5352
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2628    0.0594   21.25  <2e-16 ***
## LastAuthorFemale1 -0.0584    0.0212   -2.75  0.0060 **
## Year1997         -0.0313    0.0890   -0.35  0.7255
## Year1998          0.0634    0.0854    0.74  0.4579
## Year1999         -0.0748    0.0756   -0.99  0.3223
## Year2000         -0.1305    0.0852   -1.53  0.1259
## Year2001         -0.1967    0.1124   -1.75  0.0801 .
## Year2002         -0.0168    0.0780   -0.22  0.8296
## Year2003         -0.1352    0.0724   -1.87  0.0621 .
## Year2004         -0.2005    0.0731   -2.74  0.0061 **
## Year2005         -0.1049    0.0704   -1.49  0.1367
## Year2006         -0.1835    0.0711   -2.58  0.0099 **
```

```

## Year2007          -0.1730      0.0687   -2.52   0.0118 *
## Year2008          -0.0687      0.0684   -1.00   0.3155
## Year2009          -0.0237      0.0667   -0.35   0.7230
## Year2010          -0.0573      0.0691   -0.83   0.4072
## Year2011          -0.1472      0.0664   -2.22   0.0268 *
## Year2012          -0.0596      0.0680   -0.88   0.3809
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0252, Adjusted R-squared:  0.017
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 1830 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.185  0.862  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      4.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2015"
## [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
##   79   58   49   50   44   46   52   35   32   24   45   51   42   56   48
## 2011 2012
##   35   47
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   39   41   29   29   29   14   33   24   25   17   37   42   31   44   36
## 2011 2012

```

```
## 28 31
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 34 31 25 25 23 8 29 21 19 14 33 36 29 38 29
## 2011 2012
## 25 28
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.07
```



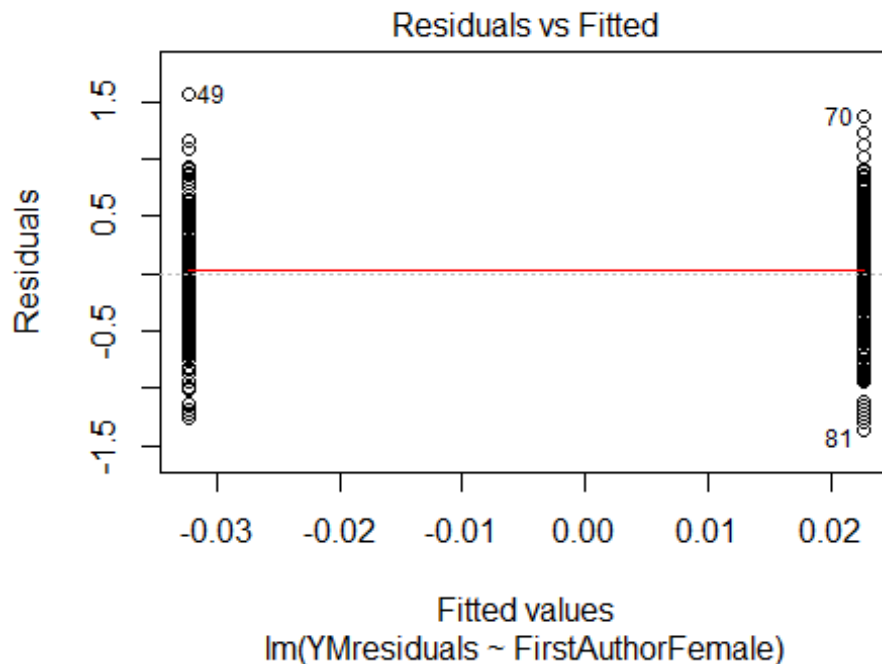
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.25, df = 1, p-value = 0.6

## [1] "Female first author team size 2018 geometric mean: 4.51955989053061"
## [1] "Male first author team size 2018 geometric mean: 2.75985952097081"

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

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 660, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.54508053653122"
## [1] "Male last author team size 2018 geometric mean: 3.16082217095398"

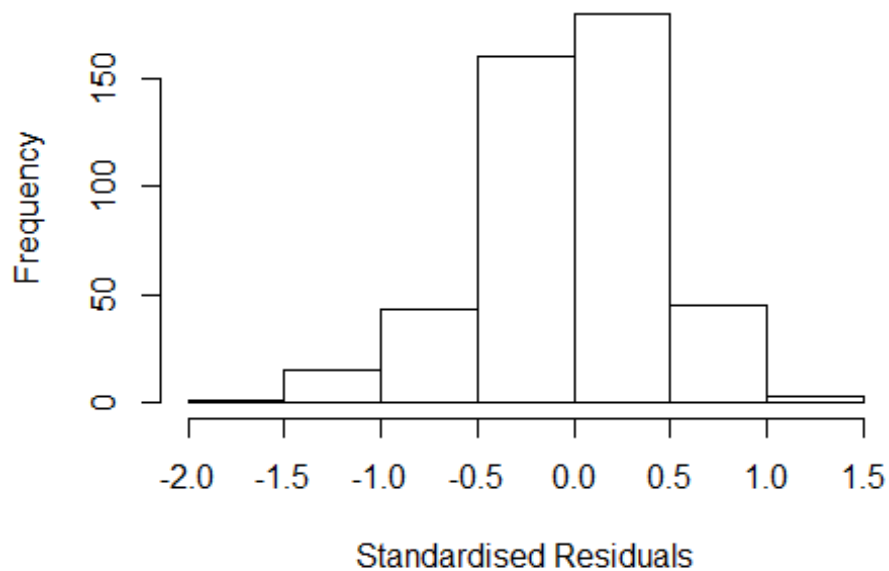
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 600, p-value = 0.06
## 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	1.707	1	1.307
LastAuthorFemale	1.648	1	1.284
UniqueAuthors	2.654	4	1.130
Year	4.803	16	1.050

Residuals from first and last author and team size



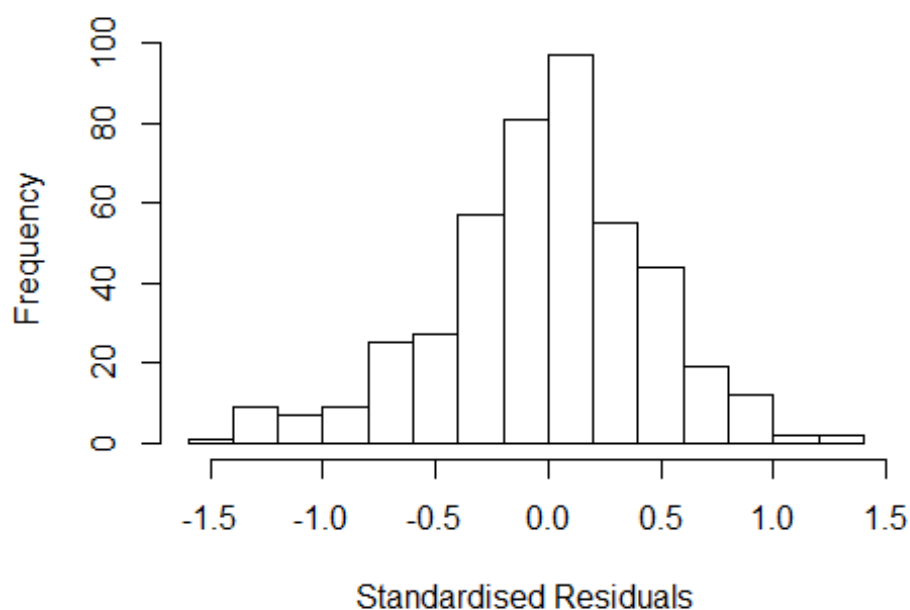
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.50502 -0.29194 0.00532 0.27088 1.35048
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2158 0.1314 9.25 <2e-16 ***
## FirstAuthorFemale1 -0.0640 0.0530 -1.21 0.2282
## LastAuthorFemale1 0.0643 0.0575 1.12 0.2640
## UniqueAuthors2 0.0805 0.0816 0.99 0.3246
## UniqueAuthors3 0.1608 0.0816 1.97 0.0495 *
## UniqueAuthors4 0.1931 0.0805 2.40 0.0168 *
## UniqueAuthors5 0.2297 0.0806 2.85 0.0046 **
## Year1997 -0.1487 0.1265 -1.18 0.2404
## Year1998 0.0297 0.1449 0.20 0.8379
## Year1999 -0.1638 0.1429 -1.15 0.2524
```

```

## Year2000          -0.1781      0.1482    -1.20    0.2301
## Year2001          -0.2852      0.1575    -1.81    0.0709 .
## Year2002          -0.0609      0.1562    -0.39    0.6970
## Year2003          -0.1668      0.1319    -1.26    0.2066
## Year2004          -0.1561      0.1604    -0.97    0.3308
## Year2005          -0.2757      0.1384    -1.99    0.0470 *
## Year2006          -0.0640      0.1339    -0.48    0.6330
## Year2007          -0.1118      0.1321    -0.85    0.3979
## Year2008          -0.2046      0.1382    -1.48    0.1396
## Year2009           0.0592      0.1242     0.48    0.6339
## Year2010          -0.1541      0.1427    -1.08    0.2807
## Year2011          -0.0364      0.1518    -0.24    0.8106
## Year2012          -0.1818      0.1561    -1.16    0.2449
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.42
## Multiple R-squared:  0.0667, Adjusted R-squared:  0.0182
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 406 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.171  0.862  0.948  0.888  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.24e-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 1.681 1      1.297
## LastAuthorFemale  1.599 1      1.265
## Year              1.925 16      1.021

```

Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.40768 -0.29005  0.00922  0.25958  1.37848
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.34852    0.11058   12.20  <2e-16 ***
## FirstAuthorFemale1 -0.06938    0.05310   -1.31    0.192
## LastAuthorFemale1  0.03927    0.05735    0.68    0.494
## Year1997          -0.15090    0.12112   -1.25    0.213
## Year1998          -0.00131    0.14658   -0.01    0.993
## Year1999          -0.17477    0.13971   -1.25    0.212
## Year2000          -0.19463    0.14409   -1.35    0.177
## Year2001          -0.26024    0.15357   -1.69    0.091 .
## Year2002          -0.08938    0.15576   -0.57    0.566
## Year2003          -0.14829    0.12917   -1.15    0.252
## Year2004          -0.13024    0.14952   -0.87    0.384
## Year2005          -0.25684    0.13489   -1.90    0.058 .
```

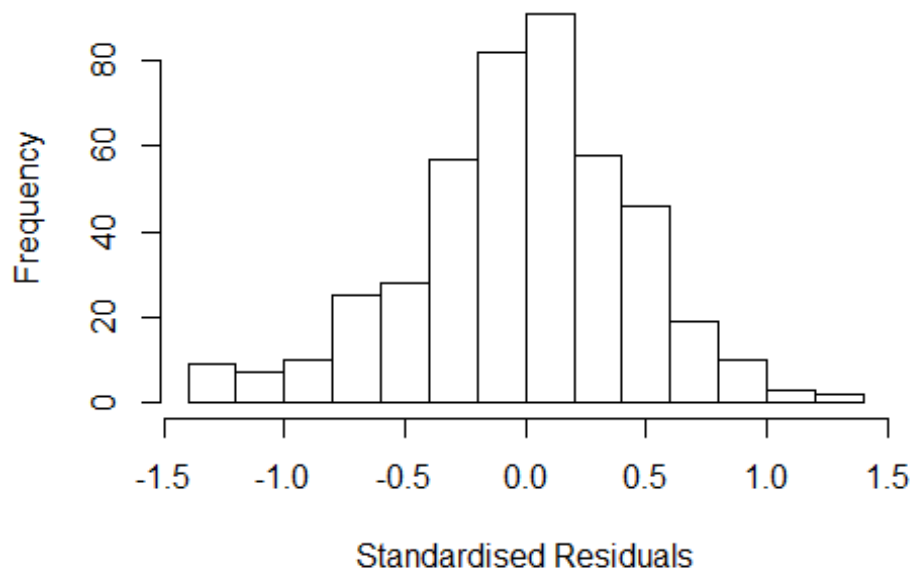


```

## Year2006          -0.04979    0.13211   -0.38    0.706
## Year2007          -0.06729    0.12728   -0.53    0.597
## Year2008          -0.14387    0.13017   -1.11    0.270
## Year2009           0.08927    0.12167    0.73    0.464
## Year2010          -0.11729    0.14028   -0.84    0.404
## Year2011           0.00188    0.15614    0.01    0.990
## Year2012          -0.13612    0.15037   -0.91    0.366
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.42
## Multiple R-squared:  0.0406, Adjusted R-squared:  0.00022
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 401 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.239  0.852  0.949  0.885  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      2.24e-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 1.414 1      1.189
## Year              1.414 16      1.011

```

Residuals from first author



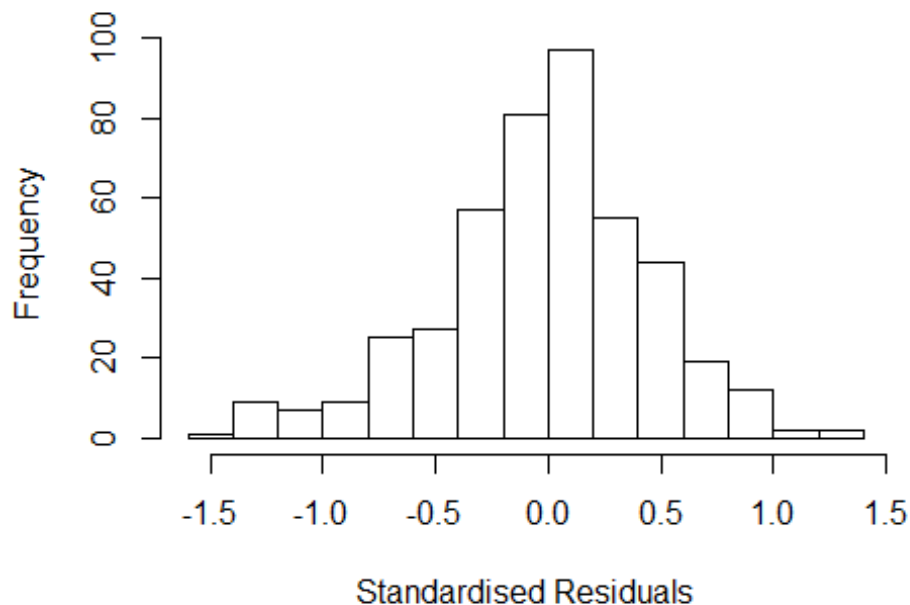
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3882 -0.2907 0.0129 0.2603 1.3683
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.35872 0.10789 12.59 <2e-16 ***
## FirstAuthorFemale1 -0.05897 0.04928 -1.20 0.232
## Year1997 -0.15522 0.12004 -1.29 0.197
## Year1998 -0.00751 0.14417 -0.05 0.959
## Year1999 -0.18306 0.13818 -1.32 0.186
## Year2000 -0.19996 0.14372 -1.39 0.165
## Year2001 -0.26622 0.15066 -1.77 0.078 .
## Year2002 -0.08942 0.15551 -0.58 0.566
## Year2003 -0.14683 0.12957 -1.13 0.258
## Year2004 -0.13654 0.14797 -0.92 0.357
## Year2005 -0.25874 0.13589 -1.90 0.058 .
## Year2006 -0.05382 0.13200 -0.41 0.684
```

```

## Year2007          -0.07568    0.12497   -0.61    0.545
## Year2008          -0.15318    0.12805   -1.20    0.232
## Year2009           0.08846    0.12203    0.72    0.469
## Year2010          -0.12665    0.13787   -0.92    0.359
## Year2011           0.00530    0.15911    0.03    0.973
## Year2012          -0.13771    0.15012   -0.92    0.359
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.419
## Multiple R-squared:  0.0392, Adjusted R-squared:  0.00112
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 409 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.251  0.852  0.951  0.886  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      2.24e-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 1.381 1          1.175
## Year              1.381 16          1.010

```

Residuals from last author



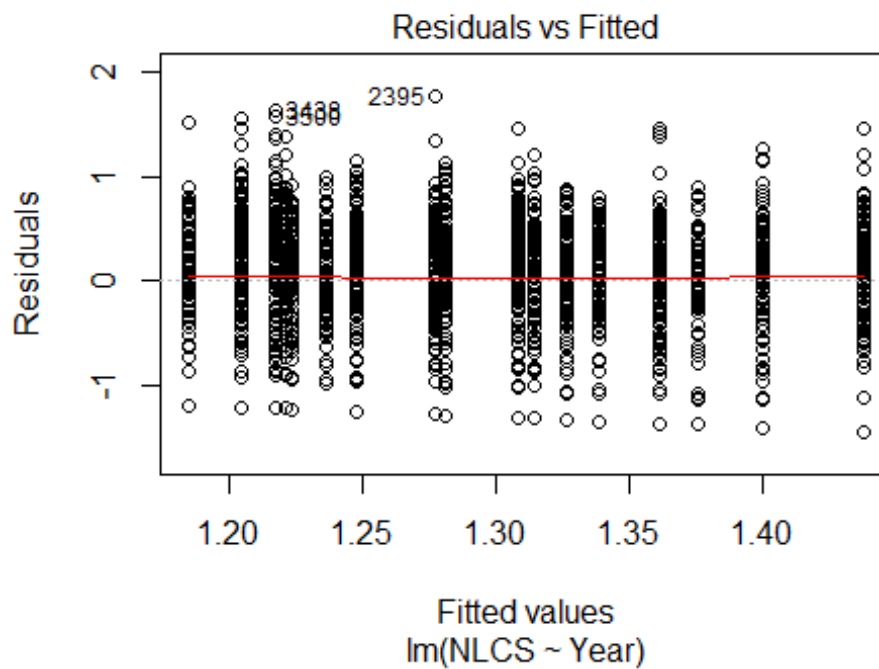
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4268 -0.2953 0.0169 0.2764 1.4065
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.32047 0.11052 11.95 <2e-16 ***
## LastAuthorFemale1 0.01730 0.05366 0.32 0.747
## Year1997 -0.14069 0.12330 -1.14 0.254
## Year1998 0.01446 0.14532 0.10 0.921
## Year1999 -0.15742 0.14106 -1.12 0.265
## Year2000 -0.17805 0.14532 -1.23 0.221
## Year2001 -0.22982 0.15344 -1.50 0.135
## Year2002 -0.07165 0.15572 -0.46 0.646
## Year2003 -0.13271 0.12969 -1.02 0.307
## Year2004 -0.12210 0.15049 -0.81 0.418
## Year2005 -0.25217 0.13687 -1.84 0.066 .
## Year2006 -0.05533 0.13263 -0.42 0.677
```

```

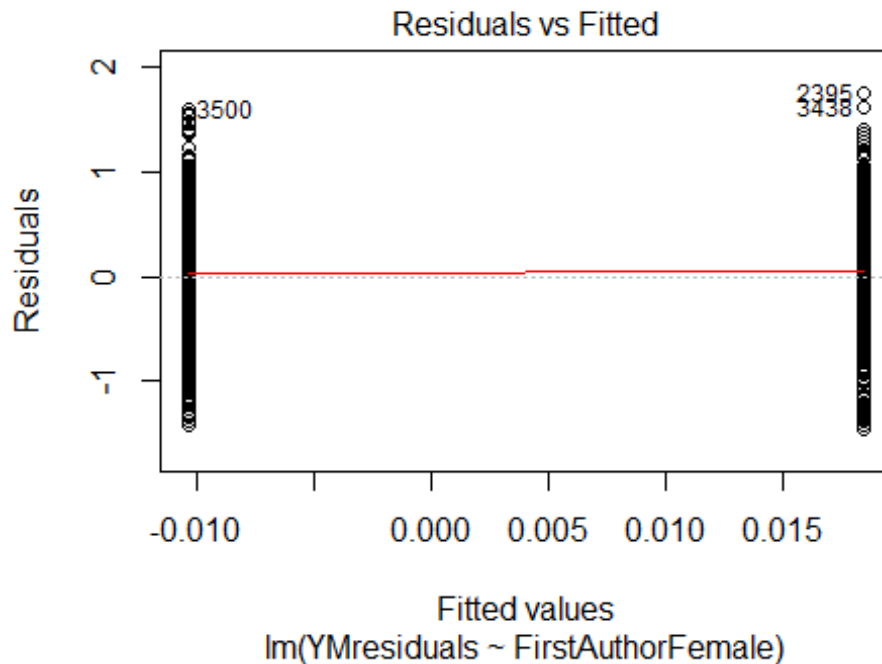
## Year2007          -0.07506      0.12857    -0.58      0.560
## Year2008          -0.15289      0.13230    -1.16      0.248
## Year2009           0.08904      0.12300      0.72      0.470
## Year2010          -0.11217      0.14267    -0.79      0.432
## Year2011           0.00399      0.15870      0.03      0.980
## Year2012          -0.13598      0.14933    -0.91      0.363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.42
## Multiple R-squared:  0.0359, Adjusted R-squared:  -0.00231
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 401 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.224  0.853  0.950  0.884  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      2.24e-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: 447"
## [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
## 177 185 157 151 158 164 161 116 141 160 191 176 198 224 224
## 2011 2012
## 219 223
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 113 124 97 101 85 101 124 88 106 116 144 144 148 163 170
## 2011 2012

```

```
## 159 165
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 96 113 81 86 72 92 116 77 88 100 120 123 124 142 151
## 2011 2012
## 136 144
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 62, df = 16, p-value = 2e-07
```

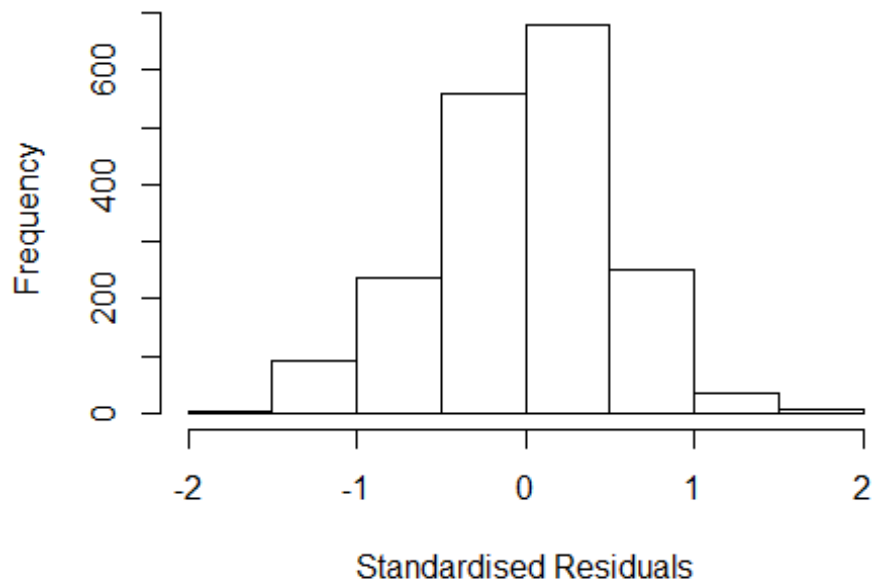


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.1161149265"
## [1] "Male first author team size 2018 geometric mean: 4.82574492644029"
##
## 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: 4.06737952318129"
## [1] "Male last author team size 2018 geometric mean: 4.76822960986811"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, 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 1.085 1          1.041
## LastAuthorFemale  1.131 1          1.064
## UniqueAuthors     1.272 4          1.031
## Year              1.409 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
## -1.6696 -0.3432  0.0239  0.3336  1.7040
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03236    0.07456   13.85 < 2e-16 ***
## FirstAuthorFemale1 -0.01336    0.02712   -0.49  0.62244
## LastAuthorFemale1  0.11314    0.02843    3.98  7.2e-05 ***
## UniqueAuthors2     0.32501    0.05811    5.59  2.6e-08 ***
## UniqueAuthors3     0.43629    0.05580    7.82  8.9e-15 ***
## UniqueAuthors4     0.52017    0.05592    9.30 < 2e-16 ***
## UniqueAuthors5     0.58524    0.05377   10.88 < 2e-16 ***
## Year1997          -0.06330    0.07283   -0.87  0.38487
## Year1998           0.00388    0.08398    0.05  0.96318
## Year1999          -0.07270    0.08112   -0.90  0.37032
```

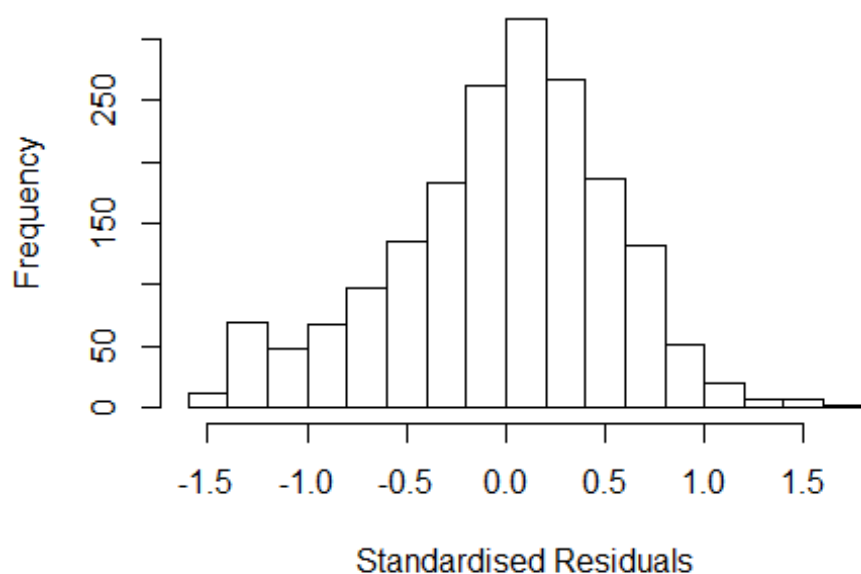


```

## Year2000      -0.25662    0.07741   -3.31  0.00093 ***
## Year2001      -0.27065    0.07718   -3.51  0.00046 ***
## Year2002      -0.20932    0.07224   -2.90  0.00380 **
## Year2003      -0.12545    0.07877   -1.59  0.11140
## Year2004      -0.13079    0.07809   -1.67  0.09413 .
## Year2005      -0.17956    0.07495   -2.40  0.01669 *
## Year2006      -0.17838    0.07513   -2.37  0.01769 *
## Year2007      -0.18995    0.07972   -2.38  0.01729 *
## Year2008      -0.23216    0.07458   -3.11  0.00188 **
## Year2009      -0.27090    0.07882   -3.44  0.00060 ***
## Year2010      -0.32614    0.08406   -3.88  0.00011 ***
## Year2011      -0.24978    0.07221   -3.46  0.00055 ***
## Year2012      -0.31218    0.08376   -3.73  0.00020 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.125, Adjusted R-squared:  0.114
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 164 weights are ~= 1. The remaining 1697 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.218  0.855  0.949   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          5.37e-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 1.054 1 1.027
## LastAuthorFemale 1.118 1 1.057
## Year 1.159 16 1.005

```

Residuals from first and last author



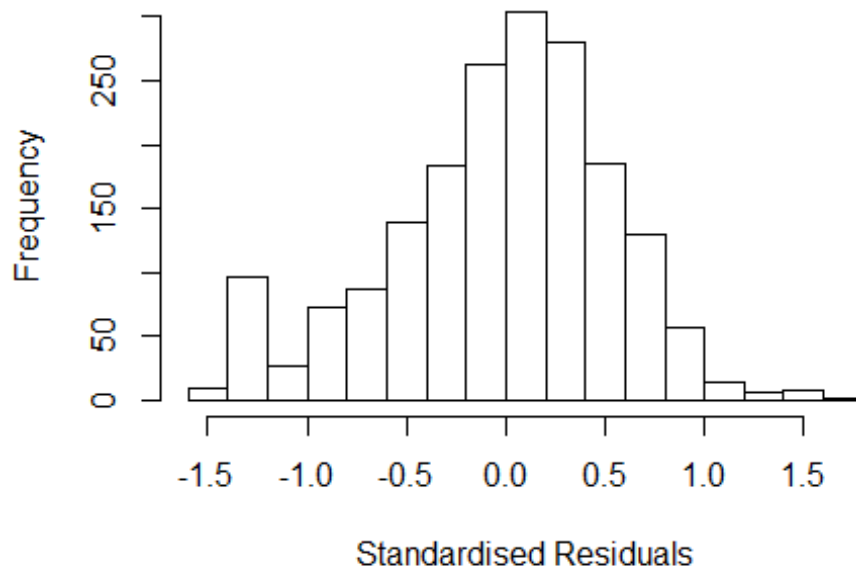
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5637 -0.3467  0.0364  0.3550  1.7956
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3986    0.0644   21.72 < 2e-16 ***
## FirstAuthorFemale1  0.0159    0.0280    0.57  0.5709
## LastAuthorFemale1  0.1181    0.0295    4.00 6.6e-05 ***
## Year1997         -0.0736    0.0784   -0.94  0.3481
## Year1998          0.0312    0.0890    0.35  0.7258
## Year1999         -0.0674    0.0876   -0.77  0.4420
## Year2000         -0.2205    0.0837   -2.64  0.0085 **
## Year2001         -0.2175    0.0836   -2.60  0.0093 **
## Year2002         -0.1835    0.0770   -2.38  0.0173 *
## Year2003         -0.0388    0.0795   -0.49  0.6258
## Year2004         -0.0487    0.0810   -0.60  0.5483
## Year2005         -0.1004    0.0814   -1.23  0.2180
```

```

## Year2006          -0.1048      0.0796   -1.32    0.1883
## Year2007          -0.1199      0.0846   -1.42    0.1566
## Year2008          -0.1620      0.0812   -2.00    0.0460 *
## Year2009          -0.1548      0.0834   -1.86    0.0636 .
## Year2010          -0.2096      0.0895   -2.34    0.0193 *
## Year2011          -0.1838      0.0752   -2.44    0.0147 *
## Year2012          -0.2129      0.0905   -2.35    0.0187 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.0267, Adjusted R-squared:  0.0172
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 1693 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.199  0.854   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      5.37e-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 1.042 1          1.021
## Year              1.042 16          1.001

```

Residuals from first author



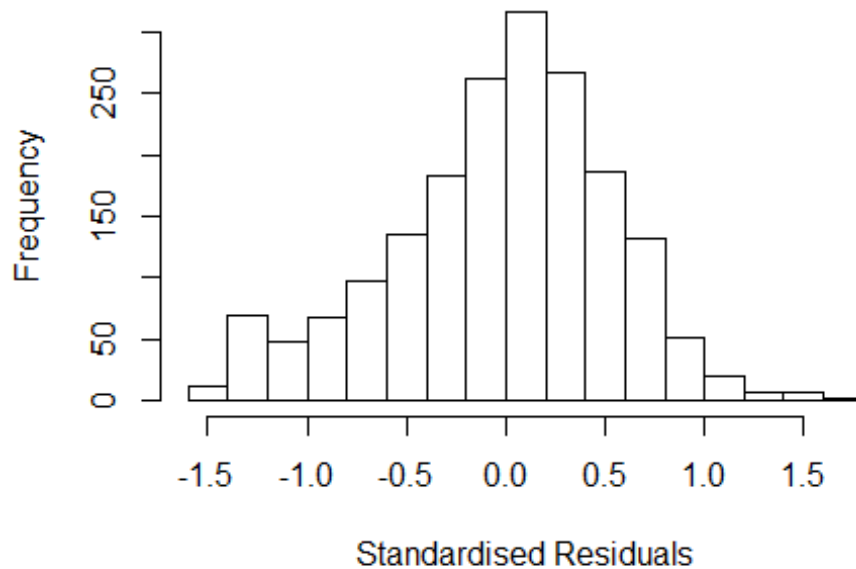
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4832 -0.3532 0.0322 0.3481 1.7717
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4344 0.0638 22.49 <2e-16 ***
## FirstAuthorFemale1 0.0305 0.0279 1.09 0.2738
## Year1997 -0.0813 0.0795 -1.02 0.3066
## Year1998 0.0182 0.0893 0.20 0.8384
## Year1999 -0.0832 0.0883 -0.94 0.3460
## Year2000 -0.2297 0.0842 -2.73 0.0064 **
## Year2001 -0.2217 0.0835 -2.65 0.0080 **
## Year2002 -0.1916 0.0772 -2.48 0.0132 *
## Year2003 -0.0456 0.0803 -0.57 0.5702
## Year2004 -0.0678 0.0814 -0.83 0.4054
## Year2005 -0.1021 0.0813 -1.26 0.2091
## Year2006 -0.1222 0.0800 -1.53 0.1269
```

```

## Year2007          -0.1135      0.0853   -1.33    0.1837
## Year2008          -0.1886      0.0809   -2.33    0.0198 *
## Year2009          -0.1606      0.0840   -1.91    0.0561 .
## Year2010          -0.2190      0.0899   -2.43    0.0150 *
## Year2011          -0.2029      0.0754   -2.69    0.0072 **
## Year2012          -0.2129      0.0919   -2.32    0.0206 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.0181, Adjusted R-squared:  0.00901
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 183 weights are ~= 1. The remaining 1678 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.213  0.852  0.947  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      5.37e-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 1.109 1          1.053
## Year            1.109 16          1.003

```

Residuals from last author



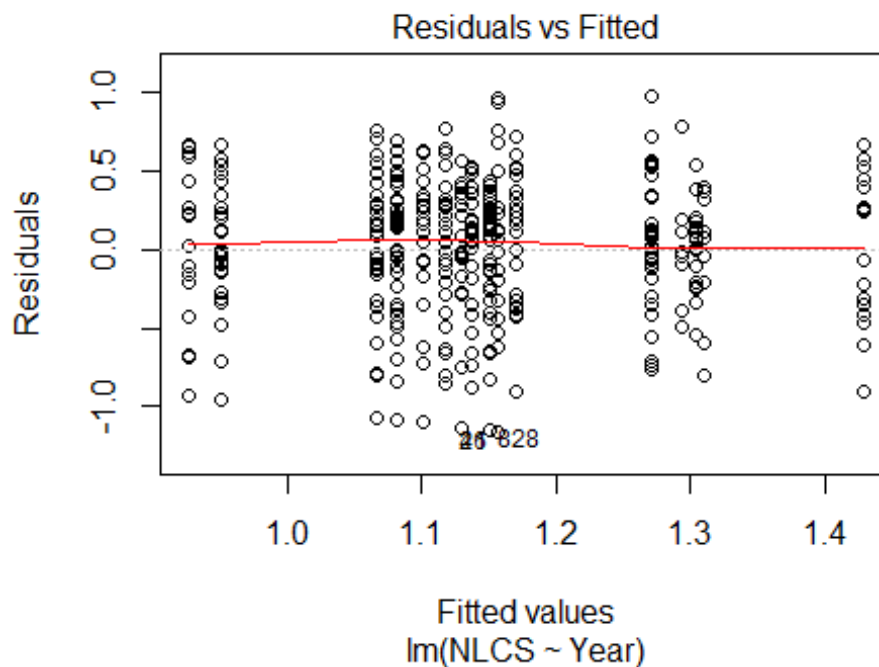
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5542 -0.3505 0.0346 0.3540 1.8057
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4011 0.0645 21.74 < 2e-16 ***
## LastAuthorFemale1 0.1203 0.0294 4.09 4.5e-05 ***
## Year1997 -0.0712 0.0784 -0.91 0.3635
## Year1998 0.0327 0.0893 0.37 0.7142
## Year1999 -0.0642 0.0877 -0.73 0.4642
## Year2000 -0.2186 0.0838 -2.61 0.0092 **
## Year2001 -0.2165 0.0838 -2.58 0.0098 **
## Year2002 -0.1825 0.0772 -2.36 0.0181 *
## Year2003 -0.0356 0.0793 -0.45 0.6541
## Year2004 -0.0482 0.0811 -0.59 0.5528
## Year2005 -0.0977 0.0814 -1.20 0.2301
## Year2006 -0.1017 0.0796 -1.28 0.2019
```

```

## Year2007          -0.1171      0.0843   -1.39   0.1648
## Year2008          -0.1589      0.0811   -1.96   0.0502 .
## Year2009          -0.1518      0.0835   -1.82   0.0693 .
## Year2010          -0.2078      0.0895   -2.32   0.0204 *
## Year2011          -0.1797      0.0752   -2.39   0.0170 *
## Year2012          -0.2094      0.0903   -2.32   0.0205 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.0265, Adjusted R-squared:  0.0175
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 159 weights are ~= 1. The remaining 1702 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.194  0.856  0.948  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      5.37e-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: 1861"
## [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
##   32   32   28   27   38   32   47   32   27   34   28   31   38   45   41
## 2011 2012
##   49   48
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19   16    9   17   19   11   32   24   19   22   18   24   23   33   26
## 2011 2012

```

```
## 34 39
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 16 12 8 12 18 7 28 20 15 19 16 21 21 31 23
## 2011 2012
## 31 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 = 19, df = 16, p-value = 0.3
```



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

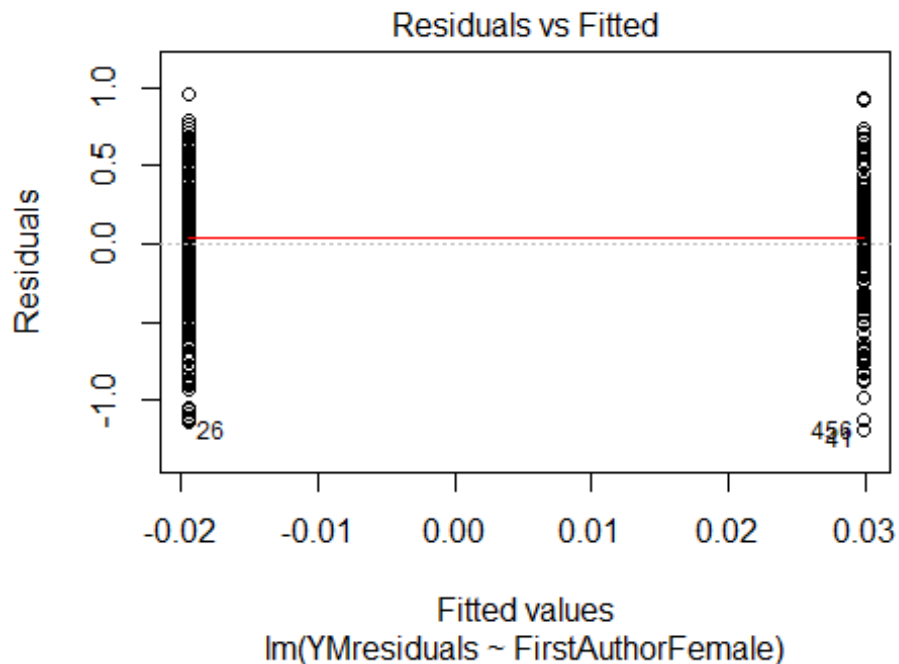
## [1] "Female first author team size 2018 geometric mean: 5.61398239352162"
## [1] "Male first author team size 2018 geometric mean: 5.66184518161356"

## Warning in wilcox.test.default(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 = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.47213595499958"
## [1] "Male last author team size 2018 geometric mean: 6.42406458936213"

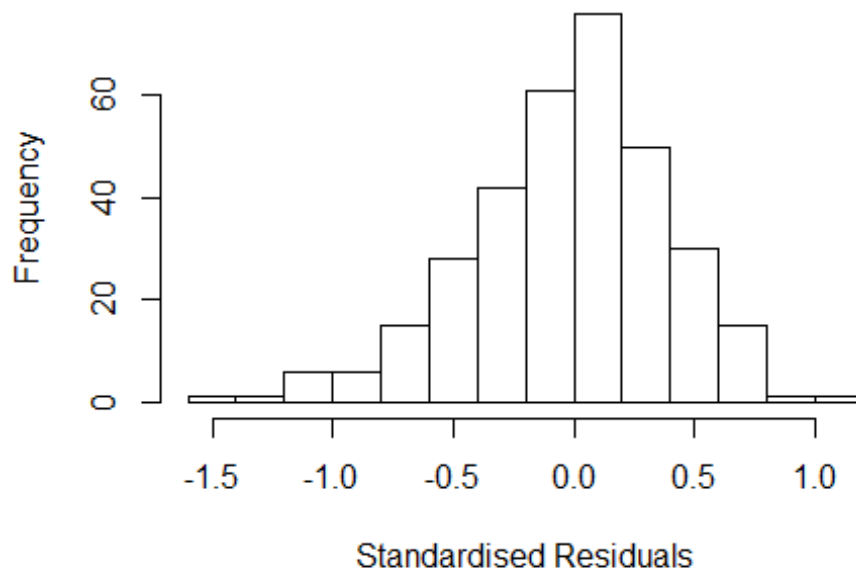
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 40, p-value = 0.3
## 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	1.237	1	1.112
LastAuthorFemale	1.280	1	1.131
UniqueAuthors	2.605	4	1.127
Year	3.790	16	1.043

Residuals from first and last author and team size



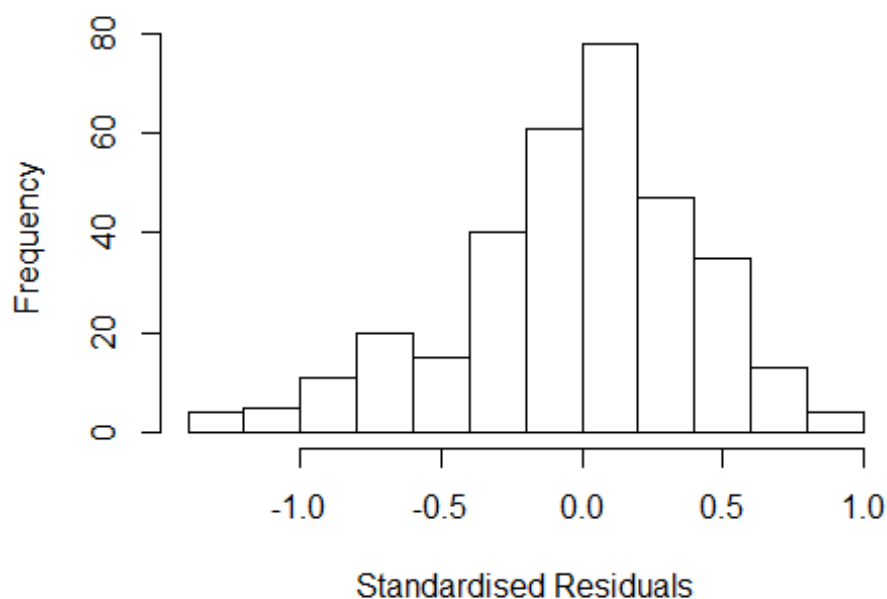
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4100 -0.2848 0.0248 0.2545 1.1059
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.19e-01 2.05e-01 3.50 0.00053 ***
## FirstAuthorFemale1 1.06e-02 4.71e-02 0.23 0.82137
## LastAuthorFemale1 8.77e-02 5.10e-02 1.72 0.08622 .
## UniqueAuthors2 3.17e-01 1.59e-01 2.00 0.04687 *
## UniqueAuthors3 4.46e-01 1.49e-01 2.99 0.00298 **
## UniqueAuthors4 4.55e-01 1.42e-01 3.21 0.00144 **
## UniqueAuthors5 5.84e-01 1.38e-01 4.24 3e-05 ***
## Year1997 2.18e-01 1.93e-01 1.13 0.25980
## Year1998 7.87e-03 1.83e-01 0.04 0.96580
## Year1999 1.99e-02 1.79e-01 0.11 0.91155
```

```

## Year2000      1.07e-01  1.58e-01  0.67  0.50187
## Year2001      4.37e-05  2.55e-01  0.00  0.99986
## Year2002     -2.41e-01  1.53e-01 -1.57  0.11702
## Year2003     -2.13e-01  1.67e-01 -1.27  0.20508
## Year2004      1.29e-02  1.80e-01  0.07  0.94281
## Year2005     -1.14e-01  1.89e-01 -0.60  0.54717
## Year2006     -2.21e-01  1.83e-01 -1.21  0.22826
## Year2007     -6.37e-02  1.53e-01 -0.42  0.67669
## Year2008     -6.46e-02  1.84e-01 -0.35  0.72579
## Year2009     -1.27e-01  1.54e-01 -0.82  0.41219
## Year2010     -1.60e-01  1.64e-01 -0.98  0.33022
## Year2011      3.62e-02  1.55e-01  0.23  0.81511
## Year2012      8.44e-03  1.51e-01  0.06  0.95540
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared:  0.177, Adjusted R-squared:  0.119
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 303 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.148  0.863  0.945  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      3.00e-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 1.229 1      1.109
## LastAuthorFemale  1.243 1      1.115
## Year              1.519 16      1.013

```

Residuals from first and last author



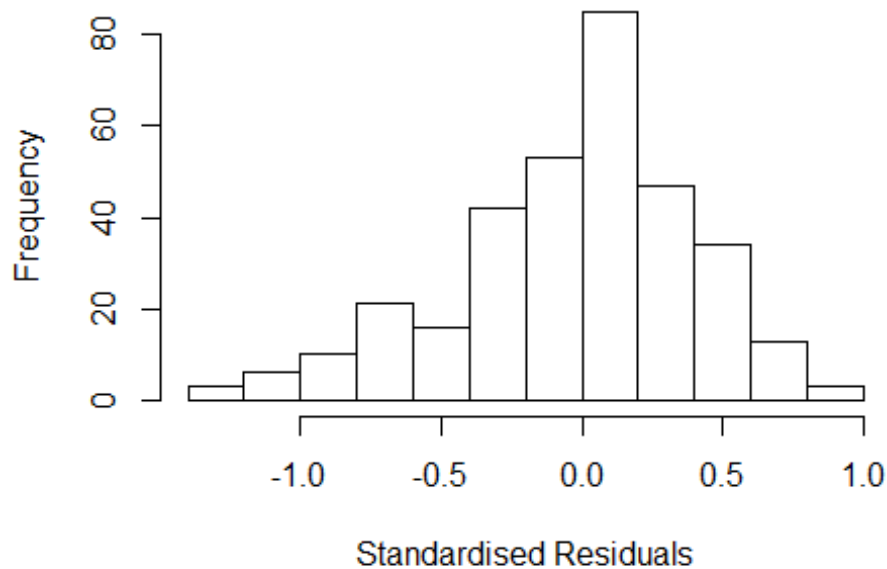
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2845 -0.2896 0.0287 0.2476 0.9962
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.185253 0.161342 7.35 1.8e-12 ***
## FirstAuthorFemale1 0.033921 0.050065 0.68 0.499
## LastAuthorFemale1 0.094098 0.056099 1.68 0.094 .
## Year1997 0.213278 0.210240 1.01 0.311
## Year1998 0.047849 0.197689 0.24 0.809
## Year1999 -0.019461 0.188413 -0.10 0.918
## Year2000 0.065295 0.167149 0.39 0.696
## Year2001 -0.023014 0.221853 -0.10 0.917
## Year2002 -0.243731 0.172493 -1.41 0.159
## Year2003 -0.185867 0.183663 -1.01 0.312
## Year2004 0.005166 0.186249 0.03 0.978
## Year2005 -0.079209 0.195452 -0.41 0.686
```

```

## Year2006          -0.248898    0.210631    -1.18    0.238
## Year2007          -0.089058    0.173583    -0.51    0.608
## Year2008          -0.054198    0.192654    -0.28    0.779
## Year2009          -0.110033    0.175076    -0.63    0.530
## Year2010          -0.150915    0.185861    -0.81    0.417
## Year2011           0.024639    0.173732     0.14    0.887
## Year2012          -0.000517    0.163995     0.00    0.997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.0711, Adjusted R-squared:  0.0179
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 305 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.288  0.863   0.948   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      3.00e-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 1.217 1      1.103
## Year              1.217 16      1.006

```

Residuals from first author



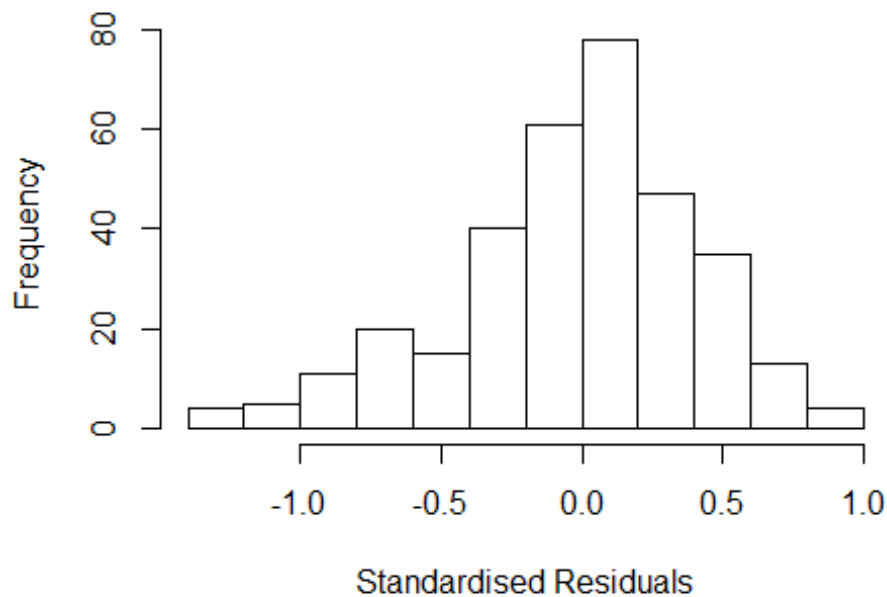
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2477 -0.2886 0.0423 0.2547 0.9606
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20603 0.16390 7.36 1.6e-12 ***
## FirstAuthorFemale1 0.04170 0.05061 0.82 0.41
## Year1997 0.21006 0.21162 0.99 0.32
## Year1998 0.02568 0.20021 0.13 0.90
## Year1999 -0.01643 0.19604 -0.08 0.93
## Year2000 0.08883 0.17034 0.52 0.60
## Year2001 -0.00353 0.23104 -0.02 0.99
## Year2002 -0.24205 0.17701 -1.37 0.17
## Year2003 -0.17220 0.18932 -0.91 0.36
## Year2004 -0.00755 0.18778 -0.04 0.97
## Year2005 -0.08880 0.20054 -0.44 0.66
## Year2006 -0.23813 0.21275 -1.12 0.26
```

```

## Year2007      -0.08481    0.17933   -0.47    0.64
## Year2008      -0.07043    0.19738   -0.36    0.72
## Year2009      -0.12346    0.17867   -0.69    0.49
## Year2010      -0.14973    0.18906   -0.79    0.43
## Year2011       0.03170    0.17823    0.18    0.86
## Year2012      -0.00542    0.16716   -0.03    0.97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.407
## Multiple R-squared:  0.0623, Adjusted R-squared:  0.0117
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 303 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.327  0.852  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      3.00e-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 1.239 1          1.113
## Year            1.239 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
## -1.2974 -0.2894 0.0242 0.2493 1.0178
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20403 0.15644 7.70 1.8e-13 ***
## LastAuthorFemale1 0.09827 0.05595 1.76 0.08 .
## Year1997 0.20639 0.20971 0.98 0.33
## Year1998 0.03711 0.19582 0.19 0.85
## Year1999 -0.01376 0.18980 -0.07 0.94
## Year2000 0.05725 0.16608 0.34 0.73
## Year2001 -0.02721 0.22428 -0.12 0.90
## Year2002 -0.24864 0.17276 -1.44 0.15
## Year2003 -0.19228 0.18267 -1.05 0.29
## Year2004 -0.00487 0.18549 -0.03 0.98
## Year2005 -0.08522 0.19269 -0.44 0.66
## Year2006 -0.25171 0.20758 -1.21 0.23
```

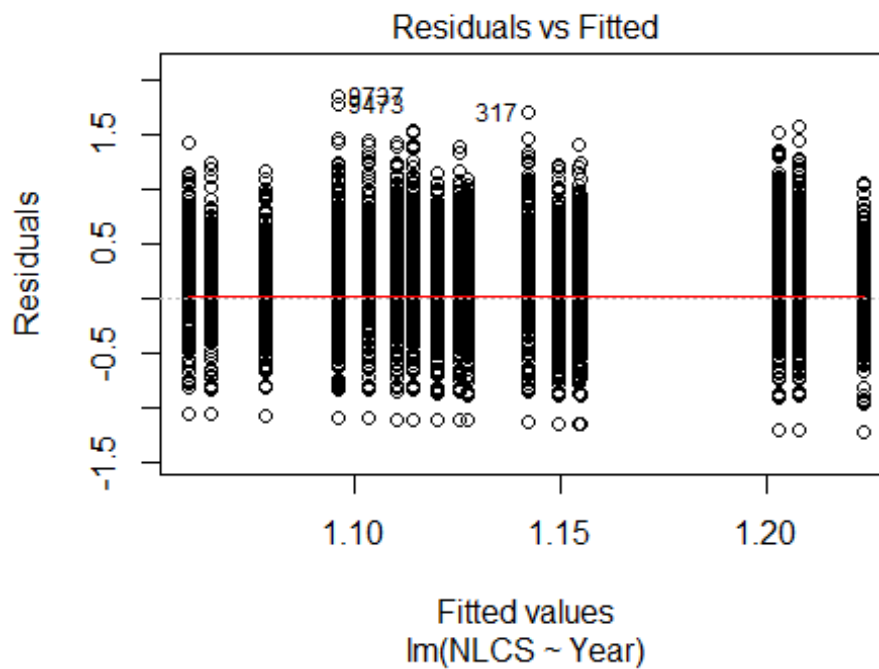


```

## Year2007          -0.09144      0.17405    -0.53      0.60
## Year2008          -0.05610      0.19268    -0.29      0.77
## Year2009          -0.11510      0.17437    -0.66      0.51
## Year2010          -0.15074      0.18676    -0.81      0.42
## Year2011           0.01814      0.17393     0.10      0.92
## Year2012          -0.01079      0.16227    -0.07      0.95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.399
## Multiple R-squared:  0.0703, Adjusted R-squared:  0.0201
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 302 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.269  0.851  0.947  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      3.00e-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: 333"
## [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
##  497  482  526  498  514  542  488  459  445  441  571  576  623  663  636
## 2011 2012
##  595  585
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  287  303  330  303  303  307  373  344  344  342  453  472  502  519  496
## 2011 2012

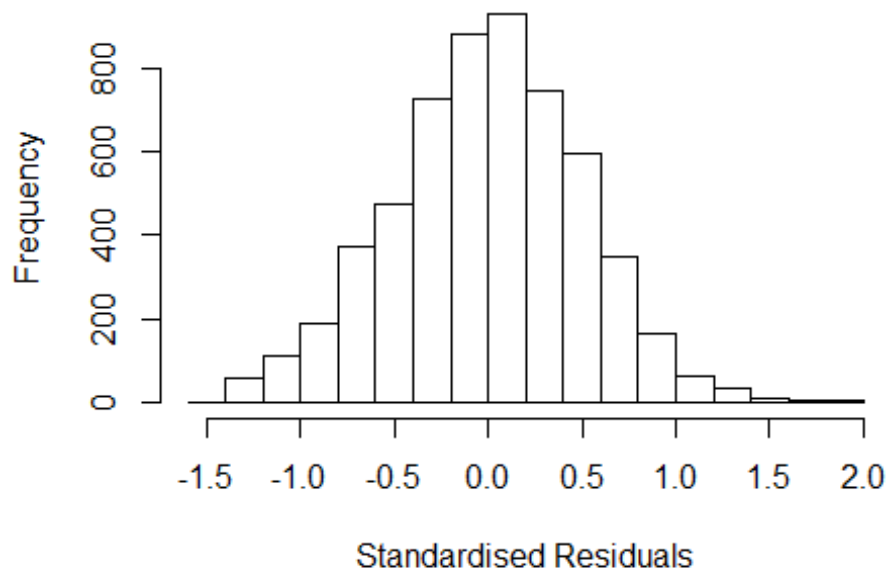
```

```
## 477 474
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 229 261 273 260 261 263 321 286 302 277 403 418 438 442 441
## 2011 2012
## 428 405
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 72, df = 16, p-value = 4e-09
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 9e-04
```


Residuals from first and last author and team size



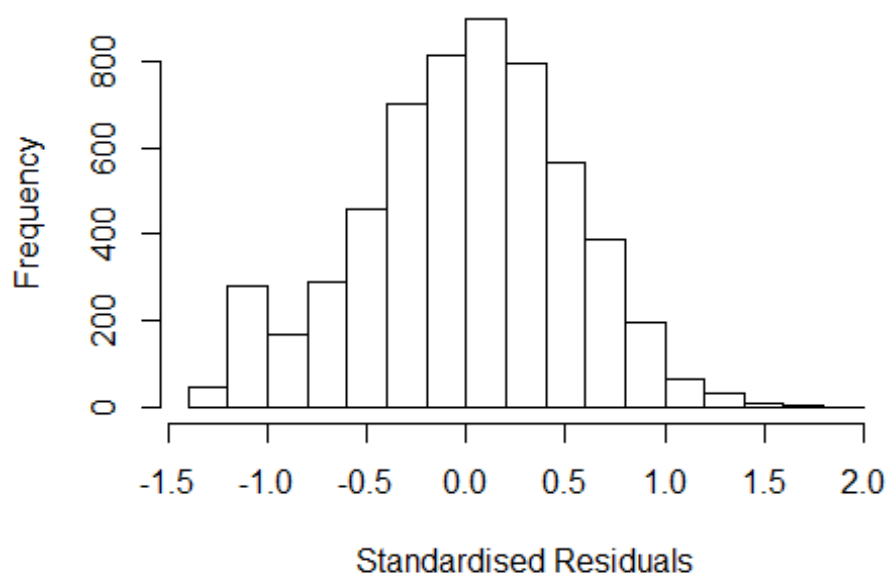
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4250 -0.3294 0.0106 0.3313 1.9650
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8310 0.0463 17.96 < 2e-16 ***
## FirstAuthorFemale1 0.0515 0.0155 3.33 0.00088 ***
## LastAuthorFemale1 0.0605 0.0184 3.30 0.00099 ***
## UniqueAuthors2 0.2395 0.0384 6.23 5.0e-10 ***
## UniqueAuthors3 0.3380 0.0354 9.55 < 2e-16 ***
## UniqueAuthors4 0.3618 0.0344 10.51 < 2e-16 ***
## UniqueAuthors5 0.5941 0.0315 18.84 < 2e-16 ***
## Year1997 -0.1711 0.0517 -3.31 0.00094 ***
## Year1998 -0.0638 0.0489 -1.30 0.19255
## Year1999 -0.0779 0.0481 -1.62 0.10528
```

```

## Year2000      -0.0859      0.0476      -1.80      0.07115 .
## Year2001      -0.0548      0.0465      -1.18      0.23870
## Year2002      -0.1420      0.0482      -2.94      0.00325 **
## Year2003      -0.2070      0.0462      -4.48      7.5e-06 ***
## Year2004      -0.2411      0.0456      -5.28      1.3e-07 ***
## Year2005      -0.2017      0.0468      -4.31      1.7e-05 ***
## Year2006      -0.1404      0.0451      -3.11      0.00187 **
## Year2007      -0.2143      0.0452      -4.74      2.1e-06 ***
## Year2008      -0.1654      0.0450      -3.68      0.00024 ***
## Year2009      -0.2023      0.0466      -4.34      1.4e-05 ***
## Year2010      -0.2138      0.0479      -4.46      8.4e-06 ***
## Year2011      -0.1501      0.0465      -3.23      0.00125 **
## Year2012      -0.1410      0.0468      -3.01      0.00260 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.125, Adjusted R-squared:  0.121
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 481 weights are ~= 1. The remaining 5227 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0807 0.8670 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          1.75e-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 1.037 1      1.019
## LastAuthorFemale  1.027 1      1.013
## Year              1.032 16      1.001

```

Residuals from first and last author



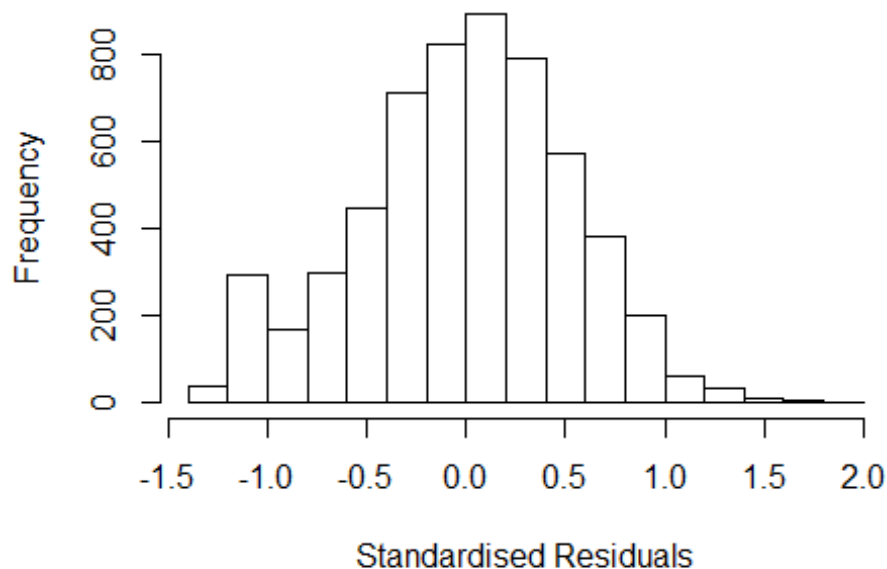
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3200 -0.3450 0.0217 0.3546 1.8592
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1927 0.0389 30.63 < 2e-16 ***
## FirstAuthorFemale1 0.0645 0.0160 4.02 5.9e-05 ***
## LastAuthorFemale1 0.0426 0.0195 2.18 0.0292 *
## Year1997 -0.1559 0.0520 -3.00 0.0027 **
## Year1998 -0.0520 0.0519 -1.00 0.3167
## Year1999 -0.0488 0.0507 -0.96 0.3358
## Year2000 -0.0551 0.0497 -1.11 0.2680
## Year2001 0.0201 0.0474 0.42 0.6712
## Year2002 -0.0807 0.0503 -1.60 0.1090
## Year2003 -0.1292 0.0493 -2.62 0.0088 **
## Year2004 -0.1440 0.0479 -3.01 0.0026 **
## Year2005 -0.0885 0.0487 -1.82 0.0689 .
```

```

## Year2006          -0.0489      0.0465   -1.05   0.2924
## Year2007          -0.1110      0.0472   -2.35   0.0187 *
## Year2008          -0.0760      0.0467   -1.63   0.1042
## Year2009          -0.0941      0.0484   -1.94   0.0520 .
## Year2010          -0.1039      0.0491   -2.12   0.0344 *
## Year2011          -0.0182      0.0476   -0.38   0.7015
## Year2012          -0.0105      0.0483   -0.22   0.8276
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.012, Adjusted R-squared:  0.00891
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 455 weights are ~= 1. The remaining 5253 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.173  0.865   0.952   0.904   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.75e-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 1.022 1      1.011
## Year              1.022 16      1.001

```

Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2874 -0.3463 0.0208 0.3508 1.8533
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19609 0.03888 30.76 <2e-16 ***
## FirstAuthorFemale1 0.07020 0.01591 4.41 1e-05 ***
## Year1997 -0.15381 0.05209 -2.95 0.0032 **
## Year1998 -0.05202 0.05188 -1.00 0.3161
## Year1999 -0.04628 0.05070 -0.91 0.3614
## Year2000 -0.05243 0.04974 -1.05 0.2919
## Year2001 0.02113 0.04741 0.45 0.6558
## Year2002 -0.07852 0.05027 -1.56 0.1183
## Year2003 -0.12710 0.04937 -2.57 0.0101 *
## Year2004 -0.14162 0.04776 -2.97 0.0030 **
## Year2005 -0.08585 0.04867 -1.76 0.0778 .
## Year2006 -0.04734 0.04650 -1.02 0.3087
```

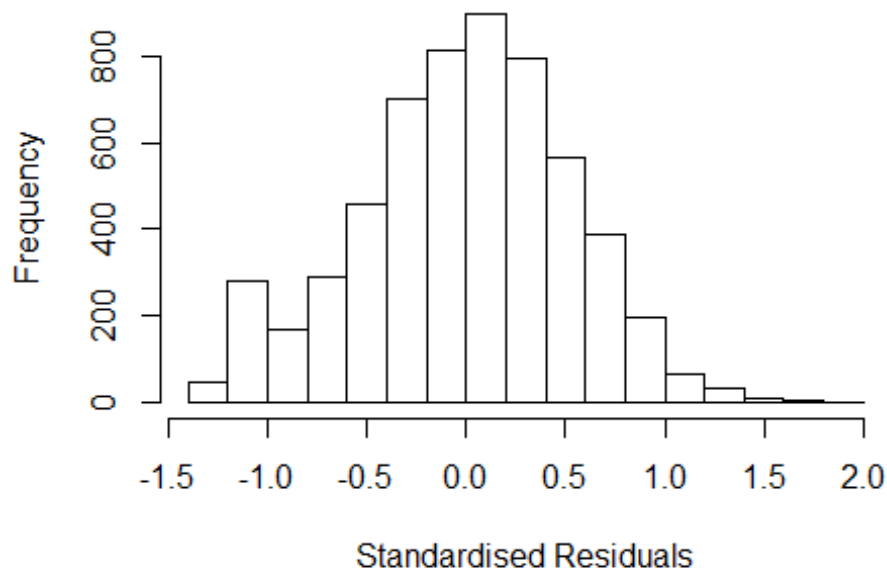


```

## Year2007          -0.11086    0.04720   -2.35    0.0189 *
## Year2008          -0.07375    0.04668   -1.58    0.1141
## Year2009          -0.09180    0.04842   -1.90    0.0580 .
## Year2010          -0.10139    0.04909   -2.07    0.0389 *
## Year2011          -0.01527    0.04756   -0.32    0.7481
## Year2012          -0.00787    0.04825   -0.16    0.8704
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.0112, Adjusted R-squared:  0.00824
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 448 weights are ~= 1. The remaining 5260 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.176  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      1.75e-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 1.012 1          1.006
## Year            1.012 16          1.000

```

Residuals from last author



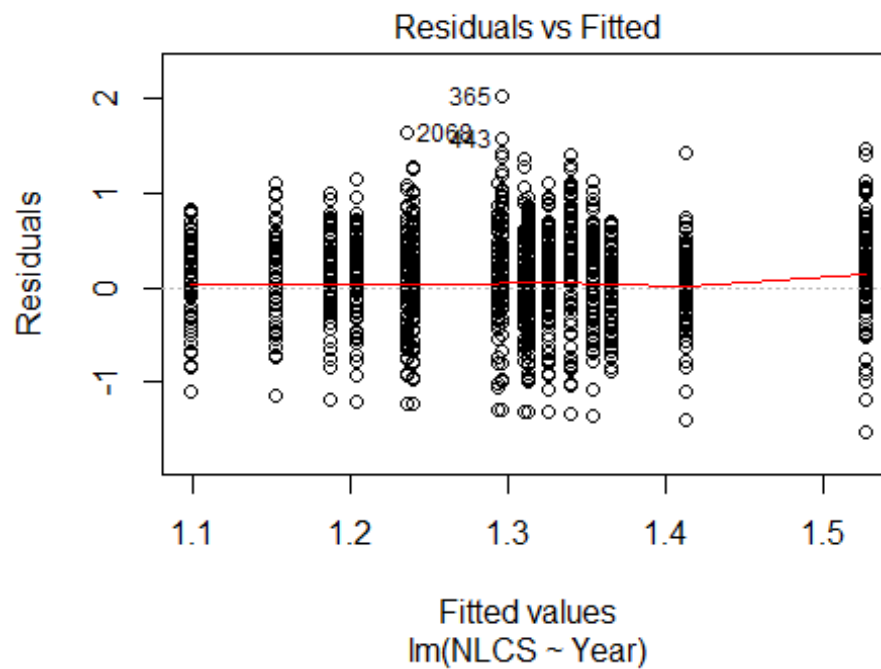
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2808 -0.3455 0.0223 0.3530 1.8423
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20391 0.03875 31.07 <2e-16 ***
## LastAuthorFemale1 0.05501 0.01932 2.85 0.0044 **
## Year1997 -0.15645 0.05203 -3.01 0.0027 **
## Year1998 -0.04893 0.05170 -0.95 0.3439
## Year1999 -0.04958 0.05063 -0.98 0.3275
## Year2000 -0.05237 0.04978 -1.05 0.2928
## Year2001 0.02183 0.04733 0.46 0.6446
## Year2002 -0.07671 0.05036 -1.52 0.1277
## Year2003 -0.12528 0.04940 -2.54 0.0112 *
## Year2004 -0.14104 0.04787 -2.95 0.0032 **
## Year2005 -0.08374 0.04860 -1.72 0.0849 .
## Year2006 -0.04444 0.04648 -0.96 0.3391
```

```

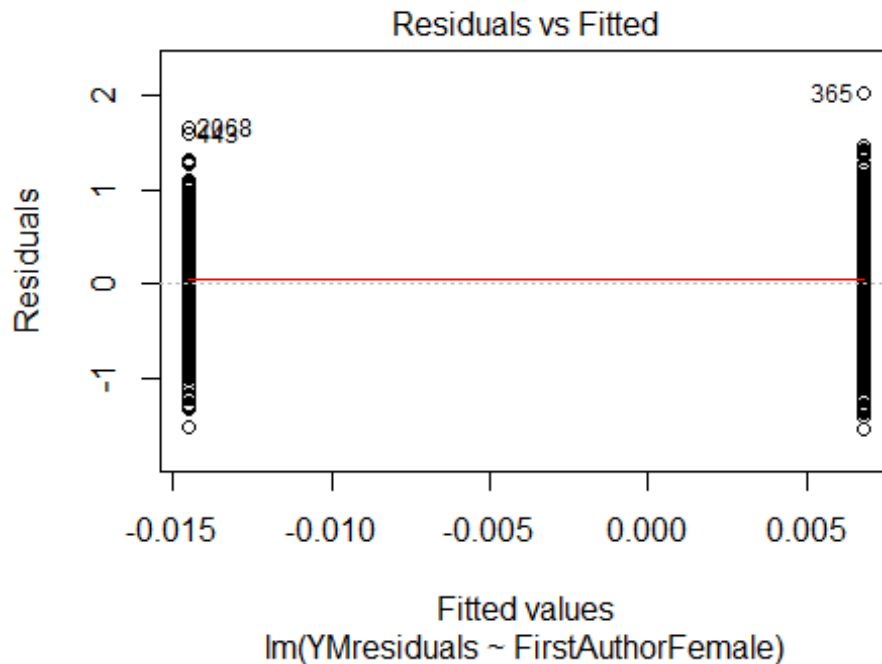
## Year2007          -0.10694      0.04724      -2.26      0.0236 *
## Year2008          -0.07069      0.04665      -1.52      0.1297
## Year2009          -0.09081      0.04835      -1.88      0.0604 .
## Year2010          -0.09822      0.04897      -2.01      0.0450 *
## Year2011          -0.00893      0.04747      -0.19      0.8507
## Year2012          -0.00349      0.04834      -0.07      0.9424
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.00929,    Adjusted R-squared:  0.00633
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 468 weights are ~= 1. The remaining 5240 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.182  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      1.75e-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: 5708"
## [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
## 177 108 141 134 108 128 131 92 73 86 109 107 124 115 110
## 2011 2012
## 120 134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 112 80 93 105 72 70 105 79 64 69 86 86 101 93 80
## 2011 2012

```

```
## 100 107
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 78 73 72 87 64 64 96 65 55 55 75 76 82 81 71
## 2011 2012
## 93 87
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 120, df = 16, p-value <2e-16
```

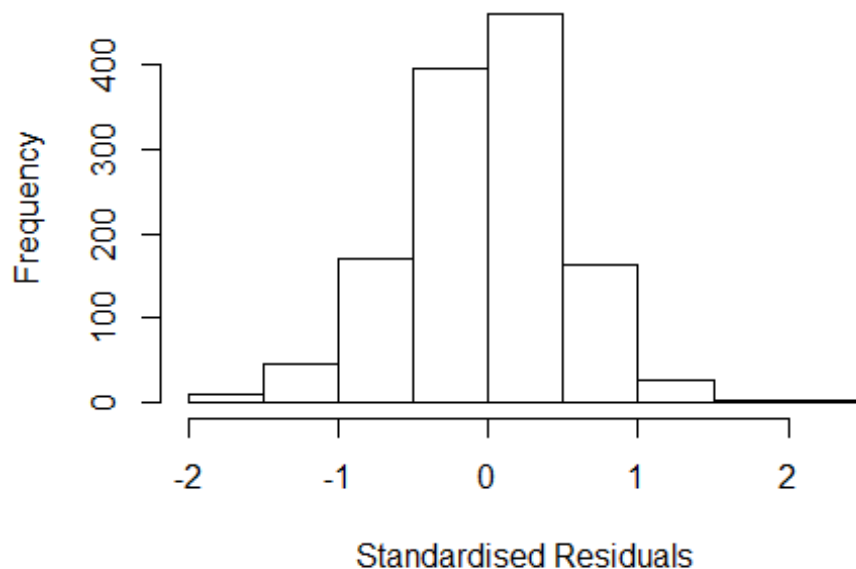


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.66, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 5.49655966256785"
## [1] "Male first author team size 2018 geometric mean: 4.66138319764132"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.20900314440306"
## [1] "Male last author team size 2018 geometric mean: 4.95410177516429"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, 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 1.206 1      1.098
## LastAuthorFemale  1.236 1      1.112
## UniqueAuthors     1.356 4      1.039
## Year              1.468 16     1.012
```

Residuals from first and last author and team size



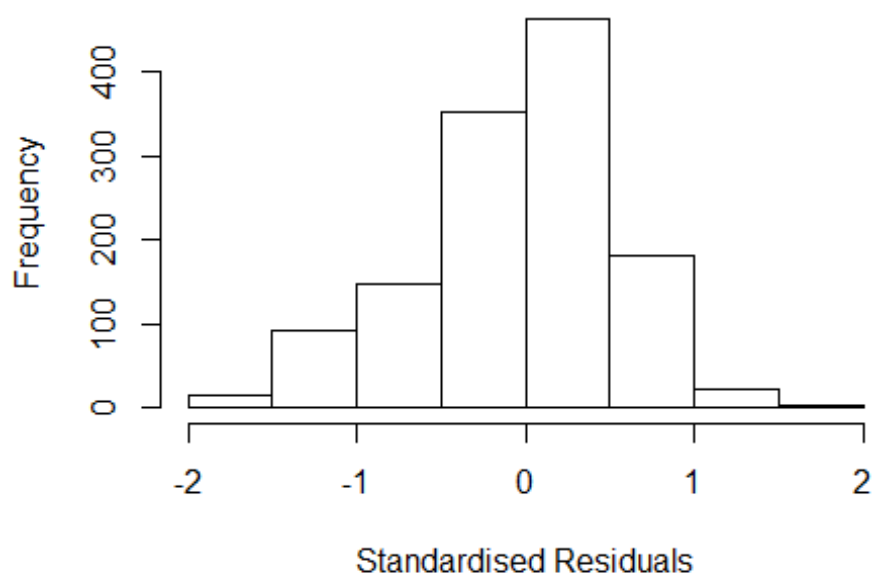
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8790 -0.3412 0.0169 0.3273 2.0497
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.83263 0.10073 8.27 3.5e-16 ***
## FirstAuthorFemale1 0.01351 0.03221 0.42 0.67508
## LastAuthorFemale1 0.00977 0.03691 0.26 0.79134
## UniqueAuthors2 0.44875 0.08602 5.22 2.1e-07 ***
## UniqueAuthors3 0.68197 0.08114 8.40 < 2e-16 ***
## UniqueAuthors4 0.75958 0.07399 10.27 < 2e-16 ***
## UniqueAuthors5 0.85937 0.06909 12.44 < 2e-16 ***
## Year1997 -0.08585 0.09801 -0.88 0.38124
## Year1998 -0.01785 0.11455 -0.16 0.87623
## Year1999 0.17347 0.09690 1.79 0.07366 .
```

```

## Year2000      -0.16898    0.09158   -1.85  0.06525 .
## Year2001      -0.23421    0.09886   -2.37  0.01798 *
## Year2002      -0.22856    0.08630   -2.65  0.00819 **
## Year2003      -0.39709    0.09165   -4.33  1.6e-05 ***
## Year2004      -0.23123    0.09592   -2.41  0.01607 *
## Year2005      -0.37801    0.10299   -3.67  0.00025 ***
## Year2006      -0.31830    0.08841   -3.60  0.00033 ***
## Year2007      -0.31451    0.09275   -3.39  0.00072 ***
## Year2008      -0.18104    0.09638   -1.88  0.06057 .
## Year2009      -0.19312    0.09314   -2.07  0.03834 *
## Year2010      -0.28325    0.09823   -2.88  0.00400 **
## Year2011      -0.15902    0.08576   -1.85  0.06395 .
## Year2012      -0.24552    0.09347   -2.63  0.00872 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.251, Adjusted R-squared:  0.238
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 1172 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0371 0.8640 0.9490 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          7.85e-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 1.097 1 1.047
## LastAuthorFemale 1.124 1 1.060
## Year 1.130 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
## -1.7224 -0.3757 0.0266 0.3648 1.8962
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.55002 0.08510 18.21 < 2e-16 ***
## FirstAuthorFemale1 0.00534 0.03523 0.15 0.8796
## LastAuthorFemale1 -0.04872 0.04255 -1.15 0.2524
## Year1997 -0.13484 0.11367 -1.19 0.2357
## Year1998 -0.07445 0.13057 -0.57 0.5686
## Year1999 0.16709 0.11084 1.51 0.1320
## Year2000 -0.15296 0.10028 -1.53 0.1274
## Year2001 -0.18520 0.10922 -1.70 0.0902 .
## Year2002 -0.22120 0.09941 -2.23 0.0262 *
## Year2003 -0.43084 0.11025 -3.91 9.8e-05 ***
## Year2004 -0.27245 0.11895 -2.29 0.0222 *
## Year2005 -0.37807 0.12224 -3.09 0.0020 **
```

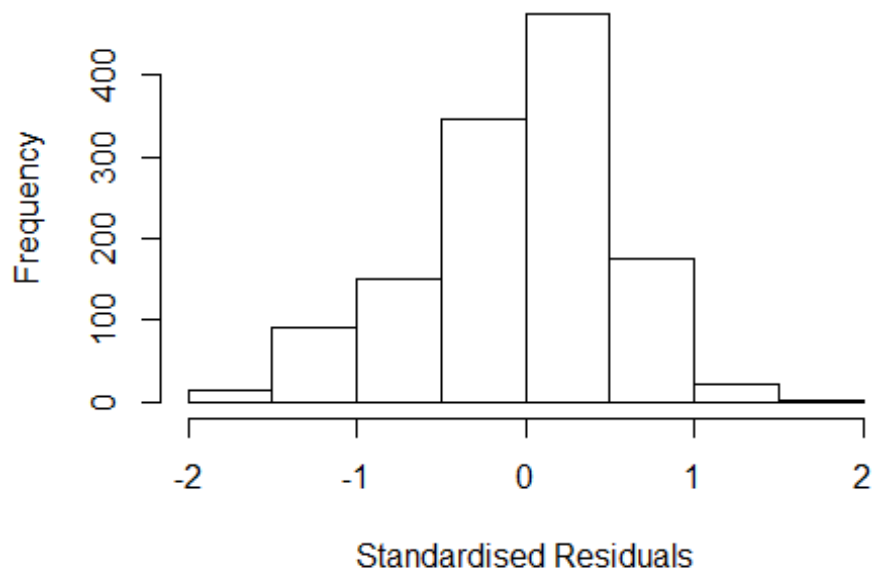


```

## Year2006      -0.31870    0.10638   -3.00   0.0028 **
## Year2007      -0.29264    0.10708   -2.73   0.0064 **
## Year2008      -0.17370    0.11022   -1.58   0.1153
## Year2009      -0.18581    0.10572   -1.76   0.0791 .
## Year2010      -0.28453    0.11293   -2.52   0.0119 *
## Year2011      -0.13701    0.09435   -1.45   0.1467
## Year2012      -0.20960    0.10206   -2.05   0.0402 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.525
## Multiple R-squared:  0.0582, Adjusted R-squared:  0.0447
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 1166 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.164  0.862  0.947   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      7.85e-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 1.053 1      1.026
## Year              1.053 16      1.002

```

Residuals from first author



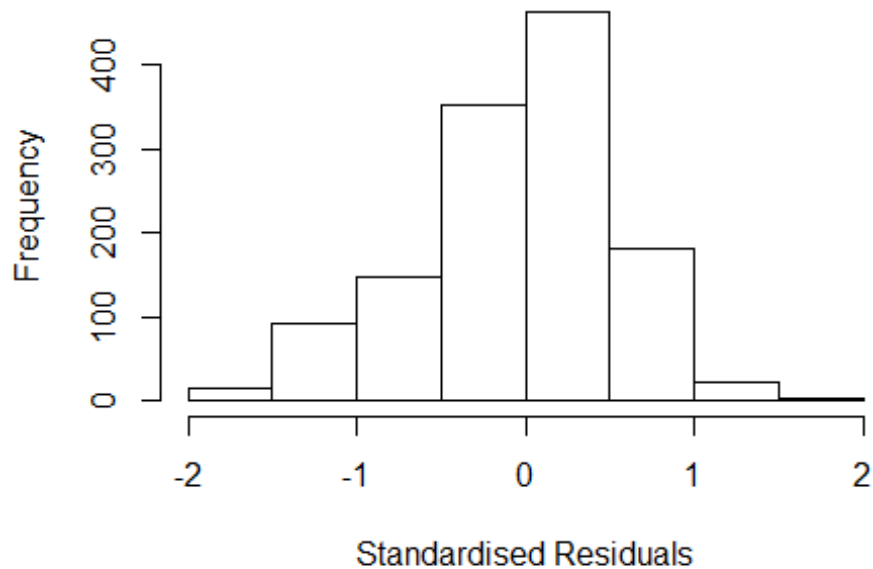
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7116 -0.3680 0.0319 0.3711 1.8538
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.54671 0.08538 18.11 < 2e-16 ***
## FirstAuthorFemale1 -0.00312 0.03473 -0.09 0.9285
## Year1997 -0.13587 0.11386 -1.19 0.2330
## Year1998 -0.07753 0.13101 -0.59 0.5541
## Year1999 0.16491 0.11132 1.48 0.1387
## Year2000 -0.15629 0.10045 -1.56 0.1200
## Year2001 -0.18358 0.10914 -1.68 0.0928 .
## Year2002 -0.22427 0.09980 -2.25 0.0248 *
## Year2003 -0.43404 0.11027 -3.94 8.7e-05 ***
## Year2004 -0.27771 0.11951 -2.32 0.0203 *
## Year2005 -0.38657 0.12274 -3.15 0.0017 **
## Year2006 -0.32322 0.10631 -3.04 0.0024 **
```

```

## Year2007          -0.30073    0.10700   -2.81    0.0050 **
## Year2008          -0.17656    0.11063   -1.60    0.1107
## Year2009          -0.19510    0.10496   -1.86    0.0633 .
## Year2010          -0.29005    0.11284   -2.57    0.0103 *
## Year2011          -0.14405    0.09439   -1.53    0.1272
## Year2012          -0.21960    0.10180   -2.16    0.0312 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.0571, Adjusted R-squared:  0.0443
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 111 weights are ~= 1. The remaining 1163 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.185  0.861  0.947  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      7.85e-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 1.079 1          1.039
## Year            1.079 16          1.002

```

Residuals from last author



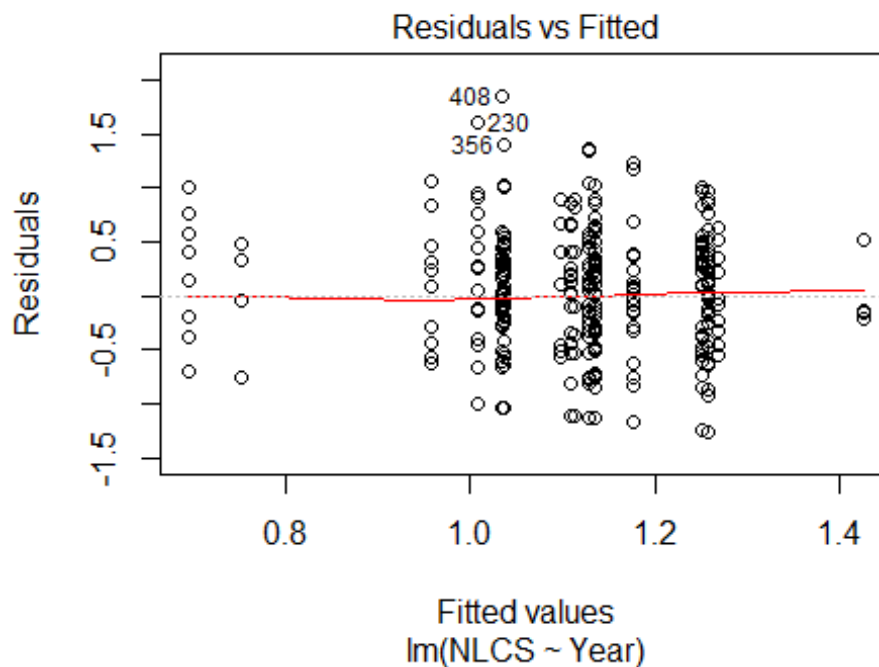
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7188 -0.3759 0.0262 0.3640 1.8943
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5514 0.0850 18.26 < 2e-16 ***
## LastAuthorFemale1 -0.0475 0.0417 -1.14 0.2542
## Year1997 -0.1350 0.1137 -1.19 0.2354
## Year1998 -0.0752 0.1305 -0.58 0.5647
## Year1999 0.1674 0.1107 1.51 0.1305
## Year2000 -0.1528 0.1002 -1.52 0.1276
## Year2001 -0.1851 0.1091 -1.70 0.0902 .
## Year2002 -0.2214 0.0994 -2.23 0.0261 *
## Year2003 -0.4307 0.1102 -3.91 9.8e-05 ***
## Year2004 -0.2718 0.1186 -2.29 0.0221 *
## Year2005 -0.3775 0.1218 -3.10 0.0020 **
## Year2006 -0.3182 0.1061 -3.00 0.0028 **
```

```

## Year2007          -0.2928      0.1071    -2.73    0.0064 **
## Year2008          -0.1734      0.1102    -1.57    0.1159
## Year2009          -0.1857      0.1057    -1.76    0.0792 .
## Year2010          -0.2838      0.1127    -2.52    0.0119 *
## Year2011          -0.1370      0.0944    -1.45    0.1468
## Year2012          -0.2092      0.1020    -2.05    0.0405 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.0582, Adjusted R-squared:  0.0454
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1165 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.164  0.862  0.947   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      7.85e-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: 1274"
## [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
##   15   13   18   16   13    8   31   20   16   27   31   35   55   34   39
## 2011 2012
##   45   34
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8    7   12   10    4    4   20   14   12   20   18   30   43   26   33
## 2011 2012

```

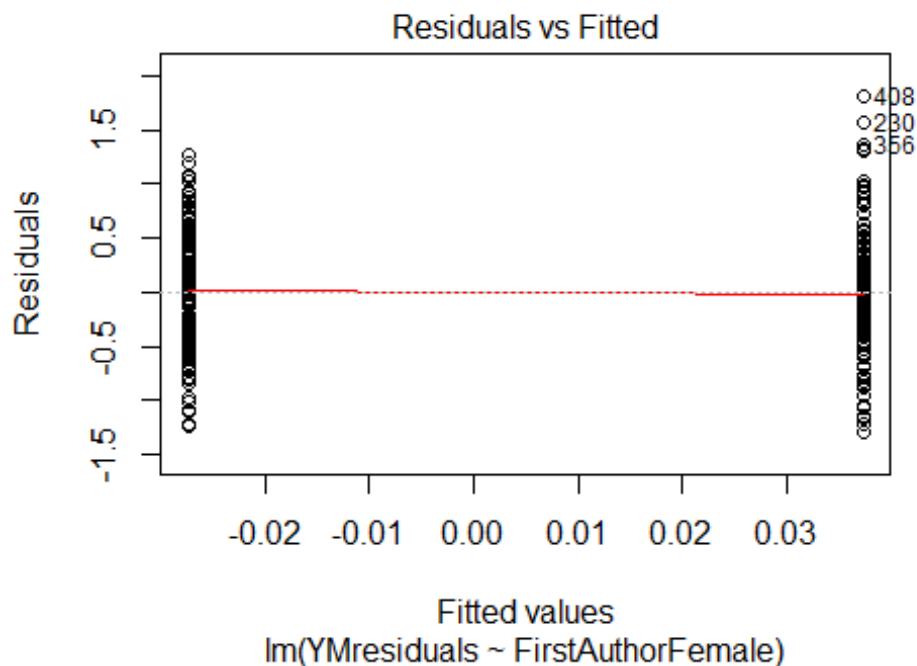
```
## 38 32
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 7 10 7 3 2 17 12 11 18 17 27 41 22 27
## 2011 2012
## 32 29
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 9.5, df = 16, p-value = 0.9
```



```
##
## 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: 3.78266777948139"
## [1] "Male first author team size 2018 geometric mean: 2.95486904972091"
## Warning in wilcox.test.default(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 = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.27312543726286"
## [1] "Male last author team size 2018 geometric mean: 3.39082414176222"

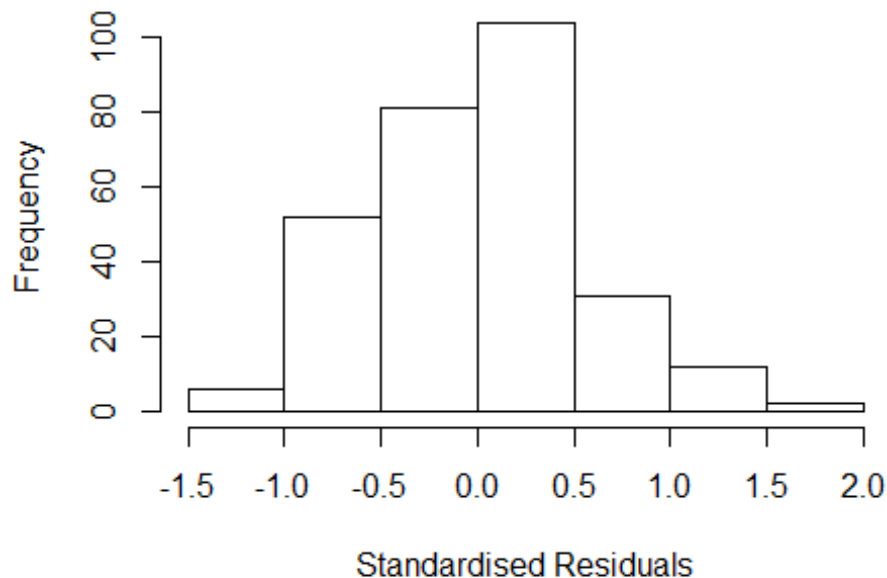
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, p-value = 0.7
## 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	1.490	1	1.221
LastAuthorFemale	1.301	1	1.141
UniqueAuthors	2.838	4	1.139
Year	3.221	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
## -1.4442 -0.3939 0.0163 0.3832 1.8711
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.86466 0.22007 3.93 0.00011 ***
## FirstAuthorFemale1 0.01067 0.08043 0.13 0.89455
## LastAuthorFemale1 0.02911 0.08065 0.36 0.71847
## UniqueAuthors2 0.08143 0.12132 0.67 0.50270
## UniqueAuthors3 0.34991 0.12163 2.88 0.00434 **
## UniqueAuthors4 0.26116 0.12685 2.06 0.04049 *
## UniqueAuthors5 0.42323 0.11147 3.80 0.00018 ***
## Year1997 0.02168 0.35293 0.06 0.95106
## Year1998 -0.35739 0.28275 -1.26 0.20735
## Year1999 0.03097 0.24769 0.13 0.90059
```

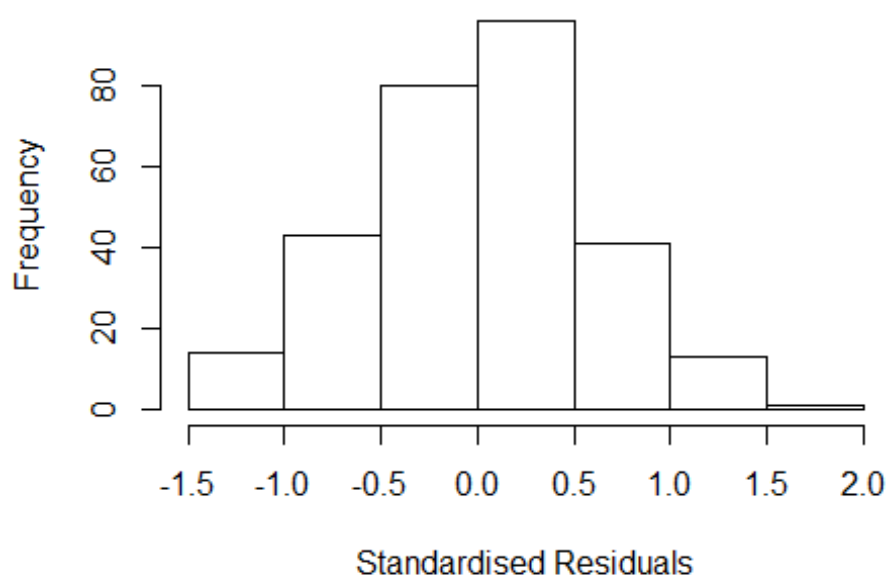


```

## Year2000      -0.39131      0.35970      -1.09      0.27763
## Year2001      0.54945      0.40297      1.36      0.17388
## Year2002      0.00153      0.26971      0.01      0.99548
## Year2003     -0.04395      0.26799     -0.16      0.86986
## Year2004     -0.11765      0.30050     -0.39      0.69574
## Year2005      0.15405      0.25805      0.60      0.55103
## Year2006     -0.17057      0.29307     -0.58      0.56105
## Year2007      0.11655      0.26168      0.45      0.65639
## Year2008     -0.07237      0.23739     -0.30      0.76073
## Year2009     -0.10641      0.24459     -0.44      0.66387
## Year2010      0.04256      0.26217      0.16      0.87115
## Year2011      0.10606      0.24960      0.42      0.67123
## Year2012      0.00449      0.24500      0.02      0.98538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.574
## Multiple R-squared:  0.144, Adjusted R-squared:  0.0725
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 264 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.266  0.867  0.951  0.912  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.47e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.340 1      1.158
## LastAuthorFemale  1.262 1      1.123
## Year              1.446 16      1.012

```

Residuals from first and last author

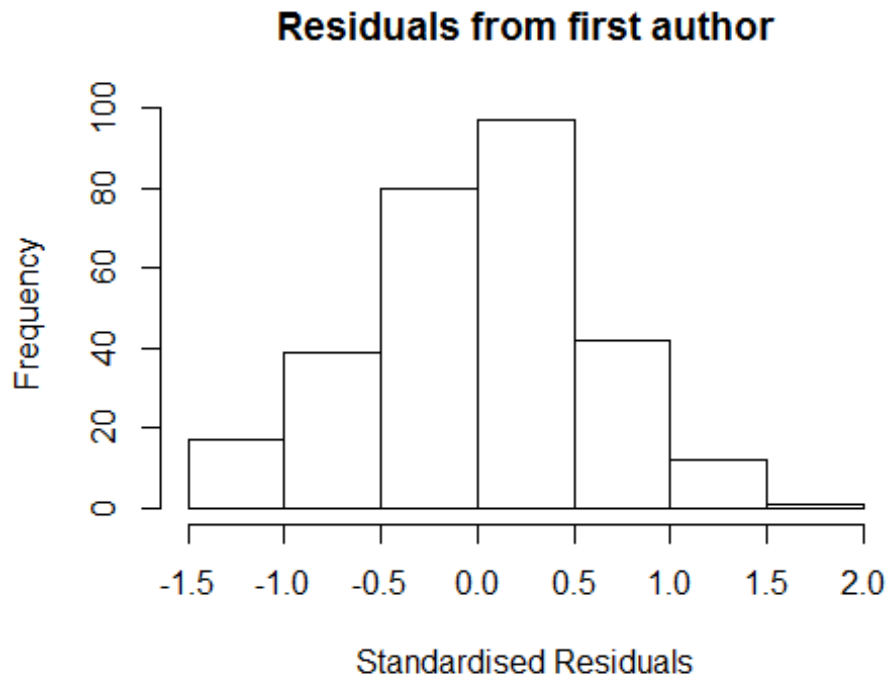


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2739 -0.3982 0.0503 0.3727 1.6770
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9094 0.2539 3.58 0.00041 ***
## FirstAuthorFemale1 0.0815 0.0813 1.00 0.31699
## LastAuthorFemale1 -0.0336 0.0846 -0.40 0.69195
## Year1997 0.2107 0.4061 0.52 0.60426
## Year1998 -0.3441 0.3190 -1.08 0.28171
## Year1999 0.1898 0.2847 0.67 0.50564
## Year2000 -0.3658 0.3842 -0.95 0.34187
## Year2001 0.7387 0.3580 2.06 0.04006 *
## Year2002 0.1612 0.2911 0.55 0.58017
## Year2003 0.1341 0.3082 0.44 0.66381
## Year2004 0.0365 0.3192 0.11 0.90911
## Year2005 0.2958 0.2915 1.01 0.31123
```

```

## Year2006          -0.0293      0.3280   -0.09   0.92892
## Year2007           0.2995      0.2873    1.04   0.29807
## Year2008           0.1034      0.2649    0.39   0.69670
## Year2009           0.0240      0.2731    0.09   0.92992
## Year2010           0.2097      0.2865    0.73   0.46483
## Year2011           0.3644      0.2789    1.31   0.19242
## Year2012           0.1922      0.2760    0.70   0.48665
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.576
## Multiple R-squared:  0.0755, Adjusted R-squared:  0.0136
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 267 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.377  0.874   0.952   0.907   0.982   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.47e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##           "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.202 1          1.096
## Year              1.202 16          1.006

```

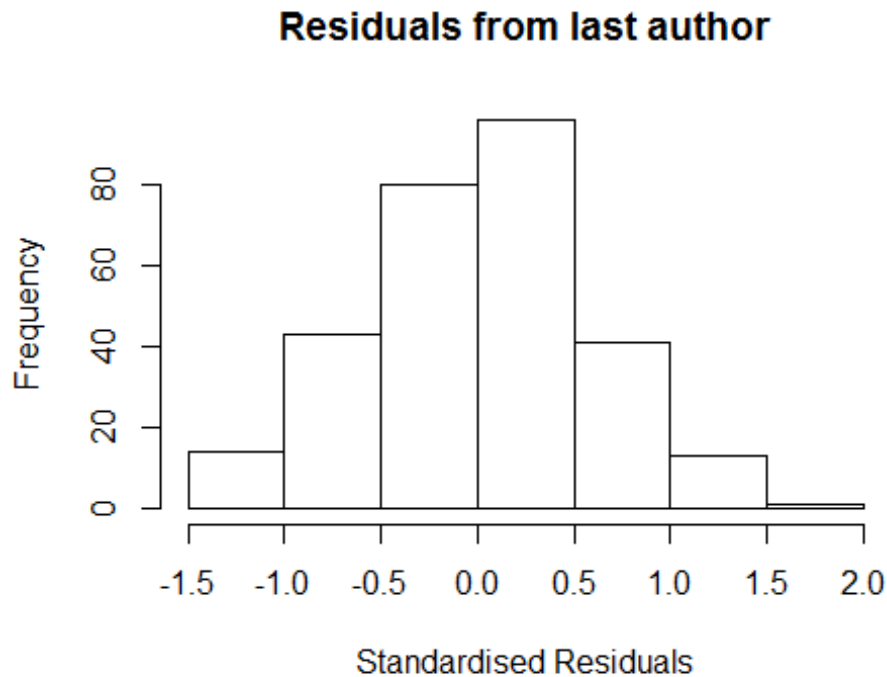


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2673 -0.3882  0.0455  0.3602  1.6626
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9074    0.2545   3.57 0.00043 ***
## FirstAuthorFemale1 0.0687    0.0776   0.88 0.37704
## Year1997        0.1984    0.3986   0.50 0.61905
## Year1998       -0.3436    0.3197  -1.07 0.28353
## Year1999        0.1904    0.2855   0.67 0.50530
## Year2000       -0.3717    0.3861  -0.96 0.33655
## Year2001        0.7071    0.3527   2.00 0.04599 *
## Year2002        0.1575    0.2915   0.54 0.58948
## Year2003        0.1332    0.3090   0.43 0.66671
## Year2004        0.0273    0.3181   0.09 0.93172
## Year2005        0.2876    0.2920   0.98 0.32562
## Year2006       -0.0337    0.3295  -0.10 0.91856
```

```

## Year2007          0.2912      0.2887      1.01  0.31410
## Year2008          0.0968      0.2666      0.36  0.71677
## Year2009          0.0206      0.2742      0.08  0.94005
## Year2010          0.2107      0.2875      0.73  0.46438
## Year2011          0.3580      0.2803      1.28  0.20256
## Year2012          0.1907      0.2770      0.69  0.49184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.585
## Multiple R-squared:  0.0743, Adjusted R-squared:  0.0161
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 264 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.399  0.877  0.953  0.909  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.47e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.144 1      1.070
## Year              1.144 16      1.004

```



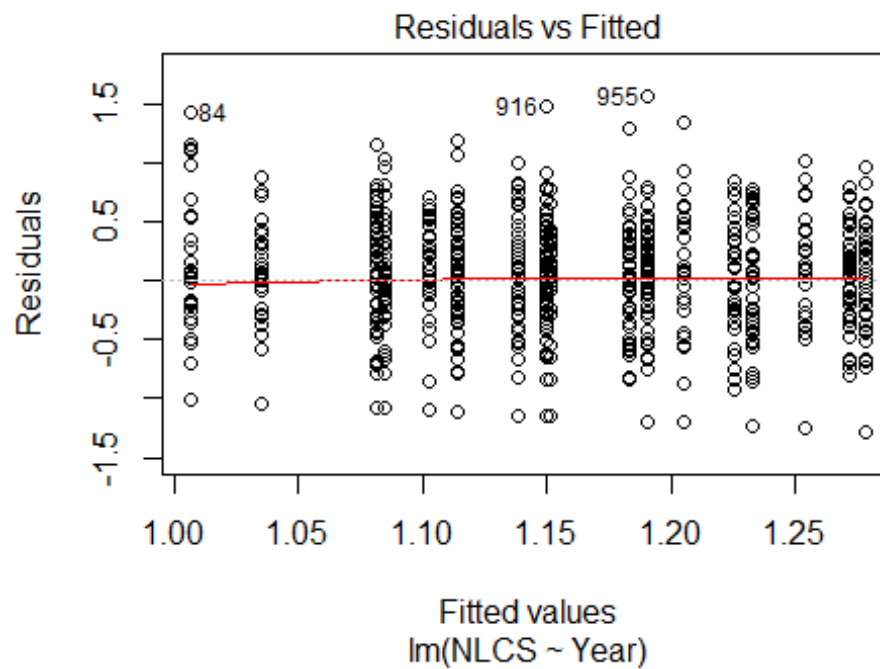
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.296 -0.406 0.032 0.367 1.699
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.916850 0.250626 3.66 0.00031 ***
## LastAuthorFemale1 0.000266 0.081032 0.00 0.99738
## Year1997 0.211325 0.404115 0.52 0.60145
## Year1998 -0.346251 0.316559 -1.09 0.27502
## Year1999 0.218996 0.285145 0.77 0.44315
## Year2000 -0.311986 0.377963 -0.83 0.40985
## Year2001 0.697384 0.353924 1.97 0.04981 *
## Year2002 0.163385 0.289057 0.57 0.57238
## Year2003 0.161503 0.303157 0.53 0.59465
## Year2004 0.043195 0.318122 0.14 0.89210
## Year2005 0.292337 0.287372 1.02 0.30993
## Year2006 -0.010806 0.329334 -0.03 0.97385
```

```

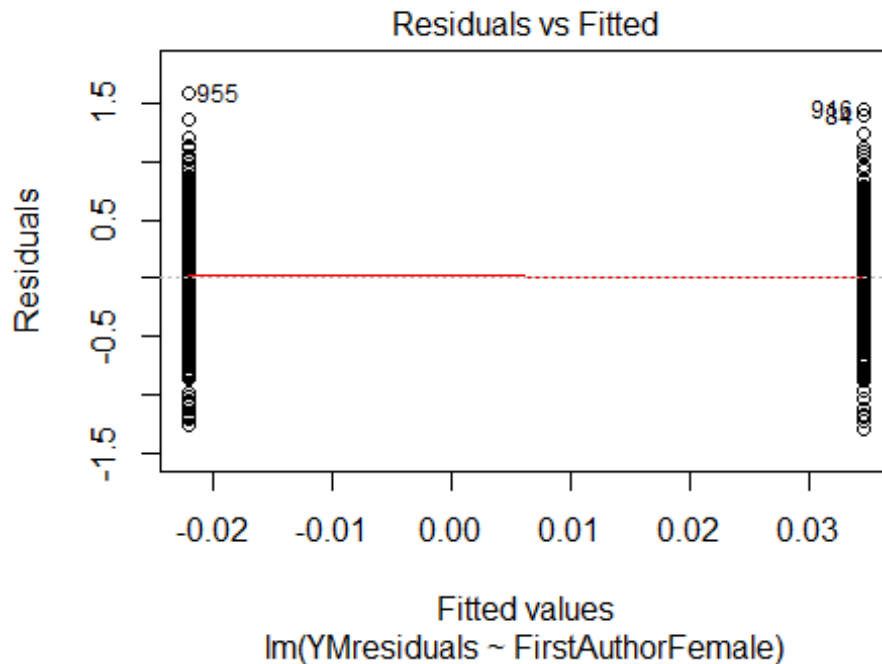
## Year2007      0.317622    0.283660    1.12  0.26383
## Year2008      0.111980    0.261442    0.43  0.66876
## Year2009      0.048926    0.268540    0.18  0.85557
## Year2010      0.226589    0.282685    0.80  0.42351
## Year2011      0.378794    0.275210    1.38  0.16984
## Year2012      0.208551    0.271586    0.77  0.44322
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.569
## Multiple R-squared:  0.0731, Adjusted R-squared:  0.0148
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 264 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.352  0.868  0.949  0.904  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.47e-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: 288"
## [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
##   58   72   70   45   62   67   51   57   58   63   50   80   85   95   75
## 2011 2012
##   80  113
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   37   41   28   42   33   33   45   39   46   30   58   66   74   61
## 2011 2012

```

```
## 58 87
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 29 35 37 24 37 30 29 31 33 38 27 47 53 65 54
## 2011 2012
## 54 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 = 22, df = 16, p-value = 0.1
```

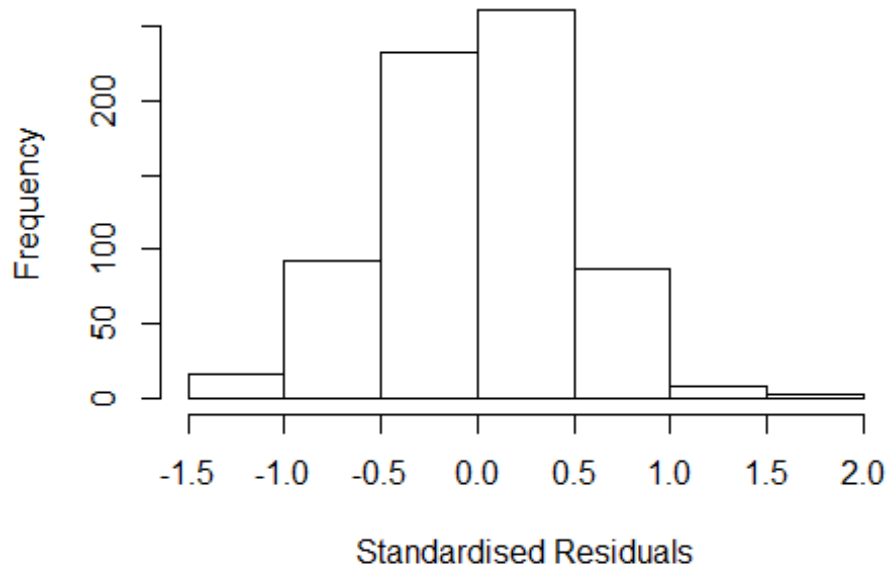


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.046, df = 1, p-value = 0.8
```

```
## [1] "Female first author team size 2018 geometric mean: 3.66743293561951"
## [1] "Male first author team size 2018 geometric mean: 3.52509957121223"
##
## 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: 3.56180711983092"
## [1] "Male last author team size 2018 geometric mean: 3.63299674902135"
##
## 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 1.137 1      1.066
## LastAuthorFemale  1.131 1      1.063
## UniqueAuthors    1.754 4      1.073
## Year              1.982 16     1.022
```

Residuals from first and last author and team size



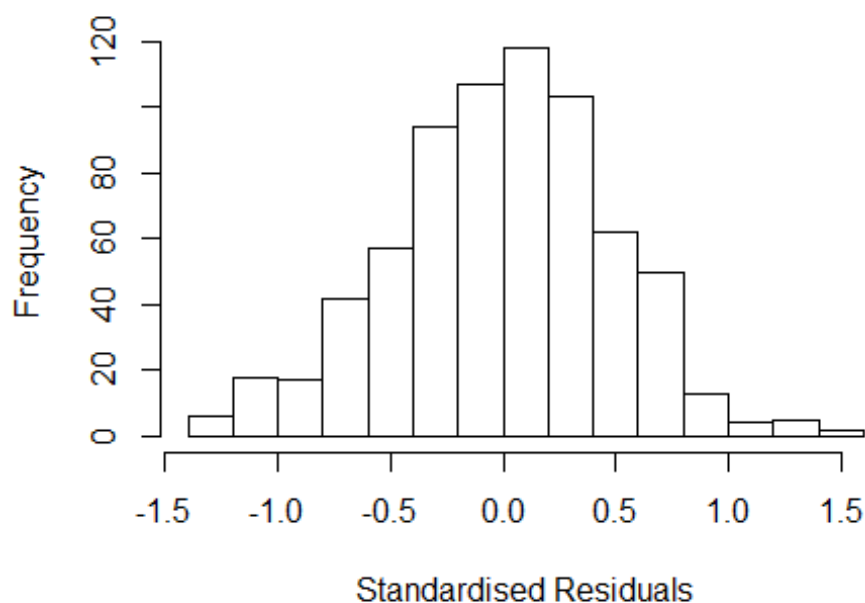
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.37609 -0.31541 0.00731 0.32307 1.87444
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.93955 0.12299 7.64 7.5e-14 ***
## FirstAuthorFemale1 0.01713 0.03923 0.44 0.66259
## LastAuthorFemale1 0.00232 0.04170 0.06 0.95570
## UniqueAuthors2 0.33673 0.08793 3.83 0.00014 ***
## UniqueAuthors3 0.26489 0.09216 2.87 0.00418 **
## UniqueAuthors4 0.30298 0.08857 3.42 0.00066 ***
## UniqueAuthors5 0.43422 0.08527 5.09 4.6e-07 ***
## Year1997 -0.25217 0.17384 -1.45 0.14735
## Year1998 -0.11411 0.14653 -0.78 0.43640
## Year1999 -0.11034 0.15208 -0.73 0.46835
```

```

## Year2000          0.03917    0.14188    0.28  0.78257
## Year2001          -0.20313    0.13597   -1.49  0.13567
## Year2002          -0.07211    0.14954   -0.48  0.62980
## Year2003          0.03421    0.13945    0.25  0.80630
## Year2004          -0.18506    0.13862   -1.33  0.18233
## Year2005          -0.11198    0.14295   -0.78  0.43367
## Year2006          0.06866    0.15809    0.43  0.66419
## Year2007          -0.06919    0.13250   -0.52  0.60173
## Year2008          -0.06299    0.12882   -0.49  0.62503
## Year2009          -0.18789    0.13488   -1.39  0.16405
## Year2010          -0.06967    0.12967   -0.54  0.59125
## Year2011          -0.08413    0.13654   -0.62  0.53800
## Year2012          -0.17548    0.13103   -1.34  0.18093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.464
## Multiple R-squared:  0.0957, Adjusted R-squared:  0.0663
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 633 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0651 0.8590 0.9490 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          1.43e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.097 1          1.047
## LastAuthorFemale 1.099 1          1.048
## Year          1.188 16          1.005

```

Residuals from first and last author



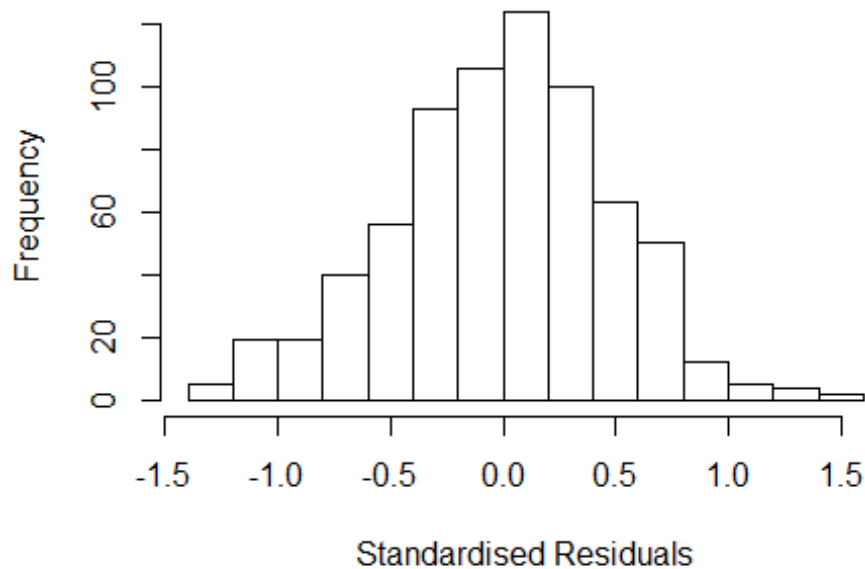
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3954 -0.3000  0.0175  0.3064  1.5660
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.185848   0.120312   9.86   <2e-16 ***
## FirstAuthorFemale1  0.052282   0.039508   1.32    0.19
## LastAuthorFemale1  0.015624   0.042362   0.37    0.71
## Year1997        -0.285327   0.178860  -1.60    0.11
## Year1998        -0.087689   0.149661  -0.59    0.56
## Year1999        -0.080225   0.150175  -0.53    0.59
## Year2000         0.065858   0.142644   0.46    0.64
## Year2001        -0.156939   0.144543  -1.09    0.28
## Year2002         0.011722   0.150144   0.08    0.94
## Year2003         0.094852   0.140796   0.67    0.50
## Year2004        -0.139722   0.142428  -0.98    0.33
## Year2005        -0.025034   0.145107  -0.17    0.86
```

```

## Year2006          0.141641    0.156561    0.90    0.37
## Year2007          -0.032813    0.140634   -0.23    0.82
## Year2008          -0.000808    0.133406   -0.01    1.00
## Year2009          -0.134773    0.135802   -0.99    0.32
## Year2010           0.002096    0.135281    0.02    0.99
## Year2011          -0.036670    0.140736   -0.26    0.79
## Year2012          -0.122679    0.134128   -0.91    0.36
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.472
## Multiple R-squared:  0.0388, Adjusted R-squared:  0.0133
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 630 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.249  0.861  0.949  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      1.43e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.084 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
## -1.3824 -0.2969  0.0161  0.3149  1.5621
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18759    0.12059   9.85  <2e-16 ***
## FirstAuthorFemale1  0.05405    0.03926   1.38    0.17
## Year1997      -0.28319    0.17864  -1.59    0.11
## Year1998      -0.08717    0.14948  -0.58    0.56
## Year1999      -0.08010    0.15011  -0.53    0.59
## Year2000       0.06400    0.14292   0.45    0.65
## Year2001      -0.15581    0.14464  -1.08    0.28
## Year2002       0.01171    0.15011   0.08    0.94
## Year2003       0.09575    0.14065   0.68    0.50
## Year2004      -0.13700    0.14180  -0.97    0.33
## Year2005      -0.02423    0.14487  -0.17    0.87
## Year2006       0.14081    0.15655   0.90    0.37
```

```

## Year2007          -0.03172    0.14037   -0.23    0.82
## Year2008          0.00127    0.13332    0.01    0.99
## Year2009         -0.13152    0.13522   -0.97    0.33
## Year2010          0.00531    0.13495    0.04    0.97
## Year2011         -0.03349    0.14004   -0.24    0.81
## Year2012         -0.11869    0.13304   -0.89    0.37
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.479
## Multiple R-squared:  0.038, Adjusted R-squared:  0.014
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 634 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.266  0.866  0.952  0.903  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.43e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.085 1          1.042
## Year              1.085 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
## -1.3651 -0.3063  0.0128  0.3243  1.5504

```

```

##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19452    0.12018    9.94  <2e-16 ***
## LastAuthorFemale1  0.02124    0.04208    0.50    0.61
## Year1997        -0.28031    0.17870   -1.57    0.12
## Year1998        -0.08572    0.15024   -0.57    0.57
## Year1999        -0.07170    0.15142   -0.47    0.64
## Year2000         0.08483    0.14209    0.60    0.55
## Year2001        -0.14789    0.14546   -1.02    0.31
## Year2002         0.02441    0.14977    0.16    0.87
## Year2003         0.11087    0.14031    0.79    0.43
## Year2004        -0.12920    0.14340   -0.90    0.37
## Year2005        -0.01781    0.14574   -0.12    0.90
## Year2006         0.14934    0.15712    0.95    0.34
## Year2007        -0.02796    0.14132   -0.20    0.84
## Year2008         0.00609    0.13362    0.05    0.96
## Year2009        -0.12791    0.13639   -0.94    0.35
## Year2010         0.01182    0.13560    0.09    0.93
## Year2011        -0.02366    0.14134   -0.17    0.87
## Year2012        -0.10995    0.13500   -0.81    0.42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.472
## Multiple R-squared:  0.0365, Adjusted R-squared:  0.0124
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 642 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.259  0.863   0.954   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           1.43e-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: 698"
## [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]"
##
## 1997 1998 2000 2002 2005 2007 2008 2010 2011 2012
##    1    3    2    2    2    3    3    2    1    1
##
## 1997 1998 2000 2002 2005 2007 2008 2010 2011 2012
##    0    3    1    1    1    3    2    2    0    1
##
## 1997 1998 2000 2002 2005 2007 2008 2010 2011 2012
##    0    2    1    1    1    3    2    2    0    1
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: NaN"
## [1] "Male first author team size 2018 geometric mean: NaN"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

## Warning in lmrob.S(x, y, control = control, mf = mf): S-estimated scale ==
## 0: Probably exact fit; check your data

## [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

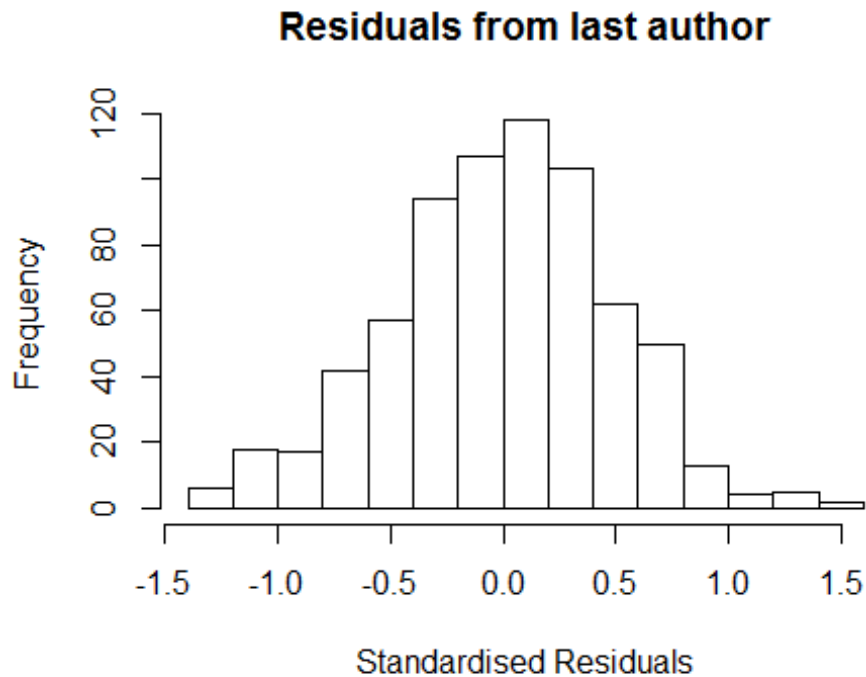
## [1] "Regression 3: First author gender, Year as factors"

## 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"

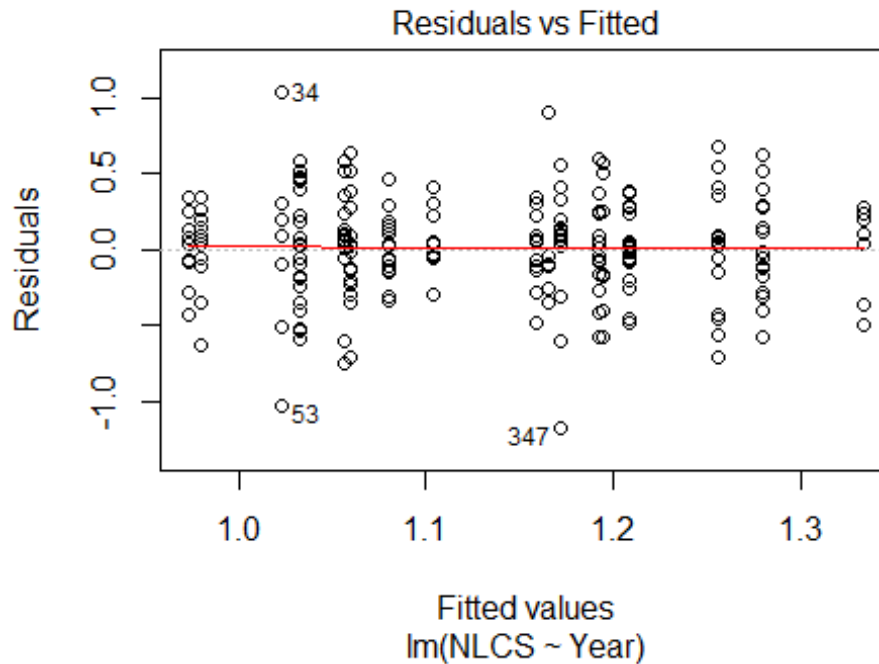
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

```



```
## [1] "Sample size for the above analysis: 13"
## [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
##    23    25    18    28    25    25    27    19    16    25    29    26    23    30    27
## 2011 2012
##    17    26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##     7     7     7    11     7    11    16    11     9    19    20    14    14    24    18
## 2011 2012
##    14    17
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##     7     7     7     7     5     7    16    11     9    18    19     9    12    21    16
## 2011 2012
##    13    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 = 29, df = 16, p-value = 0.03
```



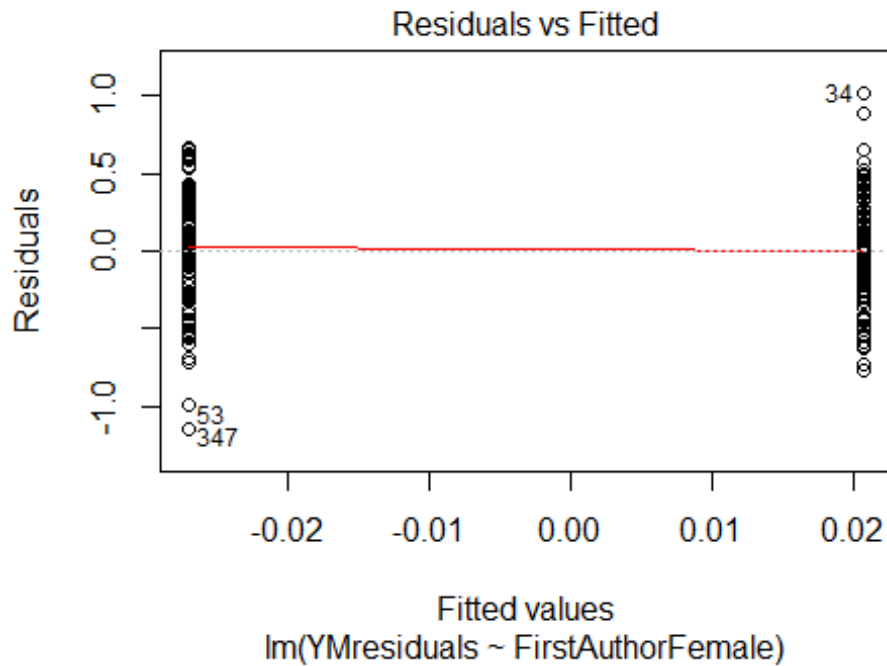
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.1, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 3.1748021039364"
## [1] "Male first author team size 2018 geometric mean: 2.21336383940064"

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

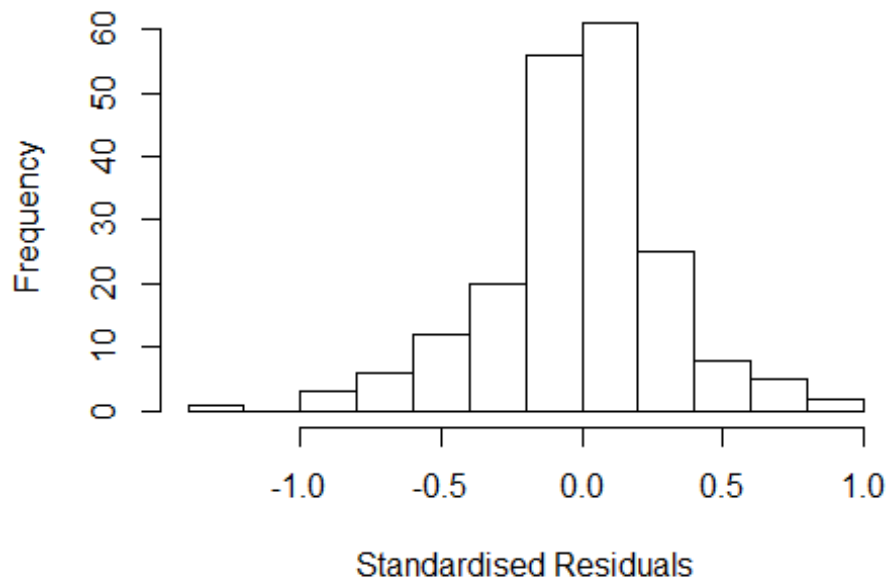
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.51984209978975"
## [1] "Male last author team size 2018 geometric mean: 2.85232327045476"

## Warning in wilcox.test.default(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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.493 1      1.579
## LastAuthorFemale  2.213 1      1.488
## UniqueAuthors    9.570 4      1.326
## Year             24.243 16      1.105
```

Residuals from first and last author and team size



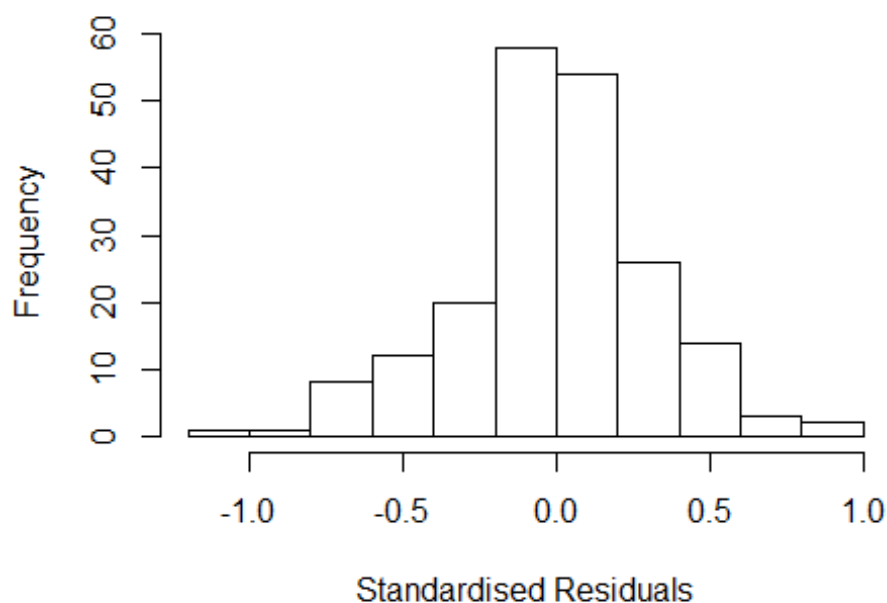
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2873 -0.1392 0.0102 0.1612 0.9390
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10157 0.30492 3.61 0.0004 ***
## FirstAuthorFemale1 -0.02524 0.05474 -0.46 0.6453
## LastAuthorFemale1 -0.10646 0.05660 -1.88 0.0617 .
## UniqueAuthors2 0.19675 0.14352 1.37 0.1722
## UniqueAuthors3 0.21820 0.14896 1.46 0.1448
## UniqueAuthors4 0.20090 0.13412 1.50 0.1359
## UniqueAuthors5 0.33666 0.13347 2.52 0.0125 *
## Year1997 -0.15091 0.30683 -0.49 0.6234
## Year1998 0.04629 0.29318 0.16 0.8747
## Year1999 -0.04587 0.28202 -0.16 0.8710
```

```

## Year2000      -0.28500    0.31506   -0.90    0.3669
## Year2001      -0.22594    0.27532   -0.82    0.4129
## Year2002      -0.24619    0.29269   -0.84    0.4014
## Year2003      -0.23579    0.27796   -0.85    0.3974
## Year2004      -0.30589    0.27918   -1.10    0.2747
## Year2005      -0.20612    0.27441   -0.75    0.4536
## Year2006      -0.10396    0.27452   -0.38    0.7054
## Year2007      -0.00872    0.27529   -0.03    0.9748
## Year2008      -0.11811    0.28239   -0.42    0.6763
## Year2009      -0.37779    0.29390   -1.29    0.2003
## Year2010       0.01868    0.28797    0.06    0.9483
## Year2011      -0.20931    0.28938   -0.72    0.4705
## Year2012      -0.02529    0.29587   -0.09    0.9320
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.243
## Multiple R-squared:  0.25,   Adjusted R-squared:  0.156
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## observation 14 is an outlier with |weight| = 0 ( < 0.0005);
## 21 weights are ~= 1. The remaining 177 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.103  0.807  0.953  0.860  0.989  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           5.03e-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 1.877 1 1.370
## LastAuthorFemale 1.409 1 1.187
## Year 2.458 16 1.028

```

Residuals from first and last author



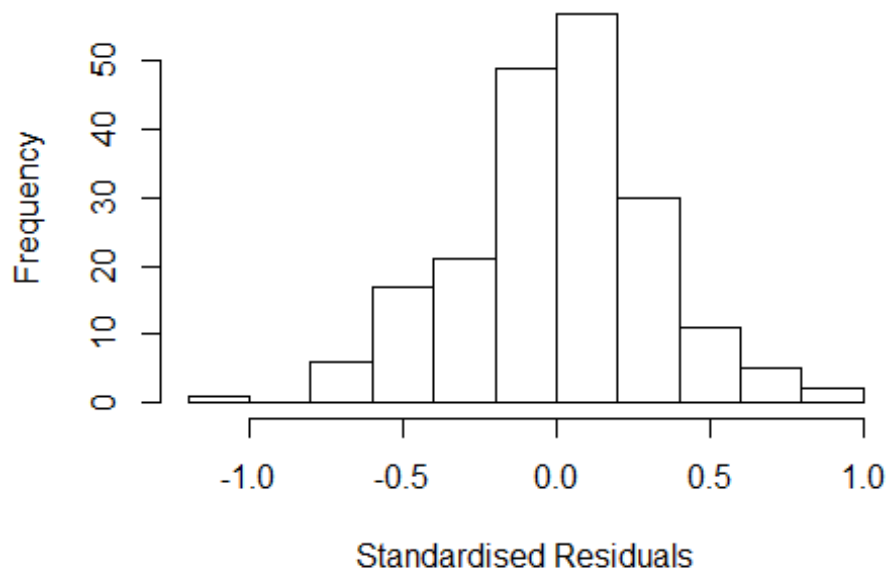
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.12e+00 -1.63e-01 -7.06e-05 1.73e-01 9.43e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27290 0.28958 4.40 1.9e-05 ***
## FirstAuthorFemale1 0.00302 0.05274 0.06 0.954
## LastAuthorFemale1 -0.11253 0.05028 -2.24 0.026 *
## Year1997 -0.15655 0.33278 -0.47 0.639
## Year1998 0.10518 0.30150 0.35 0.728
## Year1999 -0.06836 0.29604 -0.23 0.818
## Year2000 -0.15187 0.31405 -0.48 0.629
## Year2001 -0.13870 0.28500 -0.49 0.627
## Year2002 -0.17370 0.28836 -0.60 0.548
## Year2003 -0.22279 0.28780 -0.77 0.440
## Year2004 -0.26938 0.28412 -0.95 0.344
## Year2005 -0.15863 0.27736 -0.57 0.568
```

```

## Year2006      -0.02641    0.27778   -0.10    0.924
## Year2007      0.07820    0.27942    0.28    0.780
## Year2008     -0.07142    0.28760   -0.25    0.804
## Year2009     -0.29485    0.29553   -1.00    0.320
## Year2010      0.08159    0.28449    0.29    0.775
## Year2011     -0.13776    0.29504   -0.47    0.641
## Year2012      0.02648    0.29901    0.09    0.930
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.276
## Multiple R-squared:  0.161, Adjusted R-squared:  0.0767
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 172 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0658 0.8240 0.9500 0.8740 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      5.03e-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 1.855 1      1.362
## Year      1.855 16      1.019

```


Residuals from first author



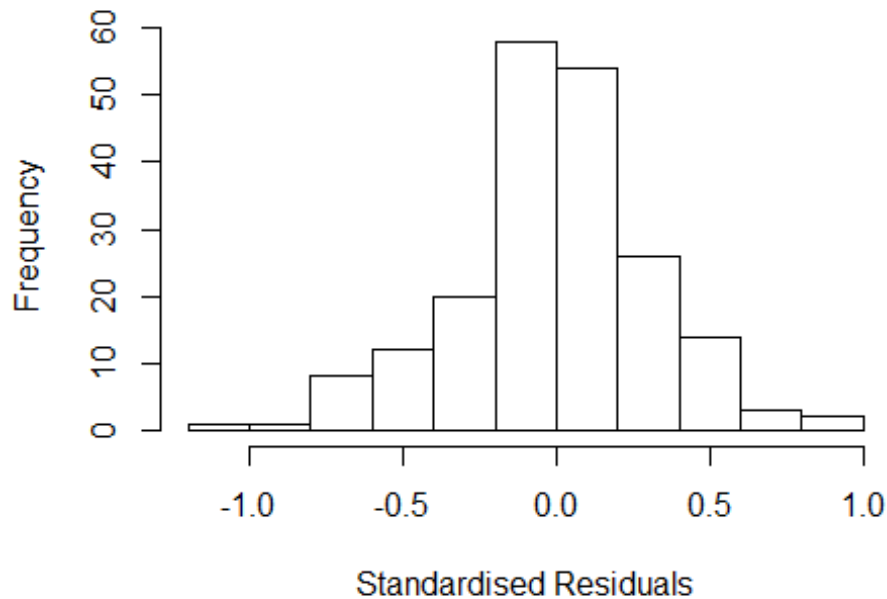
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0584 -0.1642 0.0194 0.1841 0.9992
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17835 0.23326 5.05 1.1e-06 ***
## FirstAuthorFemale1 0.00135 0.05440 0.02 0.98
## Year1997 -0.11993 0.34399 -0.35 0.73
## Year1998 0.17741 0.25579 0.69 0.49
## Year1999 -0.04486 0.25045 -0.18 0.86
## Year2000 -0.08157 0.26677 -0.31 0.76
## Year2001 -0.09543 0.24464 -0.39 0.70
## Year2002 -0.12214 0.24466 -0.50 0.62
## Year2003 -0.17488 0.24059 -0.73 0.47
## Year2004 -0.20076 0.24088 -0.83 0.41
## Year2005 -0.10256 0.23253 -0.44 0.66
## Year2006 0.01359 0.23362 0.06 0.95
```

```

## Year2007          0.14825    0.23134    0.64    0.52
## Year2008          -0.00625    0.24405   -0.03    0.98
## Year2009          -0.21277    0.24575   -0.87    0.39
## Year2010          0.11330    0.24614    0.46    0.65
## Year2011          -0.09605    0.25209   -0.38    0.70
## Year2012          0.08890    0.25675    0.35    0.73
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.288
## Multiple R-squared:  0.129, Adjusted R-squared:  0.0469
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 183 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.148  0.844  0.954  0.888  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      5.03e-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 1.4 1          1.183
## Year              1.4 16          1.011

```

Residuals from last author



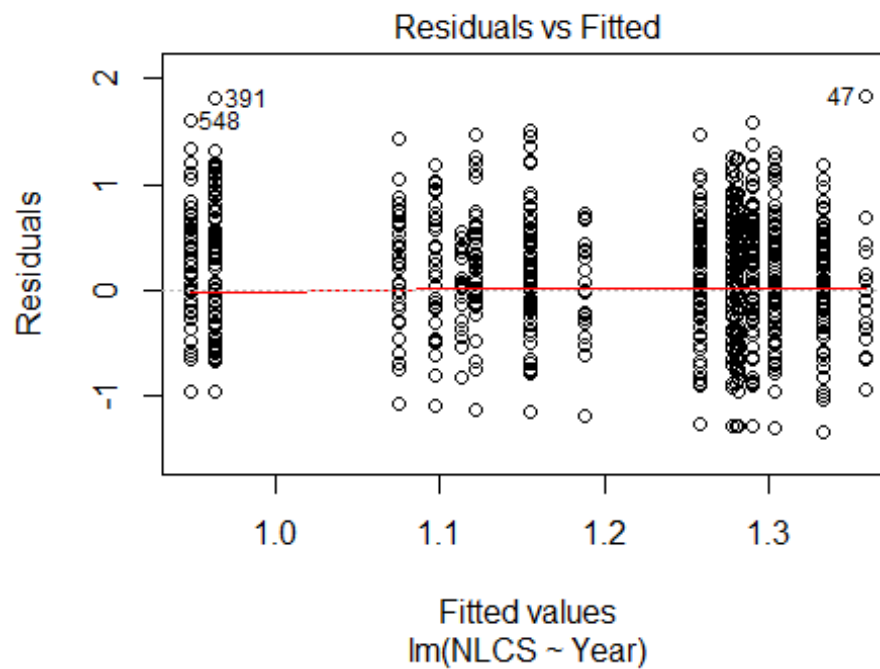
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.117702 -0.161511 0.000615 0.174752 0.946356
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2744 0.2832 4.50 1.2e-05 ***
## LastAuthorFemale1 -0.1123 0.0509 -2.21 0.029 *
## Year1997 -0.1567 0.3322 -0.47 0.638
## Year1998 0.1051 0.3019 0.35 0.728
## Year1999 -0.0680 0.2971 -0.23 0.819
## Year2000 -0.1538 0.3012 -0.51 0.610
## Year2001 -0.1389 0.2850 -0.49 0.627
## Year2002 -0.1742 0.2872 -0.61 0.545
## Year2003 -0.2228 0.2882 -0.77 0.440
## Year2004 -0.2685 0.2863 -0.94 0.350
## Year2005 -0.1577 0.2799 -0.56 0.574
## Year2006 -0.0265 0.2776 -0.10 0.924
```

```

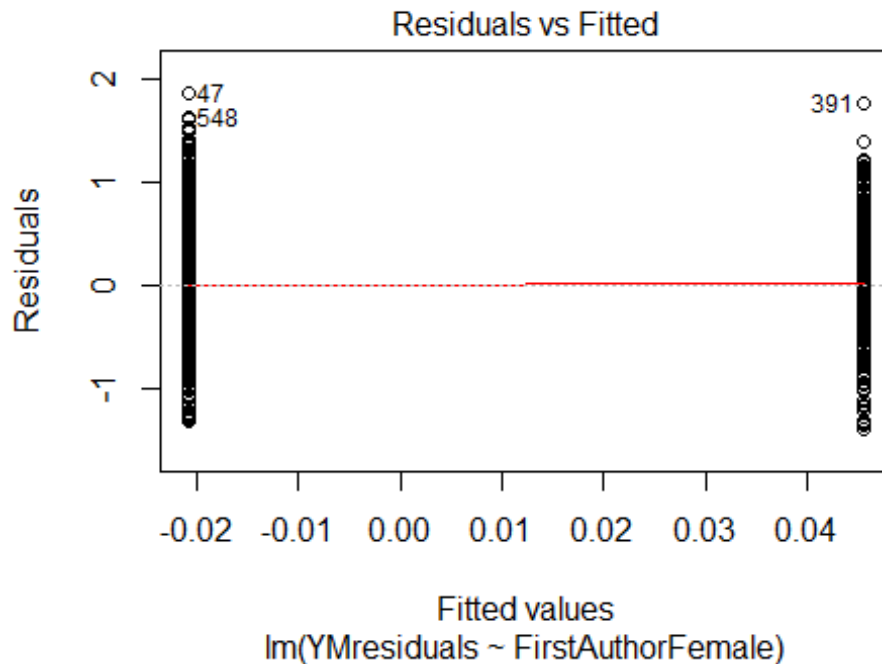
## Year2007          0.0786      0.2808      0.28      0.780
## Year2008          -0.0712     0.2882     -0.25     0.805
## Year2009          -0.2947     0.2958     -1.00     0.320
## Year2010           0.0819     0.2853      0.29     0.775
## Year2011          -0.1383     0.2944     -0.47     0.639
## Year2012           0.0273     0.3010      0.09     0.928
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.277
## Multiple R-squared:  0.161, Adjusted R-squared:  0.0818
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 171 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.066  0.825   0.949   0.874   0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.03e-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: 199"
## [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
##   42   38   34   86   79   56   82   80   66   59   93  113   89  112  105
## 2011 2012
##  119  116
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19   17   22   56   47   45   70   71   52   51   79   96   79   92   89
## 2011 2012

```

```
## 99 98
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 13 19 48 41 44 65 66 47 48 68 83 60 80 81
## 2011 2012
## 88 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 = 15, df = 16, p-value = 0.5
```

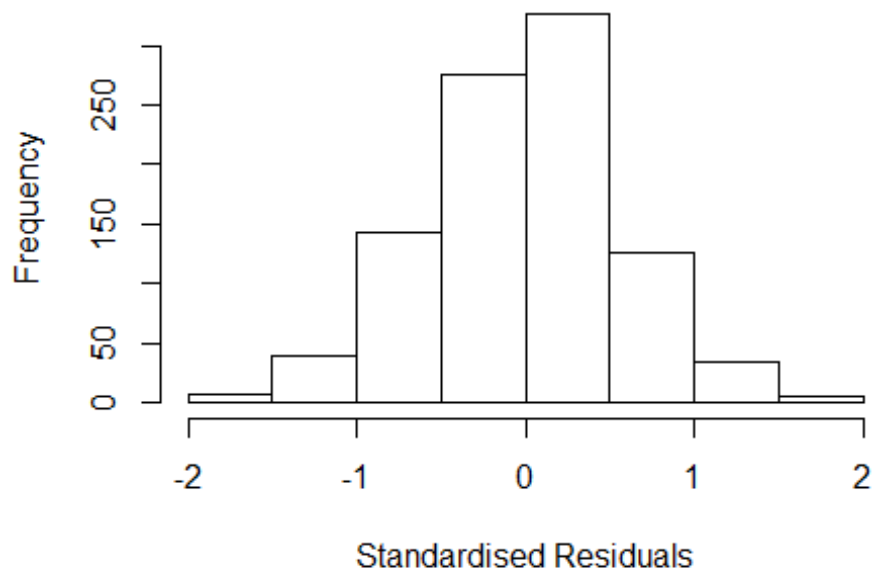


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.2, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.26565886565833"
## [1] "Male first author team size 2018 geometric mean: 4.68275943966015"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2800, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.20466626201674"
## [1] "Male last author team size 2018 geometric mean: 4.70447482139806"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2600, 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 1.182 1      1.087
## LastAuthorFemale  1.153 1      1.074
## UniqueAuthors    1.894 4      1.083
## Year              2.114 16     1.024
```

Residuals from first and last author and team size



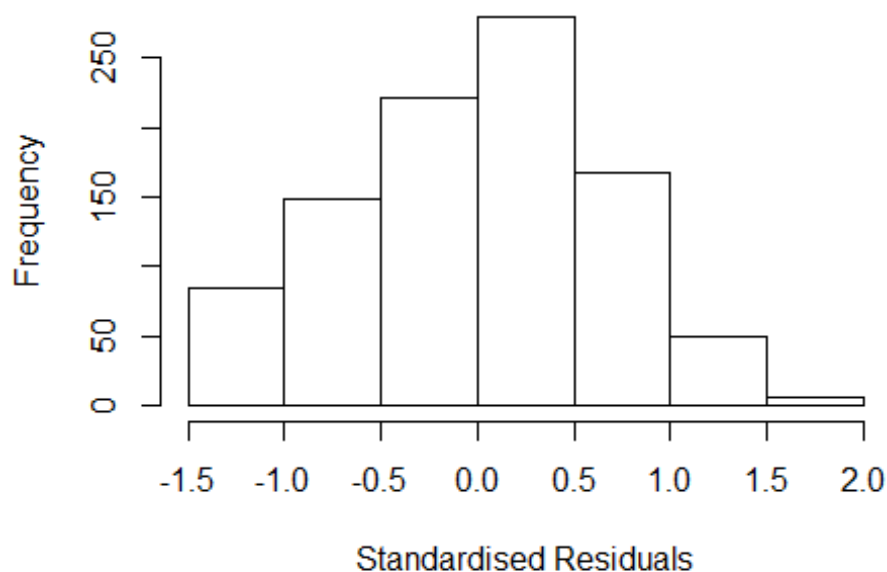
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6599 -0.4096 0.0134 0.3834 1.8692
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.42290 0.10951 3.86 0.00012 ***
## FirstAuthorFemale1 -0.00737 0.04135 -0.18 0.85865
## LastAuthorFemale1 0.02187 0.04496 0.49 0.62674
## UniqueAuthors2 0.47371 0.06812 6.95 6.7e-12 ***
## UniqueAuthors3 0.70288 0.07259 9.68 < 2e-16 ***
## UniqueAuthors4 0.86956 0.07262 11.97 < 2e-16 ***
## UniqueAuthors5 1.03376 0.06585 15.70 < 2e-16 ***
## Year1997 0.47359 0.22600 2.10 0.03639 *
## Year1998 0.05650 0.14161 0.40 0.69002
## Year1999 0.11319 0.12902 0.88 0.38053
```

```

## Year2000          0.14876      0.13472      1.10  0.26977
## Year2001         -0.00984      0.12174     -0.08  0.93558
## Year2002          0.01454      0.12323      0.12  0.90612
## Year2003         -0.07452      0.11793     -0.63  0.52759
## Year2004          0.01465      0.15306      0.10  0.92376
## Year2005          0.11910      0.12668      0.94  0.34740
## Year2006          0.18485      0.12228      1.51  0.13095
## Year2007          0.20528      0.11670      1.76  0.07888 .
## Year2008          0.20327      0.12556      1.62  0.10580
## Year2009          0.20188      0.11689      1.73  0.08449 .
## Year2010          0.09459      0.11919      0.79  0.42764
## Year2011          0.04489      0.11837      0.38  0.70458
## Year2012          0.08999      0.11498      0.78  0.43403
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.314, Adjusted R-squared:  0.298
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 875 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.864  0.942  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      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  1.086 1      1.042
## Year              1.190 16      1.005

```


Residuals from first and last author



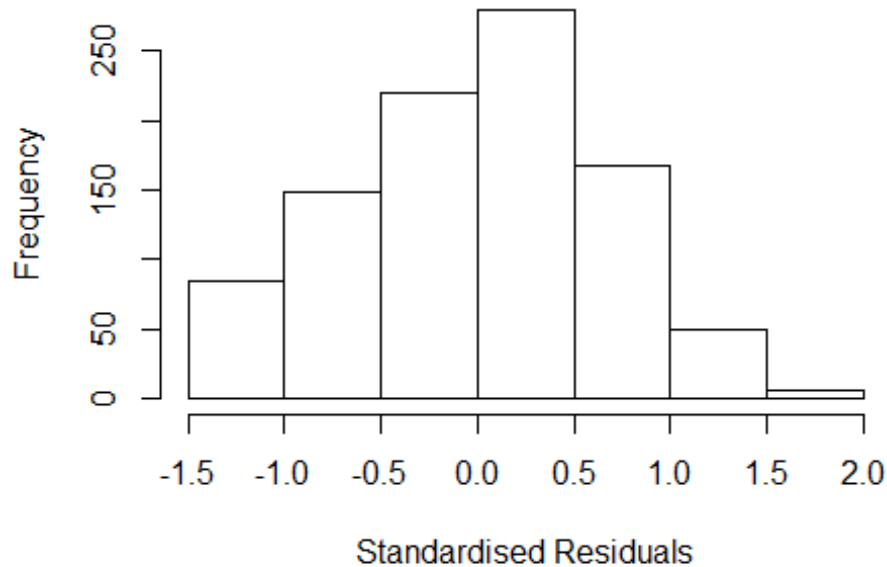
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4439 -0.4825 0.0334 0.4690 1.8728
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0272 0.1030 9.97 <2e-16 ***
## FirstAuthorFemale1 0.0639 0.0489 1.31 0.192
## LastAuthorFemale1 0.0267 0.0542 0.49 0.622
## Year1997 0.2920 0.2003 1.46 0.145
## Year1998 0.1421 0.1543 0.92 0.358
## Year1999 -0.1981 0.1395 -1.42 0.156
## Year2000 0.0134 0.1604 0.08 0.933
## Year2001 0.0308 0.1555 0.20 0.843
## Year2002 -0.0896 0.1413 -0.63 0.526
## Year2003 -0.1483 0.1352 -1.10 0.273
## Year2004 0.0693 0.1586 0.44 0.662
## Year2005 0.2499 0.1384 1.81 0.071 .
```

```

## Year2006          0.2024      0.1303      1.55      0.121
## Year2007          0.2989      0.1235      2.42      0.016 *
## Year2008          0.3261      0.1359      2.40      0.017 *
## Year2009          0.2471      0.1303      1.90      0.058 .
## Year2010          0.2367      0.1292      1.83      0.067 .
## Year2011          0.1147      0.1294      0.89      0.376
## Year2012          0.2739      0.1216      2.25      0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.672
## Multiple R-squared:  0.0586, Adjusted R-squared:  0.0406
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 868 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.418  0.866  0.946  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.05e-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 1.1 1      1.049
## Year              1.1 16      1.003

```

Residuals from first author



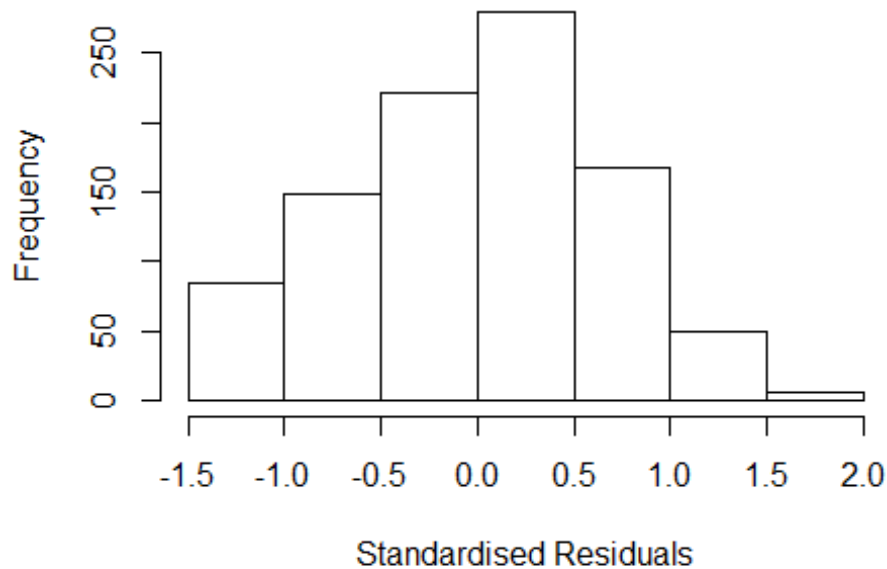
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4241 -0.4889 0.0346 0.4639 1.8686
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0300 0.1029 10.01 <2e-16 ***
## FirstAuthorFemale1 0.0673 0.0489 1.38 0.169
## Year1997 0.2934 0.1993 1.47 0.141
## Year1998 0.1475 0.1544 0.96 0.340
## Year1999 -0.1967 0.1396 -1.41 0.159
## Year2000 0.0135 0.1604 0.08 0.933
## Year2001 0.0334 0.1553 0.22 0.830
## Year2002 -0.0894 0.1413 -0.63 0.527
## Year2003 -0.1461 0.1352 -1.08 0.280
## Year2004 0.0728 0.1584 0.46 0.646
## Year2005 0.2517 0.1382 1.82 0.069 .
## Year2006 0.2046 0.1306 1.57 0.117
```

```

## Year2007          0.2982      0.1236      2.41      0.016 *
## Year2008          0.3268      0.1360      2.40      0.016 *
## Year2009          0.2486      0.1302      1.91      0.057 .
## Year2010          0.2389      0.1293      1.85      0.065 .
## Year2011          0.1208      0.1282      0.94      0.346
## Year2012          0.2780      0.1216      2.29      0.022 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.672
## Multiple R-squared:  0.0582, Adjusted R-squared:  0.0412
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 872 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.420  0.865   0.947   0.910   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.05e-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 1.083 1          1.041
## Year            1.083 16          1.003

```

Residuals from last author



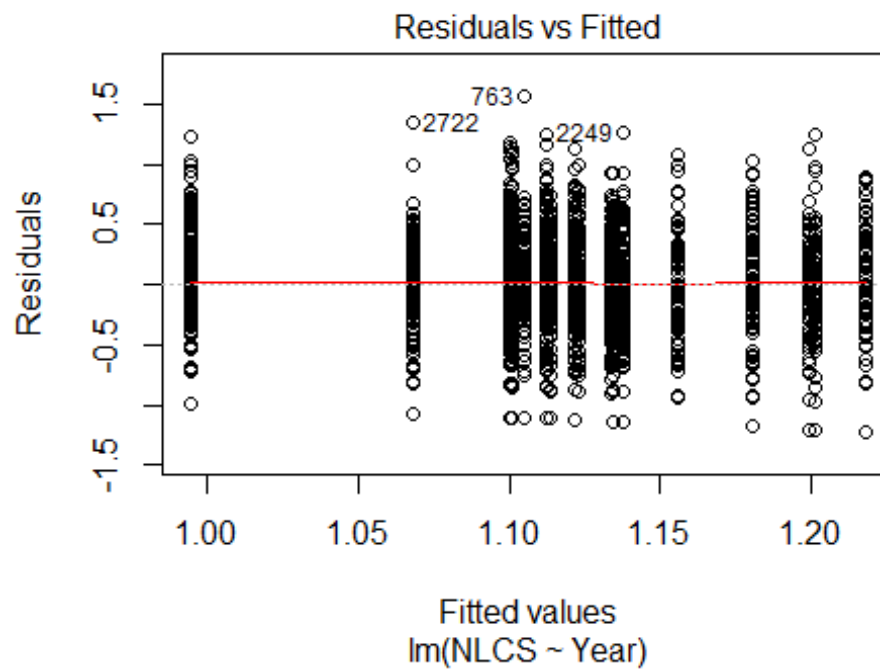
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4105 -0.4885 0.0304 0.4639 1.8651
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04431 0.10374 10.07 <2e-16 ***
## LastAuthorFemale1 0.03651 0.05374 0.68 0.497
## Year1997 0.28262 0.20182 1.40 0.162
## Year1998 0.13197 0.15710 0.84 0.401
## Year1999 -0.20527 0.14103 -1.46 0.146
## Year2000 0.00684 0.16213 0.04 0.966
## Year2001 0.03215 0.15786 0.20 0.839
## Year2002 -0.09159 0.14357 -0.64 0.524
## Year2003 -0.15599 0.13603 -1.15 0.252
## Year2004 0.05843 0.15905 0.37 0.713
## Year2005 0.24884 0.13984 1.78 0.075 .
## Year2006 0.19970 0.13160 1.52 0.129
```

```

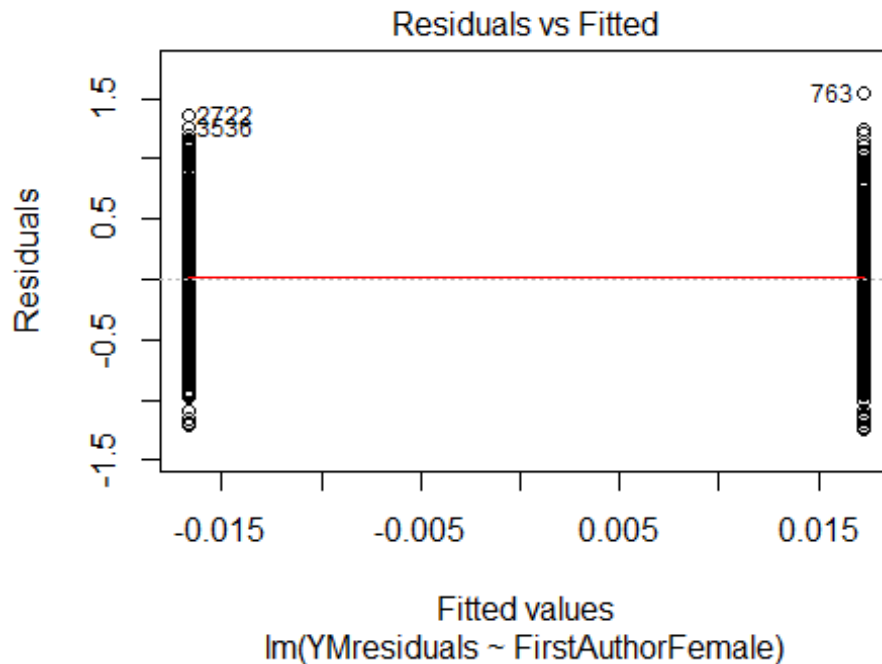
## Year2007      0.30182    0.12520    2.41    0.016 *
## Year2008      0.32973    0.13662    2.41    0.016 *
## Year2009      0.25746    0.13146    1.96    0.050 .
## Year2010      0.24309    0.13018    1.87    0.062 .
## Year2011      0.12224    0.12989    0.94    0.347
## Year2012      0.27645    0.12309    2.25    0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.672
## Multiple R-squared:  0.0571, Adjusted R-squared:  0.04
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 94 weights are ~= 1. The remaining 862 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.421  0.863  0.947  0.909  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.05e-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: 956"
## [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
## 165 186 149 142 198 197 180 180 181 202 250 273 277 306 314
## 2011 2012
## 383 372
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 83 75 82 76 97 94 110 123 124 128 158 177 200 203 213
## 2011 2012

```

```
## 250 236
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 72 65 72 62 85 78 93 105 106 106 136 158 175 183 187
## 2011 2012
## 216 214
## [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
```

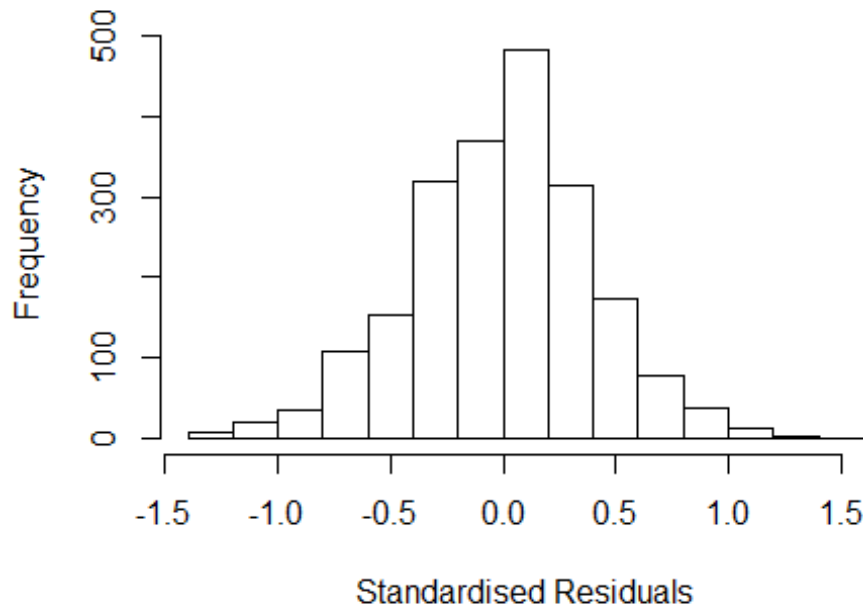


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.057, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 5.12379727354343"
## [1] "Male first author team size 2018 geometric mean: 4.77402337226638"
##
## 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] "Female last author team size 2018 geometric mean: 4.89795257961349"
## [1] "Male last author team size 2018 geometric mean: 5.02459139667299"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4800, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.046 1      1.023
## LastAuthorFemale  1.083 1      1.041
## UniqueAuthors    1.226 4      1.026
## Year              1.207 16     1.006
```


Residuals from first and last author and team size



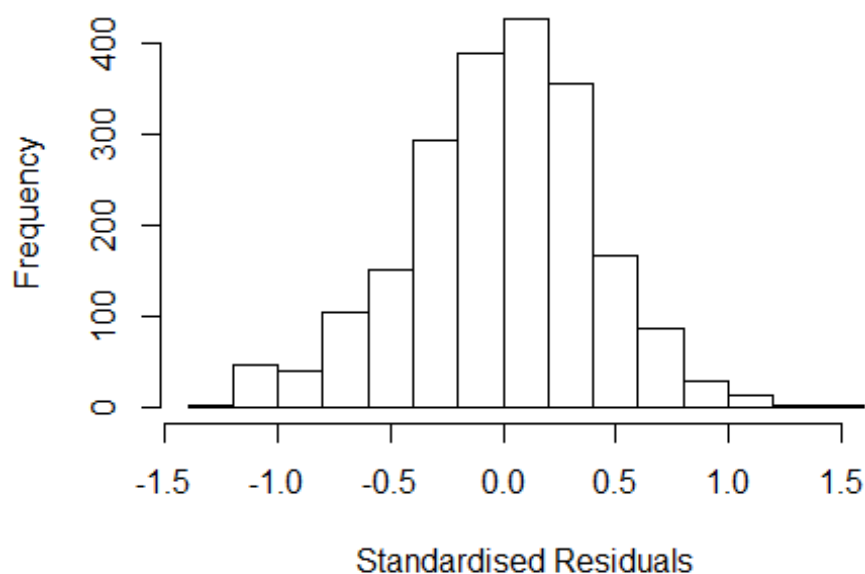
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3984 -0.2625  0.0194  0.2402  1.5127
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8730    0.0848   10.30 < 2e-16 ***
## FirstAuthorFemale1 -0.0384    0.0179   -2.15  0.032 *
## LastAuthorFemale1 -0.0480    0.0197   -2.44  0.015 *
## UniqueAuthors2     0.3157    0.0685    4.61 4.3e-06 ***
## UniqueAuthors3     0.3282    0.0662    4.96 7.8e-07 ***
## UniqueAuthors4     0.3609    0.0665    5.42 6.5e-08 ***
## UniqueAuthors5     0.4783    0.0653    7.32 3.4e-13 ***
## Year1997          0.0472    0.0853    0.55  0.580
## Year1998          0.0225    0.0756    0.30  0.766
## Year1999         -0.0816    0.0792   -1.03  0.303
```

```

## Year2000          -0.0805      0.0727   -1.11    0.268
## Year2001          -0.0895      0.0780   -1.15    0.252
## Year2002           0.0111      0.0709    0.16    0.876
## Year2003          -0.1102      0.0692   -1.59    0.112
## Year2004          -0.1415      0.0726   -1.95    0.051 .
## Year2005          -0.1037      0.0721   -1.44    0.151
## Year2006          -0.1501      0.0686   -2.19    0.029 *
## Year2007          -0.0662      0.0700   -0.95    0.344
## Year2008          -0.0934      0.0681   -1.37    0.170
## Year2009          -0.1053      0.0695   -1.52    0.130
## Year2010          -0.1169      0.0715   -1.64    0.102
## Year2011          -0.0820      0.0685   -1.20    0.231
## Year2012          -0.1871      0.0680   -2.75    0.006 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0998, Adjusted R-squared:  0.0904
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 183 weights are ~= 1. The remaining 1930 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0772 0.8610 0.9500 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.73e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.041 1 1.020
## LastAuthorFemale 1.034 1 1.017
## Year 1.043 16 1.001

```

Residuals from first and last author



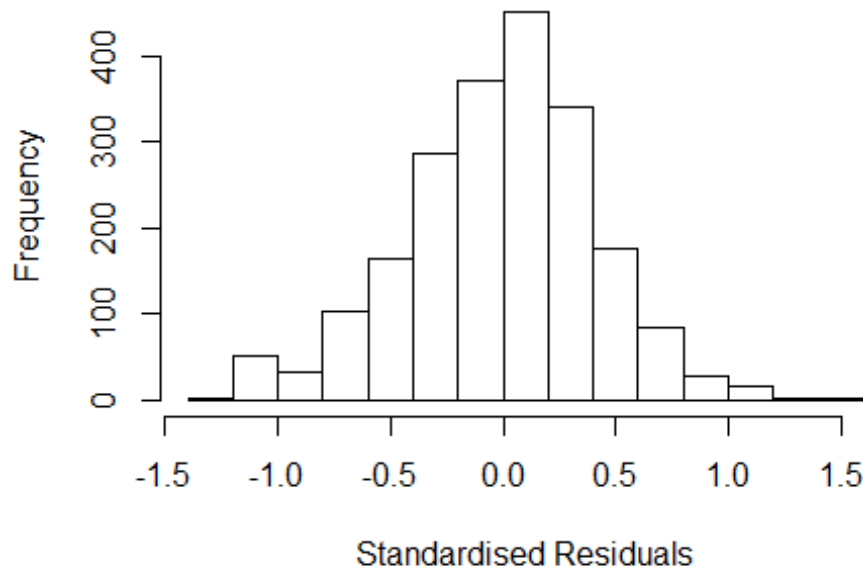
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2835 -0.2711  0.0158  0.2616  1.5206
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2004     0.0652   18.40  <2e-16 ***
## FirstAuthorFemale1 -0.0247     0.0184   -1.34   0.1806
## LastAuthorFemale1  -0.0648     0.0198   -3.27   0.0011 **
## Year1997           0.0831     0.0877    0.95   0.3431
## Year1998           0.0627     0.0778    0.81   0.4199
## Year1999          -0.0560     0.0837   -0.67   0.5036
## Year2000          -0.0247     0.0748   -0.33   0.7414
## Year2001          -0.0391     0.0807   -0.49   0.6276
## Year2002           0.0546     0.0742    0.74   0.4615
## Year2003          -0.0353     0.0719   -0.49   0.6238
## Year2004          -0.0775     0.0745   -1.04   0.2988
## Year2005          -0.0448     0.0744   -0.60   0.5470
```

```

## Year2006          -0.0922      0.0716   -1.29   0.1978
## Year2007          -0.0127      0.0724   -0.18   0.8608
## Year2008          -0.0230      0.0705   -0.33   0.7441
## Year2009          -0.0381      0.0724   -0.53   0.5989
## Year2010          -0.0504      0.0746   -0.68   0.4995
## Year2011          -0.0145      0.0711   -0.20   0.8385
## Year2012          -0.1383      0.0716   -1.93   0.0533 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.39
## Multiple R-squared:  0.0243, Adjusted R-squared:  0.0159
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 179 weights are ~= 1. The remaining 1934 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0949 0.8650 0.9500 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1      1.012
## Year              1.024 16      1.001

```

Residuals from first author



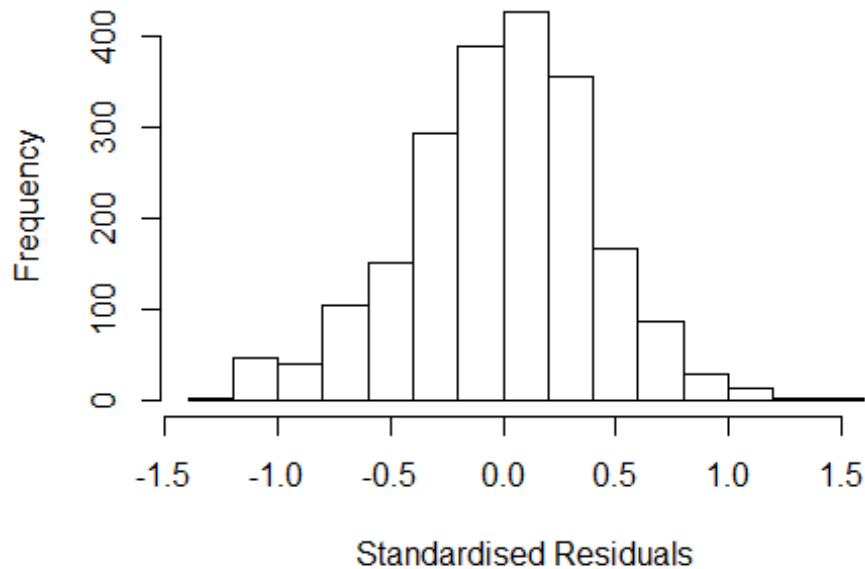
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2692 -0.2671 0.0164 0.2627 1.5339
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1869 0.0649 18.28 <2e-16 ***
## FirstAuthorFemale1 -0.0332 0.0183 -1.81 0.071 .
## Year1997 0.0823 0.0882 0.93 0.351
## Year1998 0.0616 0.0777 0.79 0.428
## Year1999 -0.0559 0.0829 -0.67 0.501
## Year2000 -0.0237 0.0747 -0.32 0.750
## Year2001 -0.0412 0.0808 -0.51 0.610
## Year2002 0.0542 0.0737 0.74 0.462
## Year2003 -0.0332 0.0717 -0.46 0.643
## Year2004 -0.0789 0.0746 -1.06 0.291
## Year2005 -0.0482 0.0745 -0.65 0.517
## Year2006 -0.0952 0.0716 -1.33 0.184
```

```

## Year2007          -0.0184      0.0721   -0.26    0.799
## Year2008          -0.0245      0.0705   -0.35    0.728
## Year2009          -0.0421      0.0722   -0.58    0.560
## Year2010          -0.0577      0.0746   -0.77    0.439
## Year2011          -0.0185      0.0709   -0.26    0.795
## Year2012          -0.1442      0.0717   -2.01    0.044 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.39
## Multiple R-squared:  0.019, Adjusted R-squared:  0.0111
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 193 weights are ~= 1. The remaining 1920 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0877 0.8610 0.9510 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.019 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
## -1.2736 -0.2738 0.0126 0.2605 1.5314
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1897 0.0646 18.41 < 2e-16 ***
## LastAuthorFemale1 -0.0685 0.0197 -3.48 0.00051 ***
## Year1997 0.0839 0.0878 0.96 0.33923
## Year1998 0.0646 0.0777 0.83 0.40585
## Year1999 -0.0561 0.0837 -0.67 0.50257
## Year2000 -0.0230 0.0749 -0.31 0.75826
## Year2001 -0.0388 0.0807 -0.48 0.63067
## Year2002 0.0569 0.0742 0.77 0.44382
## Year2003 -0.0371 0.0720 -0.52 0.60633
## Year2004 -0.0759 0.0747 -1.02 0.30965
## Year2005 -0.0459 0.0744 -0.62 0.53742
## Year2006 -0.0935 0.0716 -1.31 0.19203
```

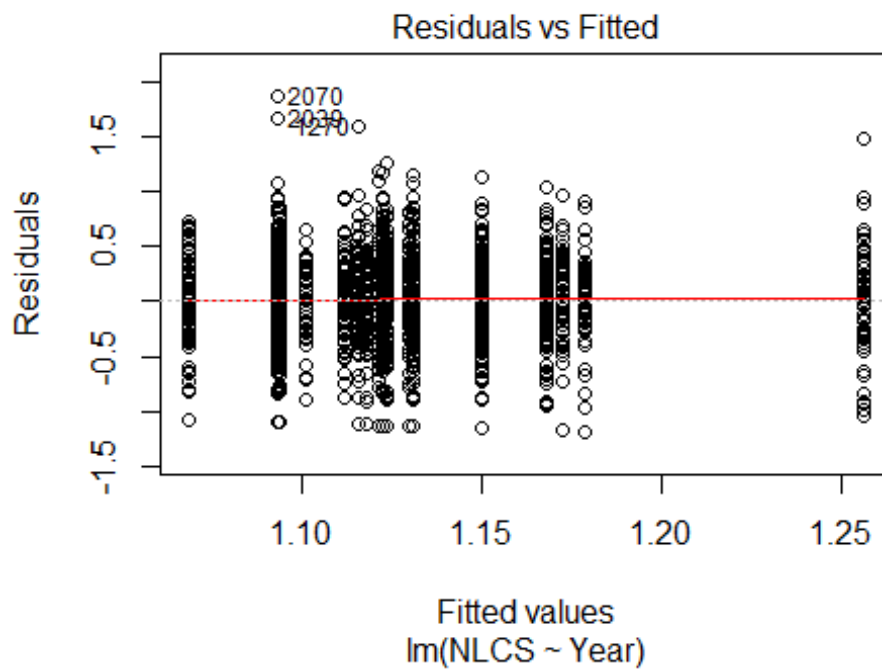
```

## Year2007          -0.0136      0.0725   -0.19  0.85091
## Year2008          -0.0246      0.0705   -0.35  0.72780
## Year2009          -0.0388      0.0724   -0.54  0.59178
## Year2010          -0.0528      0.0747   -0.71  0.47947
## Year2011          -0.0152      0.0713   -0.21  0.83125
## Year2012          -0.1408      0.0716   -1.97  0.04948 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.39
## Multiple R-squared:  0.0233, Adjusted R-squared:  0.0154
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 169 weights are ~ = 1. The remaining 1944 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0887 0.8660 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      4.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2113"
## [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
##   86   99  110   92   92  114  105   98  124  119  146  162  200  181  218
## 2011 2012
##  213  267
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   59   63   54   68   43   50   74   75   87   93  102  115  145  137  154
## 2011 2012

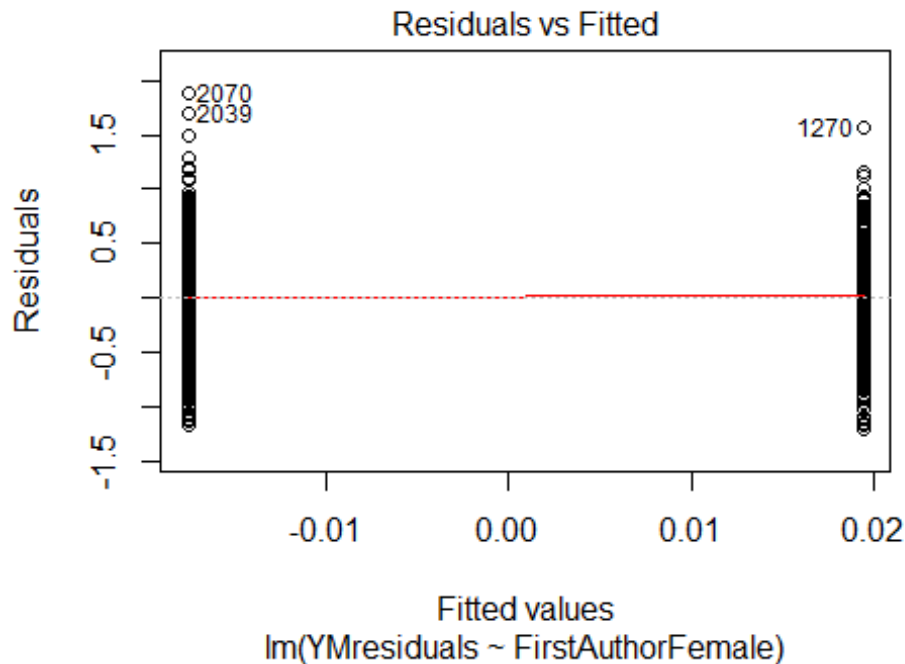
```



```
## 145 190
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 56 46 56 37 45 61 64 72 79 85 101 124 116 134
## 2011 2012
## 132 174
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.03
```

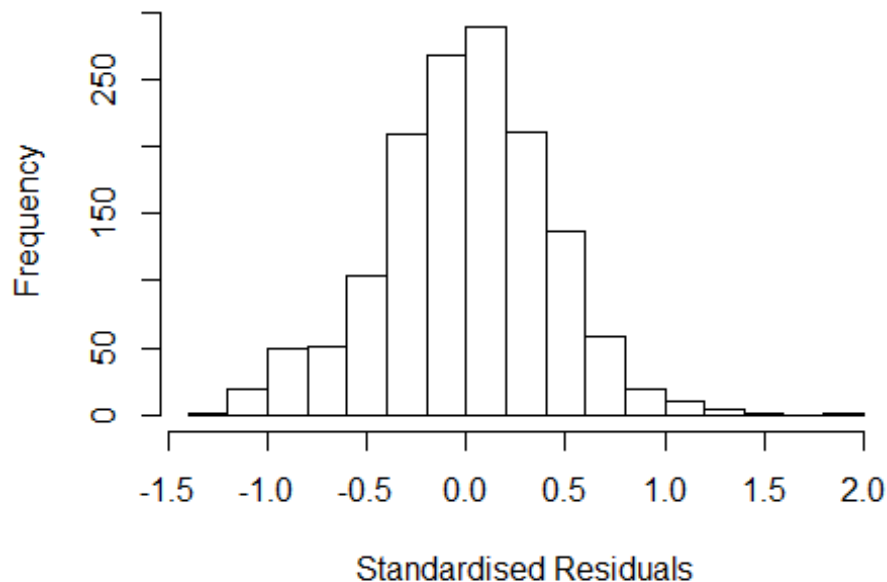


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 9e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.59168166911909"
## [1] "Male first author team size 2018 geometric mean: 4.24543554709425"
##
## 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] "Female last author team size 2018 geometric mean: 4.56271335788409"
## [1] "Male last author team size 2018 geometric mean: 4.36008083409154"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4400, 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 1.144 1      1.070
## LastAuthorFemale  1.103 1      1.050
## UniqueAuthors     1.328 4      1.036
## Year               1.481 16     1.012
```

Residuals from first and last author and team size



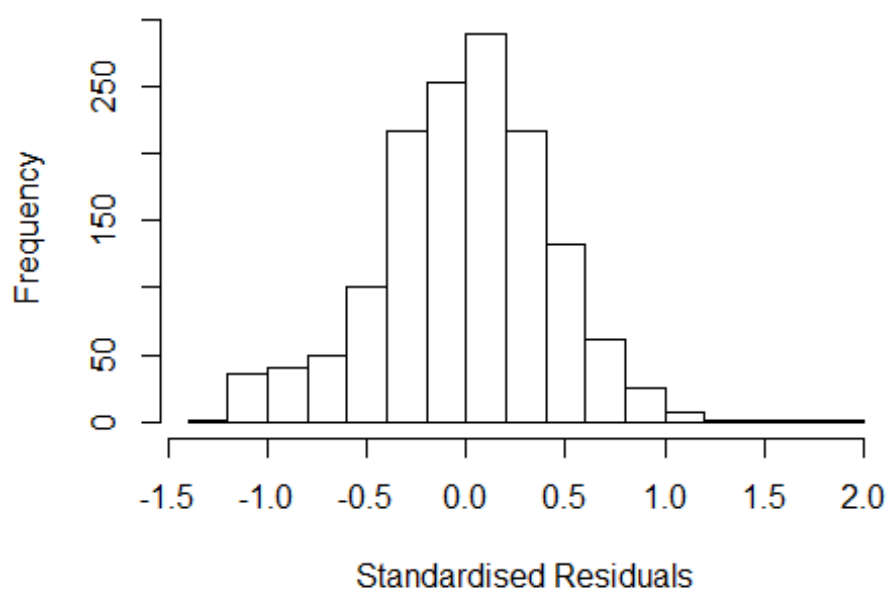
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2869 -0.2709  0.0118  0.2698  1.8933
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.05750    0.06810   15.53 < 2e-16 ***
## FirstAuthorFemale1 0.03220    0.02339    1.38  0.16874
## LastAuthorFemale1 0.00234    0.02416    0.10  0.92296
## UniqueAuthors2    0.15072    0.05723    2.63  0.00854 **
## UniqueAuthors3    0.19721    0.05398    3.65  0.00027 ***
## UniqueAuthors4    0.19972    0.05249    3.81  0.00015 ***
## UniqueAuthors5    0.22750    0.04921    4.62  4.1e-06 ***
## Year1997          0.07091    0.08737    0.81  0.41716
## Year1998         -0.04922    0.08238   -0.60  0.55030
## Year1999         -0.12831    0.07585   -1.69  0.09095 .
```

```

## Year2000      -0.15893      0.09769      -1.63      0.10398
## Year2001      -0.12976      0.07315      -1.77      0.07629 .
## Year2002      -0.16140      0.07250      -2.23      0.02615 *
## Year2003      -0.15252      0.07315      -2.08      0.03725 *
## Year2004      -0.19283      0.07412      -2.60      0.00938 **
## Year2005      -0.14395      0.06892      -2.09      0.03692 *
## Year2006      -0.09752      0.07237      -1.35      0.17799
## Year2007      -0.12006      0.07460      -1.61      0.10772
## Year2008      -0.12546      0.07039      -1.78      0.07490 .
## Year2009      -0.18881      0.06957      -2.71      0.00673 **
## Year2010      -0.07831      0.06789      -1.15      0.24890
## Year2011      -0.14875      0.06747      -2.20      0.02763 *
## Year2012      -0.11772      0.06557      -1.80      0.07282 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.399
## Multiple R-squared:  0.0437, Adjusted R-squared:  0.0289
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(892,905) are outliers with |weight| = 0 ( < 7e-05);
## 129 weights are ~= 1. The remaining 1305 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0974 0.8660 0.9500 0.8960 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           6.96e-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 1.114 1 1.056
## LastAuthorFemale 1.100 1 1.049
## Year 1.136 16 1.004

```

Residuals from first and last author



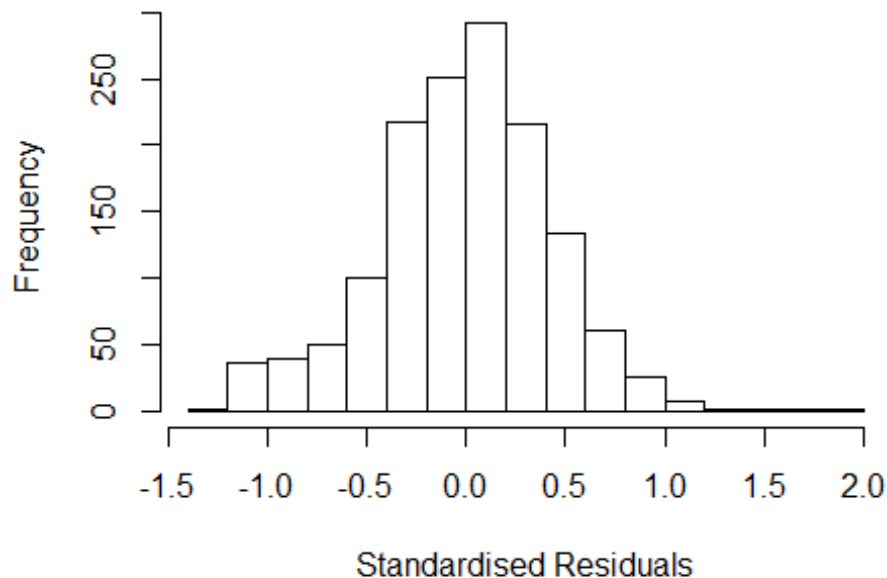
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26063 -0.27070 0.00995 0.26857 1.90908
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.211767 0.057397 21.11 <2e-16 ***
## FirstAuthorFemale1 0.048864 0.023492 2.08 0.038 *
## LastAuthorFemale1 0.000645 0.024255 0.03 0.979
## Year1997 0.069808 0.087732 0.80 0.426
## Year1998 -0.036307 0.079440 -0.46 0.648
## Year1999 -0.113467 0.077865 -1.46 0.145
## Year2000 -0.131517 0.099286 -1.32 0.186
## Year2001 -0.114178 0.075439 -1.51 0.130
## Year2002 -0.150007 0.074189 -2.02 0.043 *
## Year2003 -0.120390 0.073702 -1.63 0.103
## Year2004 -0.163223 0.074438 -2.19 0.028 *
## Year2005 -0.122970 0.069387 -1.77 0.077 .
```

```

## Year2006      -0.075666    0.073551   -1.03    0.304
## Year2007      -0.087566    0.073941   -1.18    0.237
## Year2008      -0.100481    0.071117   -1.41    0.158
## Year2009      -0.166847    0.071222   -2.34    0.019 *
## Year2010      -0.063077    0.069133   -0.91    0.362
## Year2011      -0.115936    0.068355   -1.70    0.090 .
## Year2012      -0.082123    0.065668   -1.25    0.211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared:  0.0178, Adjusted R-squared:  0.00536
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 905 is an outlier with |weight| = 0 ( < 7e-05);
## 124 weights are ~ = 1. The remaining 1311 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0275 0.8640 0.9510 0.8950 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      6.96e-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 1.079 1          1.039
## Year              1.079 16          1.002

```

Residuals from first author



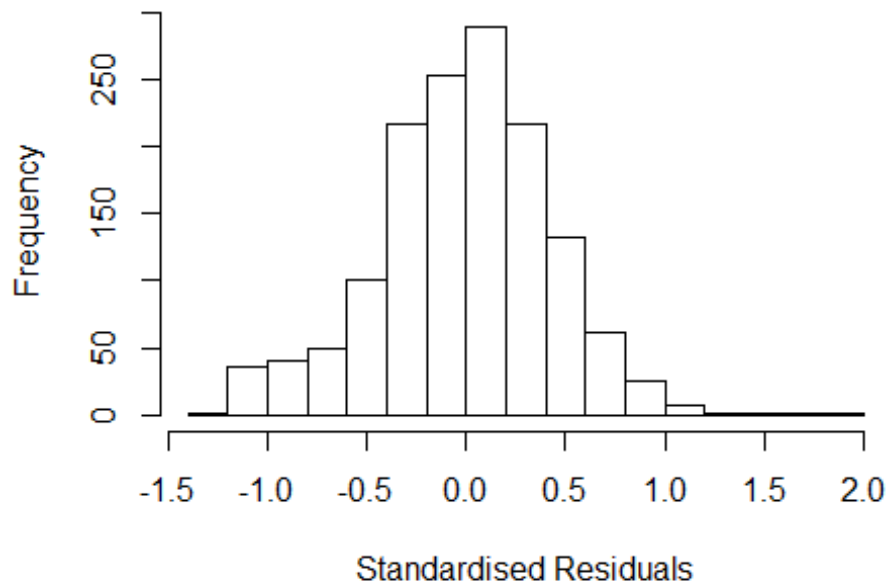
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2609 -0.2709 0.0101 0.2685 1.9089
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2119 0.0570 21.24 <2e-16 ***
## FirstAuthorFemale1 0.0490 0.0231 2.12 0.034 *
## Year1997 0.0698 0.0878 0.80 0.426
## Year1998 -0.0363 0.0794 -0.46 0.647
## Year1999 -0.1135 0.0779 -1.46 0.145
## Year2000 -0.1315 0.0993 -1.32 0.186
## Year2001 -0.1142 0.0754 -1.52 0.130
## Year2002 -0.1499 0.0742 -2.02 0.043 *
## Year2003 -0.1204 0.0737 -1.63 0.103
## Year2004 -0.1632 0.0744 -2.19 0.028 *
## Year2005 -0.1229 0.0694 -1.77 0.077 .
## Year2006 -0.0756 0.0735 -1.03 0.304
```

```

## Year2007          -0.0875      0.0738   -1.19    0.236
## Year2008          -0.1004      0.0711   -1.41    0.158
## Year2009          -0.1668      0.0712   -2.34    0.019 *
## Year2010          -0.0631      0.0691   -0.91    0.362
## Year2011          -0.1159      0.0684   -1.69    0.090 .
## Year2012          -0.0820      0.0658   -1.25    0.212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared:  0.0178, Adjusted R-squared:  0.00605
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 905 is an outlier with |weight| = 0 ( < 7e-05);
## 123 weights are ~= 1. The remaining 1312 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0277 0.8640 0.9510 0.8950 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      6.96e-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 1.061 1          1.030
## Year            1.061 16          1.002

```


Residuals from last author



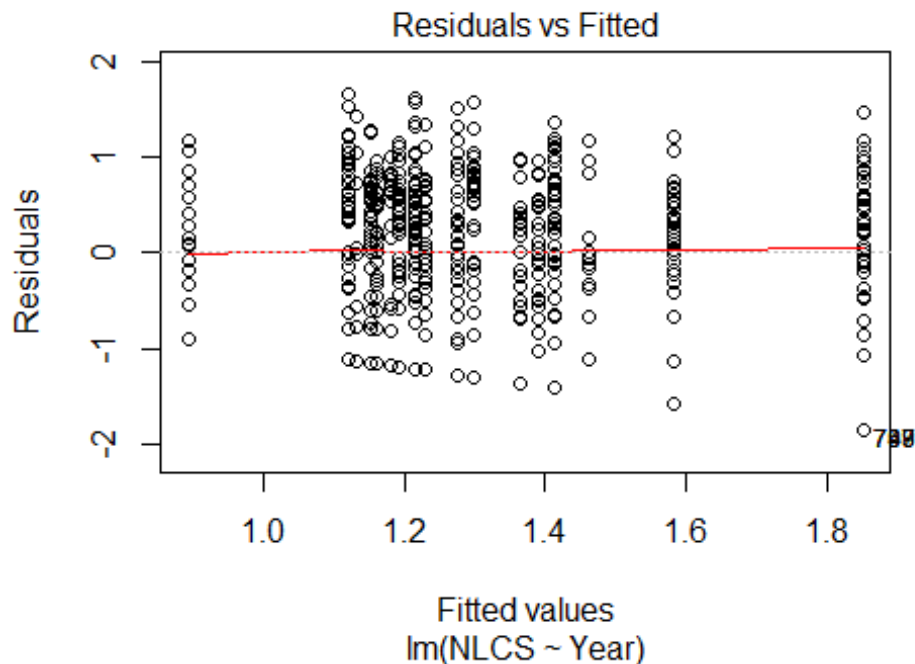
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22591 -0.26785 0.00934 0.27317 1.89196
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2259 0.0577 21.24 <2e-16 ***
## LastAuthorFemale1 0.0121 0.0239 0.51 0.611
## Year1997 0.0744 0.0879 0.85 0.397
## Year1998 -0.0381 0.0803 -0.47 0.636
## Year1999 -0.1145 0.0792 -1.45 0.148
## Year2000 -0.1254 0.1008 -1.24 0.214
## Year2001 -0.1073 0.0763 -1.41 0.160
## Year2002 -0.1452 0.0747 -1.94 0.052 .
## Year2003 -0.1150 0.0735 -1.57 0.118
## Year2004 -0.1564 0.0743 -2.10 0.036 *
## Year2005 -0.1186 0.0694 -1.71 0.087 .
## Year2006 -0.0668 0.0738 -0.90 0.366
```

```

## Year2007          -0.0818      0.0739   -1.11    0.269
## Year2008          -0.0910      0.0712   -1.28    0.202
## Year2009          -0.1639      0.0717   -2.29    0.022 *
## Year2010          -0.0585      0.0695   -0.84    0.401
## Year2011          -0.1074      0.0684   -1.57    0.117
## Year2012          -0.0737      0.0655   -1.12    0.261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.402
## Multiple R-squared:  0.0145, Adjusted R-squared:  0.00269
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 905 is an outlier with |weight| = 0 ( < 7e-05);
## 121 weights are ~= 1. The remaining 1314 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0341 0.8670 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          6.96e-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: 1436"
## [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
##   49   43   38   25   31   46   58   22   30   55   51   50   46   63   79
## 2011 2012
##   103  118
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   41   23   26   13    9   23   49   18   29   44   45   46   42   56   68

```

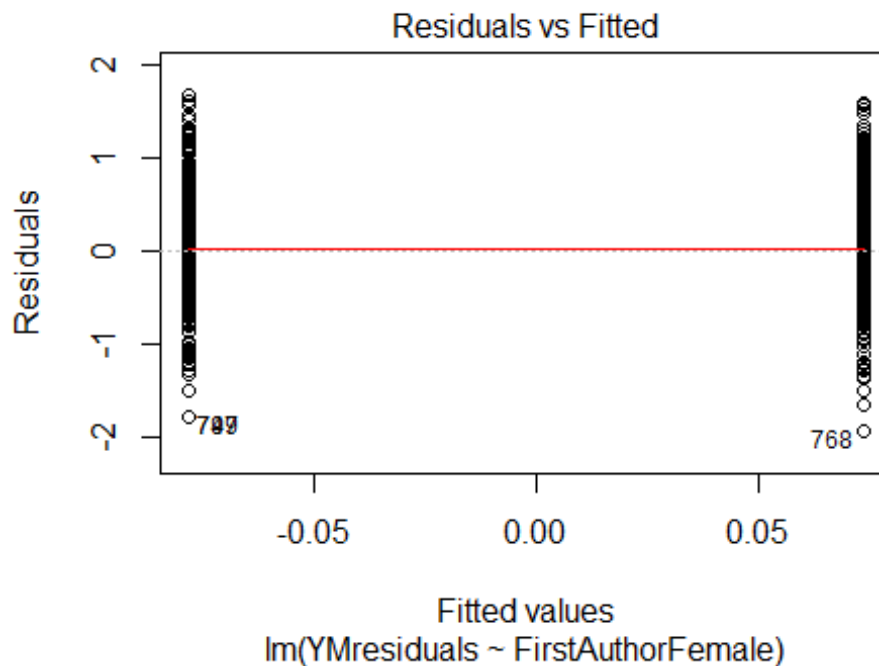
```
## 2011 2012
## 87 102
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 38 23 26 12 8 21 47 18 26 40 40 41 37 48 56
## 2011 2012
## 84 90
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.5
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.66, df = 1, p-value = 0.4
## [1] "Female first author team size 2018 geometric mean: 4.19704375705936"
## [1] "Male first author team size 2018 geometric mean: 3.57757474635906"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
```

```
## W = 860, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.67173578940366"
## [1] "Male last author team size 2018 geometric mean: 4.48290720163049"

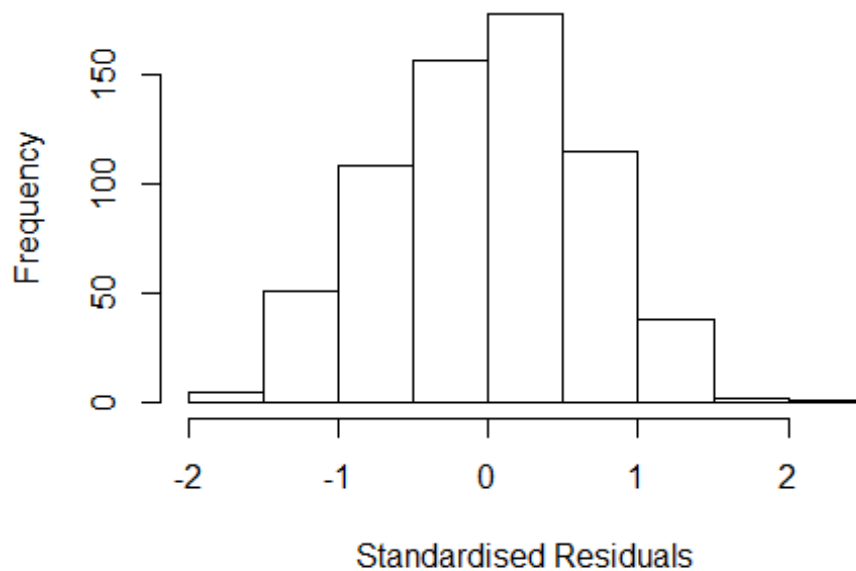
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 660, 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	1.374	1	1.172
LastAuthorFemale	1.339	1	1.157
UniqueAuthors	2.190	4	1.103
Year	2.391	16	1.028

Residuals from first and last author and team size



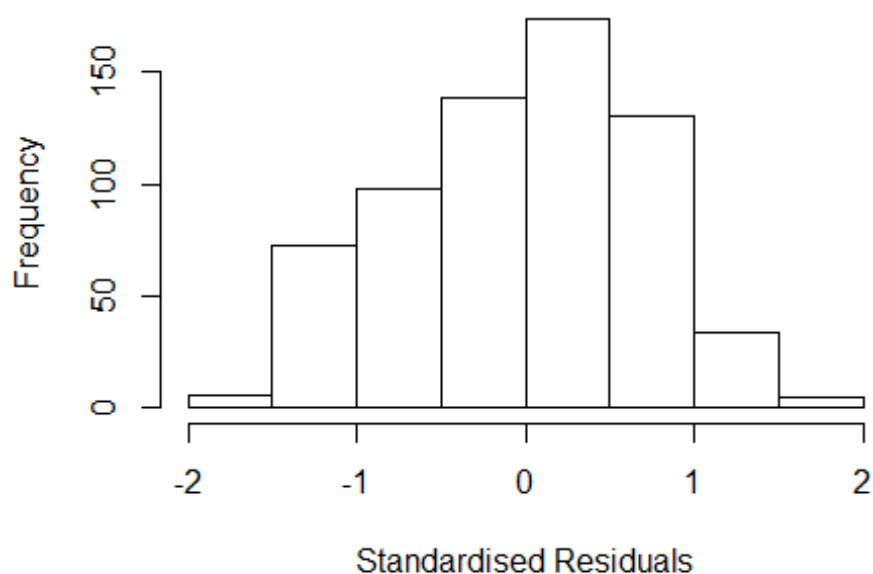
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8510 -0.5016 0.0308 0.4852 2.0312
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8857 0.1102 8.03 4.7e-15 ***
## FirstAuthorFemale1 0.0593 0.0645 0.92 0.3578
## LastAuthorFemale1 0.1523 0.0631 2.41 0.0161 *
## UniqueAuthors2 0.2653 0.0985 2.69 0.0073 **
## UniqueAuthors3 0.2897 0.1031 2.81 0.0051 **
## UniqueAuthors4 0.4471 0.1086 4.12 4.3e-05 ***
## UniqueAuthors5 0.6739 0.0900 7.49 2.3e-13 ***
## Year1997 0.2164 0.1805 1.20 0.2312
## Year1998 0.2944 0.1526 1.93 0.0541 .
## Year1999 0.2997 0.2010 1.49 0.1364
```

```

## Year2000          -0.0002      0.3708      0.00      0.9996
## Year2001          -0.3373      0.1724     -1.96      0.0508 .
## Year2002          -0.0955      0.1394     -0.69      0.4936
## Year2003          -0.0991      0.2048     -0.48      0.6285
## Year2004          -0.1312      0.1545     -0.85      0.3963
## Year2005          -0.0265      0.1458     -0.18      0.8559
## Year2006          -0.0455      0.1562     -0.29      0.7712
## Year2007          -0.1156      0.1691     -0.68      0.4944
## Year2008           0.2437      0.1602      1.52      0.1286
## Year2009           0.4415      0.1454      3.04      0.0025 **
## Year2010          -0.3464      0.1601     -2.16      0.0309 *
## Year2011           0.0798      0.1221      0.65      0.5133
## Year2012          -0.1430      0.1220     -1.17      0.2414
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.703
## Multiple R-squared:  0.184, Adjusted R-squared:  0.155
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 602 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.384  0.878  0.950  0.913  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.53e-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 1.307 1          1.143
## LastAuthorFemale  1.284 1          1.133
## Year              1.214 16          1.006

```

Residuals from first and last author



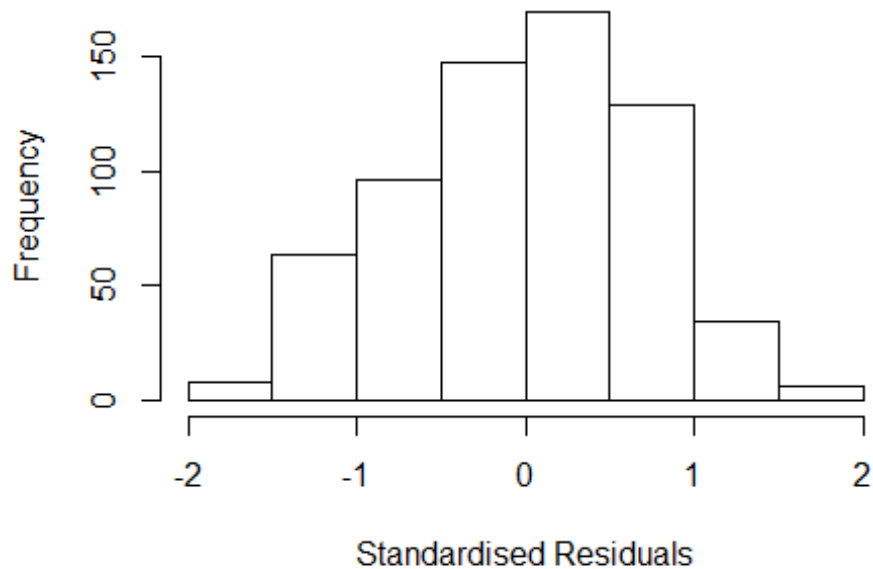
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8418 -0.5299 0.0549 0.5082 1.6825
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11246 0.10380 10.72 < 2e-16 ***
## FirstAuthorFemale1 0.12896 0.06664 1.94 0.053 .
## LastAuthorFemale1 0.15532 0.06605 2.35 0.019 *
## Year1997 0.08120 0.17915 0.45 0.651
## Year1998 0.12955 0.16322 0.79 0.428
## Year1999 0.12439 0.21766 0.57 0.568
## Year2000 -0.14290 0.34879 -0.41 0.682
## Year2001 -0.36604 0.18515 -1.98 0.048 *
## Year2002 -0.03429 0.14763 -0.23 0.816
## Year2003 -0.01849 0.24580 -0.08 0.940
## Year2004 -0.08422 0.16502 -0.51 0.610
## Year2005 0.00827 0.15471 0.05 0.957
```

```

## Year2006          0.04924      0.16017      0.31      0.759
## Year2007          -0.02388      0.19308     -0.12      0.902
## Year2008           0.35315      0.15531      2.27      0.023 *
## Year2009           0.60036      0.14749      4.07     5.3e-05 ***
## Year2010          -0.20125      0.16719     -1.20      0.229
## Year2011           0.20423      0.12696      1.61      0.108
## Year2012           0.00400      0.12835      0.03      0.975
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.735
## Multiple R-squared:  0.0978, Adjusted R-squared:  0.0723
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 54 weights are ~= 1. The remaining 601 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.510  0.865   0.942   0.913   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.53e-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 1.114 1      1.056
## Year              1.114 16      1.003

```


Residuals from first author



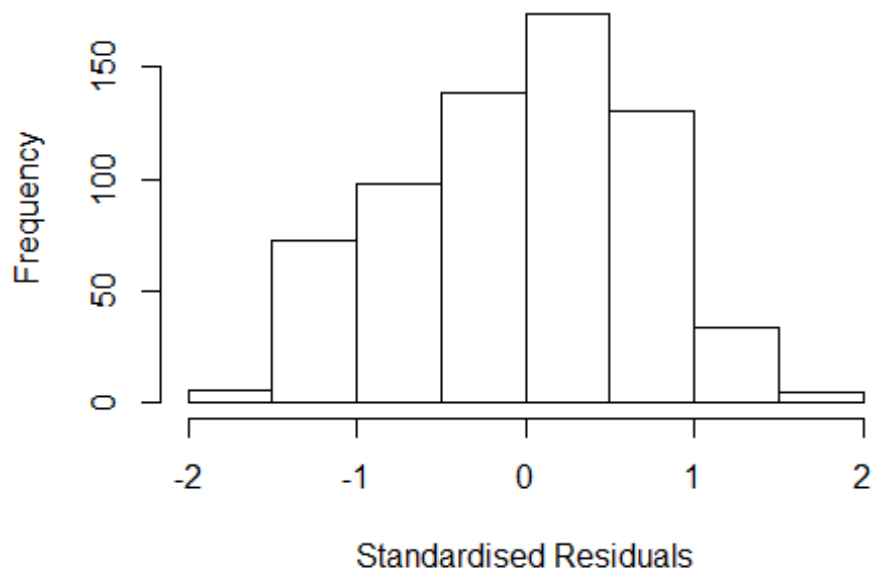
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.9493 -0.5586 0.0365 0.5162 1.6478
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13747 0.10649 10.68 < 2e-16 ***
## FirstAuthorFemale1 0.19124 0.06166 3.10 0.002 **
## Year1997 0.11482 0.17987 0.64 0.523
## Year1998 0.15113 0.16231 0.93 0.352
## Year1999 0.14211 0.21222 0.67 0.503
## Year2000 -0.11959 0.36197 -0.33 0.741
## Year2001 -0.36248 0.18394 -1.97 0.049 *
## Year2002 -0.03369 0.15094 -0.22 0.823
## Year2003 0.00302 0.25027 0.01 0.990
## Year2004 -0.07148 0.16517 -0.43 0.665
## Year2005 -0.01176 0.15762 -0.07 0.941
## Year2006 0.05668 0.16361 0.35 0.729
```

```

## Year2007          -0.03304    0.19059   -0.17    0.862
## Year2008          0.36786    0.15648    2.35    0.019 *
## Year2009          0.62055    0.15116    4.11  4.6e-05 ***
## Year2010         -0.18049    0.16716   -1.08    0.281
## Year2011          0.20073    0.12943    1.55    0.121
## Year2012          0.01378    0.13110    0.11    0.916
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.742
## Multiple R-squared:  0.09,   Adjusted R-squared:  0.0658
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 597 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.470  0.864  0.945  0.913  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.53e-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



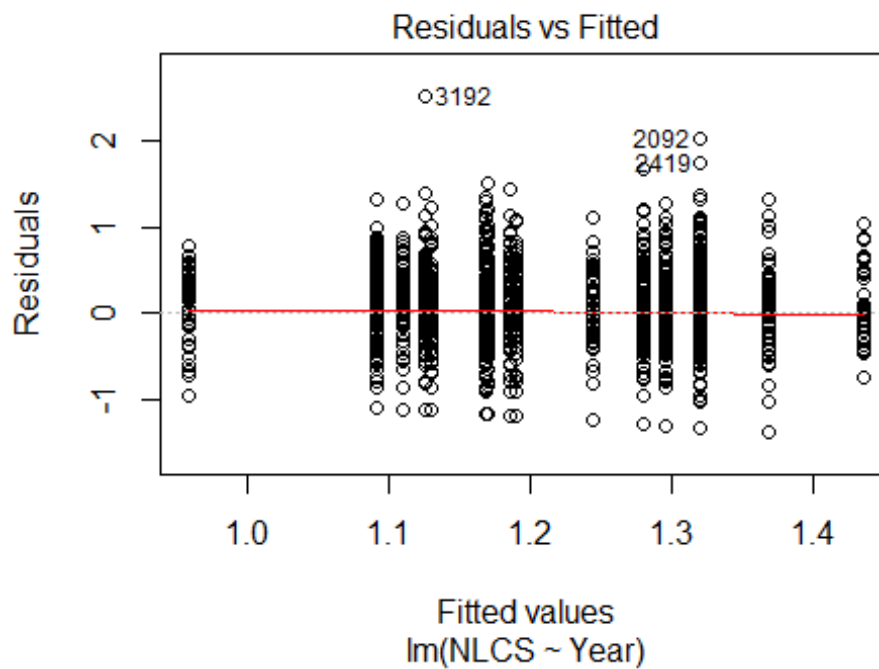
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7604 -0.5704 0.0484 0.5119 1.7030
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1427 0.1032 11.08 < 2e-16 ***
## LastAuthorFemale1 0.2081 0.0617 3.37 0.00079 ***
## Year1997 0.1066 0.1771 0.60 0.54751
## Year1998 0.1398 0.1607 0.87 0.38455
## Year1999 0.1286 0.2119 0.61 0.54421
## Year2000 -0.1587 0.3755 -0.42 0.67264
## Year2001 -0.3544 0.1854 -1.91 0.05639 .
## Year2002 -0.0256 0.1463 -0.17 0.86132
## Year2003 -0.0151 0.2435 -0.06 0.95047
## Year2004 -0.0708 0.1689 -0.42 0.67544
## Year2005 0.0441 0.1548 0.28 0.77582
## Year2006 0.0690 0.1586 0.44 0.66351
```

```

## Year2007          0.0214      0.1917      0.11  0.91104
## Year2008          0.3623      0.1569      2.31  0.02128 *
## Year2009          0.6178      0.1463      4.22  2.8e-05 ***
## Year2010         -0.1836      0.1699     -1.08  0.28028
## Year2011          0.2139      0.1279      1.67  0.09499 .
## Year2012          0.0179      0.1284      0.14  0.88909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.73
## Multiple R-squared:  0.0929, Adjusted R-squared:  0.0687
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 599 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.540  0.868  0.939  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      1.53e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 655"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
## 153 204 129 163 169 188 146 114 121 134 150 184 177 172 184
## 2011 2012
## 142 175
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 61 76 40 90 82 63 108 72 87 89 114 135 133 114 141
## 2011 2012

```

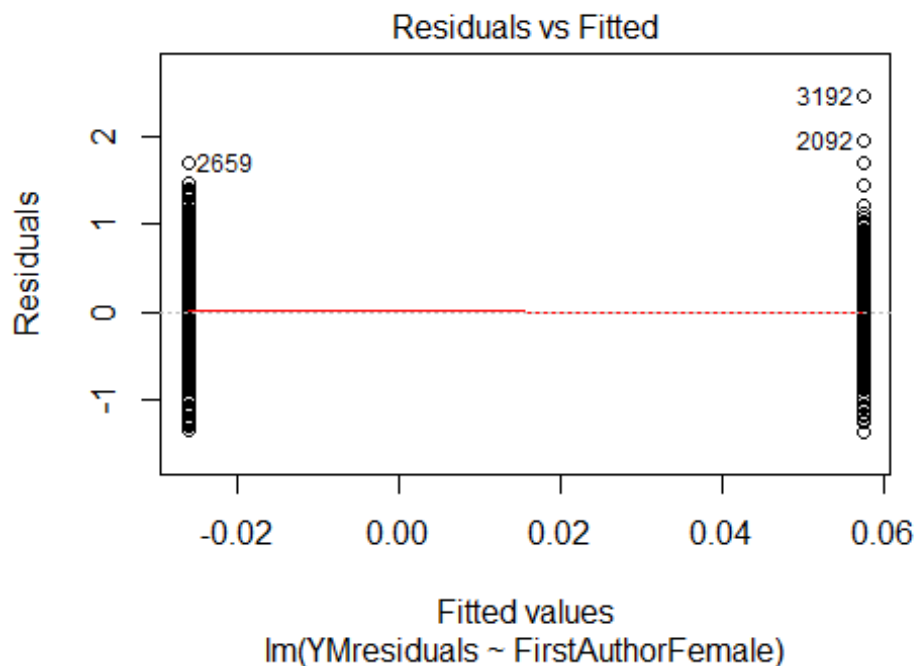
```
## 106 134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 52 72 33 75 73 58 93 58 78 76 105 118 117 89 120
## 2011 2012
## 89 122
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 16, df = 16, p-value = 0.4
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.00045, df = 1, p-value = 1
## [1] "Female first author team size 2018 geometric mean: 6.21456973995997"
## [1] "Male first author team size 2018 geometric mean: 5.88623434923291"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

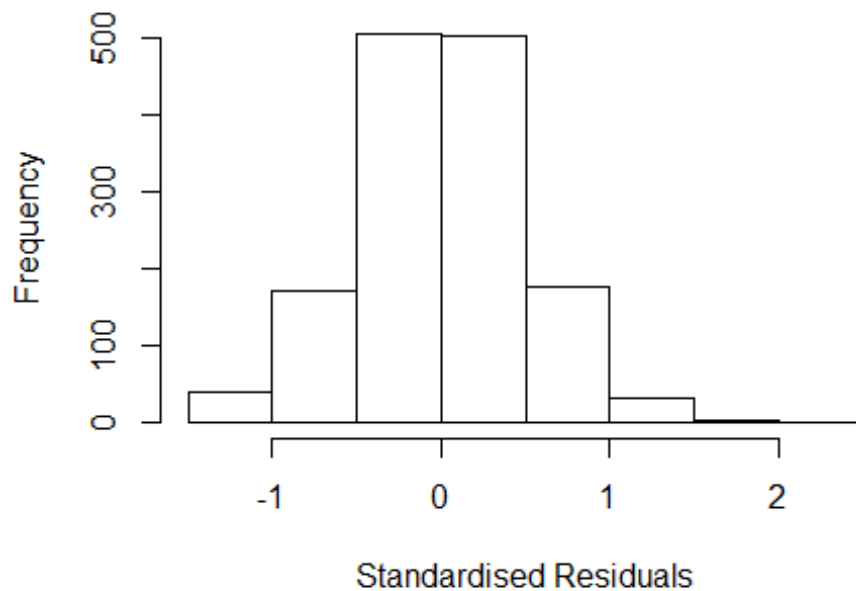
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 710, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 7.13895103466583"
## [1] "Male last author team size 2018 geometric mean: 5.49126179535578"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 810, 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 1.119  1      1.058
## LastAuthorFemale  1.124  1      1.060
## UniqueAuthors    1.313  4      1.035
## Year              1.455 16      1.012
```

Residuals from first and last author and team size



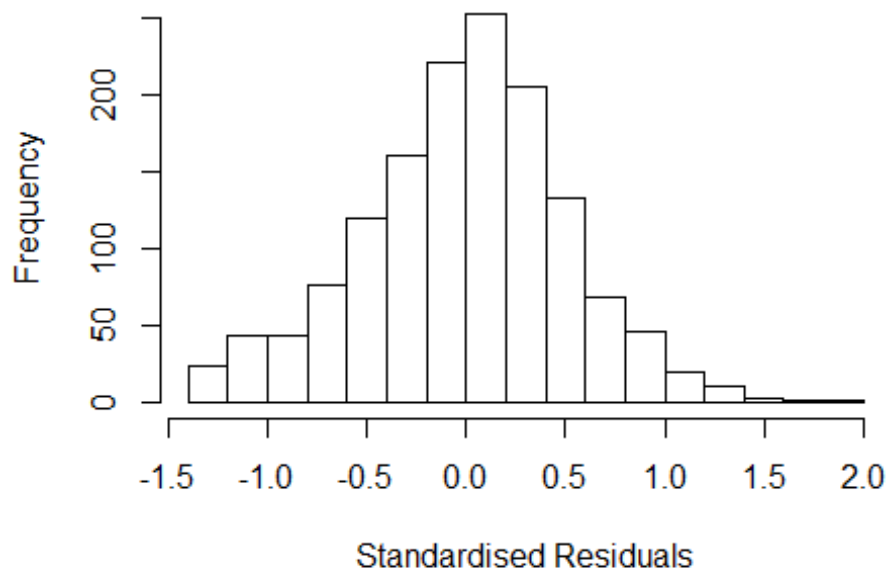
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.46864 -0.31133 -0.00232 0.32386 2.14066
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.87896 0.08915 9.86 < 2e-16 ***
## FirstAuthorFemale1 0.01894 0.02970 0.64 0.5237
## LastAuthorFemale1 0.03453 0.03553 0.97 0.3313
## UniqueAuthors2 0.13421 0.05815 2.31 0.0211 *
## UniqueAuthors3 0.24348 0.05122 4.75 2.2e-06 ***
## UniqueAuthors4 0.25298 0.04978 5.08 4.2e-07 ***
## UniqueAuthors5 0.40865 0.04561 8.96 < 2e-16 ***
## Year1997 0.25281 0.09887 2.56 0.0107 *
## Year1998 0.30988 0.10509 2.95 0.0032 **
## Year1999 -0.00250 0.09865 -0.03 0.9798
```

```

## Year2000      0.06075    0.10077    0.60    0.5467
## Year2001      0.13777    0.09825    1.40    0.1610
## Year2002     -0.02056    0.09609   -0.21    0.8306
## Year2003     -0.12664    0.10053   -1.26    0.2080
## Year2004      0.11479    0.10184    1.13    0.2599
## Year2005     -0.00301    0.09766   -0.03    0.9754
## Year2006      0.00645    0.10052    0.06    0.9488
## Year2007      0.15524    0.09382    1.65    0.0982 .
## Year2008      0.15699    0.09530    1.65    0.0997 .
## Year2009      0.14650    0.09767    1.50    0.1338
## Year2010      0.11563    0.09154    1.26    0.2067
## Year2011     -0.05272    0.09942   -0.53    0.5960
## Year2012     -0.07305    0.09353   -0.78    0.4349
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.474
## Multiple R-squared:  0.116, Adjusted R-squared:  0.103
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 105 weights are ~= 1. The remaining 1323 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0052 0.8650 0.9530 0.9000 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.00e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.089 1      1.043
## LastAuthorFemale 1.123 1      1.060
## Year      1.154 16      1.004

```


Residuals from first and last author



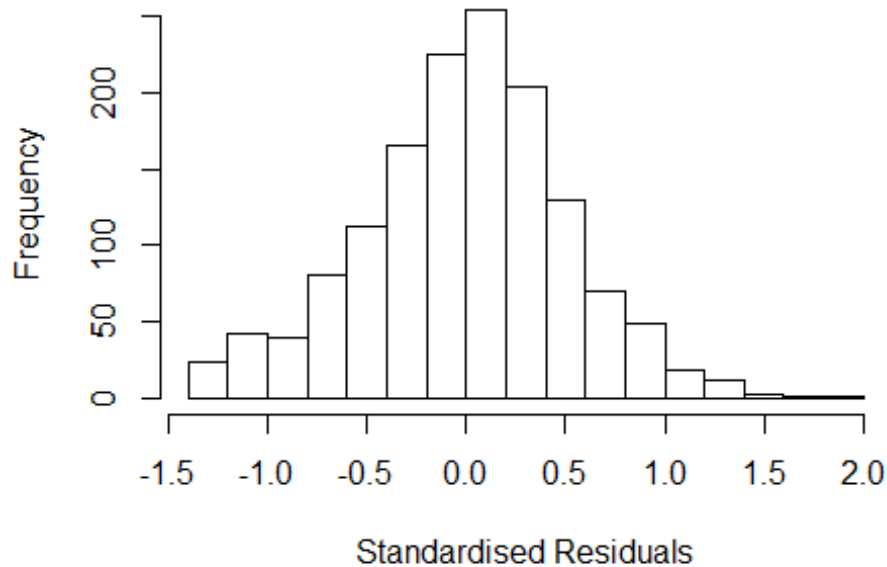
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3513 -0.3217  0.0184  0.3206  1.9664
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0618    0.0860   12.35  <2e-16 ***
## FirstAuthorFemale1  0.0556    0.0304    1.83   0.0672 .
## LastAuthorFemale1  0.0436    0.0369    1.18   0.2373
## Year1997          0.2895    0.1027    2.82   0.0049 **
## Year1998          0.3209    0.1124    2.86   0.0044 **
## Year1999          0.0523    0.1048    0.50   0.6182
## Year2000          0.1001    0.1073    0.93   0.3511
## Year2001          0.1528    0.1049    1.46   0.1455
## Year2002          0.0248    0.1013    0.25   0.8063
## Year2003         -0.0911    0.1071   -0.85   0.3948
## Year2004          0.1439    0.1054    1.37   0.1725
## Year2005          0.0216    0.1039    0.21   0.8352
```

```

## Year2006          0.0811      0.1083      0.75      0.4543
## Year2007          0.2442      0.0984      2.48      0.0132 *
## Year2008          0.1959      0.0992      1.98      0.0484 *
## Year2009          0.2124      0.1010      2.10      0.0357 *
## Year2010          0.1968      0.0956      2.06      0.0397 *
## Year2011          0.0539      0.1030      0.52      0.6010
## Year2012          0.0377      0.0970      0.39      0.6977
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.0412, Adjusted R-squared:  0.0289
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1290 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0642 0.8580 0.9490 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.054 1      1.027
## Year      1.054 16      1.002

```

Residuals from first author



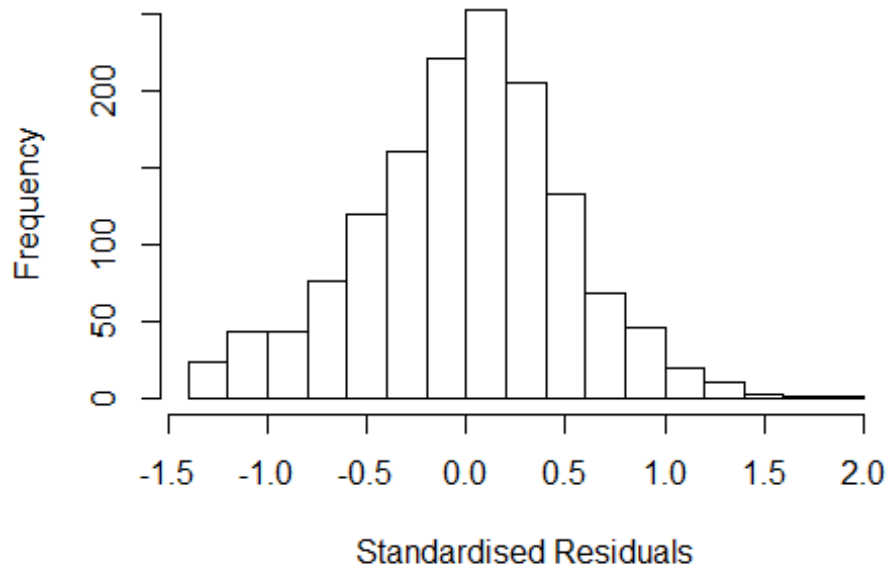
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3561 -0.3258 0.0152 0.3205 1.9516
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0681 0.0853 12.52 <2e-16 ***
## FirstAuthorFemale1 0.0630 0.0298 2.11 0.0349 *
## Year1997 0.2880 0.1027 2.80 0.0051 **
## Year1998 0.3199 0.1129 2.83 0.0047 **
## Year1999 0.0521 0.1048 0.50 0.6190
## Year2000 0.1024 0.1072 0.96 0.3395
## Year2001 0.1498 0.1047 1.43 0.1529
## Year2002 0.0225 0.1007 0.22 0.8231
## Year2003 -0.0854 0.1068 -0.80 0.4242
## Year2004 0.1422 0.1050 1.35 0.1757
## Year2005 0.0262 0.1034 0.25 0.8001
## Year2006 0.0795 0.1081 0.74 0.4621
```

```

## Year2007          0.2453      0.0982      2.50      0.0126 *
## Year2008          0.1944      0.0988      1.97      0.0492 *
## Year2009          0.2097      0.1008      2.08      0.0376 *
## Year2010          0.1957      0.0952      2.06      0.0399 *
## Year2011          0.0533      0.1027      0.52      0.6040
## Year2012          0.0400      0.0967      0.41      0.6793
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.0399, Adjusted R-squared:  0.0284
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 140 weights are ~= 1. The remaining 1288 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0696 0.8570 0.9480 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      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 1.085 1      1.042
## Year      1.085 16      1.003

```

Residuals from last author



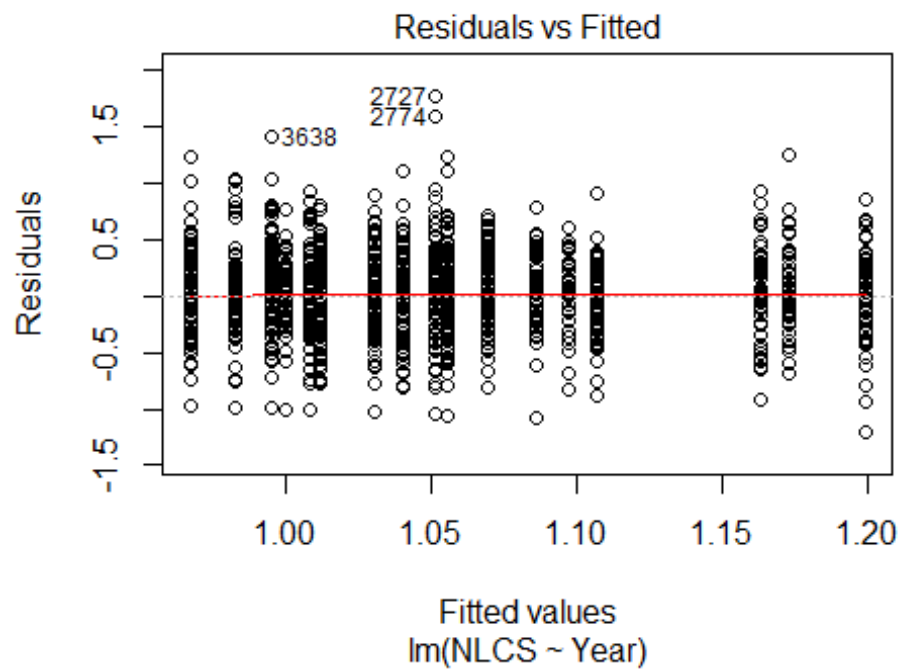
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3597 -0.3273 0.0127 0.3219 2.0059
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0693 0.0866 12.35 <2e-16 ***
## LastAuthorFemale1 0.0565 0.0362 1.56 0.1190
## Year1997 0.2904 0.1034 2.81 0.0050 **
## Year1998 0.3267 0.1130 2.89 0.0039 **
## Year1999 0.0571 0.1052 0.54 0.5874
## Year2000 0.1067 0.1074 0.99 0.3205
## Year2001 0.1633 0.1054 1.55 0.1214
## Year2002 0.0340 0.1018 0.33 0.7386
## Year2003 -0.0923 0.1077 -0.86 0.3920
## Year2004 0.1489 0.1057 1.41 0.1593
## Year2005 0.0260 0.1049 0.25 0.8042
## Year2006 0.0923 0.1086 0.85 0.3953
```

```

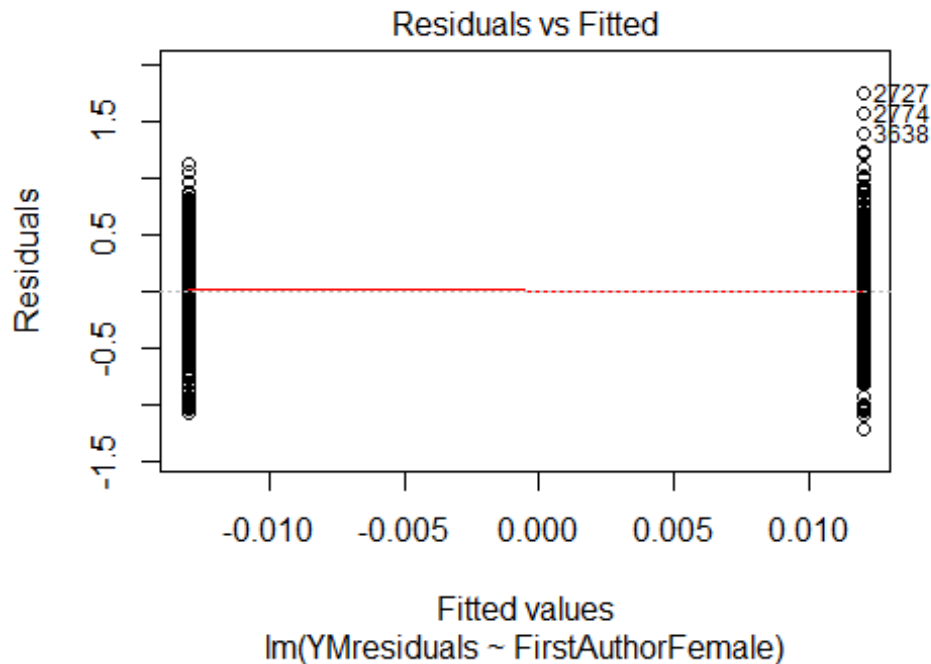
## Year2007          0.2528      0.0987      2.56      0.0105 *
## Year2008          0.2035      0.0996      2.04      0.0412 *
## Year2009          0.2239      0.1010      2.22      0.0268 *
## Year2010          0.2060      0.0961      2.15      0.0321 *
## Year2011          0.0648      0.1036      0.63      0.5318
## Year2012          0.0491      0.0974      0.50      0.6141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.0388, Adjusted R-squared:  0.0272
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 150 weights are ~= 1. The remaining 1278 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0495 0.8590 0.9480 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.00e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1428"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
## 118 112 174 126 170 169 135 131 106 162 174 164 144 147 142
## 2011 2012
## 150 178
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 63 59 51 72 35 52 78 81 63 96 88 88 98 88 87
## 2011 2012

```

```
## 100 124
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 56 55 44 65 27 43 64 64 47 75 74 72 85 73 73
## 2011 2012
## 92 102
## [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
```

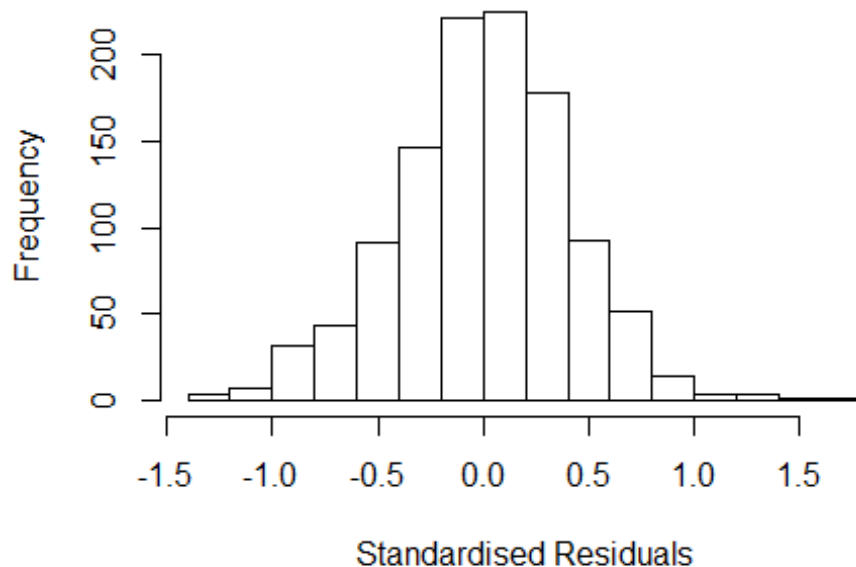


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.6, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 5.30932793781708"
## [1] "Male first author team size 2018 geometric mean: 4.9945741006044"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.63950963147328"
## [1] "Male last author team size 2018 geometric mean: 5.40119048900859"
##
## 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] "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  1.101 1      1.049
## UniqueAuthors     1.397 4      1.043
## Year              1.432 16     1.011
```


Residuals from first and last author and team size



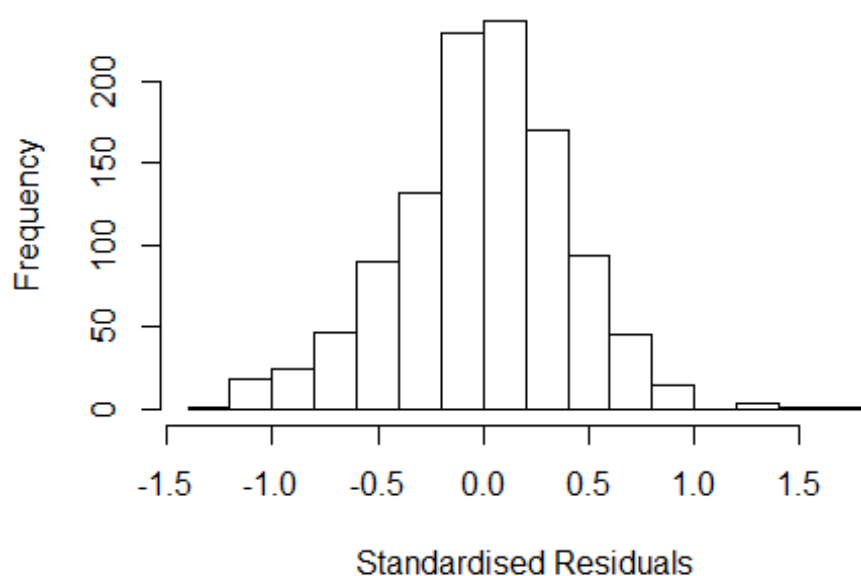
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24787 -0.25438 0.00911 0.25666 1.64387
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06615 0.07256 14.69 < 2e-16 ***
## FirstAuthorFemale1 -0.00225 0.02540 -0.09 0.9296
## LastAuthorFemale1 -0.05770 0.02611 -2.21 0.0273 *
## UniqueAuthors2 0.16260 0.06025 2.70 0.0071 **
## UniqueAuthors3 0.16181 0.06027 2.68 0.0074 **
## UniqueAuthors4 0.15762 0.06105 2.58 0.0100 **
## UniqueAuthors5 0.23833 0.05611 4.25 2.3e-05 ***
## Year1997 0.01913 0.07321 0.26 0.7939
## Year1998 -0.02823 0.08016 -0.35 0.7247
## Year1999 -0.09845 0.06642 -1.48 0.1386
```

```

## Year2000      -0.09570    0.09155   -1.05    0.2961
## Year2001      -0.06944    0.07478   -0.93    0.3533
## Year2002      -0.21451    0.07398   -2.90    0.0038 **
## Year2003      -0.19655    0.07620   -2.58    0.0100 *
## Year2004      -0.21404    0.07037   -3.04    0.0024 **
## Year2005      -0.20354    0.07225   -2.82    0.0049 **
## Year2006      -0.21720    0.07674   -2.83    0.0047 **
## Year2007      -0.19154    0.07672   -2.50    0.0127 *
## Year2008      -0.20203    0.07031   -2.87    0.0041 **
## Year2009      -0.13435    0.08017   -1.68    0.0941 .
## Year2010      -0.15726    0.07129   -2.21    0.0276 *
## Year2011      -0.19413    0.06945   -2.80    0.0053 **
## Year2012      -0.22455    0.07326   -3.07    0.0022 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.382
## Multiple R-squared:  0.0648, Adjusted R-squared:  0.0459
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 1023 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.025  0.861  0.952  0.898  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.00e-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  1.077 1          1.038
## Year              1.102 16          1.003

```

Residuals from first and last author



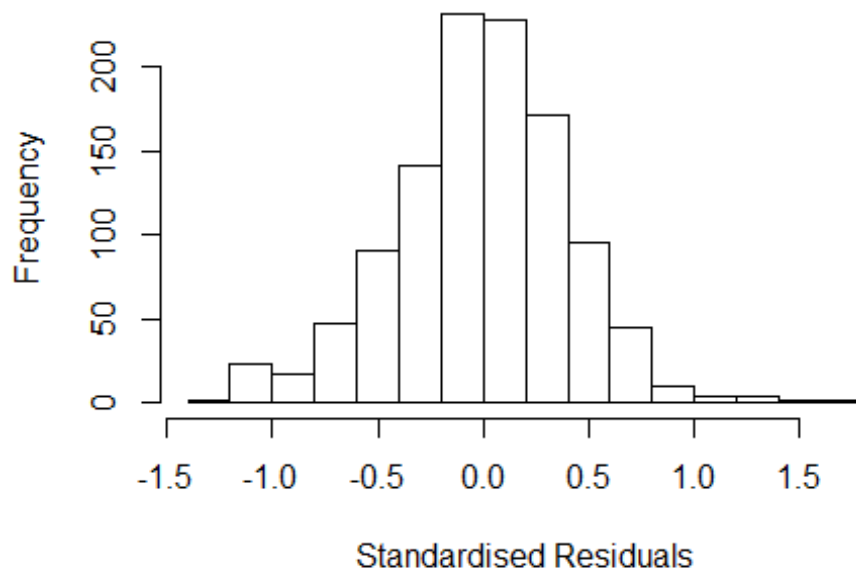
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24999 -0.23950  0.00929  0.25642  1.67806
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2240     0.0546   22.43  <2e-16 ***
## FirstAuthorFemale1  0.0150     0.0253    0.59  0.5540
## LastAuthorFemale1 -0.0556     0.0261   -2.13  0.0334 *
## Year1997          0.0260     0.0720    0.36  0.7186
## Year1998         -0.0601     0.0805   -0.75  0.4554
## Year1999         -0.0883     0.0660   -1.34  0.1814
## Year2000         -0.0955     0.0901   -1.06  0.2892
## Year2001         -0.0568     0.0731   -0.78  0.4374
## Year2002         -0.2189     0.0729   -3.00  0.0027 **
## Year2003         -0.1831     0.0776   -2.36  0.0184 *
## Year2004         -0.1940     0.0690   -2.81  0.0050 **
## Year2005         -0.1818     0.0703   -2.59  0.0098 **
```

```

## Year2006          -0.2443      0.0758   -3.23   0.0013 **
## Year2007          -0.1732      0.0741   -2.34   0.0196 *
## Year2008          -0.1729      0.0688   -2.51   0.0121 *
## Year2009          -0.0881      0.0790   -1.12   0.2651
## Year2010          -0.1371      0.0725   -1.89   0.0588 .
## Year2011          -0.1615      0.0683   -2.36   0.0182 *
## Year2012          -0.1967      0.0737   -2.67   0.0077 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0363, Adjusted R-squared:  0.0204
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 94 weights are ~= 1. The remaining 1017 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0132 0.8640 0.9510 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.00e-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 1.053 1      1.026
## Year      1.053 16      1.002

```

Residuals from first author



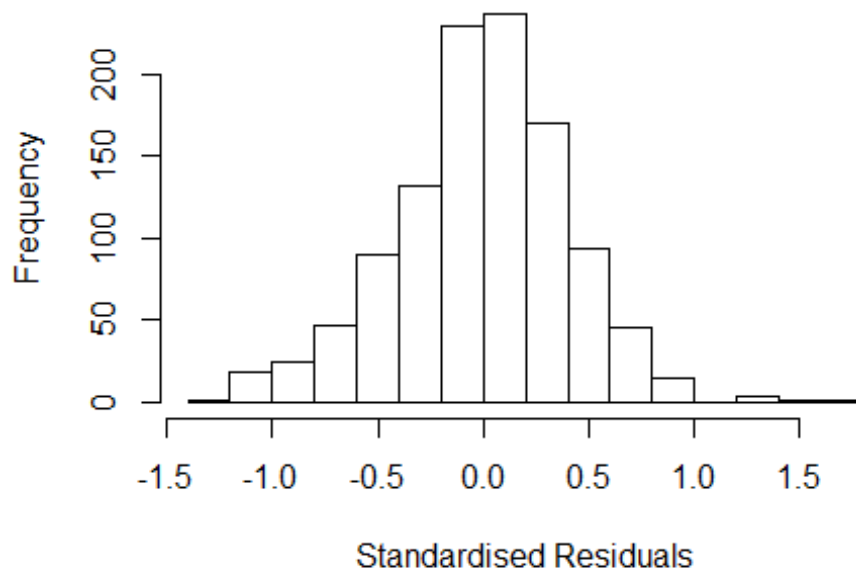
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23918 -0.24797 0.00497 0.25529 1.69249
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20881 0.05346 22.61 <2e-16 ***
## FirstAuthorFemale1 0.00256 0.02486 0.10 0.9179
## Year1997 0.03037 0.07191 0.42 0.6728
## Year1998 -0.06032 0.07976 -0.76 0.4496
## Year1999 -0.08408 0.06542 -1.29 0.1990
## Year2000 -0.09604 0.08924 -1.08 0.2821
## Year2001 -0.05178 0.07311 -0.71 0.4789
## Year2002 -0.22204 0.07226 -3.07 0.0022 **
## Year2003 -0.17878 0.07719 -2.32 0.0207 *
## Year2004 -0.18939 0.06859 -2.76 0.0059 **
## Year2005 -0.18559 0.06978 -2.66 0.0079 **
## Year2006 -0.23892 0.07541 -3.17 0.0016 **
```

```

## Year2007          -0.17284    0.07386   -2.34    0.0195 *
## Year2008          -0.17197    0.06805   -2.53    0.0116 *
## Year2009          -0.08730    0.07836   -1.11    0.2655
## Year2010          -0.13539    0.07197   -1.88    0.0602 .
## Year2011          -0.16118    0.06806   -2.37    0.0181 *
## Year2012          -0.19132    0.07309   -2.62    0.0090 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0322, Adjusted R-squared:  0.0172
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 106 weights are ~= 1. The remaining 1005 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0099 0.8590 0.9500 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.00e-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 1.036 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
## -1.25625 -0.23893 0.00878 0.25488 1.67095
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2286 0.0540 22.77 <2e-16 ***
## LastAuthorFemale1 -0.0519 0.0256 -2.02 0.0431 *
## Year1997 0.0276 0.0716 0.39 0.6994
## Year1998 -0.0597 0.0808 -0.74 0.4597
## Year1999 -0.0878 0.0659 -1.33 0.1831
## Year2000 -0.0925 0.0893 -1.04 0.3004
## Year2001 -0.0564 0.0734 -0.77 0.4424
## Year2002 -0.2188 0.0728 -3.01 0.0027 **
## Year2003 -0.1804 0.0772 -2.34 0.0197 *
## Year2004 -0.1924 0.0689 -2.79 0.0053 **
## Year2005 -0.1809 0.0702 -2.58 0.0101 *
## Year2006 -0.2433 0.0756 -3.22 0.0013 **
```

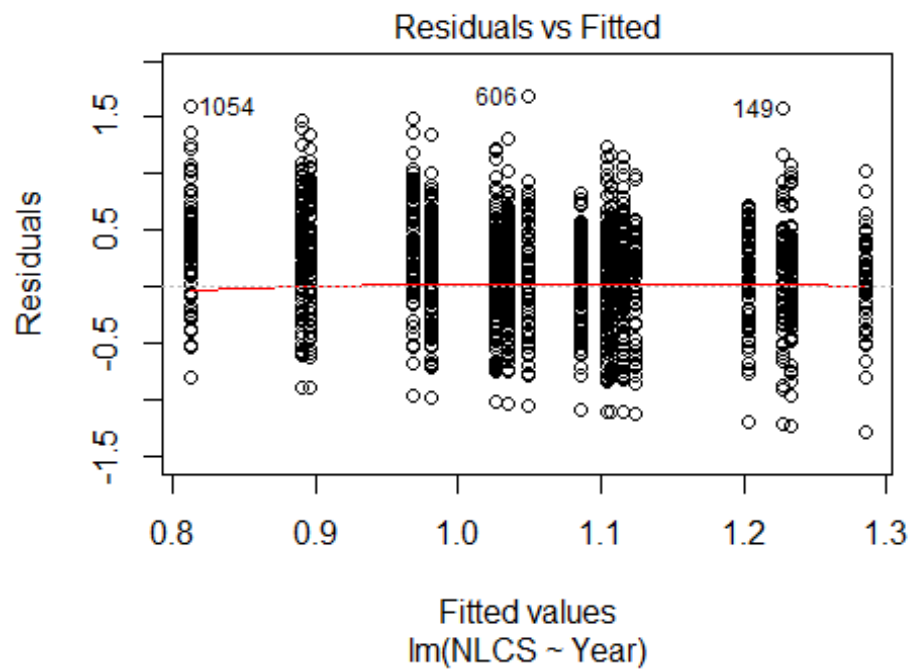
```

## Year2007          -0.1716      0.0740   -2.32   0.0207 *
## Year2008          -0.1725      0.0687   -2.51   0.0122 *
## Year2009          -0.0856      0.0786   -1.09   0.2763
## Year2010          -0.1356      0.0723   -1.87   0.0611 .
## Year2011          -0.1586      0.0678   -2.34   0.0195 *
## Year2012          -0.1934      0.0732   -2.64   0.0083 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.036, Adjusted R-squared:  0.021
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 93 weights are ~= 1. The remaining 1018 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0145 0.8610 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      9.00e-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: 1111"
## [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
##   90  100   72   77  125  136  136  152  148  164  160  175  166  186  178
## 2011 2012
##  191  198
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   77   82   53   54   67   69  116  133  121  140  127  153  147  150  155
## 2011 2012

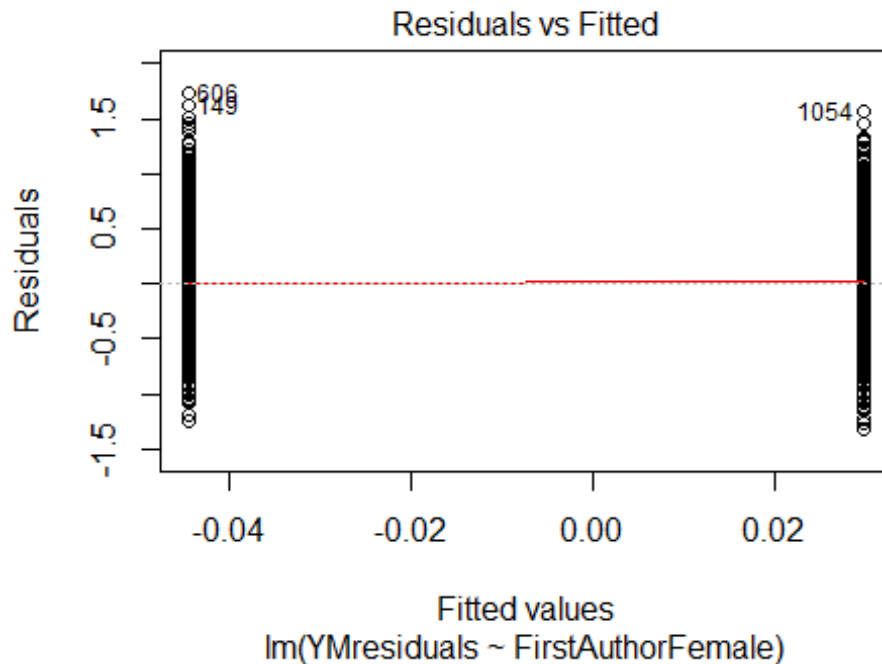
```



```
## 157 168
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 71 76 50 50 63 62 104 121 114 126 112 135 136 137 142
## 2011 2012
## 142 153
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 85, df = 16, p-value = 2e-11
```

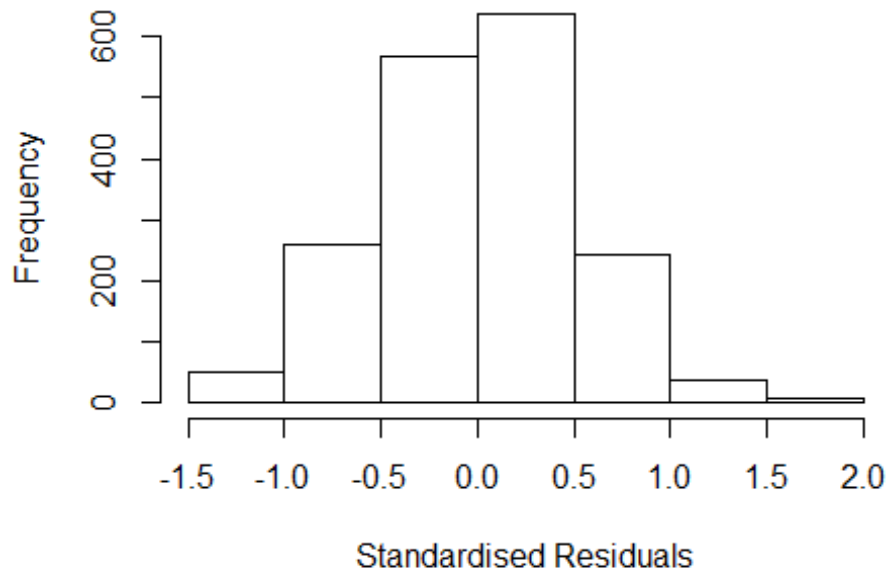


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 1e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.45172591584897"
## [1] "Male first author team size 2018 geometric mean: 4.5130620676879"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3400, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.25031136604255"
## [1] "Male last author team size 2018 geometric mean: 4.76042765230148"
##
## 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 1.124 1      1.060
## LastAuthorFemale  1.128 1      1.062
## UniqueAuthors    1.341 4      1.037
## Year              1.345 16     1.009
```

Residuals from first and last author and team size



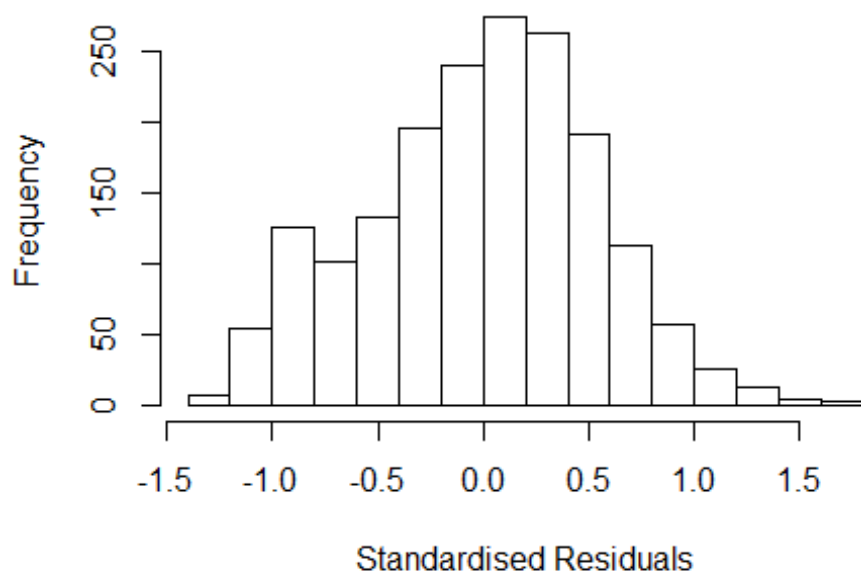
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4653 -0.3474 0.0156 0.3542 1.9167
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.89831 0.07294 12.32 < 2e-16 ***
## FirstAuthorFemale1 0.06202 0.02757 2.25 0.02462 *
## LastAuthorFemale1 0.03653 0.02645 1.38 0.16746
## UniqueAuthors2 0.26402 0.04810 5.49 4.6e-08 ***
## UniqueAuthors3 0.38962 0.04614 8.44 < 2e-16 ***
## UniqueAuthors4 0.42954 0.04791 8.97 < 2e-16 ***
## UniqueAuthors5 0.50843 0.04538 11.20 < 2e-16 ***
## Year1997 -0.00802 0.08681 -0.09 0.92642
## Year1998 -0.06504 0.09244 -0.70 0.48177
## Year1999 0.02202 0.08555 0.26 0.79695
```

```

## Year2000      -0.06064      0.08749      -0.69      0.48832
## Year2001      -0.29989      0.08953      -3.35      0.00083 ***
## Year2002      -0.16970      0.08035      -2.11      0.03482 *
## Year2003      -0.43658      0.08552      -5.11      3.7e-07 ***
## Year2004      -0.29081      0.09285      -3.13      0.00176 **
## Year2005      -0.37837      0.08344      -4.53      6.2e-06 ***
## Year2006      -0.41168      0.08983      -4.58      4.9e-06 ***
## Year2007      -0.30447      0.08044      -3.78      0.00016 ***
## Year2008      -0.21388      0.07334      -2.92      0.00359 **
## Year2009      -0.27749      0.07843      -3.54      0.00041 ***
## Year2010      -0.20500      0.07480      -2.74      0.00619 **
## Year2011      -0.25986      0.07586      -3.43      0.00063 ***
## Year2012      -0.22446      0.07265      -3.09      0.00204 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.152, Adjusted R-squared:  0.142
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 147 weights are ~= 1. The remaining 1647 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.104  0.864  0.945   0.902   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          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 1.081 1      1.040
## LastAuthorFemale  1.071 1      1.035
## Year              1.056 16      1.002

```

Residuals from first and last author



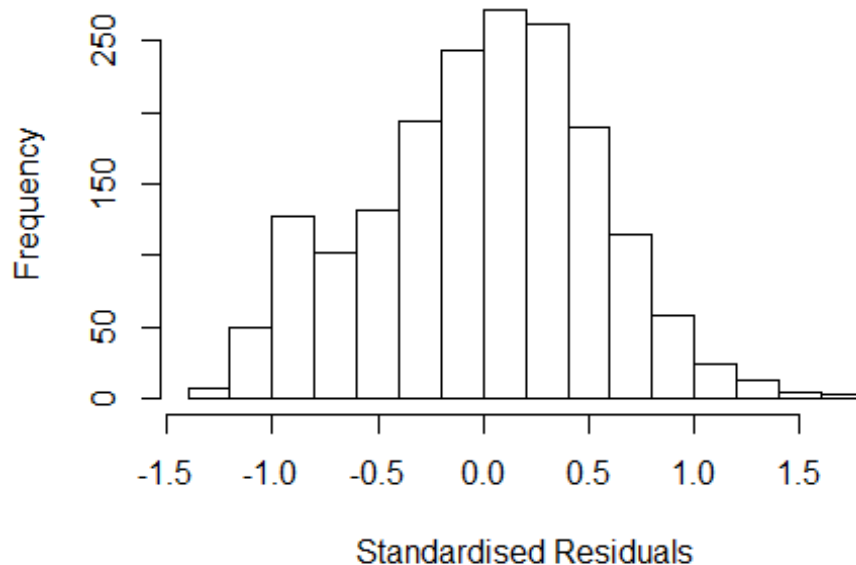
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3193 -0.3657  0.0305  0.3701  1.7645
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20197    0.06408   18.76 < 2e-16 ***
## FirstAuthorFemale1 0.08704    0.02942    2.96  0.0031 **
## LastAuthorFemale1 0.00857    0.02738    0.31  0.7544
## Year1997        -0.05030    0.08592   -0.59  0.5583
## Year1998        -0.11709    0.09574   -1.22  0.2215
## Year1999         0.03032    0.08475    0.36  0.7206
## Year2000        -0.02435    0.08558   -0.28  0.7760
## Year2001        -0.23348    0.09235   -2.53  0.0115 *
## Year2002        -0.16048    0.08024   -2.00  0.0457 *
## Year2003        -0.47412    0.09471   -5.01 6.1e-07 ***
## Year2004        -0.30546    0.10784   -2.83  0.0047 **
## Year2005        -0.36673    0.08938   -4.10 4.3e-05 ***
```

```

## Year2006      -0.39034    0.09892   -3.95  8.3e-05 ***
## Year2007      -0.25605    0.08077   -3.17  0.0016 **
## Year2008      -0.15724    0.07271   -2.16  0.0307 *
## Year2009      -0.20835    0.07699   -2.71  0.0069 **
## Year2010      -0.14192    0.07493   -1.89  0.0584 .
## Year2011      -0.20913    0.07441   -2.81  0.0050 **
## Year2012      -0.13782    0.07117   -1.94  0.0529 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.513
## Multiple R-squared:  0.0605, Adjusted R-squared:  0.0509
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 150 weights are ~= 1. The remaining 1644 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.212  0.845   0.945   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      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 1.036 1      1.018
## Year      1.036 16      1.001

```

Residuals from first author



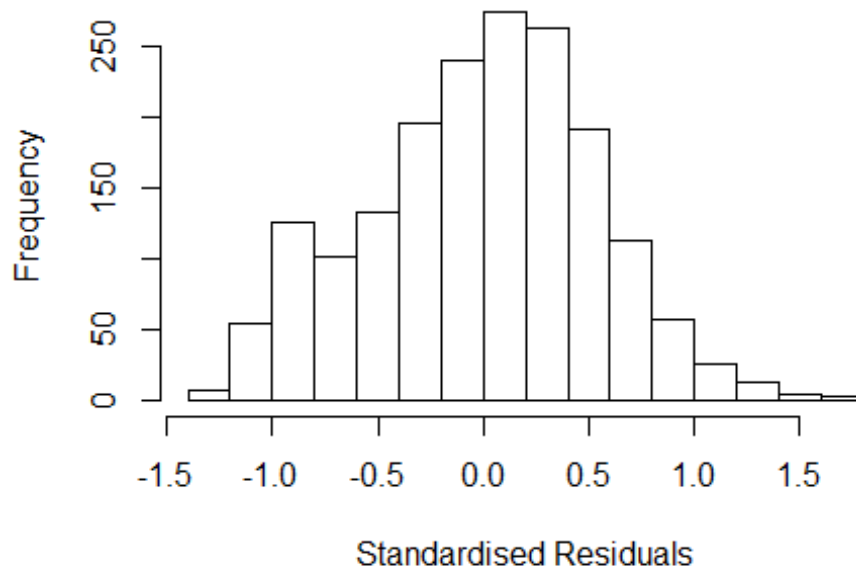
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3243 -0.3631 0.0289 0.3679 1.7620
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2041 0.0637 18.91 < 2e-16 ***
## FirstAuthorFemale1 0.0897 0.0289 3.11 0.0019 **
## Year1997 -0.0494 0.0856 -0.58 0.5641
## Year1998 -0.1171 0.0955 -1.23 0.2205
## Year1999 0.0305 0.0846 0.36 0.7181
## Year2000 -0.0246 0.0855 -0.29 0.7736
## Year2001 -0.2331 0.0923 -2.53 0.0116 *
## Year2002 -0.1604 0.0801 -2.00 0.0455 *
## Year2003 -0.4727 0.0937 -5.04 5.0e-07 ***
## Year2004 -0.3057 0.1074 -2.85 0.0045 **
## Year2005 -0.3659 0.0891 -4.11 4.2e-05 ***
## Year2006 -0.3899 0.0987 -3.95 8.1e-05 ***
```

```

## Year2007          -0.2553      0.0804   -3.18   0.0015 **
## Year2008          -0.1562      0.0724   -2.16   0.0312 *
## Year2009          -0.2081      0.0768   -2.71   0.0068 **
## Year2010          -0.1411      0.0746   -1.89   0.0587 .
## Year2011          -0.2084      0.0741   -2.81   0.0050 **
## Year2012          -0.1374      0.0710   -1.94   0.0530 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.516
## Multiple R-squared:  0.0601, Adjusted R-squared:  0.0511
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 150 weights are ~= 1. The remaining 1644 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.220  0.847   0.946   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      5.57e-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 1.028 1          1.014
## Year              1.028 16          1.001

```


Residuals from last author



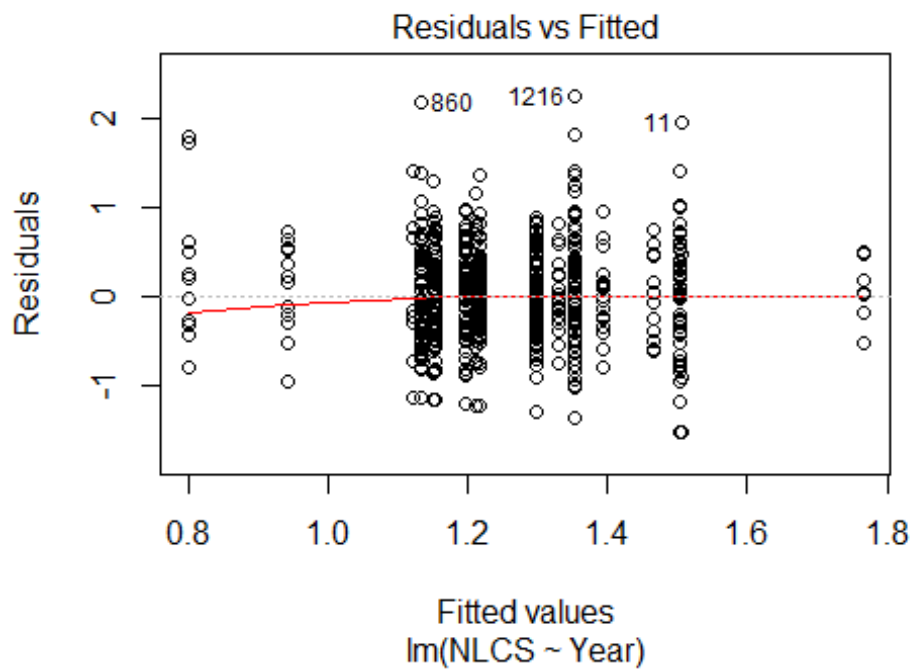
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3063 -0.3575 0.0342 0.3710 1.7276
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2345 0.0639 19.32 < 2e-16 ***
## LastAuthorFemale1 0.0341 0.0271 1.26 0.20851
## Year1997 -0.0372 0.0865 -0.43 0.66702
## Year1998 -0.1076 0.0966 -1.11 0.26559
## Year1999 0.0378 0.0844 0.45 0.65464
## Year2000 -0.0193 0.0868 -0.22 0.82413
## Year2001 -0.2291 0.0928 -2.47 0.01367 *
## Year2002 -0.1504 0.0808 -1.86 0.06272 .
## Year2003 -0.4770 0.0967 -4.93 8.9e-07 ***
## Year2004 -0.2999 0.1092 -2.75 0.00607 **
## Year2005 -0.3666 0.0915 -4.01 6.4e-05 ***
## Year2006 -0.3801 0.1007 -3.78 0.00017 ***
```

```

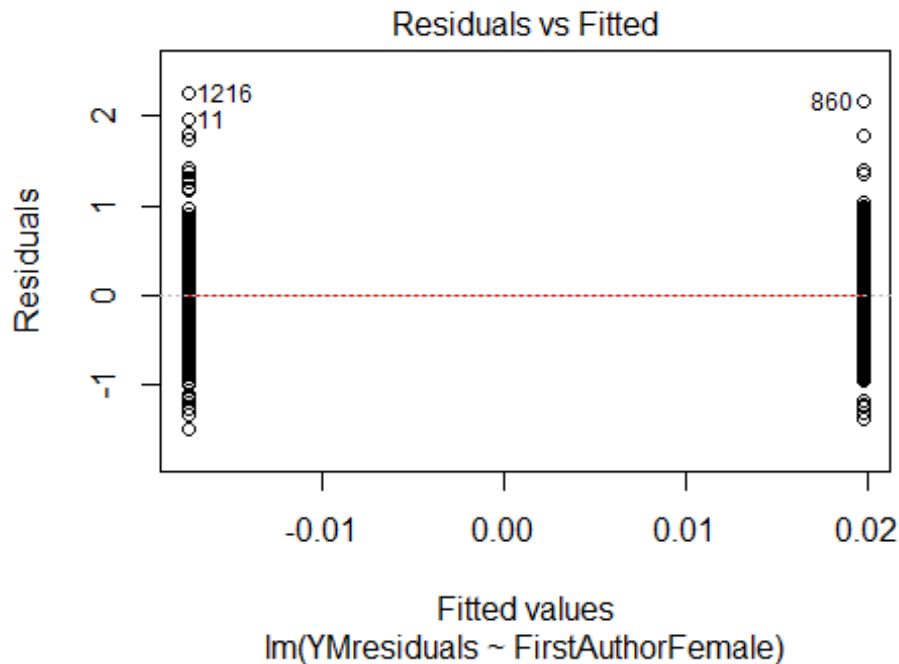
## Year2007          -0.2480      0.0819   -3.03  0.00251 **
## Year2008          -0.1498      0.0737   -2.03  0.04213 *
## Year2009          -0.1983      0.0779   -2.55  0.01096 *
## Year2010          -0.1258      0.0758   -1.66  0.09725 .
## Year2011          -0.1957      0.0754   -2.60  0.00948 **
## Year2012          -0.1239      0.0720   -1.72  0.08555 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.513
## Multiple R-squared:  0.055, Adjusted R-squared:  0.0459
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 146 weights are ~ = 1. The remaining 1648 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.234  0.846  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      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: 1794"
## [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
##   11   26   20   19   62   27   29   31   29   45   78   80  101  100   92
## 2011 2012
##  121  139
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4   15    8    9   16   14   20   21   24   32   55   64   79   74   75
## 2011 2012

```

```
## 96 108
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 14 3 8 12 11 20 19 20 30 48 49 60 62 67
## 2011 2012
## 87 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 = 64, df = 16, p-value = 1e-07
```

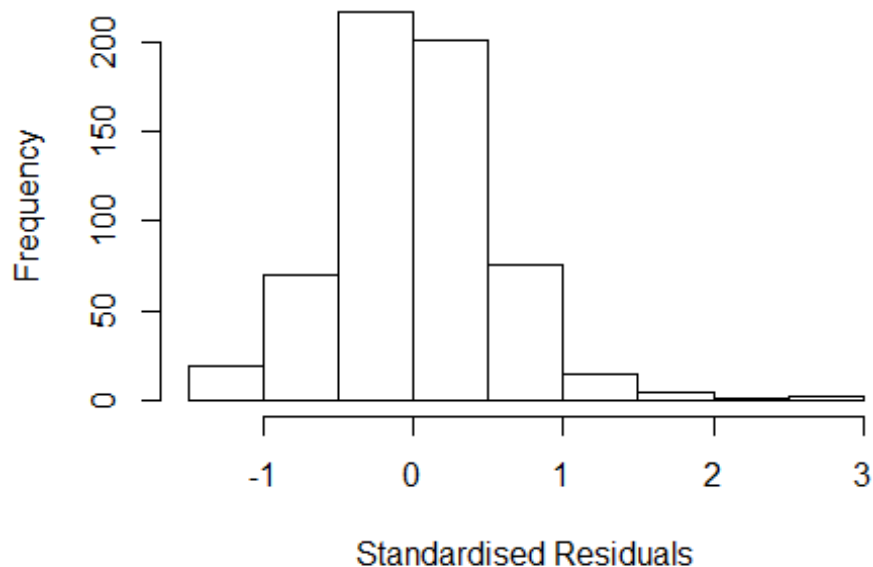


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.3, df = 1, p-value = 0.07
```



```
## [1] "Female first author team size 2018 geometric mean: 4.98556729378089"
## [1] "Male first author team size 2018 geometric mean: 3.98918574831221"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3200, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.75637150606388"
## [1] "Male last author team size 2018 geometric mean: 4.33655200361026"
##
## 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 1.273 1          1.128
## LastAuthorFemale  1.170 1          1.082
## UniqueAuthors    1.947 4          1.087
## Year              2.387 16         1.028
```

Residuals from first and last author and team size



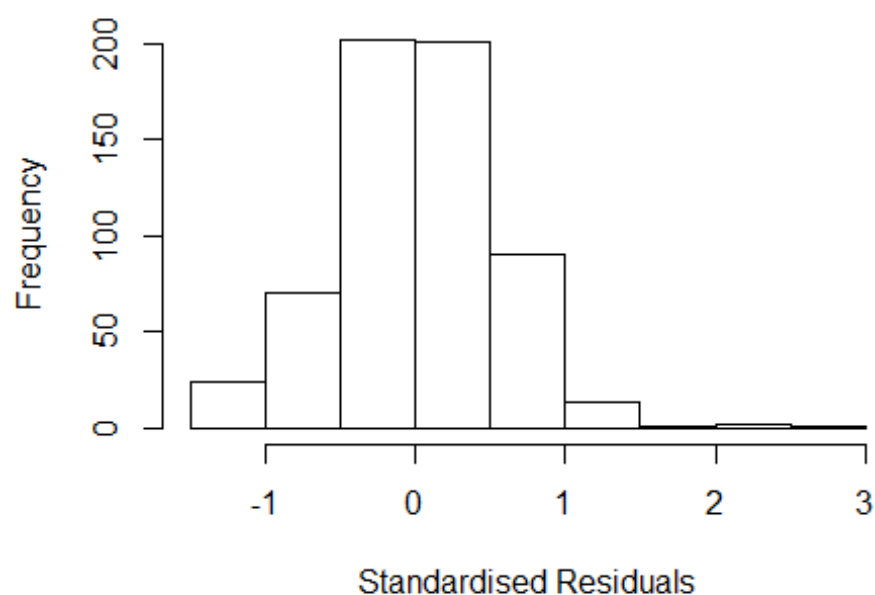
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 11      0030153642 3.451 1996      1704      5      2.777
## 1216 84859897031 3.590 2011      2718      1      2.639
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4750 -0.3494 -0.0138  0.3493  2.7770
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.4302     0.7238   0.59   0.553
## FirstAuthorFemale1 0.0393     0.0463   0.85   0.396
## LastAuthorFemale1 -0.0165     0.0456  -0.36   0.717
## UniqueAuthors2     0.2439     0.0946   2.58   0.010 *
## UniqueAuthors3     0.2145     0.0906   2.37   0.018 *
## UniqueAuthors4     0.3736     0.0893   4.18 3.3e-05 ***
## UniqueAuthors5     0.4289     0.0795   5.40 9.9e-08 ***
## Year1997           0.4177     0.7334   0.57   0.569
## Year1998           0.9360     0.7377   1.27   0.205
```

```

## Year1999      0.4205      0.7934      0.53      0.596
## Year2000     -0.2248      0.7407     -0.30      0.762
## Year2001      0.7569      0.7368      1.03      0.305
## Year2002      0.7216      0.7256      0.99      0.320
## Year2003      0.5533      0.7233      0.76      0.445
## Year2004      0.4123      0.7295      0.57      0.572
## Year2005      0.5362      0.7279      0.74      0.462
## Year2006      0.8010      0.7294      1.10      0.273
## Year2007      0.5107      0.7255      0.70      0.482
## Year2008      0.4843      0.7223      0.67      0.503
## Year2009      0.3877      0.7242      0.54      0.593
## Year2010      0.2977      0.7232      0.41      0.681
## Year2011      0.5204      0.7245      0.72      0.473
## Year2012      0.5050      0.7217      0.70      0.484
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.152, Adjusted R-squared:  0.12
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 2 observations c(4,512) are outliers with |weight| = 0 ( < 0.00017);
## 55 weights are ~= 1. The remaining 548 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0379 0.8810 0.9500 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      1.65e-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 1.198 1      1.094
## LastAuthorFemale 1.129 1      1.062
## Year      1.233 16      1.007

```

Residuals from first and last author



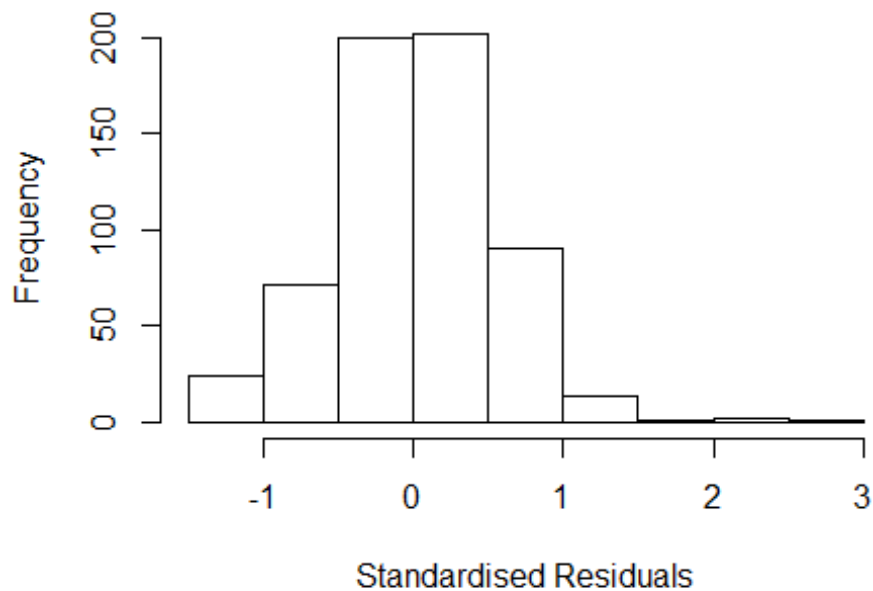
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 11 0030153642 3.451 1996      1704      5      2.757
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4808 -0.3438  0.0107  0.3411  2.7565
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.69447    0.70504   0.99   0.33
## FirstAuthorFemale1 0.06030    0.04694   1.28   0.20
## LastAuthorFemale1 -0.00772    0.04660  -0.17   0.87
## Year1997         0.33746    0.72033   0.47   0.64
## Year1998         0.97310    0.72804   1.34   0.18
## Year1999         0.34441    0.79830   0.43   0.67
## Year2000        -0.26422    0.72609  -0.36   0.72
## Year2001         0.65791    0.72355   0.91   0.36
## Year2002         0.68142    0.71050   0.96   0.34
## Year2003         0.57478    0.70951   0.81   0.42
## Year2004         0.41208    0.71663   0.58   0.57
## Year2005         0.51746    0.71523   0.72   0.47
```

```

## Year2006          0.78628    0.71304    1.10    0.27
## Year2007          0.53881    0.71102    0.76    0.45
## Year2008          0.49173    0.70718    0.70    0.49
## Year2009          0.44801    0.70937    0.63    0.53
## Year2010          0.35149    0.70891    0.50    0.62
## Year2011          0.60357    0.71014    0.85    0.40
## Year2012          0.56810    0.70746    0.80    0.42
##
## Robust residual standard error: 0.523
## Multiple R-squared:  0.092, Adjusted R-squared:  0.0641
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## observation 4 is an outlier with |weight| = 0 ( < 0.00017);
## 48 weights are ~= 1. The remaining 556 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0156 0.8730 0.9520 0.9000 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.65e-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 1.155 1      1.075
## Year              1.155 16      1.005

```


Residuals from first author



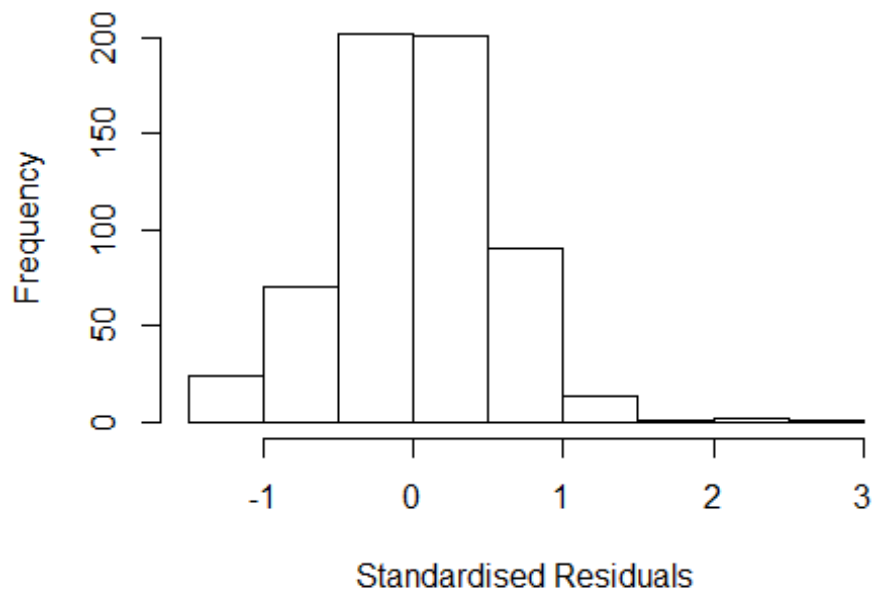
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 11 0030153642 3.451 1996      1704      5      2.757
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.47953 -0.34494  0.00951  0.34251  2.75803
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.6930      0.7061   0.98   0.33
## FirstAuthorFemale1 0.0579      0.0465   1.25   0.21
## Year1997          0.3391      0.7214   0.47   0.64
## Year1998          0.9746      0.7291   1.34   0.18
## Year1999          0.3430      0.7987   0.43   0.67
## Year2000         -0.2657      0.7268  -0.37   0.71
## Year2001          0.6572      0.7243   0.91   0.36
## Year2002          0.6814      0.7115   0.96   0.34
## Year2003          0.5751      0.7105   0.81   0.42
## Year2004          0.4096      0.7178   0.57   0.57
## Year2005          0.5166      0.7164   0.72   0.47
## Year2006          0.7866      0.7141   1.10   0.27
```

```

## Year2007          0.5381      0.7120      0.76      0.45
## Year2008          0.4920      0.7082      0.69      0.49
## Year2009          0.4474      0.7102      0.63      0.53
## Year2010          0.3507      0.7097      0.49      0.62
## Year2011          0.6031      0.7111      0.85      0.40
## Year2012          0.5676      0.7085      0.80      0.42
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.0922, Adjusted R-squared:  0.066
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## observation 4 is an outlier with |weight| = 0 ( < 0.00017);
## 46 weights are ~= 1. The remaining 558 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0142 0.8730 0.9530 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          1.65e-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 1.092 1          1.045
## Year            1.092 16          1.003

```

Residuals from last author



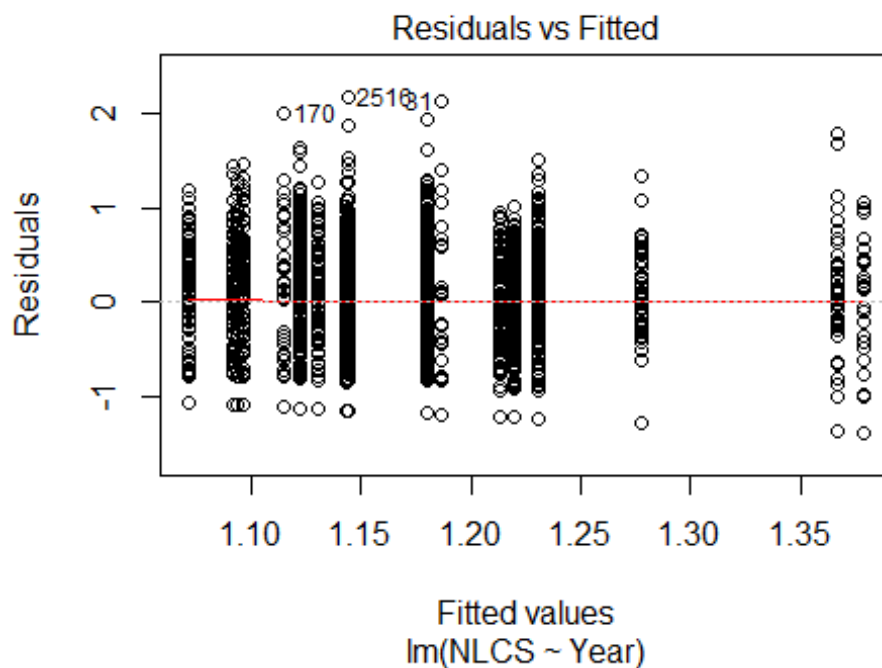
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 11 0030153642 3.451 1996    1704      5      2.757
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.508 -0.347  0.015  0.346  2.758
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6931    0.7060    0.98   0.33
## LastAuthorFemale1 0.0116    0.0462    0.25   0.80
## Year1997        0.3389    0.7213    0.47   0.64
## Year1998        0.9745    0.7290    1.34   0.18
## Year1999        0.3526    0.7966    0.44   0.66
## Year2000       -0.2535    0.7282   -0.35   0.73
## Year2001        0.6694    0.7227    0.93   0.35
## Year2002        0.6896    0.7117    0.97   0.33
## Year2003        0.5934    0.7104    0.84   0.40
## Year2004        0.4247    0.7176    0.59   0.55
## Year2005        0.5385    0.7156    0.75   0.45
## Year2006        0.8145    0.7138    1.14   0.25
```

```

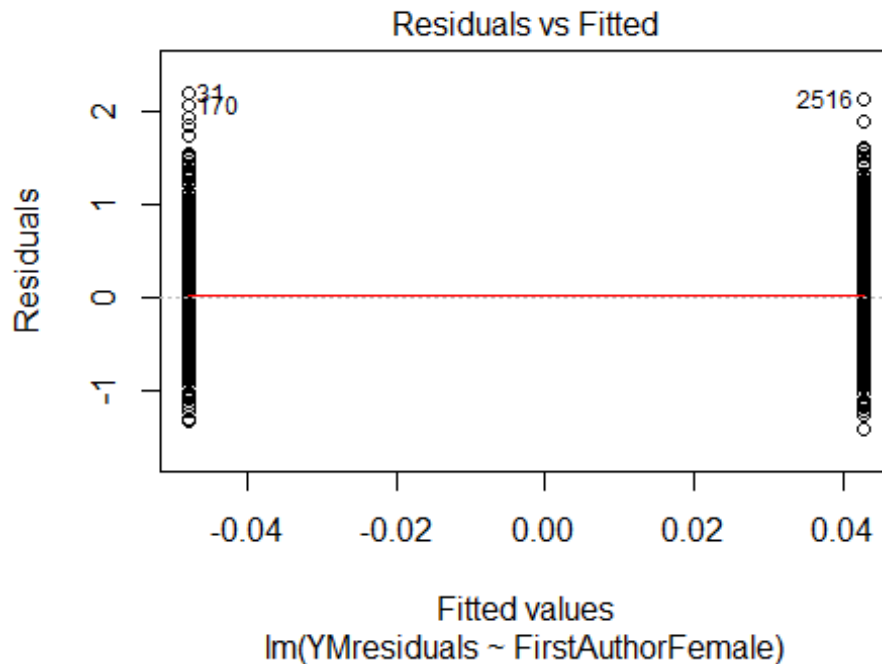
## Year2007          0.5583      0.7119      0.78      0.43
## Year2008          0.5178      0.7077      0.73      0.46
## Year2009          0.4757      0.7099      0.67      0.50
## Year2010          0.3771      0.7096      0.53      0.60
## Year2011          0.6305      0.7105      0.89      0.38
## Year2012          0.5996      0.7080      0.85      0.40
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.09,   Adjusted R-squared:  0.0637
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## observation 4 is an outlier with |weight| = 0 ( < 0.00017);
## 50 weights are ~= 1. The remaining 554 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0198 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          1.65e-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: 605"
## [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
##   60   66   79   92  100  113  108  125  126  136  159  197  251  318  382
## 2011 2012
##  380  428
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   33   39   51   48   79   80  101  109  113  125  167  210  260  331
## 2011 2012
##  327  366

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   27   30   32   47   45   73   75   94   96  102  111  151  182  238  302
## 2011 2012
##  305  330
## [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 = 8e-08
```

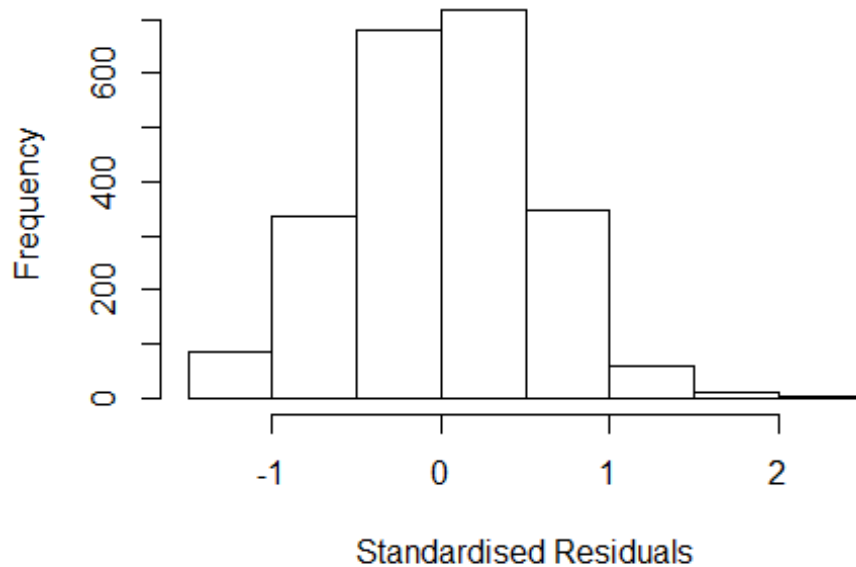


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.025, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 3.82334105408737"
## [1] "Male first author team size 2018 geometric mean: 2.80746877096922"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26000, p-value = 3e-05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.45958658881438"
## [1] "Male last author team size 2018 geometric mean: 3.32824238407285"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 22000, 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 1.265 1          1.125
## LastAuthorFemale  1.194 1          1.093
## UniqueAuthors    1.235 4          1.027
## Year              1.229 16         1.006
```

Residuals from first and last author and team size



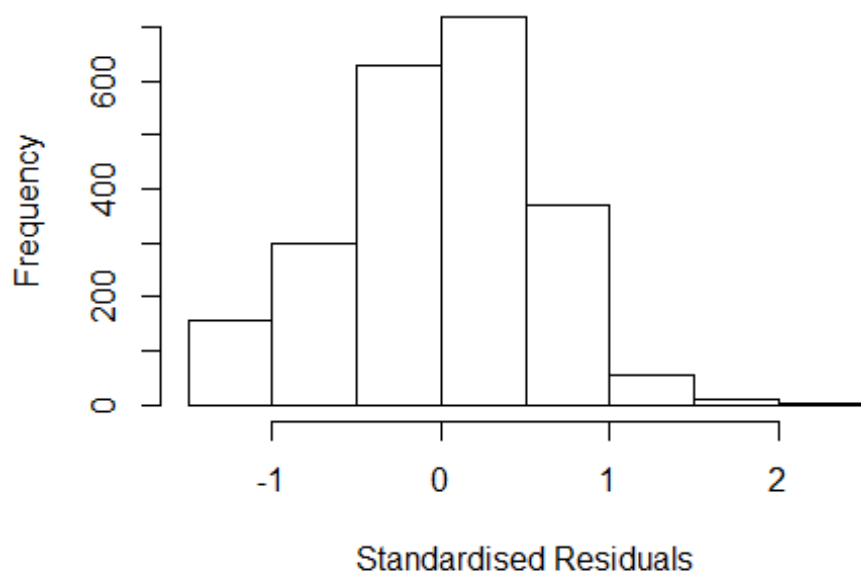
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4470 -0.3800 0.0134 0.3822 2.2187
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8102 0.1523 5.32 1.1e-07 ***
## FirstAuthorFemale1 0.0528 0.0273 1.94 0.053 .
## LastAuthorFemale1 -0.0216 0.0268 -0.81 0.420
## UniqueAuthors2 0.1694 0.0419 4.05 5.4e-05 ***
## UniqueAuthors3 0.2921 0.0447 6.53 8.3e-11 ***
## UniqueAuthors4 0.3944 0.0440 8.97 < 2e-16 ***
## UniqueAuthors5 0.3858 0.0346 11.14 < 2e-16 ***
## Year1997 0.4673 0.1884 2.48 0.013 *
## Year1998 0.1305 0.2396 0.54 0.586
## Year1999 0.2487 0.1697 1.47 0.143
```

```

## Year2000          0.3085      0.1818      1.70      0.090 .
## Year2001          0.0928      0.1684      0.55      0.582
## Year2002          0.1137      0.1740      0.65      0.514
## Year2003          0.0636      0.1688      0.38      0.706
## Year2004          0.0606      0.1635      0.37      0.711
## Year2005          0.0307      0.1606      0.19      0.849
## Year2006          0.1483      0.1575      0.94      0.346
## Year2007          0.0856      0.1561      0.55      0.583
## Year2008          0.1352      0.1546      0.87      0.382
## Year2009          0.1495      0.1552      0.96      0.335
## Year2010          0.0518      0.1539      0.34      0.737
## Year2011          0.0290      0.1544      0.19      0.851
## Year2012          0.0821      0.1534      0.54      0.593
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.564
## Multiple R-squared:  0.0871, Adjusted R-squared:  0.078
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 169 weights are ~= 1. The remaining 2071 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0872 0.8710 0.9500 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.46e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.168 1          1.081
## LastAuthorFemale 1.146 1          1.070
## Year          1.079 16          1.002

```


Residuals from first and last author



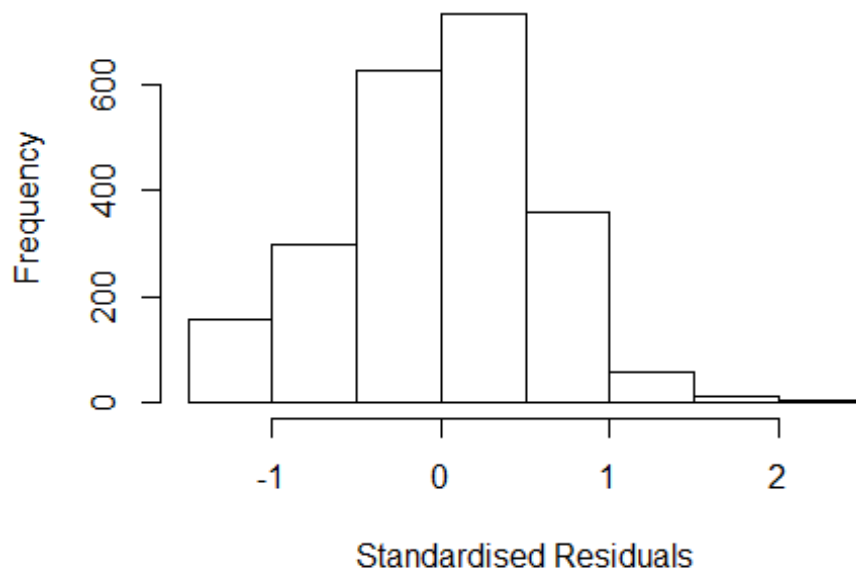
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4383 -0.3879 0.0283 0.3952 2.3068
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0142 0.1707 5.94 3.3e-09 ***
## FirstAuthorFemale1 0.1150 0.0273 4.21 2.6e-05 ***
## LastAuthorFemale1 -0.0314 0.0273 -1.15 0.250
## Year1997 0.3851 0.2088 1.84 0.065 .
## Year1998 0.0459 0.2484 0.18 0.853
## Year1999 0.1914 0.1880 1.02 0.309
## Year2000 0.3091 0.2000 1.55 0.122
## Year2001 0.0528 0.1900 0.28 0.781
## Year2002 0.0763 0.1903 0.40 0.689
## Year2003 0.0346 0.1894 0.18 0.855
## Year2004 0.0559 0.1830 0.31 0.760
## Year2005 0.0545 0.1811 0.30 0.764
```

```

## Year2006          0.1706      0.1774      0.96      0.336
## Year2007          0.1095      0.1769      0.62      0.536
## Year2008          0.1564      0.1746      0.90      0.371
## Year2009          0.1575      0.1755      0.90      0.370
## Year2010          0.0570      0.1743      0.33      0.744
## Year2011          0.0481      0.1749      0.28      0.783
## Year2012          0.1159      0.1742      0.67      0.506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.584
## Multiple R-squared:  0.0196, Adjusted R-squared:  0.0116
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 184 weights are ~= 1. The remaining 2056 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.084  0.871  0.950  0.909  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.053 1      1.026
## Year              1.053 16      1.002

```

Residuals from first author



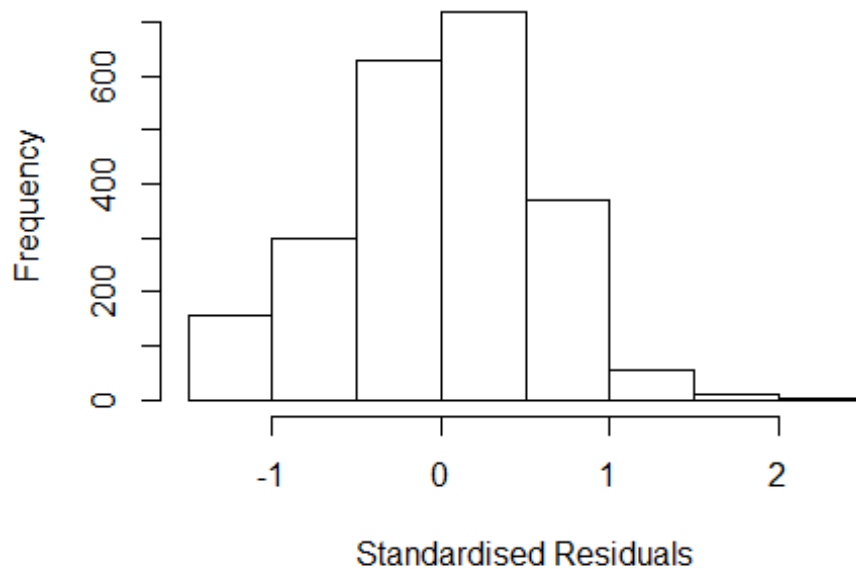
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4158 -0.3904 0.0271 0.3961 2.3130
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0080 0.1702 5.92 3.6e-09 ***
## FirstAuthorFemale1 0.1029 0.0260 3.96 7.6e-05 ***
## Year1997 0.3860 0.2085 1.85 0.064 .
## Year1998 0.0454 0.2481 0.18 0.855
## Year1999 0.1916 0.1875 1.02 0.307
## Year2000 0.3049 0.1996 1.53 0.127
## Year2001 0.0515 0.1894 0.27 0.786
## Year2002 0.0754 0.1896 0.40 0.691
## Year2003 0.0327 0.1887 0.17 0.862
## Year2004 0.0551 0.1823 0.30 0.763
## Year2005 0.0555 0.1805 0.31 0.758
## Year2006 0.1730 0.1769 0.98 0.328
```

```

## Year2007          0.1090      0.1763      0.62      0.537
## Year2008          0.1549      0.1739      0.89      0.373
## Year2009          0.1567      0.1748      0.90      0.370
## Year2010          0.0560      0.1736      0.32      0.747
## Year2011          0.0467      0.1742      0.27      0.789
## Year2012          0.1127      0.1734      0.65      0.516
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.584
## Multiple R-squared:  0.019, Adjusted R-squared:  0.0115
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 188 weights are ~= 1. The remaining 2052 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0819 0.8690 0.9490 0.9080 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      4.46e-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 1.029 1      1.014
## Year              1.029 16      1.001

```

Residuals from last author



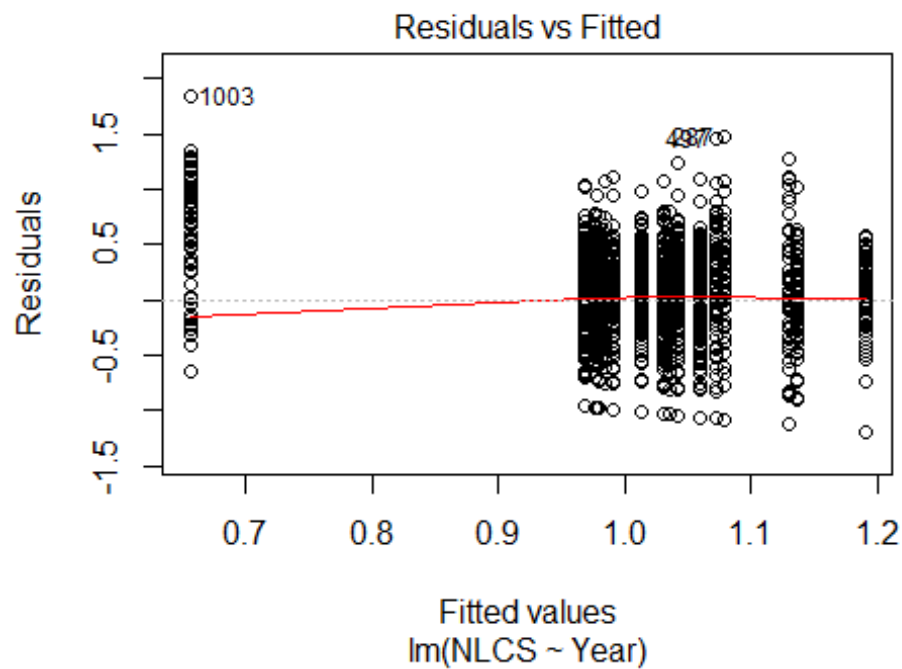
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4310 -0.3856 0.0203 0.4073 2.2881
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0329 0.1652 6.25 4.9e-10 ***
## LastAuthorFemale1 0.0132 0.0260 0.51 0.612
## Year1997 0.3981 0.2051 1.94 0.052 .
## Year1998 0.0665 0.2462 0.27 0.787
## Year1999 0.2155 0.1830 1.18 0.239
## Year2000 0.3050 0.1955 1.56 0.119
## Year2001 0.0646 0.1847 0.35 0.726
## Year2002 0.0930 0.1852 0.50 0.615
## Year2003 0.0512 0.1850 0.28 0.782
## Year2004 0.0732 0.1779 0.41 0.681
## Year2005 0.0741 0.1756 0.42 0.673
## Year2006 0.1960 0.1722 1.14 0.255
```

```

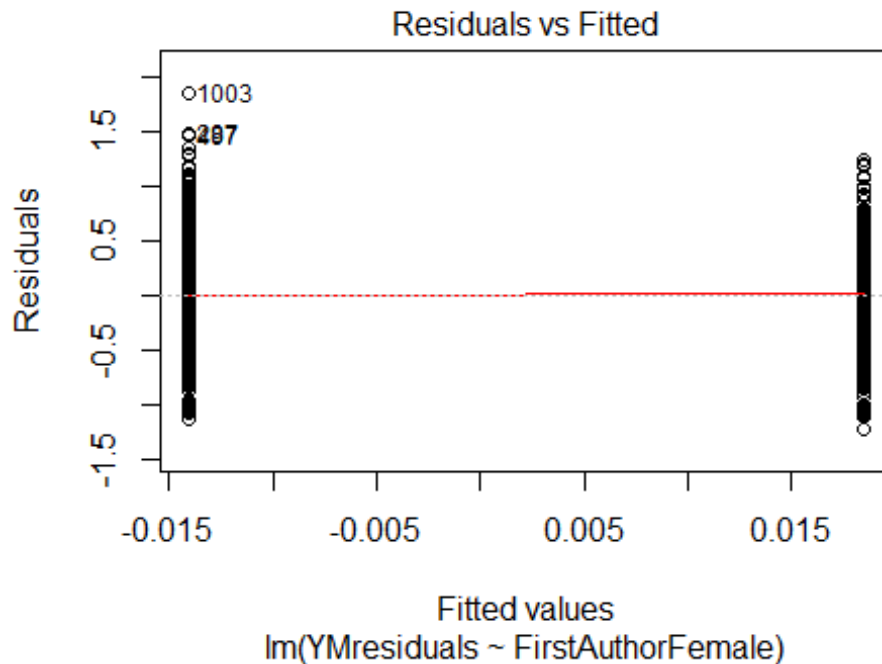
## Year2007          0.1377      0.1716      0.80      0.422
## Year2008          0.1759      0.1692      1.04      0.299
## Year2009          0.1847      0.1700      1.09      0.277
## Year2010          0.0848      0.1690      0.50      0.616
## Year2011          0.0762      0.1694      0.45      0.653
## Year2012          0.1390      0.1687      0.82      0.410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0116, Adjusted R-squared:  0.00405
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 179 weights are ~= 1. The remaining 2061 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0944 0.8670 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      4.46e-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: 2240"
## [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
## 166 143 190 130 269 144 208 114 124 154 141 188 201 195 209
## 2011 2012
## 240 242
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 55 62 76 59 154 64 128 69 85 101 85 126 142 132 145
## 2011 2012

```

```
## 167 162
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 48 54 65 54 130 54 108 55 69 88 74 109 120 118 132
## 2011 2012
## 139 141
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 160, df = 16, p-value <2e-16
```

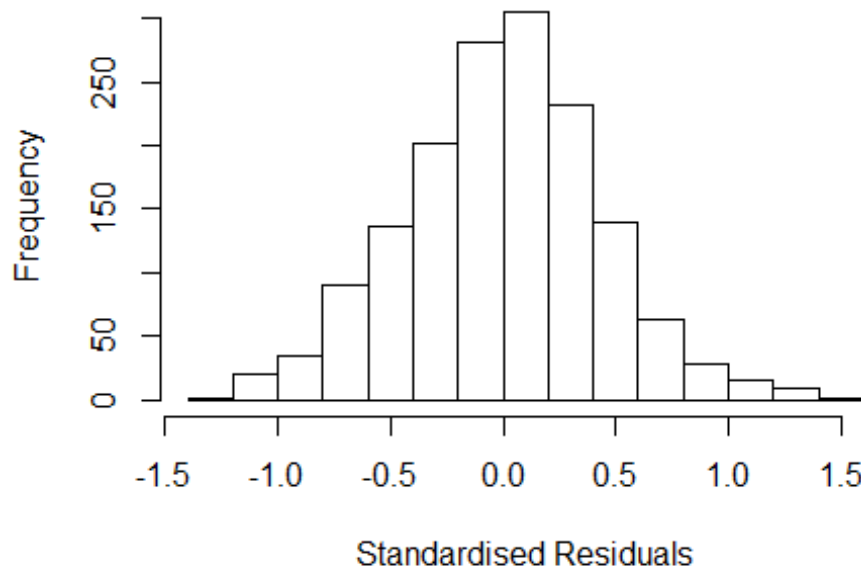


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 1e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 5.16748843835721"
## [1] "Male first author team size 2018 geometric mean: 5.80836543242541"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.55994590987724"
## [1] "Male last author team size 2018 geometric mean: 5.94579936740234"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, 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 1.094 1      1.046
## LastAuthorFemale  1.081 1      1.040
## UniqueAuthors    1.356 4      1.039
## Year              1.530 16     1.013
```


Residuals from first and last author and team size



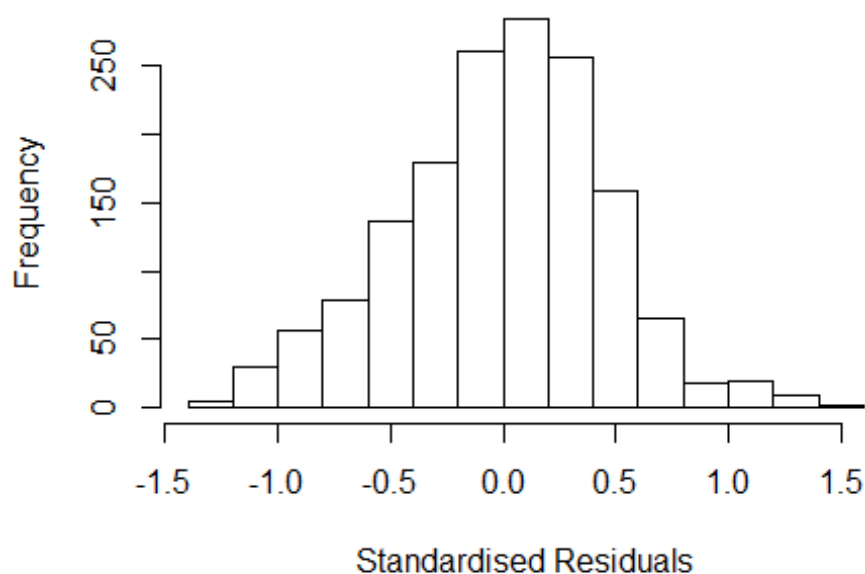
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3262 -0.2823 0.0113 0.2777 1.4031
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9131 0.0964 9.47 < 2e-16 ***
## FirstAuthorFemale1 0.0324 0.0232 1.40 0.16295
## LastAuthorFemale1 0.0527 0.0248 2.13 0.03341 *
## UniqueAuthors2 0.1280 0.0668 1.92 0.05559 .
## UniqueAuthors3 0.2739 0.0624 4.39 1.2e-05 ***
## UniqueAuthors4 0.2720 0.0600 4.53 6.3e-06 ***
## UniqueAuthors5 0.4131 0.0548 7.54 8.0e-14 ***
## Year1997 -0.0533 0.1210 -0.44 0.65951
## Year1998 -0.0252 0.1018 -0.25 0.80470
## Year1999 -0.1116 0.0952 -1.17 0.24160
```

```

## Year2000          -0.6218      0.1360    -4.57  5.2e-06 ***
## Year2001          -0.0293      0.0916    -0.32  0.74960
## Year2002          -0.1890      0.0853    -2.21  0.02695 *
## Year2003          -0.2744      0.0969    -2.83  0.00470 **
## Year2004          -0.2936      0.0857    -3.43  0.00063 ***
## Year2005          -0.2494      0.0871    -2.86  0.00424 **
## Year2006          -0.2564      0.0936    -2.74  0.00625 **
## Year2007          -0.1874      0.0881    -2.13  0.03354 *
## Year2008          -0.1632      0.0874    -1.87  0.06209 .
## Year2009          -0.2351      0.0878    -2.68  0.00746 **
## Year2010          -0.2703      0.0867    -3.12  0.00185 **
## Year2011          -0.2381      0.0866    -2.75  0.00604 **
## Year2012          -0.2801      0.0862    -3.25  0.00117 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.169, Adjusted R-squared:  0.157
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 1450 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.178  0.844  0.947  0.888  0.986  0.999
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           6.42e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.065 1           1.032
## LastAuthorFemale  1.071 1           1.035
## Year              1.140 16           1.004

```

Residuals from first and last author



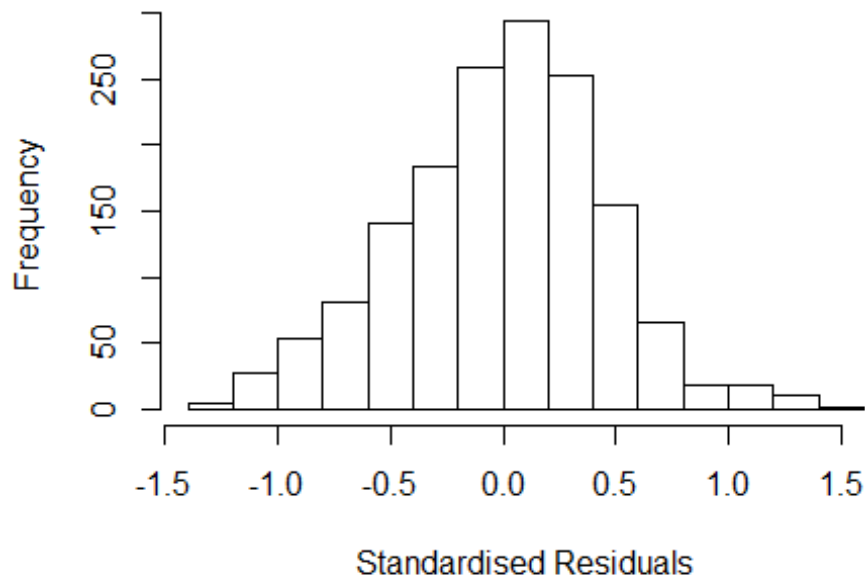
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2758 -0.2938  0.0304  0.2976  1.4297
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1683     0.0952   12.27 < 2e-16 ***
## FirstAuthorFemale1  0.0525     0.0239    2.19  0.028 *
## LastAuthorFemale1  0.0305     0.0253    1.21  0.228
## Year1997          -0.0168     0.1337   -0.13  0.900
## Year1998          -0.0295     0.1189   -0.25  0.804
## Year1999          -0.0684     0.1097   -0.62  0.533
## Year2000          -0.5999     0.1439   -4.17 3.2e-05 ***
## Year2001           0.0245     0.1033    0.24  0.812
## Year2002          -0.1264     0.1008   -1.25  0.210
## Year2003          -0.2016     0.1112   -1.81  0.070 .
## Year2004          -0.2458     0.1022   -2.41  0.016 *
## Year2005          -0.2185     0.1016   -2.15  0.032 *
```

```

## Year2006          -0.2007      0.1077    -1.86     0.063 .
## Year2007          -0.1332      0.1042    -1.28     0.201
## Year2008          -0.0998      0.1016    -0.98     0.326
## Year2009          -0.1764      0.1025    -1.72     0.086 .
## Year2010          -0.1975      0.1044    -1.89     0.059 .
## Year2011          -0.1831      0.1007    -1.82     0.069 .
## Year2012          -0.2218      0.1020    -2.17     0.030 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.09,    Adjusted R-squared:  0.0794
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 1456 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.198  0.845   0.945   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      6.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.065 1      1.032
## Year              1.065 16      1.002

```

Residuals from first author



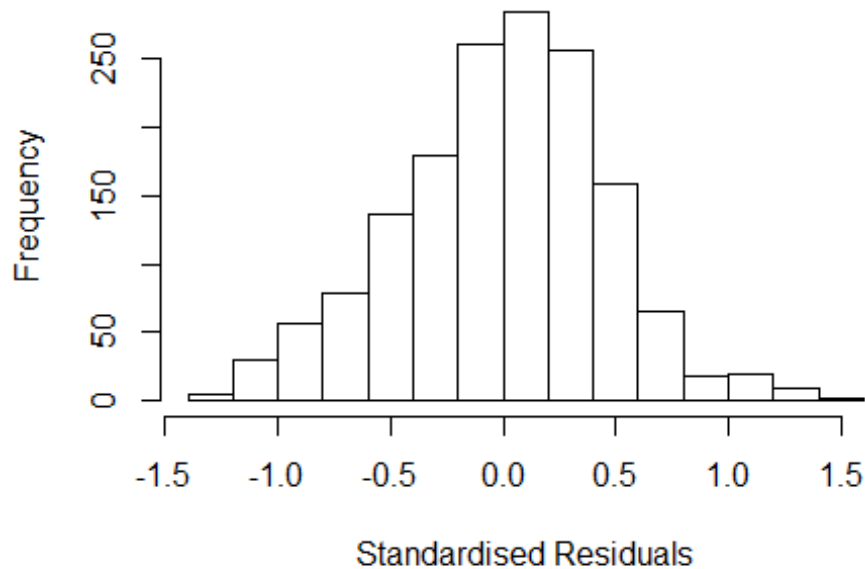
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2565 -0.2982 0.0284 0.2995 1.4227
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1789 0.0939 12.56 <2e-16 ***
## FirstAuthorFemale1 0.0564 0.0241 2.33 0.020 *
## Year1997 -0.0194 0.1336 -0.15 0.885
## Year1998 -0.0328 0.1190 -0.28 0.783
## Year1999 -0.0694 0.1095 -0.63 0.526
## Year2000 -0.6036 0.1443 -4.18 3e-05 ***
## Year2001 0.0212 0.1027 0.21 0.836
## Year2002 -0.1303 0.1004 -1.30 0.195
## Year2003 -0.2075 0.1105 -1.88 0.061 .
## Year2004 -0.2514 0.1017 -2.47 0.014 *
## Year2005 -0.2218 0.1013 -2.19 0.029 *
## Year2006 -0.2063 0.1072 -1.93 0.054 .
```

```

## Year2007          -0.1369      0.1038   -1.32    0.188
## Year2008          -0.1038      0.1012   -1.03    0.305
## Year2009          -0.1805      0.1021   -1.77    0.077 .
## Year2010          -0.1988      0.1043   -1.91    0.057 .
## Year2011          -0.1862      0.1003   -1.86    0.064 .
## Year2012          -0.2243      0.1018   -2.20    0.028 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.0889, Adjusted R-squared:  0.0788
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 111 weights are ~= 1. The remaining 1447 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.202  0.842  0.945  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      6.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.073 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
## -1.2535 -0.3026 0.0293 0.2981 1.4038
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1759 0.0954 12.32 < 2e-16 ***
## LastAuthorFemale1 0.0383 0.0255 1.50 0.133
## Year1997 -0.0155 0.1335 -0.12 0.908
## Year1998 -0.0206 0.1185 -0.17 0.862
## Year1999 -0.0499 0.1095 -0.46 0.649
## Year2000 -0.5817 0.1433 -4.06 5.2e-05 ***
## Year2001 0.0394 0.1035 0.38 0.704
## Year2002 -0.1124 0.1007 -1.12 0.265
## Year2003 -0.1890 0.1111 -1.70 0.089 .
## Year2004 -0.2358 0.1023 -2.31 0.021 *
## Year2005 -0.2053 0.1016 -2.02 0.044 *
## Year2006 -0.1857 0.1076 -1.73 0.085 .
```

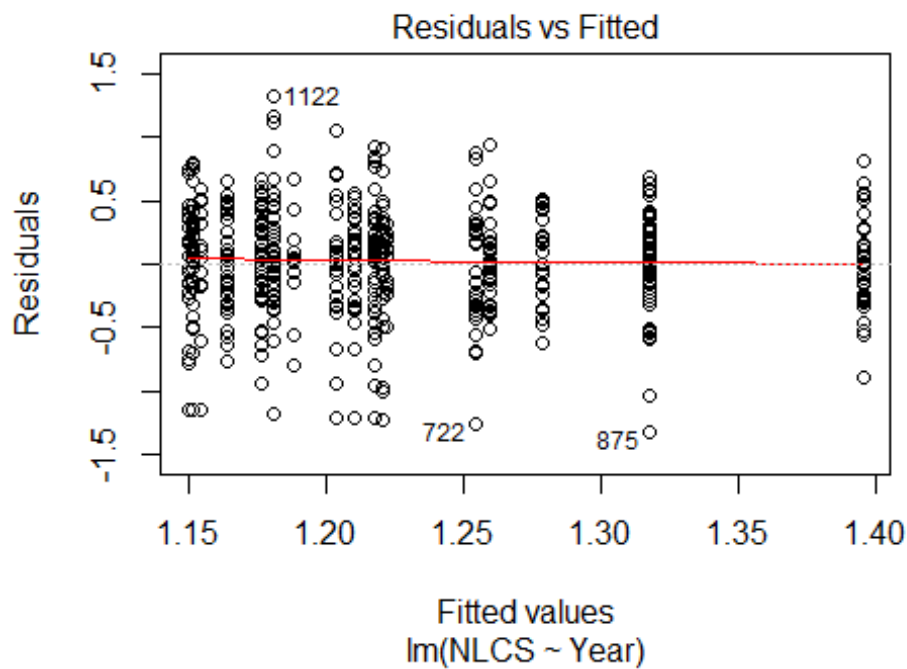
```

## Year2007          -0.1174      0.1039    -1.13      0.259
## Year2008          -0.0880      0.1018    -0.86      0.387
## Year2009          -0.1581      0.1023    -1.55      0.122
## Year2010          -0.1845      0.1044    -1.77      0.078 .
## Year2011          -0.1650      0.1005    -1.64      0.101
## Year2012          -0.2026      0.1016    -1.99      0.046 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0862, Adjusted R-squared:  0.0761
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 125 weights are ~= 1. The remaining 1433 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.216  0.842  0.942  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      6.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1558"
## [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
##   51   56   41   63   64   50   35   27   31   36   34   62   41   71   69
## 2011 2012
##   63   63
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   29   14   38   38   26   25   16   23   15   25   40   31   55   51
## 2011 2012

```



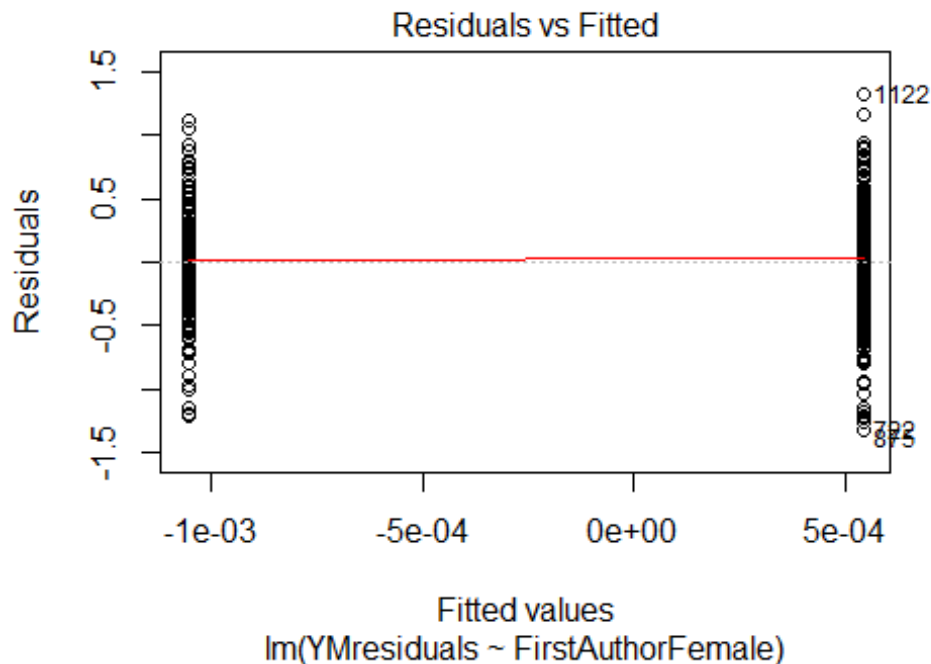
```
## 42 42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 33 27 9 31 37 23 22 14 21 13 22 35 24 43 47
## 2011 2012
## 33 40
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 33, df = 16, p-value = 0.007
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.022, df = 1, p-value = 0.9
## [1] "Female first author team size 2018 geometric mean: 7.93581062004139"
## [1] "Male first author team size 2018 geometric mean: 6.88629377838631"
## Warning in wilcox.test.default(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 = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.92601893386042"
## [1] "Male last author team size 2018 geometric mean: 7.56751907620557"

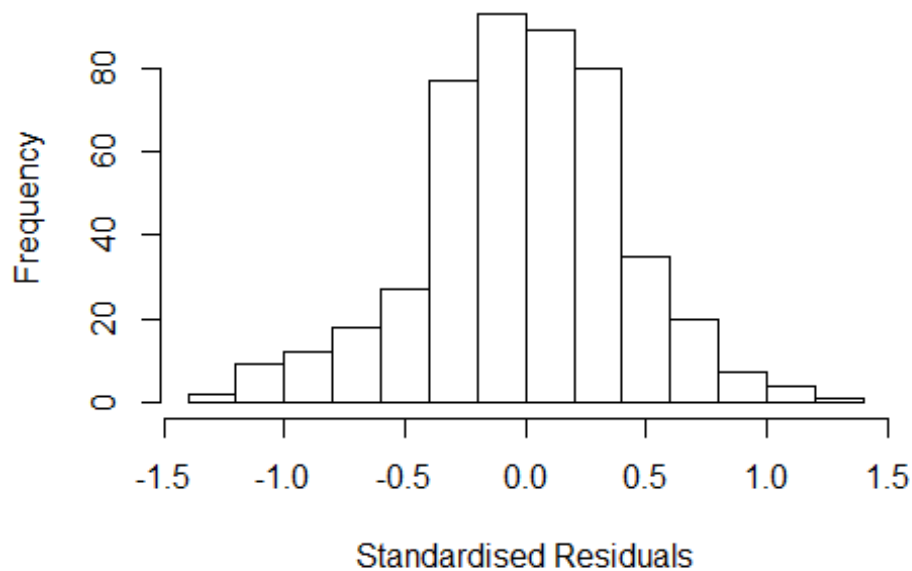
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 80, p-value = 0.8
## 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	1.294	1	1.138
LastAuthorFemale	1.330	1	1.153
UniqueAuthors	2.816	4	1.138
Year	3.610	16	1.041

Residuals from first and last author and team size



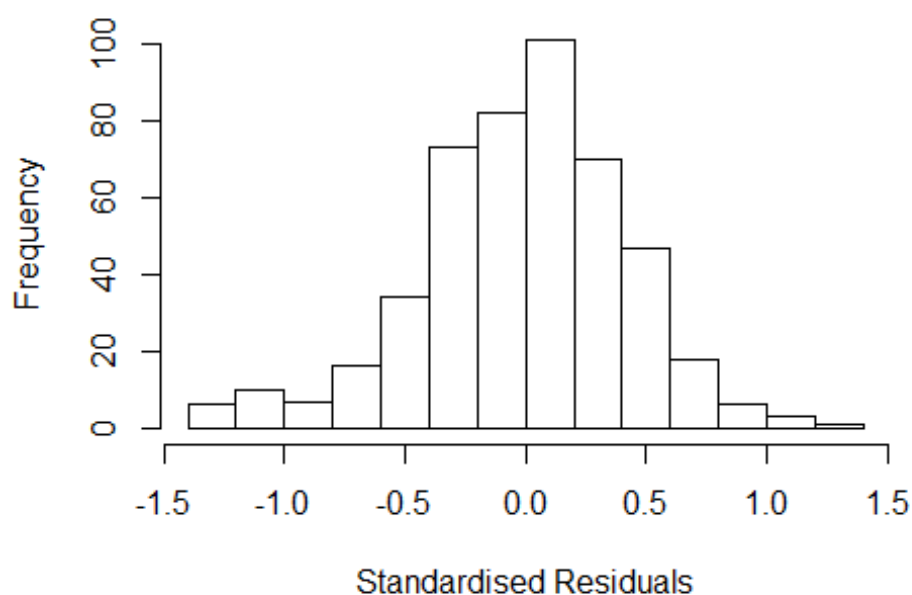
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26425 -0.27076 -0.00538  0.26511  1.22449
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11263    0.12255   9.08  <2e-16 ***
## FirstAuthorFemale1 -0.03001    0.04458  -0.67   0.501
## LastAuthorFemale1 -0.01507    0.05417  -0.28   0.781
## UniqueAuthors2     0.06781    0.12368   0.55   0.584
## UniqueAuthors3     0.12636    0.11957   1.06   0.291
## UniqueAuthors4     0.07466    0.11631   0.64   0.521
## UniqueAuthors5     0.25136    0.11213   2.24   0.025 *
## Year1997          0.02956    0.09570   0.31   0.758
## Year1998         -0.04764    0.09577  -0.50   0.619
## Year1999          0.01143    0.10349   0.11   0.912
```

```

## Year2000      -0.07376    0.09126   -0.81    0.419
## Year2001      0.07404    0.10031    0.74    0.461
## Year2002     -0.00544    0.12859   -0.04    0.966
## Year2003     -0.07902    0.10758   -0.73    0.463
## Year2004     -0.07759    0.12680   -0.61    0.541
## Year2005     -0.15455    0.14473   -1.07    0.286
## Year2006     -0.06975    0.11292   -0.62    0.537
## Year2007      0.11310    0.09097    1.24    0.214
## Year2008      0.04353    0.10826    0.40    0.688
## Year2009      0.02978    0.08845    0.34    0.737
## Year2010     -0.00620    0.08746   -0.07    0.944
## Year2011     -0.12722    0.10617   -1.20    0.231
## Year2012     -0.08549    0.11141   -0.77    0.443
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.0711, Adjusted R-squared:  0.0258
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 443 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.283  0.883  0.955  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      2.11e-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 1.186 1      1.089
## LastAuthorFemale  1.324 1      1.151
## Year              1.543 16      1.014

```

Residuals from first and last author



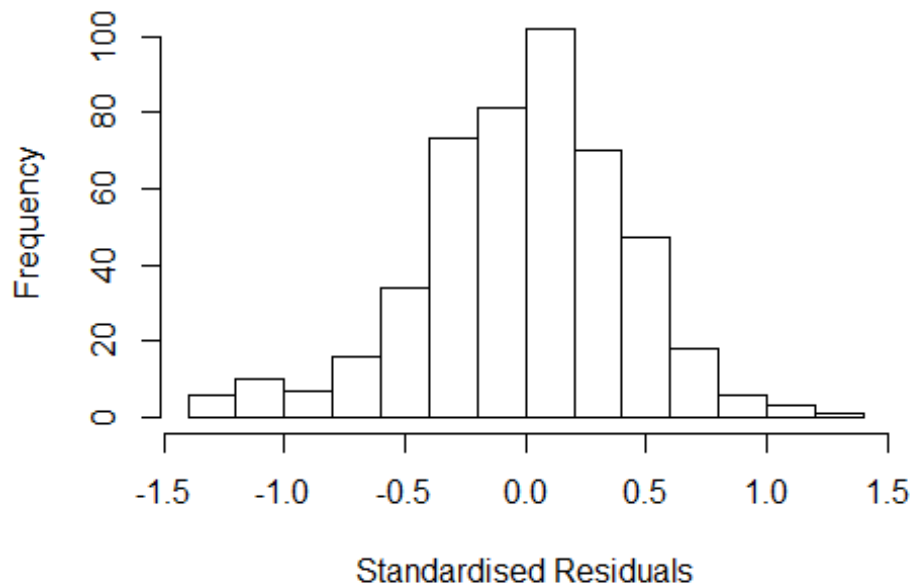
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2744 -0.2857  0.0159  0.2726  1.2947
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.25070    0.05856   21.36  <2e-16 ***
## FirstAuthorFemale1 -0.03034    0.04330   -0.70   0.484
## LastAuthorFemale1 -0.00191    0.05645   -0.03   0.973
## Year1997         0.00165    0.09570    0.02   0.986
## Year1998        -0.06318    0.09861   -0.64   0.522
## Year1999         0.01338    0.10062    0.13   0.894
## Year2000        -0.06869    0.08942   -0.77   0.443
## Year2001         0.05027    0.09808    0.51   0.609
## Year2002         0.02562    0.13183    0.19   0.846
## Year2003        -0.05597    0.10404   -0.54   0.591
## Year2004        -0.04391    0.12601   -0.35   0.728
## Year2005        -0.07901    0.14930   -0.53   0.597
```

```

## Year2006      -0.02953    0.12285   -0.24    0.810
## Year2007      0.14927    0.08606    1.73    0.083 .
## Year2008      0.07443    0.11191    0.67    0.506
## Year2009      0.08280    0.08110    1.02    0.308
## Year2010      0.01884    0.08368    0.23    0.822
## Year2011     -0.10178    0.10220   -1.00    0.320
## Year2012     -0.04244    0.10784   -0.39    0.694
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0262, Adjusted R-squared:  -0.0124
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 431 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.298  0.878   0.950   0.899   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.11e-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 1.177 1      1.085
## Year              1.177 16      1.005

```

Residuals from first author



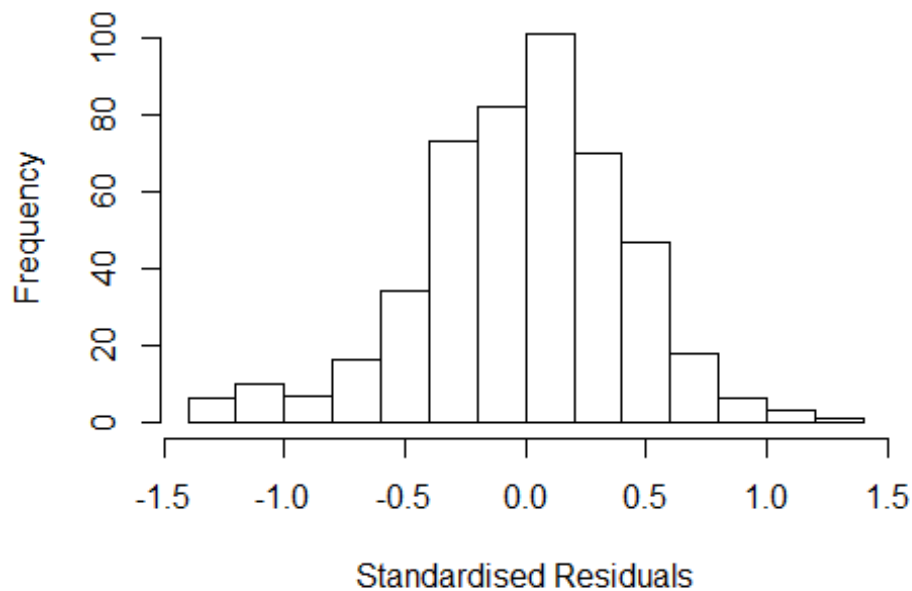
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2766 -0.2858 0.0158 0.2726 1.2952
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25034 0.05808 21.53 <2e-16 ***
## FirstAuthorFemale1 -0.03037 0.04320 -0.70 0.482
## Year1997 0.00182 0.09578 0.02 0.985
## Year1998 -0.06281 0.09821 -0.64 0.523
## Year1999 0.01319 0.09944 0.13 0.895
## Year2000 -0.06870 0.08933 -0.77 0.442
## Year2001 0.05039 0.09777 0.52 0.607
## Year2002 0.02622 0.12950 0.20 0.840
## Year2003 -0.05595 0.10405 -0.54 0.591
## Year2004 -0.04362 0.12607 -0.35 0.730
## Year2005 -0.07861 0.14938 -0.53 0.599
## Year2006 -0.02954 0.12263 -0.24 0.810
```

```

## Year2007          0.14902      0.08606      1.73      0.084 .
## Year2008          0.07437      0.11203      0.66      0.507
## Year2009          0.08311      0.08088      1.03      0.305
## Year2010          0.01893      0.08374      0.23      0.821
## Year2011         -0.10147      0.10228     -0.99      0.322
## Year2012         -0.04256      0.10735     -0.40      0.692
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.0262, Adjusted R-squared:  -0.0101
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 432 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.296  0.878  0.950  0.899  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.11e-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 1.312 1      1.145
## Year              1.312 16      1.009

```


Residuals from last author



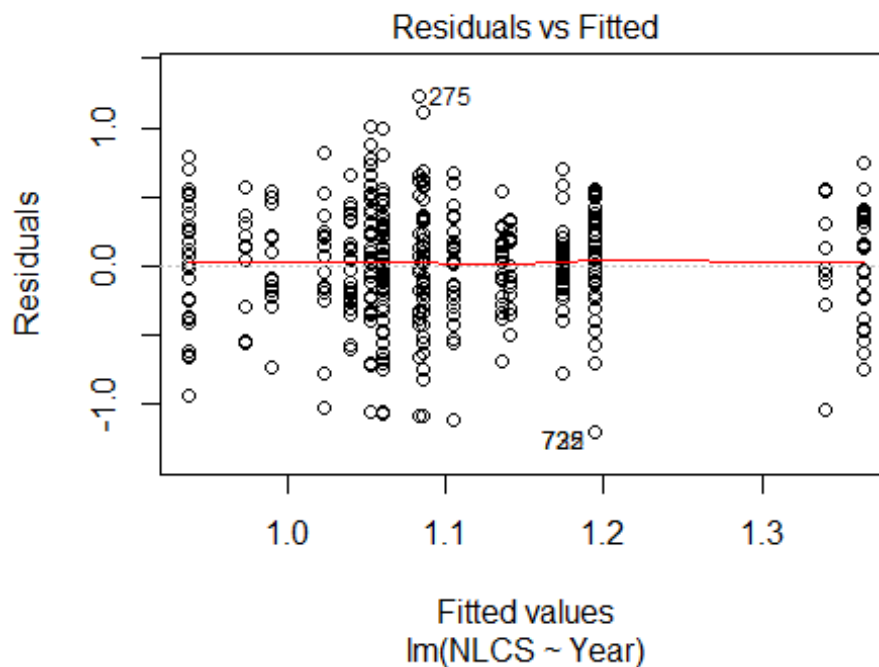
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2576 -0.2952  0.0169  0.2782  1.3124
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.24524    0.05770   21.58  <2e-16 ***
## LastAuthorFemale1 -0.00275    0.05610   -0.05    0.961
## Year1997        -0.00327    0.09567   -0.03    0.973
## Year1998        -0.06727    0.09913   -0.68    0.498
## Year1999         0.01187    0.09984    0.12    0.905
## Year2000        -0.07289    0.08909   -0.82    0.414
## Year2001         0.04692    0.09661    0.49    0.627
## Year2002         0.01512    0.13142    0.12    0.908
## Year2003        -0.06402    0.10298   -0.62    0.534
## Year2004        -0.04635    0.12512   -0.37    0.711
## Year2005        -0.08180    0.14728   -0.56    0.579
## Year2006        -0.02867    0.12182   -0.24    0.814
```

```

## Year2007          0.14308      0.08504      1.68      0.093 .
## Year2008          0.06948      0.11244      0.62      0.537
## Year2009          0.07741      0.08076      0.96      0.338
## Year2010          0.01484      0.08276      0.18      0.858
## Year2011         -0.10385      0.10267     -1.01      0.312
## Year2012         -0.05467      0.10445     -0.52      0.601
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0251, Adjusted R-squared:  -0.0112
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 429 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.285  0.876  0.950  0.899  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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: 474"
## [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
##   65   50   27   37   44   40   43   30   27   20   28   56   49   60   61
## 2011 2012
##   46   71
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   27   27   12   19   17   12   24   17   15   15   15   37   32   45   49
## 2011 2012

```

```
## 29 42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 24 24 10 14 13 10 20 13 12 13 12 30 29 39 47
## 2011 2012
## 26 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 = 31, df = 16, p-value = 0.01
```



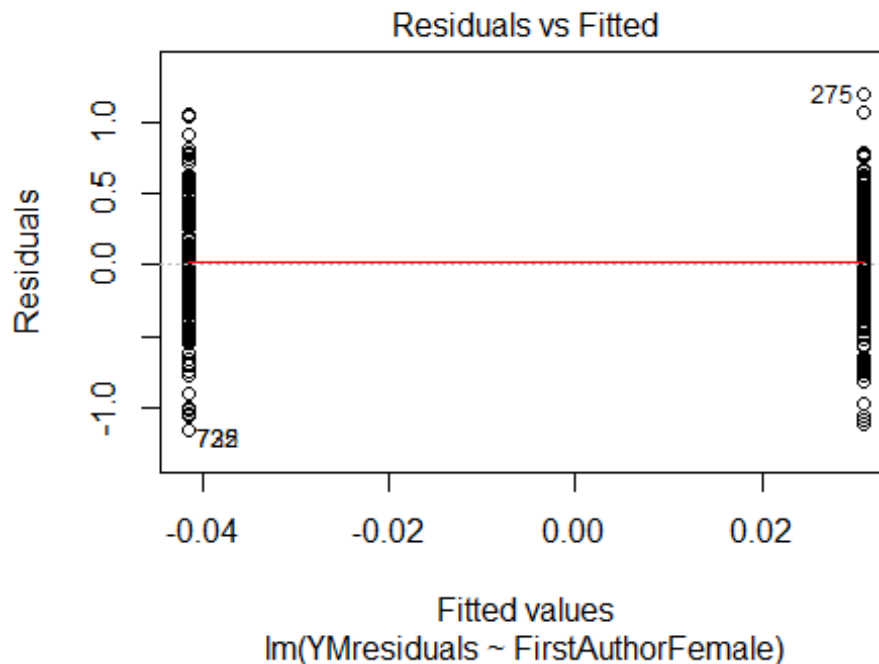
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 4.22628041018091"
## [1] "Male first author team size 2018 geometric mean: 2.30205338054865"

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

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 500, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.36289489886567"
## [1] "Male last author team size 2018 geometric mean: 2.82441087065747"

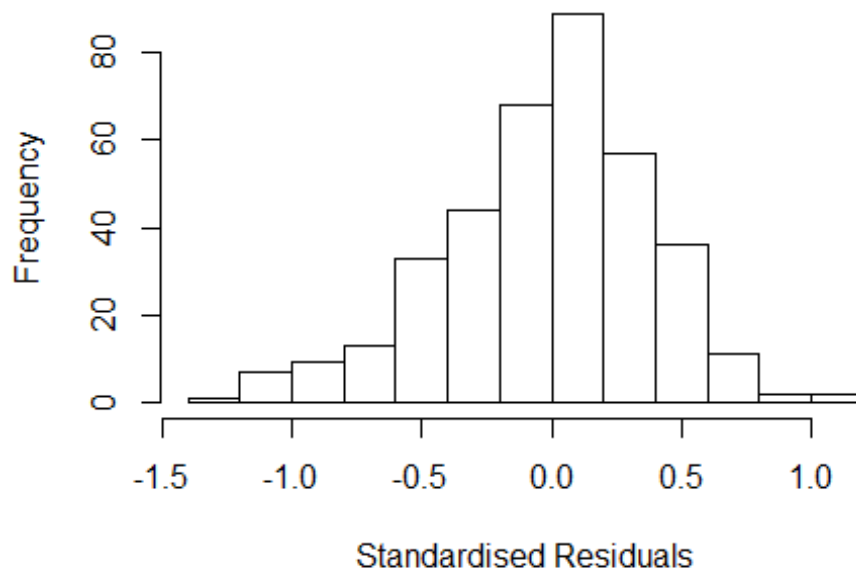
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 250, p-value = 0.6
## 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	1.342	1	1.159
LastAuthorFemale	1.190	1	1.091
UniqueAuthors	3.139	4	1.154
Year	4.146	16	1.045

Residuals from first and last author and team size



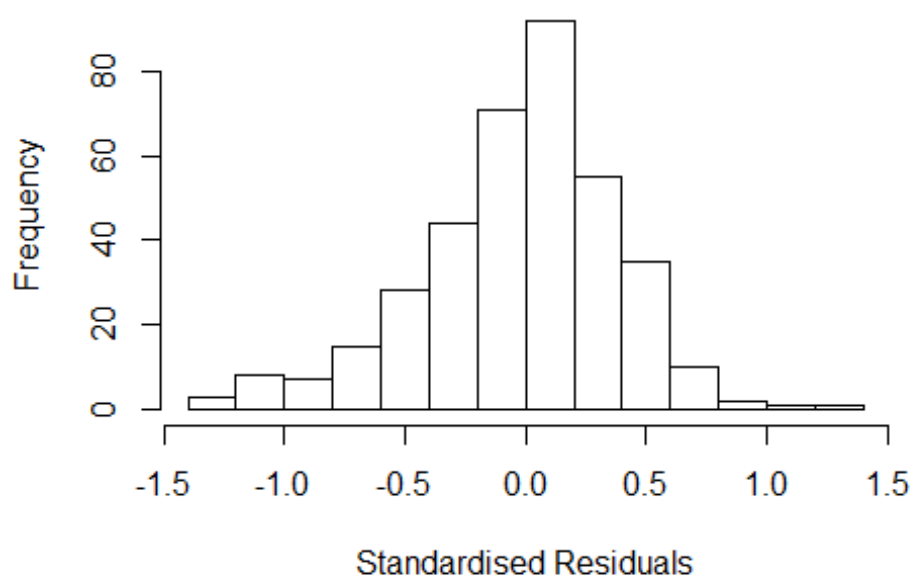
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2154 -0.2422 0.0198 0.2448 1.1452
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8718 0.1655 5.27 2.4e-07 ***
## FirstAuthorFemale1 -0.0767 0.0454 -1.69 0.092 .
## LastAuthorFemale1 -0.0507 0.0524 -0.97 0.334
## UniqueAuthors2 0.3224 0.1411 2.28 0.023 *
## UniqueAuthors3 0.3370 0.1498 2.25 0.025 *
## UniqueAuthors4 0.2794 0.1470 1.90 0.058 .
## UniqueAuthors5 0.3539 0.1455 2.43 0.015 *
## Year1997 -0.1513 0.1013 -1.49 0.136
## Year1998 0.2511 0.1558 1.61 0.108
## Year1999 0.0716 0.1332 0.54 0.591
```

```

## Year2000          0.0213      0.1804      0.12      0.906
## Year2001         -0.2653      0.1318     -2.01      0.045 *
## Year2002          0.2962      0.1341      2.21      0.028 *
## Year2003          0.0271      0.1217      0.22      0.824
## Year2004         -0.0435      0.1597     -0.27      0.786
## Year2005         -0.1761      0.1167     -1.51      0.132
## Year2006          0.0783      0.1242      0.63      0.529
## Year2007         -0.0569      0.1043     -0.55      0.585
## Year2008          0.0160      0.0905      0.18      0.860
## Year2009          0.1217      0.0993      1.23      0.221
## Year2010         -0.0326      0.0992     -0.33      0.743
## Year2011         -0.1271      0.1291     -0.99      0.325
## Year2012          0.0631      0.1137      0.55      0.579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.374
## Multiple R-squared:  0.12,   Adjusted R-squared:  0.0643
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 339 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.270  0.862  0.953   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      2.69e-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 1.257 1      1.121
## LastAuthorFemale  1.159 1      1.077
## Year              1.423 16      1.011

```

Residuals from first and last author



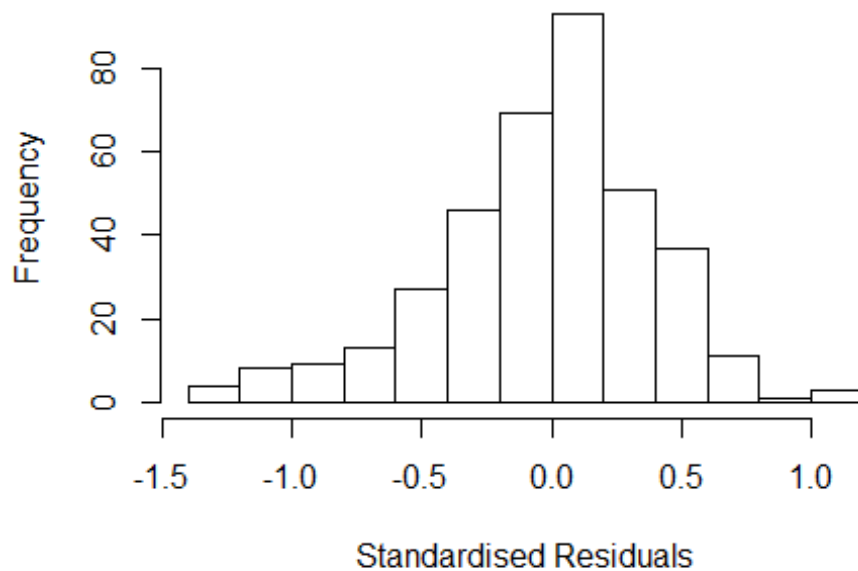
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2658 -0.2412 0.0186 0.2349 1.2362
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20112 0.07557 15.89 <2e-16 ***
## FirstAuthorFemale1 -0.07041 0.04473 -1.57 0.116
## LastAuthorFemale1 -0.06658 0.05378 -1.24 0.217
## Year1997 -0.15775 0.10413 -1.51 0.131
## Year1998 0.20559 0.16739 1.23 0.220
## Year1999 -0.03808 0.10845 -0.35 0.726
## Year2000 -0.06070 0.18429 -0.33 0.742
## Year2001 -0.26645 0.13519 -1.97 0.050 *
## Year2002 0.22697 0.12526 1.81 0.071 .
## Year2003 0.01168 0.12354 0.09 0.925
## Year2004 -0.03476 0.15902 -0.22 0.827
## Year2005 -0.19162 0.11869 -1.61 0.107
```

```

## Year2006          0.02725    0.10708    0.25    0.799
## Year2007          -0.05806    0.10656   -0.54    0.586
## Year2008           0.00262    0.08940    0.03    0.977
## Year2009           0.13508    0.09746    1.39    0.167
## Year2010          -0.03033    0.09931   -0.31    0.760
## Year2011          -0.13304    0.12795   -1.04    0.299
## Year2012           0.04750    0.11929    0.40    0.691
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.371
## Multiple R-squared:  0.0898, Adjusted R-squared:  0.0434
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 345 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.222  0.867   0.956   0.891   0.988   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.69e-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 1.245 1      1.116
## Year              1.245 16      1.007

```


Residuals from first author



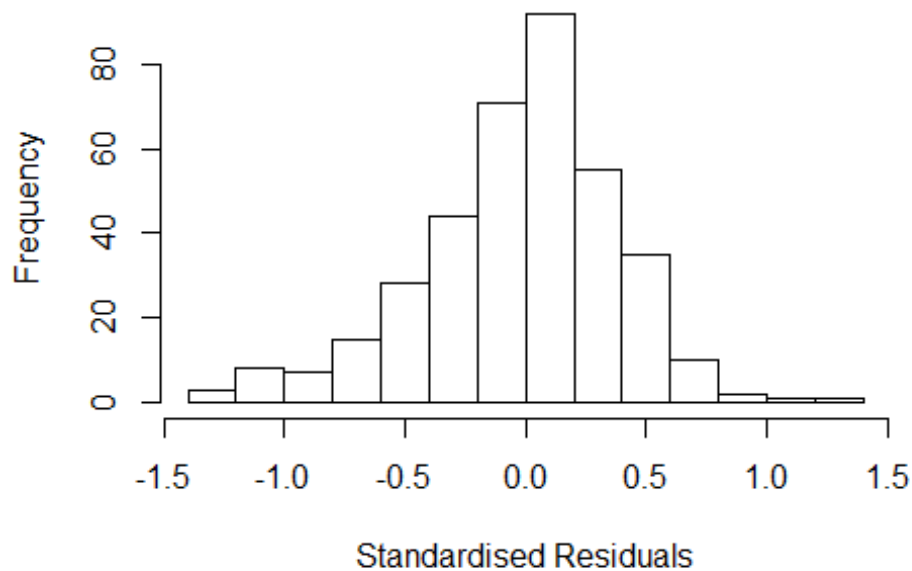
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2425 -0.2452 0.0167 0.2408 1.1713
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19597 0.07456 16.04 <2e-16 ***
## FirstAuthorFemale1 -0.08032 0.04461 -1.80 0.073 .
## Year1997 -0.16128 0.10354 -1.56 0.120
## Year1998 0.20177 0.16584 1.22 0.225
## Year1999 -0.03413 0.10854 -0.31 0.753
## Year2000 -0.05724 0.18653 -0.31 0.759
## Year2001 -0.27110 0.14015 -1.93 0.054 .
## Year2002 0.22137 0.12451 1.78 0.076 .
## Year2003 0.00768 0.12190 0.06 0.950
## Year2004 -0.03072 0.15539 -0.20 0.843
## Year2005 -0.19374 0.11654 -1.66 0.097 .
## Year2006 0.02606 0.10509 0.25 0.804
```

```

## Year2007      -0.06836    0.10797   -0.63    0.527
## Year2008      -0.00758    0.08838   -0.09    0.932
## Year2009       0.12686    0.09675    1.31    0.191
## Year2010      -0.03054    0.09852   -0.31    0.757
## Year2011      -0.15612    0.12585   -1.24    0.216
## Year2012       0.04058    0.12077    0.34    0.737
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.37
## Multiple R-squared:  0.0867, Adjusted R-squared:  0.0428
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 341 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.236  0.861  0.954  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      2.69e-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 1.136 1          1.066
## Year              1.136 16          1.004

```

Residuals from last author



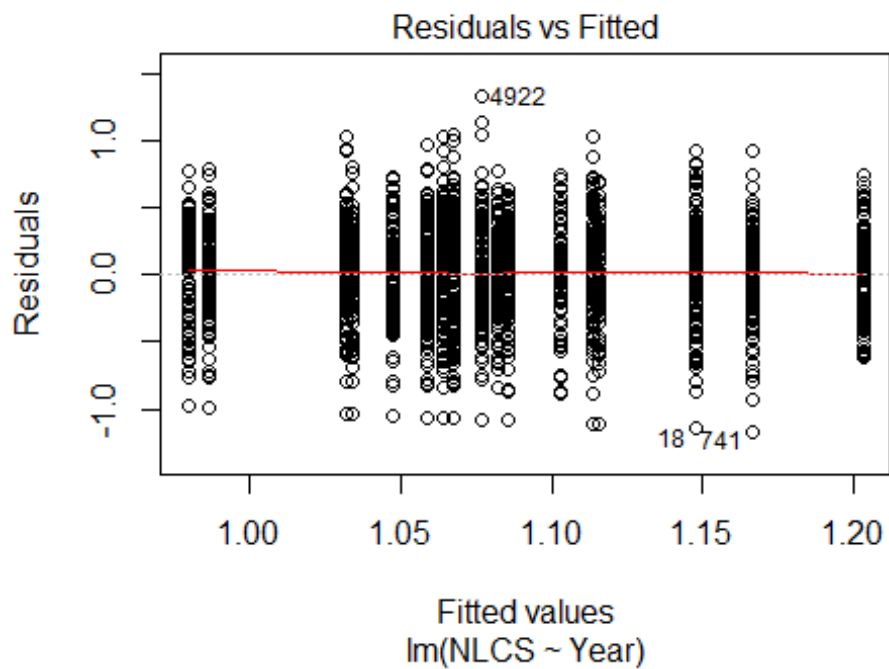
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.302 -0.246 0.033 0.243 1.280
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17765 0.07830 15.04 <2e-16 ***
## LastAuthorFemale1 -0.08119 0.05312 -1.53 0.127
## Year1997 -0.14795 0.10695 -1.38 0.167
## Year1998 0.22036 0.16119 1.37 0.172
## Year1999 -0.04153 0.11204 -0.37 0.711
## Year2000 -0.06696 0.19319 -0.35 0.729
## Year2001 -0.26526 0.13612 -1.95 0.052 .
## Year2002 0.23954 0.12672 1.89 0.060 .
## Year2003 0.00864 0.12631 0.07 0.945
## Year2004 -0.03406 0.15897 -0.21 0.830
## Year2005 -0.19465 0.12014 -1.62 0.106
## Year2006 0.01708 0.10718 0.16 0.873
```

```

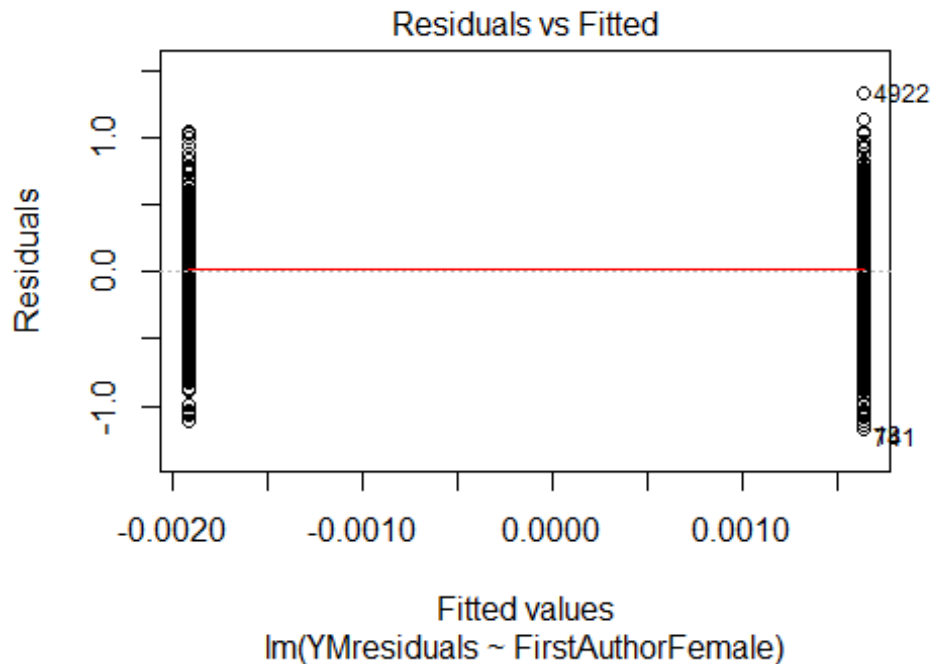
## Year2007          -0.06578      0.10504    -0.63      0.532
## Year2008          -0.00444      0.09023    -0.05      0.961
## Year2009           0.12401      0.09745     1.27      0.204
## Year2010          -0.04052      0.10021    -0.40      0.686
## Year2011          -0.11950      0.13002    -0.92      0.359
## Year2012           0.03870      0.11775     0.33      0.743
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0829, Adjusted R-squared:  0.0388
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 352 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.206  0.868  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      2.69e-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: 372"
## [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
## 217 185 194 169 205 203 231 159 147 203 174 197 200 244 264
## 2011 2012
## 284 255
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 129 111 117 105 96 84 173 106 94 124 111 121 131 149 180
## 2011 2012

```

```
## 203 181
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 117 97 104 86 80 64 142 91 86 91 94 101 109 132 152
## 2011 2012
## 170 151
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 30, df = 16, p-value = 0.02
```

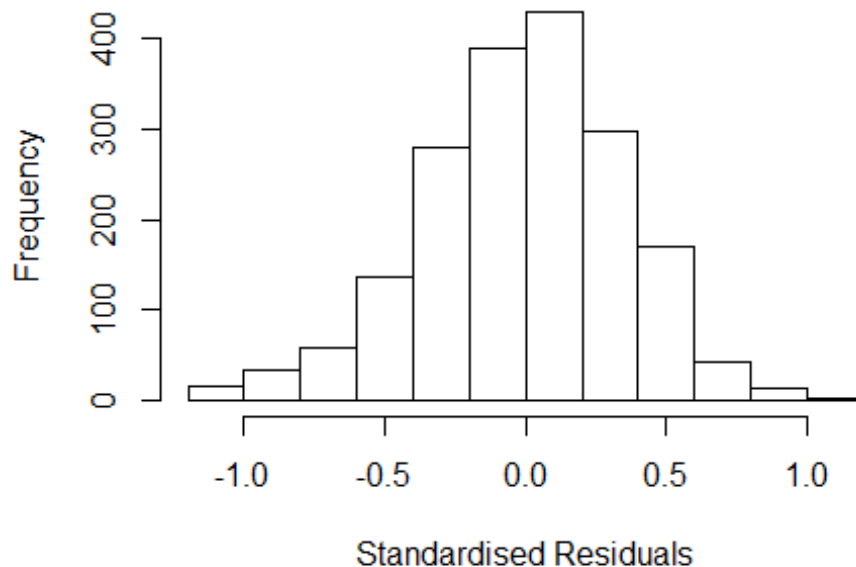


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7, df = 1, p-value = 0.008
```



```
## [1] "Female first author team size 2018 geometric mean: 5.48104979535042"
## [1] "Male first author team size 2018 geometric mean: 4.84542303345411"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5000, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.44416289231631"
## [1] "Male last author team size 2018 geometric mean: 5.03183700687403"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4800, 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 1.074 1          1.037
## LastAuthorFemale  1.106 1          1.052
## UniqueAuthors    1.244 4          1.028
## Year              1.346 16         1.009
```

Residuals from first and last author and team size



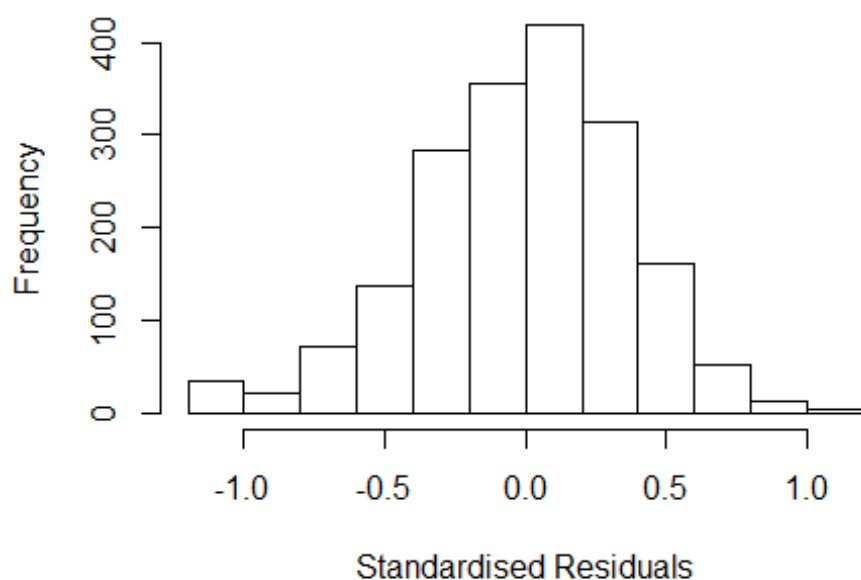
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17330 -0.23443  0.00853  0.22635  1.11980
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91126    0.05923   15.38 < 2e-16 ***
## FirstAuthorFemale1 -0.02733    0.01713   -1.60  0.11082
## LastAuthorFemale1 -0.00488    0.01967   -0.25  0.80407
## UniqueAuthors2     0.19124    0.05471    3.50  0.00048 ***
## UniqueAuthors3     0.19770    0.05224    3.78  0.00016 ***
## UniqueAuthors4     0.30551    0.05294    5.77  9.2e-09 ***
## UniqueAuthors5     0.39888    0.05066    7.87  5.8e-15 ***
## Year1997          -0.06682    0.06102   -1.10  0.27360
## Year1998           0.02713    0.04823    0.56  0.57384
## Year1999           0.01341    0.05281    0.25  0.79955
```

```

## Year2000      -0.02362    0.05478   -0.43  0.66642
## Year2001      -0.05571    0.06122   -0.91  0.36292
## Year2002      -0.12931    0.04820   -2.68  0.00737 **
## Year2003      -0.16237    0.05006   -3.24  0.00120 **
## Year2004      -0.19002    0.05782   -3.29  0.00103 **
## Year2005      -0.16974    0.05166   -3.29  0.00104 **
## Year2006      -0.10887    0.05279   -2.06  0.03934 *
## Year2007      -0.18915    0.05545   -3.41  0.00066 ***
## Year2008      -0.14781    0.05395   -2.74  0.00621 **
## Year2009      -0.13300    0.04764   -2.79  0.00529 **
## Year2010      -0.13684    0.04819   -2.84  0.00457 **
## Year2011      -0.14891    0.05022   -2.97  0.00306 **
## Year2012      -0.16925    0.05044   -3.36  0.00081 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.345
## Multiple R-squared:  0.111, Adjusted R-squared:  0.1
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 147 weights are ~= 1. The remaining 1720 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.223  0.870  0.951  0.901  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.36e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.052 1          1.026
## LastAuthorFemale 1.095 1          1.046
## Year 1.120 16          1.004

```


Residuals from first and last author



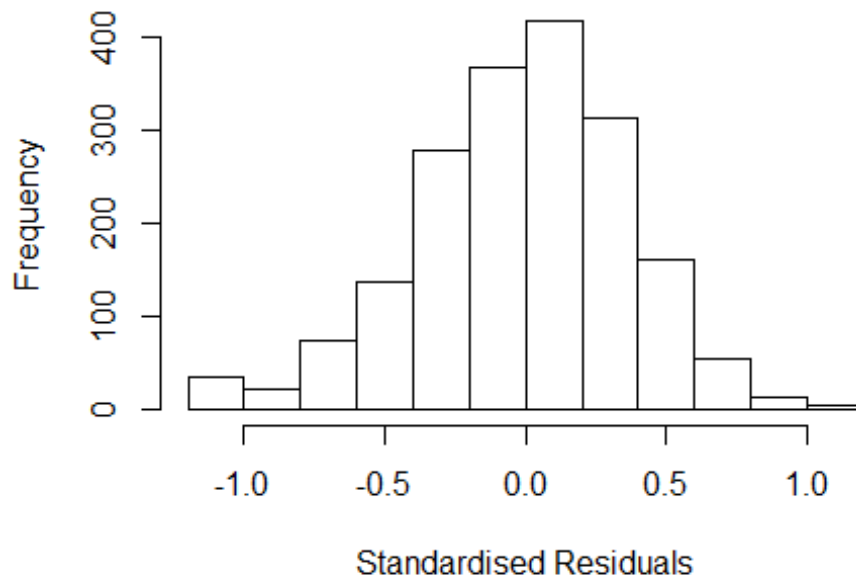
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1887 -0.2436  0.0136  0.2434  1.1680
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17471    0.03928   29.91  <2e-16 ***
## FirstAuthorFemale1 -0.00820    0.01783   -0.46  0.6456
## LastAuthorFemale1 -0.00821    0.02115   -0.39  0.6978
## Year1997        -0.06017    0.06104   -0.99  0.3244
## Year1998         0.01399    0.05003    0.28  0.7798
## Year1999         0.02704    0.05301    0.51  0.6101
## Year2000        -0.01452    0.05770   -0.25  0.8013
## Year2001        -0.03668    0.06411   -0.57  0.5672
## Year2002        -0.07116    0.04811   -1.48  0.1392
## Year2003        -0.11780    0.05255   -2.24  0.0251 *
## Year2004        -0.17056    0.05896   -2.89  0.0039 **
## Year2005        -0.13216    0.05569   -2.37  0.0177 *
```

```

## Year2006          -0.05845    0.05208   -1.12    0.2618
## Year2007          -0.15102    0.05658   -2.67    0.0077 **
## Year2008          -0.11095    0.05531   -2.01    0.0450 *
## Year2009          -0.09360    0.04792   -1.95    0.0510 .
## Year2010          -0.08943    0.04917   -1.82    0.0691 .
## Year2011          -0.08295    0.05076   -1.63    0.1024
## Year2012          -0.11747    0.04966   -2.37    0.0181 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.0208, Adjusted R-squared:  0.0112
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1729 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.253  0.871  0.950  0.901  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000      0
##      psi            subsampling            cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.039 1      1.019
## Year              1.039 16      1.001

```

Residuals from first author



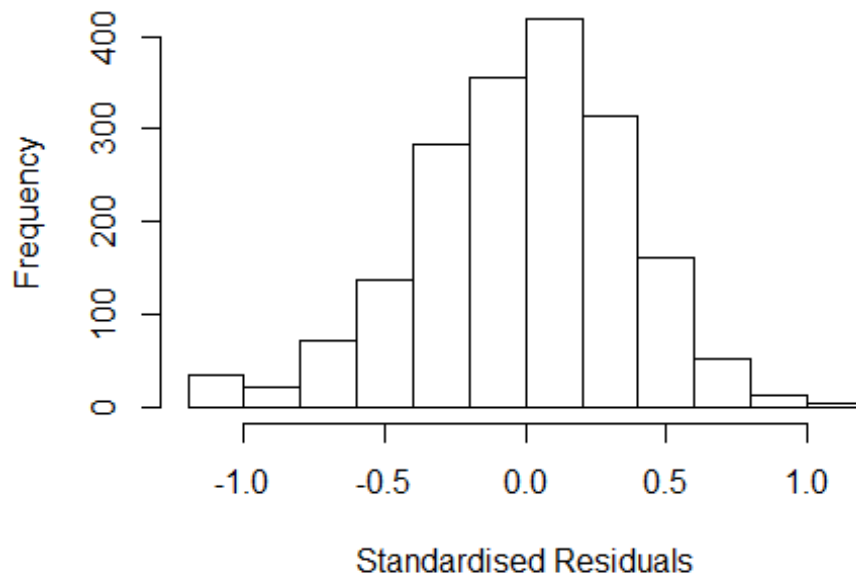
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1875 -0.2436 0.0133 0.2442 1.1622
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17298 0.03899 30.08 <2e-16 ***
## FirstAuthorFemale1 -0.00888 0.01772 -0.50 0.6164
## Year1997 -0.05986 0.06099 -0.98 0.3265
## Year1998 0.01452 0.04998 0.29 0.7714
## Year1999 0.02775 0.05292 0.52 0.6000
## Year2000 -0.01379 0.05759 -0.24 0.8107
## Year2001 -0.03568 0.06378 -0.56 0.5759
## Year2002 -0.07066 0.04808 -1.47 0.1418
## Year2003 -0.11749 0.05252 -2.24 0.0254 *
## Year2004 -0.17023 0.05887 -2.89 0.0039 **
## Year2005 -0.13245 0.05565 -2.38 0.0174 *
## Year2006 -0.05766 0.05190 -1.11 0.2667
```

```

## Year2007          -0.15138    0.05657   -2.68    0.0075 **
## Year2008          -0.11106    0.05521   -2.01    0.0444 *
## Year2009          -0.09417    0.04790   -1.97    0.0494 *
## Year2010          -0.09039    0.04904   -1.84    0.0654 .
## Year2011          -0.08410    0.05064   -1.66    0.0969 .
## Year2012          -0.11813    0.04956   -2.38    0.0172 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.0207, Adjusted R-squared:  0.0117
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 142 weights are ~= 1. The remaining 1725 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.254  0.871  0.951  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      5.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.081 1          1.040
## Year            1.081 16          1.002

```

Residuals from last author



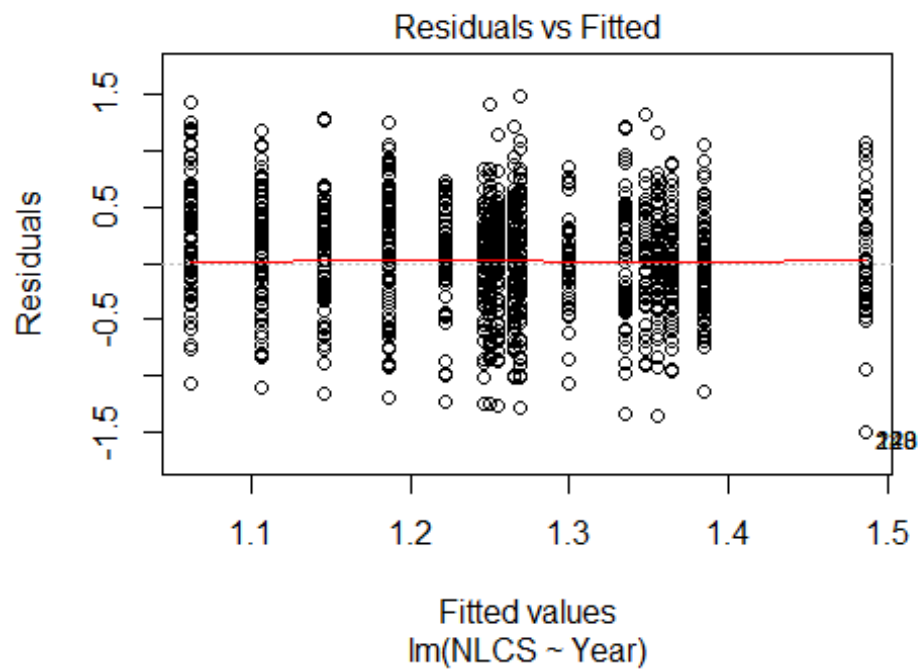
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1859 -0.2443 0.0117 0.2437 1.1730
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1725 0.0390 30.06 <2e-16 ***
## LastAuthorFemale1 -0.0091 0.0210 -0.43 0.6650
## Year1997 -0.0611 0.0610 -1.00 0.3166
## Year1998 0.0134 0.0500 0.27 0.7881
## Year1999 0.0259 0.0530 0.49 0.6255
## Year2000 -0.0158 0.0575 -0.27 0.7835
## Year2001 -0.0383 0.0641 -0.60 0.5502
## Year2002 -0.0725 0.0481 -1.51 0.1317
## Year2003 -0.1197 0.0523 -2.29 0.0223 *
## Year2004 -0.1719 0.0589 -2.92 0.0036 **
## Year2005 -0.1334 0.0556 -2.40 0.0164 *
## Year2006 -0.0610 0.0518 -1.18 0.2392
```

```

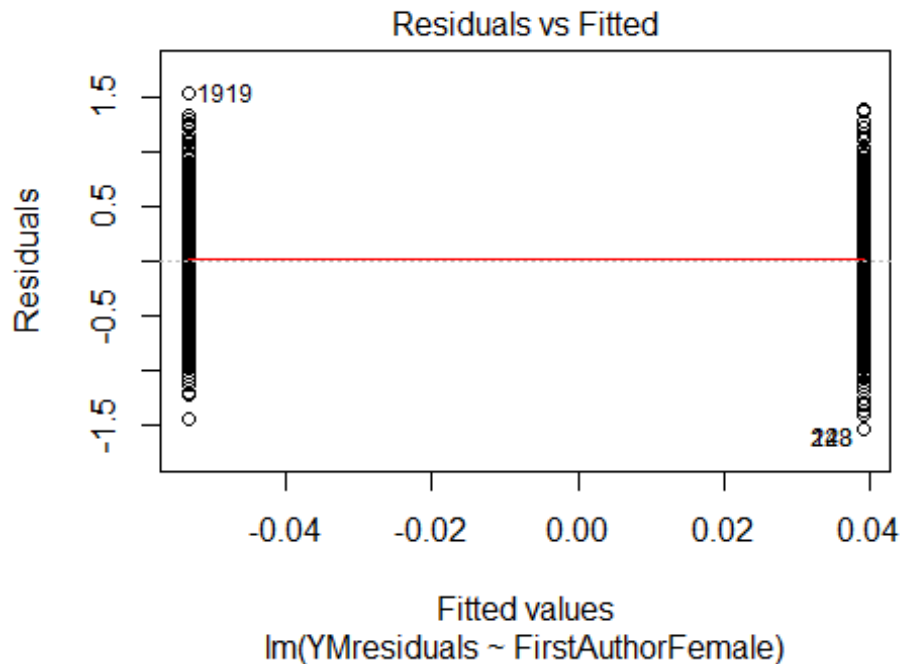
## Year2007          -0.1523      0.0565    -2.69    0.0071 **
## Year2008          -0.1124      0.0552    -2.03    0.0421 *
## Year2009          -0.0954      0.0478    -2.00    0.0459 *
## Year2010          -0.0912      0.0491    -1.86    0.0636 .
## Year2011          -0.0849      0.0506    -1.68    0.0937 .
## Year2012          -0.1194      0.0495    -2.41    0.0160 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.0206, Adjusted R-squared:  0.0116
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1726 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.257  0.872  0.951  0.901  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1867"
## [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
##   86  108   86   78   93  126   92   78   90  116  125  125  109  127  143
## 2011 2012
##  151  177
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   57   45   56   48   44   72   70   55   57   81   80   84   77   87   92
## 2011 2012

```

```
## 105 113
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 51 39 48 44 35 62 53 48 49 64 71 74 67 77 82
## 2011 2012
## 97 106
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 66, df = 16, p-value = 6e-08
```

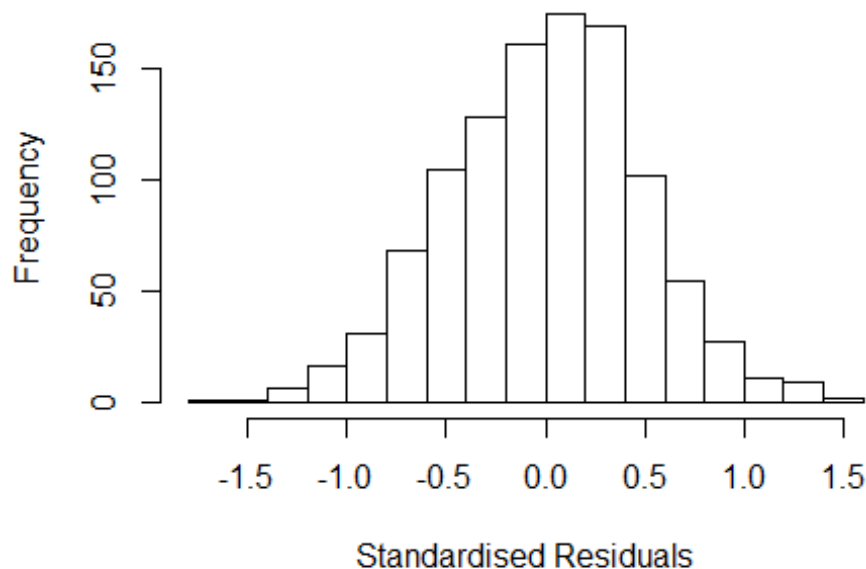


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3, df = 1, p-value = 0.08
```



```
## [1] "Female first author team size 2018 geometric mean: 4.95687789316122"
## [1] "Male first author team size 2018 geometric mean: 4.28525895025229"
##
## 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] "Female last author team size 2018 geometric mean: 4.95329519612581"
## [1] "Male last author team size 2018 geometric mean: 4.51638618009416"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.110 1      1.054
## LastAuthorFemale  1.189 1      1.090
## UniqueAuthors    1.429 4      1.046
## Year              1.570 16     1.014
```


Residuals from first and last author and team size



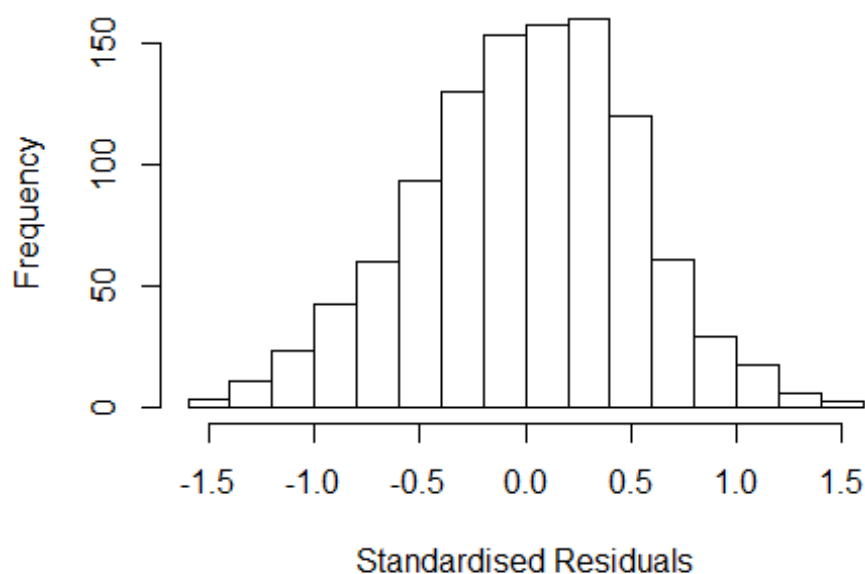
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7665 -0.3383  0.0219  0.3118  1.5111
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.81473    0.10936   7.45 2.0e-13 ***
## FirstAuthorFemale1 -0.12446    0.03223  -3.86 0.00012 ***
## LastAuthorFemale1 -0.08651    0.03717  -2.33 0.02014 *
## UniqueAuthors2    0.41037    0.07576   5.42 7.5e-08 ***
## UniqueAuthors3    0.51083    0.07644   6.68 3.8e-11 ***
## UniqueAuthors4    0.55386    0.07379   7.51 1.3e-13 ***
## UniqueAuthors5    0.72500    0.06785  10.69 < 2e-16 ***
## Year1997          0.22672    0.12451   1.82 0.06891 .
## Year1998          0.04198    0.11332   0.37 0.71115
## Year1999          0.00581    0.10821   0.05 0.95717
```

```

## Year2000      -0.03144      0.12035      -0.26      0.79398
## Year2001      0.06365      0.10066      0.63      0.52730
## Year2002      0.04374      0.10824      0.40      0.68624
## Year2003      -0.07171      0.10130      -0.71      0.47918
## Year2004      -0.07232      0.10730      -0.67      0.50047
## Year2005      -0.06505      0.09894      -0.66      0.51098
## Year2006      -0.20066      0.10295      -1.95      0.05155 .
## Year2007      -0.06405      0.10633      -0.60      0.54703
## Year2008      -0.06702      0.10579      -0.63      0.52653
## Year2009      -0.05367      0.10655      -0.50      0.61458
## Year2010      -0.14034      0.11189      -1.25      0.21003
## Year2011      -0.15481      0.09908      -1.56      0.11848
## Year2012      -0.17529      0.11163      -1.57      0.11666
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.482
## Multiple R-squared:  0.206, Adjusted R-squared:  0.189
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 86 weights are ~= 1. The remaining 981 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.151  0.876  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           9.37e-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 1.088 1           1.043
## LastAuthorFemale  1.071 1           1.035
## Year              1.157 16           1.005

```

Residuals from first and last author



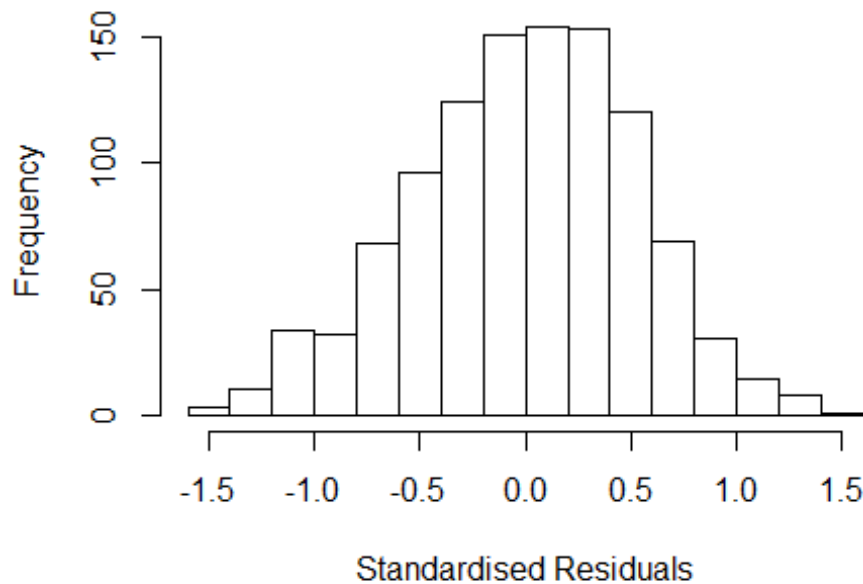
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5850 -0.3588  0.0156  0.3518  1.5979
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3892     0.0981   14.16 < 2e-16 ***
## FirstAuthorFemale1 -0.0917     0.0348   -2.64  0.00851 **
## LastAuthorFemale1  -0.1362     0.0395   -3.45  0.00058 ***
## Year1997           0.1958     0.1297    1.51  0.13134
## Year1998           0.0609     0.1229    0.50  0.62066
## Year1999          -0.0185     0.1221   -0.15  0.87926
## Year2000          -0.0478     0.1250   -0.38  0.70239
## Year2001           0.0286     0.1088    0.26  0.79264
## Year2002           0.0438     0.1166    0.38  0.70766
## Year2003          -0.0570     0.1147   -0.50  0.61953
## Year2004          -0.0477     0.1150   -0.42  0.67809
## Year2005          -0.0750     0.1099   -0.68  0.49497
```

```

## Year2006          -0.1888      0.1123    -1.68   0.09307 .
## Year2007          -0.0372      0.1166    -0.32   0.74982
## Year2008          -0.0688      0.1135    -0.61   0.54420
## Year2009          -0.0172      0.1179    -0.15   0.88430
## Year2010          -0.1150      0.1229    -0.94   0.34972
## Year2011          -0.1806      0.1123    -1.61   0.10816
## Year2012          -0.1977      0.1269    -1.56   0.11969
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.523
## Multiple R-squared:  0.0641, Adjusted R-squared:  0.048
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 992 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.330  0.880   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      9.37e-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 1.069 1      1.034
## Year              1.069 16      1.002

```

Residuals from first author



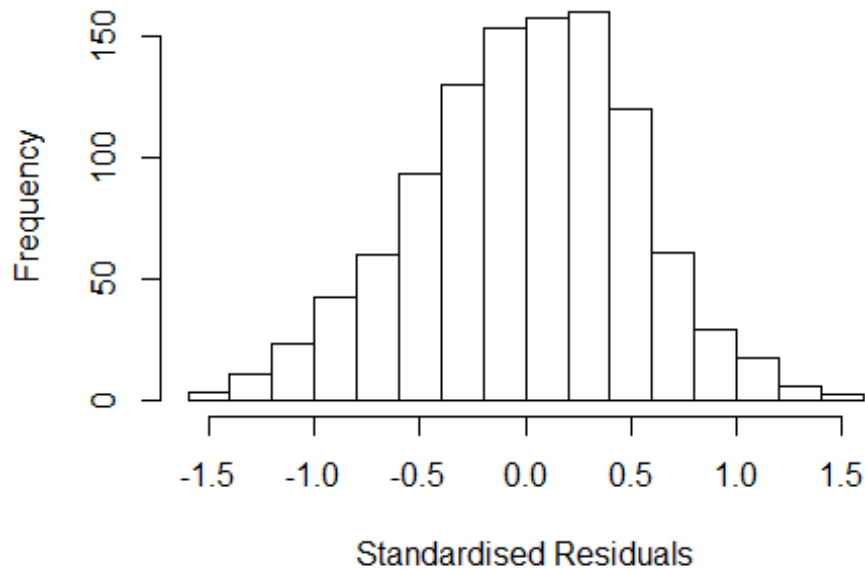
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5753 -0.3441 0.0121 0.3594 1.5099
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3750 0.0978 14.06 <2e-16 ***
## FirstAuthorFemale1 -0.1105 0.0351 -3.15 0.0017 **
## Year1997 0.2003 0.1290 1.55 0.1210
## Year1998 0.0529 0.1208 0.44 0.6615
## Year1999 -0.0215 0.1207 -0.18 0.8589
## Year2000 -0.0600 0.1243 -0.48 0.6297
## Year2001 0.0203 0.1086 0.19 0.8515
## Year2002 0.0205 0.1157 0.18 0.8595
## Year2003 -0.0622 0.1140 -0.55 0.5853
## Year2004 -0.0530 0.1149 -0.46 0.6448
## Year2005 -0.0957 0.1098 -0.87 0.3836
## Year2006 -0.2101 0.1123 -1.87 0.0617 .
```

```

## Year2007          -0.0533      0.1151   -0.46   0.6435
## Year2008          -0.0789      0.1141   -0.69   0.4898
## Year2009          -0.0324      0.1187   -0.27   0.7847
## Year2010          -0.1513      0.1234   -1.23   0.2203
## Year2011          -0.1922      0.1124   -1.71   0.0875 .
## Year2012          -0.2278      0.1266   -1.80   0.0723 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.0517, Adjusted R-squared:  0.0363
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 89 weights are ~= 1. The remaining 978 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.350  0.875   0.949   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      9.37e-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 1.063 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
## -1.5655 -0.3609  0.0179  0.3553  1.5562
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3677     0.0957   14.30  < 2e-16 ***
## LastAuthorFemale1 -0.1520     0.0399   -3.81  0.00015 ***
## Year1997          0.1978     0.1300    1.52  0.12848
## Year1998          0.0574     0.1214    0.47  0.63623
## Year1999         -0.0151     0.1212   -0.12  0.90091
## Year2000         -0.0589     0.1238   -0.48  0.63448
## Year2001          0.0229     0.1071    0.21  0.83047
## Year2002          0.0449     0.1159    0.39  0.69835
## Year2003         -0.0674     0.1139   -0.59  0.55429
## Year2004         -0.0641     0.1143   -0.56  0.57490
## Year2005         -0.0857     0.1088   -0.79  0.43075
## Year2006         -0.2047     0.1113   -1.84  0.06624 .
```

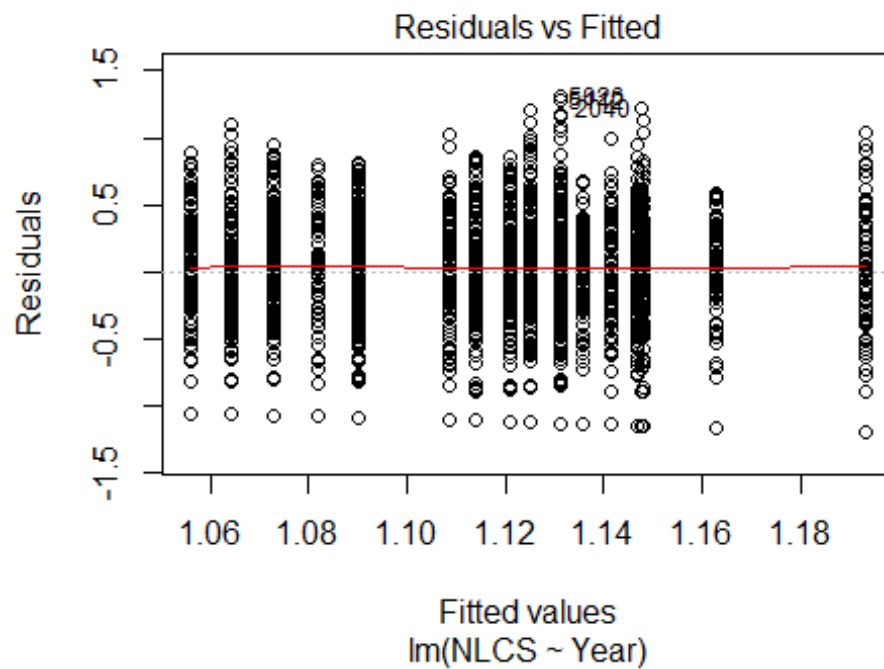
```

## Year2007          -0.0485      0.1155   -0.42  0.67485
## Year2008          -0.0904      0.1117   -0.81  0.41828
## Year2009          -0.0299      0.1159   -0.26  0.79619
## Year2010          -0.1287      0.1222   -1.05  0.29240
## Year2011          -0.2017      0.1113   -1.81  0.07022 .
## Year2012          -0.2285      0.1251   -1.83  0.06812 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.525
## Multiple R-squared:  0.0571, Adjusted R-squared:  0.0418
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 986 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.354  0.877  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      9.37e-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: 1067"
## [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
##  143  144  174  181  168  196  201  156  165  187  191  248  275  307  332
## 2011 2012
##  349  380
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   80   81  106  120   93   71  151  108  111  124  129  167  192  223  233
## 2011 2012

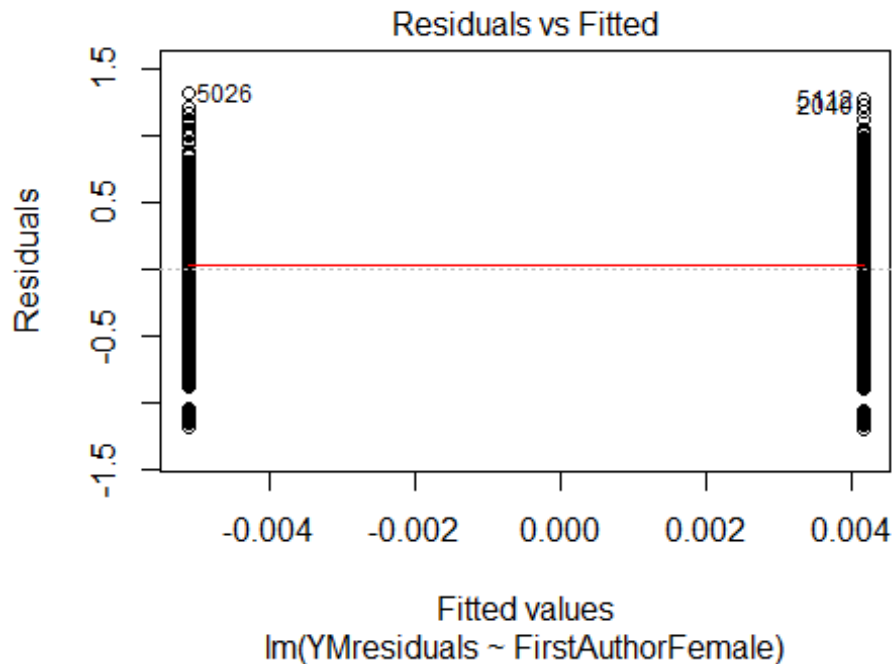
```



```
## 265 281
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 72 72 88 104 79 65 124 90 88 105 106 132 163 195 204
## 2011 2012
## 237 247
## [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
```

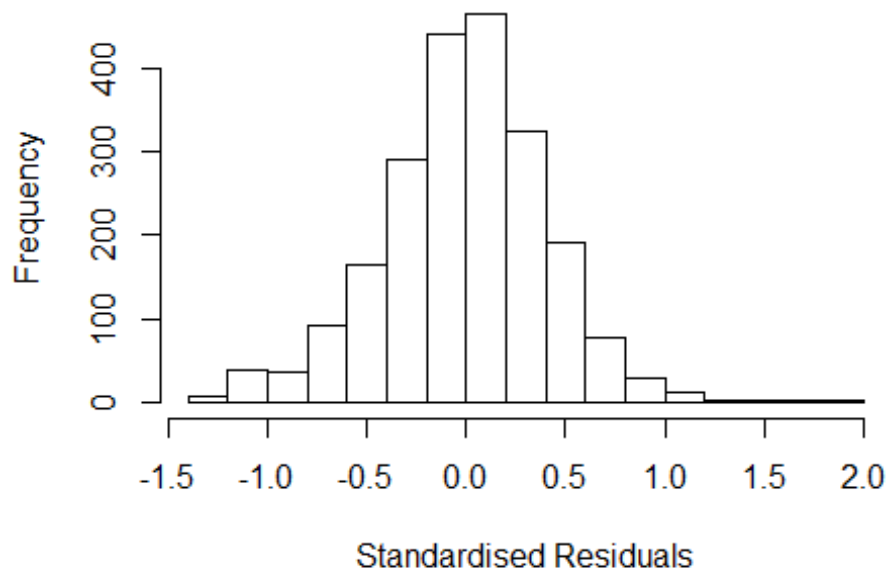


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 5.76113219767856"
## [1] "Male first author team size 2018 geometric mean: 5.40004935315016"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6500, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.45492264692042"
## [1] "Male last author team size 2018 geometric mean: 5.62451535552314"
##
## 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 1.071 1      1.035
## LastAuthorFemale  1.080 1      1.039
## UniqueAuthors    1.303 4      1.034
## Year              1.381 16     1.010
```

Residuals from first and last author and team size



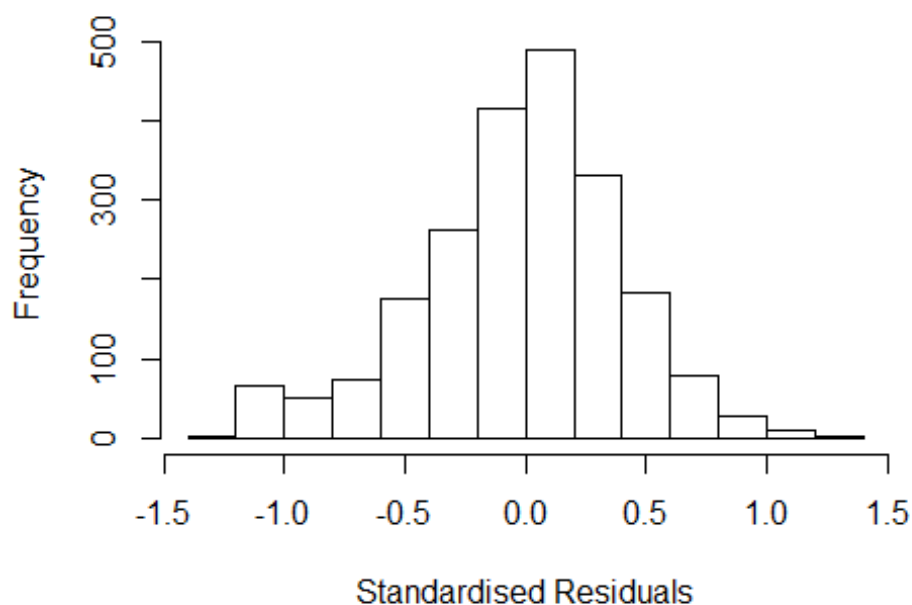
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27201 -0.24963 0.00531 0.24347 1.83176
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.80463 0.07544 10.67 < 2e-16 ***
## FirstAuthorFemale1 -0.02638 0.01771 -1.49 0.13648
## LastAuthorFemale1 -0.07459 0.02146 -3.48 0.00052 ***
## UniqueAuthors2 0.33728 0.06545 5.15 2.8e-07 ***
## UniqueAuthors3 0.39087 0.06319 6.19 7.4e-10 ***
## UniqueAuthors4 0.45838 0.06368 7.20 8.4e-13 ***
## UniqueAuthors5 0.52430 0.06010 8.72 < 2e-16 ***
## Year1997 0.00899 0.07580 0.12 0.90557
## Year1998 -0.09278 0.06524 -1.42 0.15514
## Year1999 -0.08885 0.06370 -1.39 0.16319
```

```

## Year2000      -0.05134      0.06660      -0.77      0.44086
## Year2001      -0.08733      0.07718      -1.13      0.25800
## Year2002      -0.06900      0.06095      -1.13      0.25773
## Year2003      -0.09539      0.05943      -1.60      0.10865
## Year2004      -0.06685      0.06630      -1.01      0.31346
## Year2005      -0.07931      0.06271      -1.26      0.20609
## Year2006      -0.17262      0.06336      -2.72      0.00650 **
## Year2007      -0.16432      0.06568      -2.50      0.01243 *
## Year2008      -0.09783      0.06095      -1.61      0.10858
## Year2009      -0.12374      0.06061      -2.04      0.04133 *
## Year2010      -0.10659      0.06071      -1.76      0.07928 .
## Year2011      -0.09043      0.06053      -1.49      0.13528
## Year2012      -0.09643      0.05922      -1.63      0.10361
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.371
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0928
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 2037 is an outlier with |weight| = 0 ( < 4.6e-05);
## 184 weights are ~= 1. The remaining 1986 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8590 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      4.61e-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 1.049 1      1.024
## LastAuthorFemale 1.056 1      1.028
## Year      1.107 16      1.003

```

Residuals from first and last author



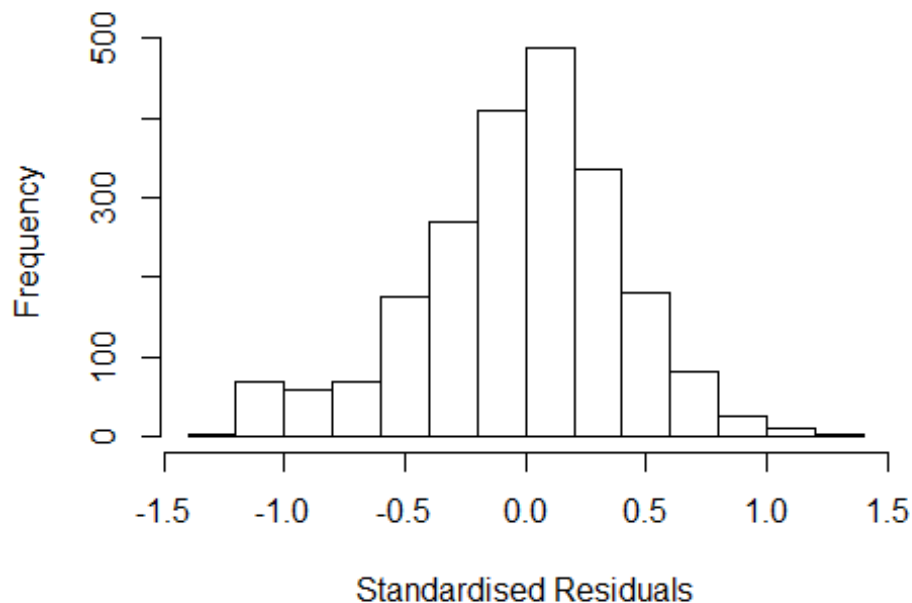
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2295 -0.2623  0.0126  0.2500  1.3475
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2142    0.0564   21.52  <2e-16 ***
## FirstAuthorFemale1 -0.0048    0.0181   -0.26   0.7913
## LastAuthorFemale1 -0.0744    0.0222   -3.36   0.0008 ***
## Year1997          0.0154    0.0804    0.19   0.8484
## Year1998         -0.0928    0.0701   -1.32   0.1857
## Year1999         -0.0946    0.0692   -1.37   0.1720
## Year2000         -0.0525    0.0735   -0.71   0.4755
## Year2001         -0.0742    0.0819   -0.91   0.3648
## Year2002         -0.0312    0.0644   -0.48   0.6278
## Year2003         -0.0420    0.0634   -0.66   0.5080
## Year2004         -0.0217    0.0675   -0.32   0.7481
## Year2005         -0.0404    0.0661   -0.61   0.5407
```

```

## Year2006          -0.1427      0.0666   -2.14   0.0322 *
## Year2007          -0.1222      0.0683   -1.79   0.0735 .
## Year2008          -0.0577      0.0637   -0.91   0.3656
## Year2009          -0.0909      0.0644   -1.41   0.1581
## Year2010          -0.0550      0.0642   -0.86   0.3917
## Year2011          -0.0455      0.0642   -0.71   0.4783
## Year2012          -0.0434      0.0631   -0.69   0.4917
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.00522
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 190 weights are ~= 1. The remaining 1981 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.181  0.859  0.950  0.888  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.61e-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 1.049 1      1.024
## Year              1.049 16      1.001

```

Residuals from first author



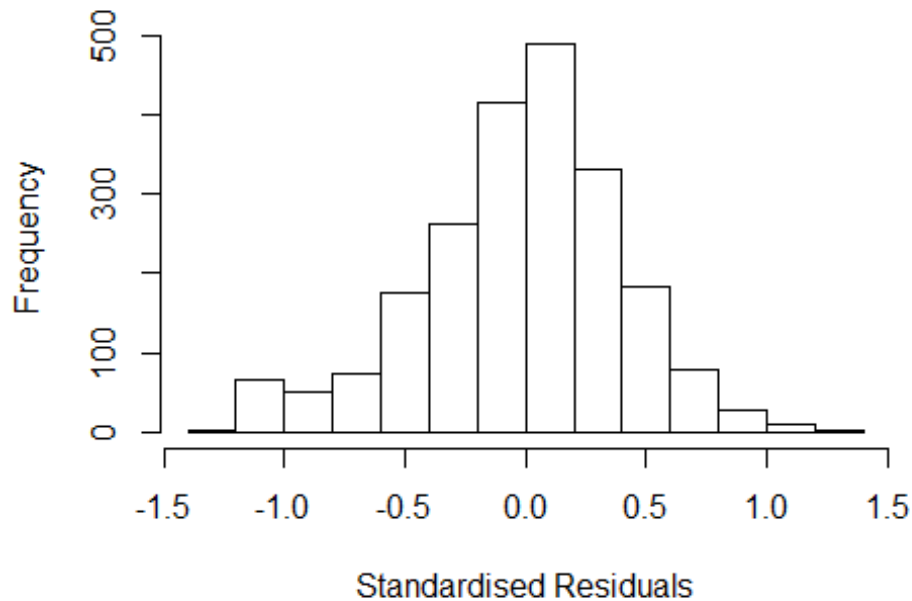
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2080 -0.2562  0.0165  0.2540  1.2972
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20030    0.05641   21.28  <2e-16 ***
## FirstAuthorFemale1 -0.00832    0.01819   -0.46    0.648
## Year1997         0.00766    0.08153    0.09    0.925
## Year1998        -0.09629    0.06979   -1.38    0.168
## Year1999        -0.08842    0.06915   -1.28    0.201
## Year2000        -0.04791    0.07410   -0.65    0.518
## Year2001        -0.06603    0.08130   -0.81    0.417
## Year2002        -0.03418    0.06445   -0.53    0.596
## Year2003        -0.04414    0.06328   -0.70    0.486
## Year2004        -0.02546    0.06740   -0.38    0.706
## Year2005        -0.04339    0.06615   -0.66    0.512
## Year2006        -0.14101    0.06654   -2.12    0.034 *
```

```

## Year2007          -0.12181    0.06829   -1.78    0.075 .
## Year2008          -0.06266    0.06351   -0.99    0.324
## Year2009          -0.09479    0.06426   -1.48    0.140
## Year2010          -0.05649    0.06416   -0.88    0.379
## Year2011          -0.05049    0.06420   -0.79    0.432
## Year2012          -0.05018    0.06298   -0.80    0.426
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.379
## Multiple R-squared:  0.00726,    Adjusted R-squared:  -0.000579
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 202 weights are ~= 1. The remaining 1969 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.219  0.851  0.949   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      4.61e-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 1.056 1      1.027
## Year              1.056 16      1.002

```


Residuals from last author



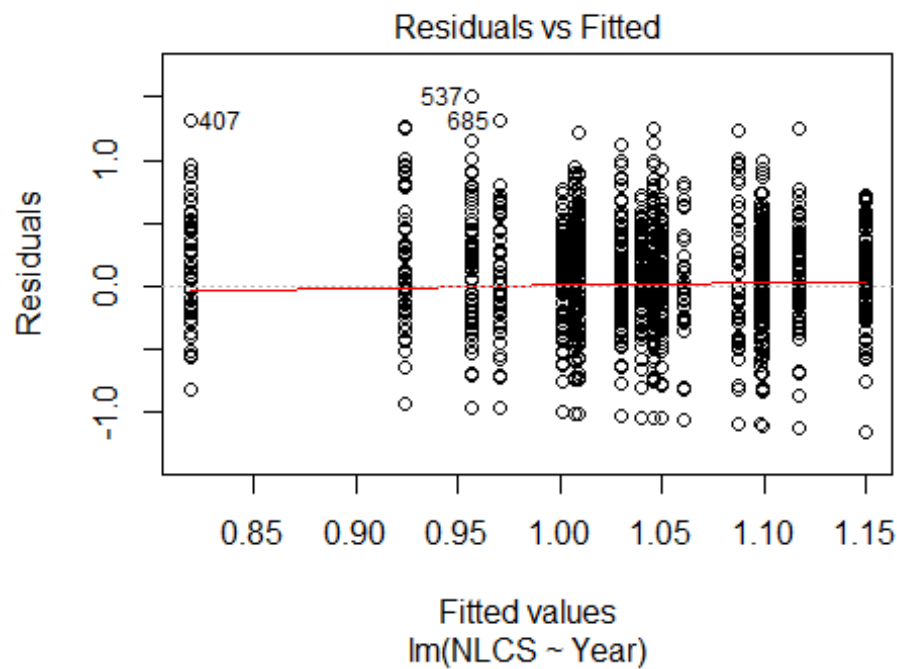
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2280 -0.2633  0.0113  0.2507  1.3452
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2127     0.0561  21.63  < 2e-16 ***
## LastAuthorFemale1 -0.0747     0.0222  -3.37  0.00077 ***
## Year1997          0.0153     0.0804   0.19  0.84935
## Year1998         -0.0927     0.0700  -1.32  0.18550
## Year1999         -0.0950     0.0692  -1.37  0.16956
## Year2000         -0.0530     0.0734  -0.72  0.47053
## Year2001         -0.0747     0.0819  -0.91  0.36169
## Year2002         -0.0319     0.0644  -0.50  0.61994
## Year2003         -0.0431     0.0632  -0.68  0.49569
## Year2004         -0.0222     0.0675  -0.33  0.74272
## Year2005         -0.0411     0.0660  -0.62  0.53297
## Year2006         -0.1437     0.0665  -2.16  0.03085 *
```

```

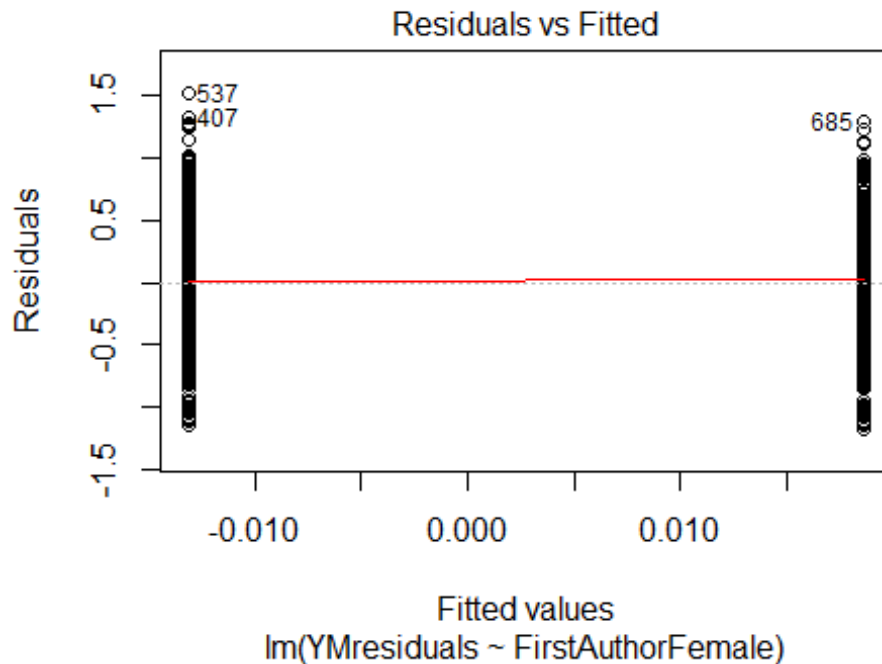
## Year2007          -0.1224      0.0682    -1.79   0.07291 .
## Year2008          -0.0581      0.0637    -0.91   0.36152
## Year2009          -0.0916      0.0642    -1.43   0.15405
## Year2010          -0.0560      0.0640    -0.88   0.38141
## Year2011          -0.0464      0.0640    -0.73   0.46816
## Year2012          -0.0442      0.0630    -0.70   0.48328
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0134, Adjusted R-squared:  0.00563
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 191 weights are ~= 1. The remaining 1980 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.183  0.859  0.949  0.888  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.61e-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: 2171"
## [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
## 129 112 117 141 134 146 124 120 109 97 128 143 150 163 151
## 2011 2012
## 169 172
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 70 54 70 92 64 36 93 87 81 76 84 102 104 135 111
## 2011 2012

```

```
## 137 133
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 60 45 62 77 61 36 77 77 69 63 66 85 89 116 98
## 2011 2012
## 124 120
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 45, df = 16, p-value = 1e-04
```

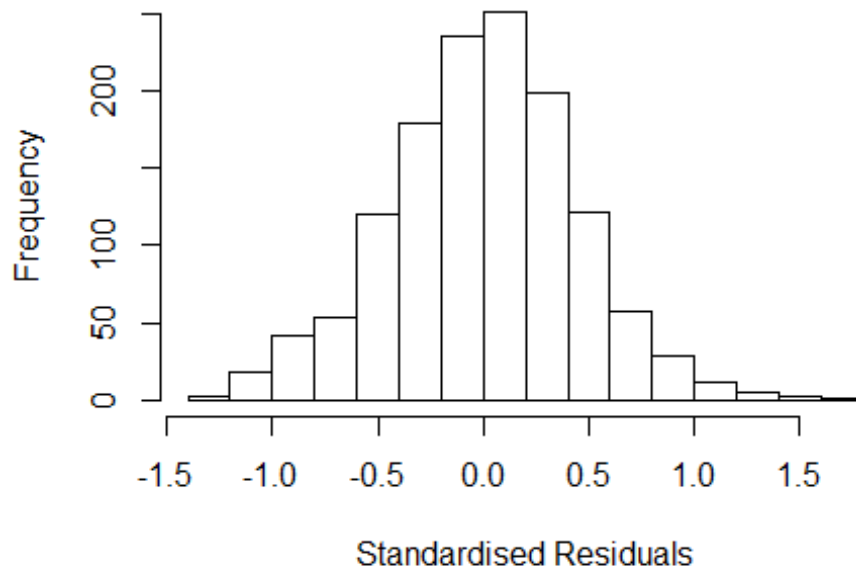


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 5.71041684928405"
## [1] "Male first author team size 2018 geometric mean: 4.93325313248762"
##
## 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: 5.02280380212968"
## [1] "Male last author team size 2018 geometric mean: 5.37474402610771"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, 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 1.103 1      1.050
## LastAuthorFemale  1.103 1      1.050
## UniqueAuthors    1.694 4      1.068
## Year              1.843 16     1.019
```

Residuals from first and last author and team size



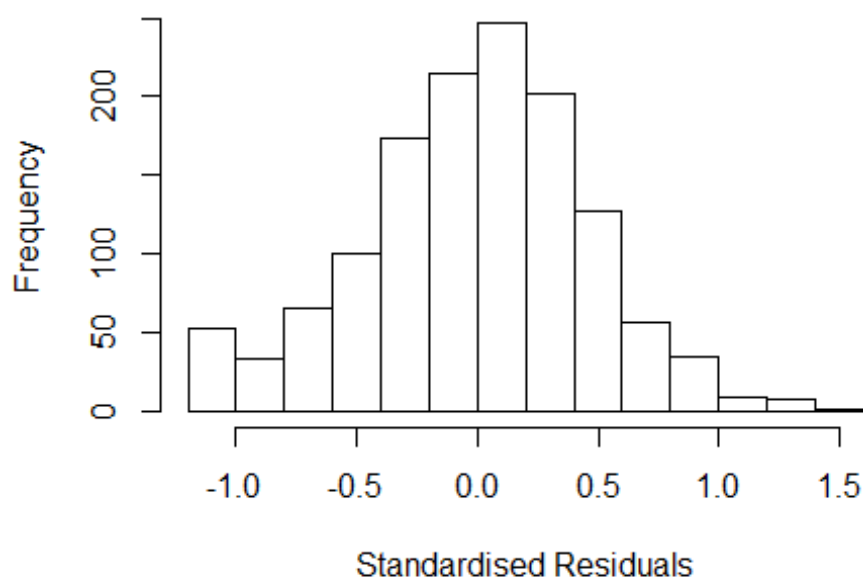
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24361 -0.27672 0.00778 0.27999 1.72143
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4159 0.0999 4.16 3.3e-05 ***
## FirstAuthorFemale1 0.0121 0.0263 0.46 0.646
## LastAuthorFemale1 -0.0720 0.0289 -2.49 0.013 *
## UniqueAuthors2 0.3318 0.0733 4.53 6.6e-06 ***
## UniqueAuthors3 0.4189 0.0718 5.84 6.7e-09 ***
## UniqueAuthors4 0.5337 0.0692 7.71 2.4e-14 ***
## UniqueAuthors5 0.6328 0.0637 9.94 < 2e-16 ***
## Year1997 0.2562 0.1356 1.89 0.059 .
## Year1998 -0.0571 0.1058 -0.54 0.589
## Year1999 0.1289 0.0979 1.32 0.188
```

```

## Year2000          0.1234      0.0998      1.24      0.217
## Year2001          0.2234      0.1295      1.73      0.085 .
## Year2002          0.1124      0.0881      1.28      0.202
## Year2003          0.1236      0.0895      1.38      0.168
## Year2004          0.2197      0.0929      2.37      0.018 *
## Year2005          0.1950      0.0973      2.00      0.045 *
## Year2006          0.1703      0.0934      1.82      0.068 .
## Year2007          0.0591      0.0941      0.63      0.530
## Year2008          0.1409      0.0888      1.59      0.113
## Year2009          0.0993      0.0882      1.13      0.260
## Year2010          0.1477      0.0918      1.61      0.108
## Year2011          0.0947      0.0887      1.07      0.286
## Year2012          0.1270      0.0880      1.44      0.149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.169, Adjusted R-squared:  0.155
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1212 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0432 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          7.55e-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 1.091 1      1.045
## LastAuthorFemale 1.076 1      1.037
## Year              1.166 16      1.005

```

Residuals from first and last author



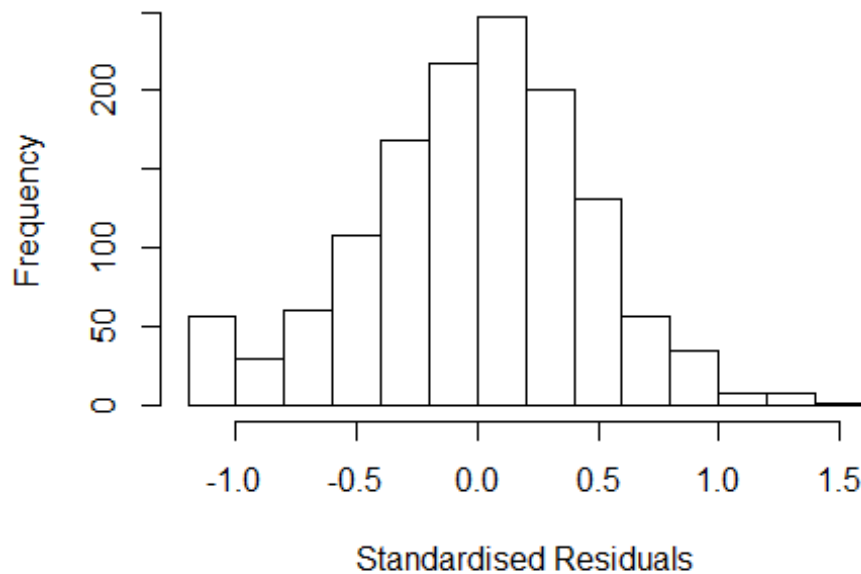
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1689 -0.3081 0.0157 0.2946 1.5310
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8545 0.0802 10.66 <2e-16 ***
## FirstAuthorFemale1 0.0312 0.0275 1.13 0.2569
## LastAuthorFemale1 -0.0518 0.0298 -1.74 0.0821 .
## Year1997 0.2388 0.1391 1.72 0.0862 .
## Year1998 -0.0944 0.1118 -0.84 0.3986
## Year1999 0.0744 0.1049 0.71 0.4782
## Year2000 0.1397 0.1075 1.30 0.1940
## Year2001 0.2300 0.1237 1.86 0.0631 .
## Year2002 0.1633 0.0901 1.81 0.0702 .
## Year2003 0.1818 0.0907 2.01 0.0451 *
## Year2004 0.2818 0.0947 2.97 0.0030 **
## Year2005 0.2797 0.0975 2.87 0.0042 **
```

```

## Year2006          0.2831      0.0941      3.01      0.0027 **
## Year2007          0.1481      0.0966      1.53      0.1253
## Year2008          0.2318      0.0892      2.60      0.0095 **
## Year2009          0.1726      0.0902      1.91      0.0559 .
## Year2010          0.2556      0.0923      2.77      0.0057 **
## Year2011          0.1968      0.0905      2.17      0.0299 *
## Year2012          0.2046      0.0898      2.28      0.0228 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.441
## Multiple R-squared:  0.0402, Adjusted R-squared:  0.0269
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 1217 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.203  0.864  0.949  0.895  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.55e-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 1.084 1      1.041
## Year              1.084 16      1.003

```


Residuals from first author



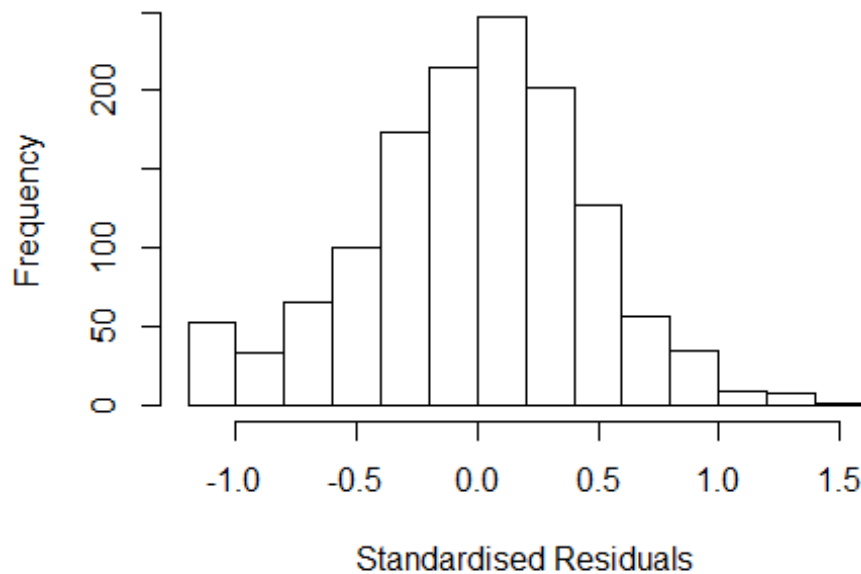
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1537 -0.3041 0.0221 0.3008 1.5469
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8391 0.0801 10.47 <2e-16 ***
## FirstAuthorFemale1 0.0296 0.0276 1.07 0.2832
## Year1997 0.2326 0.1399 1.66 0.0966 .
## Year1998 -0.0910 0.1114 -0.82 0.4142
## Year1999 0.0740 0.1049 0.71 0.4808
## Year2000 0.1369 0.1096 1.25 0.2115
## Year2001 0.2407 0.1219 1.98 0.0485 *
## Year2002 0.1634 0.0909 1.80 0.0727 .
## Year2003 0.1831 0.0914 2.00 0.0454 *
## Year2004 0.2828 0.0955 2.96 0.0031 **
## Year2005 0.2830 0.0984 2.88 0.0041 **
## Year2006 0.2849 0.0945 3.01 0.0026 **
```

```

## Year2007          0.1524      0.0971      1.57      0.1169
## Year2008          0.2307      0.0898      2.57      0.0103 *
## Year2009          0.1741      0.0909      1.92      0.0556 .
## Year2010          0.2600      0.0930      2.80      0.0053 **
## Year2011          0.1980      0.0913      2.17      0.0302 *
## Year2012          0.2048      0.0903      2.27      0.0235 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.443
## Multiple R-squared:  0.038, Adjusted R-squared:  0.0255
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 112 weights are ~= 1. The remaining 1213 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.861  0.949  0.896  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.55e-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 1.069 1      1.034
## Year              1.069 16      1.002

```

Residuals from last author



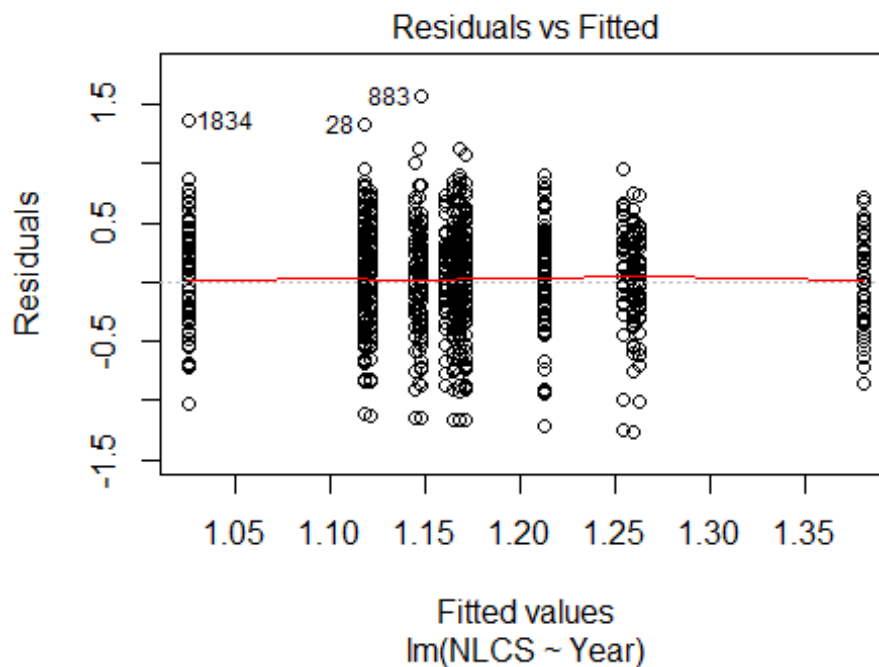
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1522 -0.3032 0.0173 0.3020 1.5199
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8621 0.0801 10.77 <2e-16 ***
## LastAuthorFemale1 -0.0506 0.0298 -1.70 0.0896 .
## Year1997 0.2451 0.1386 1.77 0.0773 .
## Year1998 -0.0942 0.1122 -0.84 0.4013
## Year1999 0.0780 0.1054 0.74 0.4597
## Year2000 0.1416 0.1078 1.31 0.1894
## Year2001 0.2321 0.1236 1.88 0.0607 .
## Year2002 0.1691 0.0899 1.88 0.0602 .
## Year2003 0.1873 0.0910 2.06 0.0397 *
## Year2004 0.2830 0.0952 2.97 0.0030 **
## Year2005 0.2857 0.0971 2.94 0.0033 **
## Year2006 0.2901 0.0940 3.09 0.0021 **
```

```

## Year2007          0.1509      0.0969      1.56      0.1198
## Year2008          0.2346      0.0894      2.62      0.0088 **
## Year2009          0.1750      0.0905      1.93      0.0534 .
## Year2010          0.2627      0.0922      2.85      0.0044 **
## Year2011          0.2045      0.0903      2.26      0.0237 *
## Year2012          0.2143      0.0894      2.40      0.0167 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.442
## Multiple R-squared:  0.0394, Adjusted R-squared:  0.0269
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1212 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.212  0.862  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      7.55e-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: 1325"
## [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
##   90   67   59   81   76  102   89   56   52   54   84  108  114   98   99
## 2011 2012
##  127  123
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   45   38   31   51   43   52   71   43   42   41   51   83   84   77   73
## 2011 2012

```

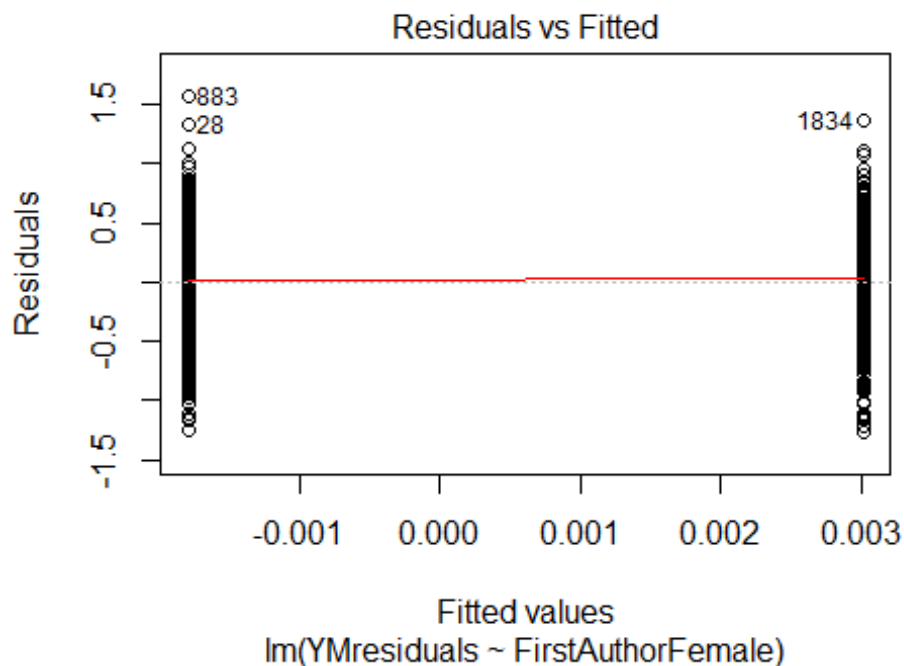
```
## 97 92
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 41 36 29 42 38 46 65 37 36 34 44 73 72 68 64
## 2011 2012
## 93 77
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 17, df = 16, p-value = 0.4
```



```
##
## 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: 5.15117815142941"
## [1] "Male first author team size 2018 geometric mean: 4.83159333374312"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

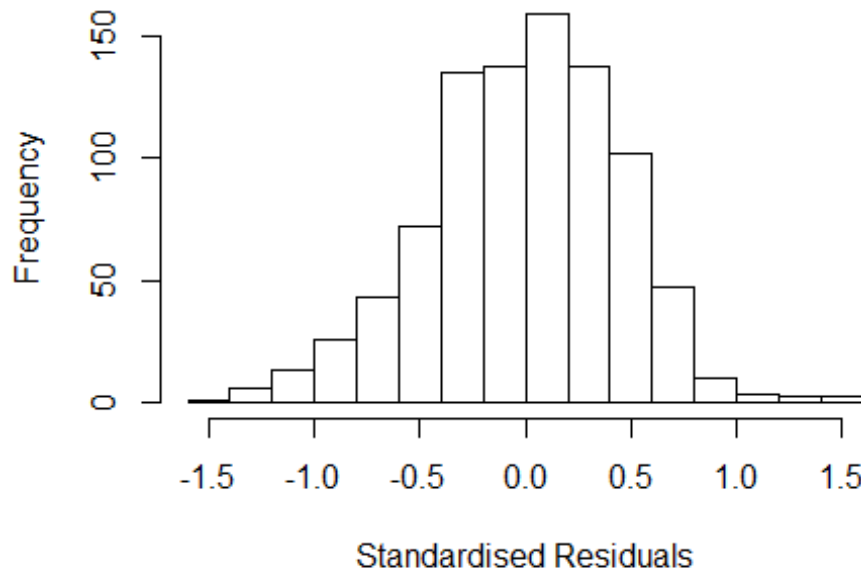
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 600, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.59603876937269"
## [1] "Male last author team size 2018 geometric mean: 4.6665633695142"

## Warning in wilcox.test.default(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] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.151 1      1.073
## LastAuthorFemale  1.087 1      1.043
## UniqueAuthors    1.566 4      1.058
## Year              1.660 16     1.016
```

Residuals from first and last author and team size



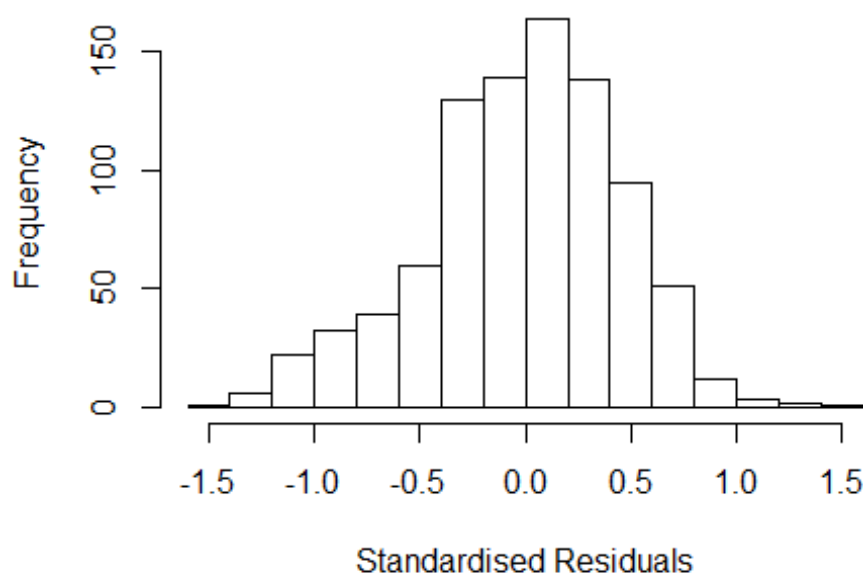
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5113 -0.3057  0.0107  0.3128  1.4499
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9377    0.0878   10.68 < 2e-16 ***
## FirstAuthorFemale1 -0.0080    0.0331   -0.24  0.80907
## LastAuthorFemale1  0.0690    0.0351    1.96  0.05000 .
## UniqueAuthors2    0.1698    0.0705    2.41  0.01622 *
## UniqueAuthors3    0.2409    0.0688    3.50  0.00049 ***
## UniqueAuthors4    0.2417    0.0629    3.84  0.00013 ***
## UniqueAuthors5    0.3998    0.0566    7.06  3.4e-12 ***
## Year1997         -0.0557    0.1040   -0.54  0.59222
## Year1998          0.1738    0.1266    1.37  0.17012
## Year1999          0.1244    0.1005    1.24  0.21583
```

```

## Year2000          0.1627      0.1024      1.59  0.11254
## Year2001          0.0210      0.1009      0.21  0.83547
## Year2002         -0.0584      0.0926     -0.63  0.52825
## Year2003         -0.0692      0.1076     -0.64  0.51994
## Year2004          0.0461      0.0996      0.46  0.64377
## Year2005         -0.1050      0.0976     -1.08  0.28217
## Year2006         -0.0744      0.1036     -0.72  0.47272
## Year2007         -0.0239      0.0958     -0.25  0.80278
## Year2008         -0.1346      0.0957     -1.41  0.16000
## Year2009         -0.0185      0.0937     -0.20  0.84393
## Year2010         -0.1362      0.0936     -1.46  0.14585
## Year2011         -0.1263      0.0897     -1.41  0.15961
## Year2012         -0.1416      0.0927     -1.53  0.12713
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.449
## Multiple R-squared:  0.107, Adjusted R-squared:  0.0847
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 820 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.235  0.885  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          1.12e-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 1.145 1          1.070
## LastAuthorFemale  1.092 1          1.045
## Year              1.148 16          1.004

```


Residuals from first and last author



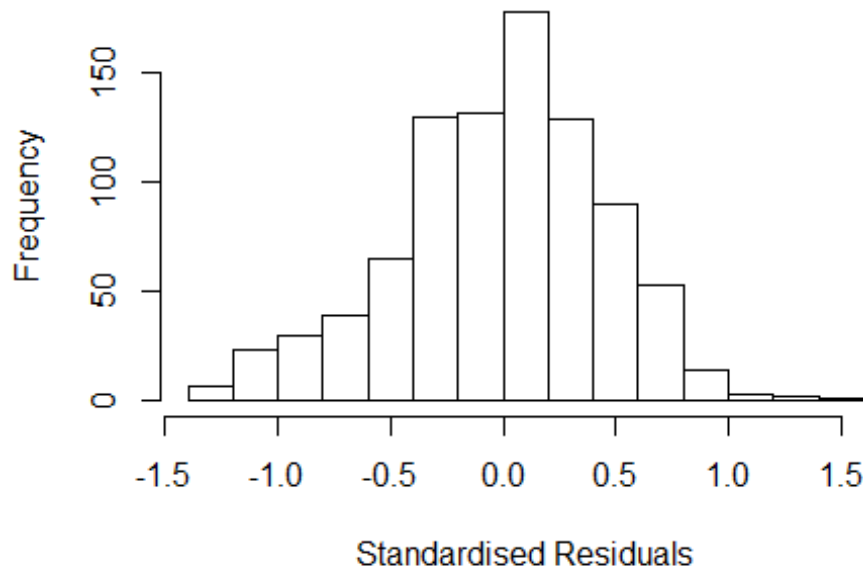
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4168 -0.2994  0.0188  0.3097  1.5742
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.139991   0.077009   14.80  <2e-16 ***
## FirstAuthorFemale1 0.012014   0.034118    0.35   0.725
## LastAuthorFemale1 0.057586   0.036357    1.58   0.114
## Year1997        -0.038145   0.113060   -0.34   0.736
## Year1998         0.219192   0.129728    1.69   0.091 .
## Year1999         0.147151   0.097559    1.51   0.132
## Year2000         0.229415   0.106382    2.16   0.031 *
## Year2001        -0.000638   0.111263   -0.01   0.995
## Year2002         0.013827   0.094972    0.15   0.884
## Year2003         0.029796   0.113452    0.26   0.793
## Year2004         0.114630   0.100188    1.14   0.253
## Year2005         0.009350   0.102698    0.09   0.927
```

```

## Year2006      -0.001225    0.102340   -0.01    0.990
## Year2007      0.034266    0.098494    0.35    0.728
## Year2008     -0.042057    0.098462   -0.43    0.669
## Year2009      0.059165    0.099224    0.60    0.551
## Year2010     -0.015305    0.092915   -0.16    0.869
## Year2011     -0.011633    0.092068   -0.13    0.899
## Year2012     -0.080387    0.092437   -0.87    0.385
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.456
## Multiple R-squared:  0.03,   Adjusted R-squared:  0.0101
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 824 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.208  0.872  0.952  0.904  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.1 1         1.049
## Year              1.1 16         1.003

```

Residuals from first author



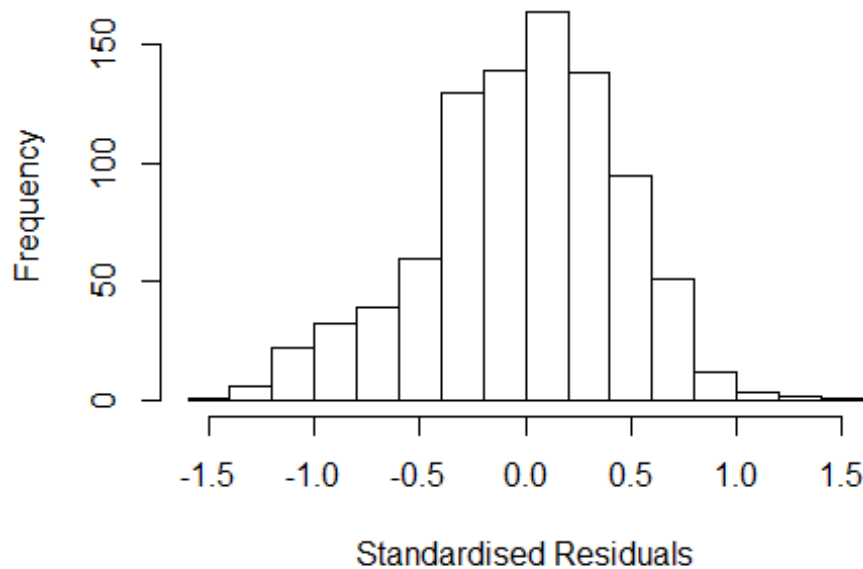
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3727 -0.3084 0.0236 0.3027 1.5612
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14782 0.07831 14.66 <2e-16 ***
## FirstAuthorFemale1 0.02285 0.03345 0.68 0.495
## Year1997 -0.03367 0.11458 -0.29 0.769
## Year1998 0.22489 0.13053 1.72 0.085 .
## Year1999 0.15198 0.09916 1.53 0.126
## Year2000 0.22711 0.10616 2.14 0.033 *
## Year2001 -0.00266 0.11232 -0.02 0.981
## Year2002 0.01961 0.09611 0.20 0.838
## Year2003 0.03117 0.11522 0.27 0.787
## Year2004 0.12199 0.10158 1.20 0.230
## Year2005 0.00812 0.10379 0.08 0.938
## Year2006 0.00401 0.10363 0.04 0.969
```

```

## Year2007          0.03729    0.09913    0.38    0.707
## Year2008         -0.03566    0.09912   -0.36    0.719
## Year2009          0.05782    0.10006    0.58    0.563
## Year2010         -0.01160    0.09396   -0.12    0.902
## Year2011         -0.00721    0.09292   -0.08    0.938
## Year2012         -0.07642    0.09391   -0.81    0.416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.455
## Multiple R-squared:  0.0274, Adjusted R-squared:  0.00857
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 822 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.215  0.873  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      1.12e-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 1.051 1      1.025
## Year              1.051 16      1.002

```

Residuals from last author



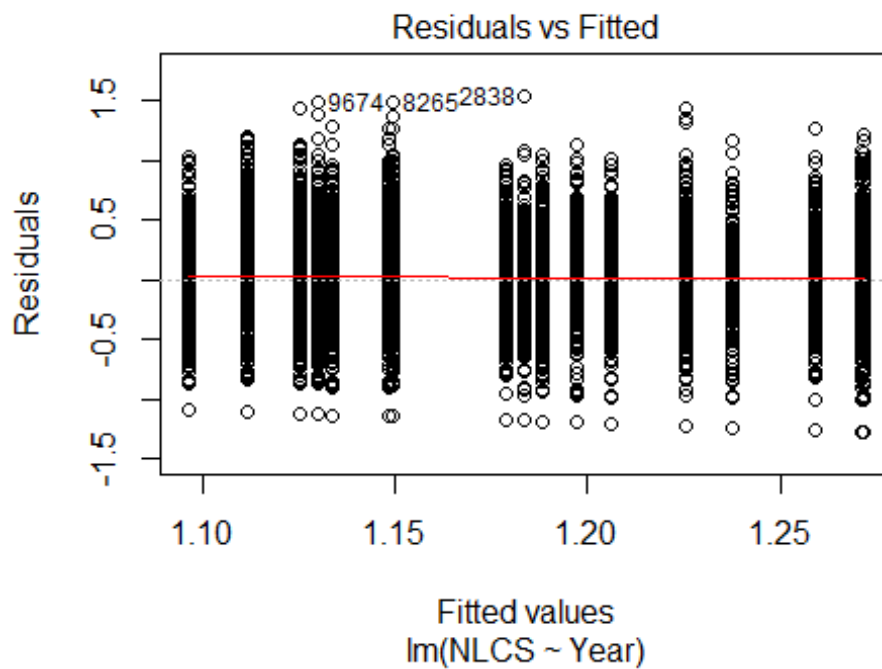
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4216 -0.3045 0.0261 0.3096 1.5699
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.142319 0.076832 14.87 <2e-16 ***
## LastAuthorFemale1 0.060266 0.035638 1.69 0.091 .
## Year1997 -0.038055 0.112994 -0.34 0.736
## Year1998 0.219022 0.129325 1.69 0.091 .
## Year1999 0.149080 0.097311 1.53 0.126
## Year2000 0.230794 0.106380 2.17 0.030 *
## Year2001 0.000816 0.110947 0.01 0.994
## Year2002 0.014649 0.094751 0.15 0.877
## Year2003 0.030541 0.113133 0.27 0.787
## Year2004 0.114624 0.099968 1.15 0.252
## Year2005 0.010881 0.102497 0.11 0.915
## Year2006 0.000782 0.102360 0.01 0.994
```

```

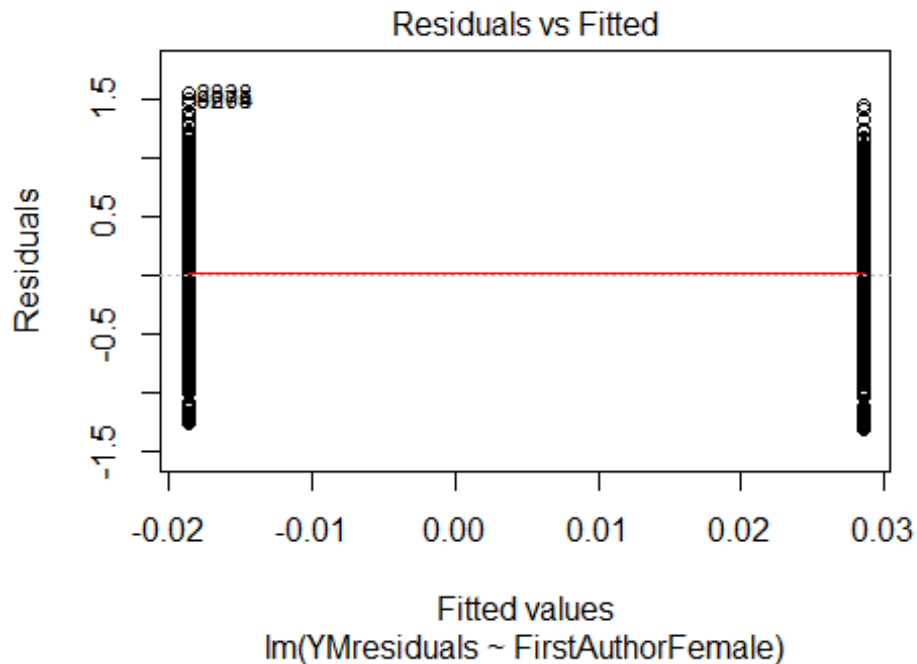
## Year2007          0.036973    0.098623    0.37    0.708
## Year2008          -0.041826    0.098575   -0.42    0.671
## Year2009          0.059925    0.099326    0.60    0.546
## Year2010          -0.013384    0.092661   -0.14    0.885
## Year2011          -0.009608    0.091471   -0.11    0.916
## Year2012          -0.079442    0.092403   -0.86    0.390
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.0298, Adjusted R-squared:  0.011
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 825 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.214  0.874  0.951  0.905  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.12e-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: 895"
## [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
## 423 431 390 441 391 448 395 373 419 454 514 552 624 702 689
## 2011 2012
## 685 641
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 274 272 234 286 214 211 277 268 293 328 344 395 457 511 495
## 2011 2012

```

```
## 500 481
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 242 245 201 252 192 173 244 226 242 270 291 325 389 452 435
## 2011 2012
## 440 418
## [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
```

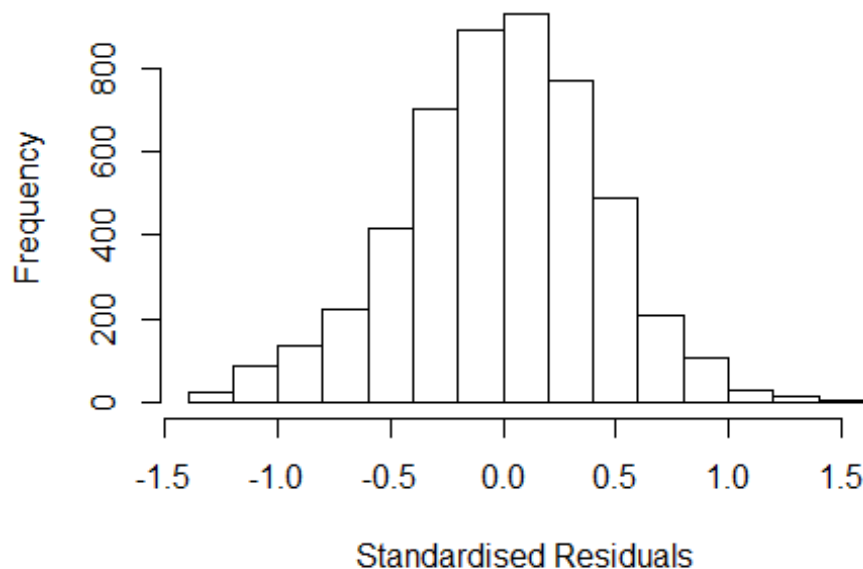


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9, df = 1, p-value = 0.003
```



```
## [1] "Female first author team size 2018 geometric mean: 5.00307170816362"
## [1] "Male first author team size 2018 geometric mean: 4.46707791442608"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 42000, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.95452408194205"
## [1] "Male last author team size 2018 geometric mean: 4.65073467020316"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 35000, 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 1.062 1          1.031
## LastAuthorFemale  1.050 1          1.025
## UniqueAuthors     1.127 4          1.015
## Year              1.168 16         1.005
```


Residuals from first and last author and team size



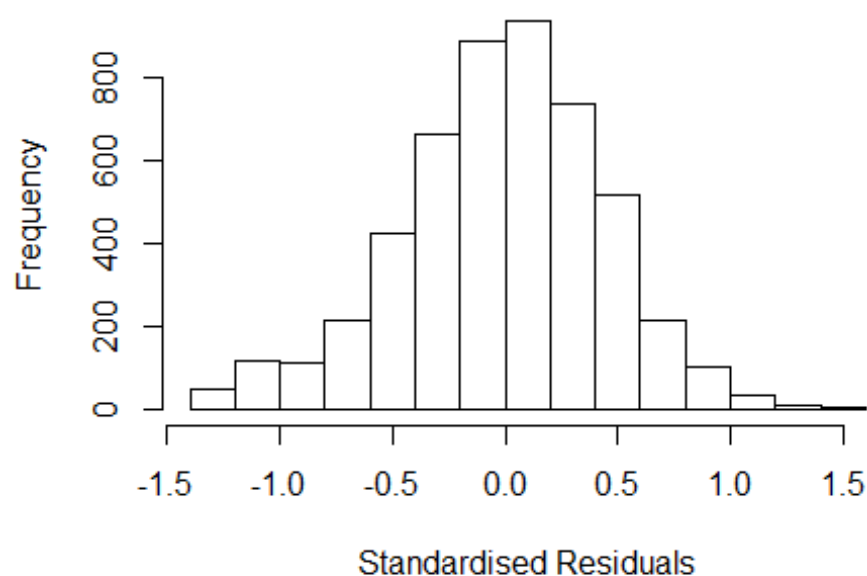
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.38382 -0.27885  0.00665  0.28321  1.53800
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9388    0.0432   21.73 < 2e-16 ***
## FirstAuthorFemale1 0.0313    0.0130    2.41 0.01586 *
## LastAuthorFemale1 0.0709    0.0143    4.96 7.4e-07 ***
## UniqueAuthors2    0.2206    0.0385    5.73 1.1e-08 ***
## UniqueAuthors3    0.2510    0.0374    6.70 2.2e-11 ***
## UniqueAuthors4    0.3095    0.0378    8.18 3.5e-16 ***
## UniqueAuthors5    0.3888    0.0366   10.63 < 2e-16 ***
## Year1997          0.0390    0.0417    0.93 0.34997
## Year1998          0.0400    0.0414    0.97 0.33295
## Year1999          0.0532    0.0372    1.43 0.15292
```

```

## Year2000          0.0249      0.0412      0.60  0.54542
## Year2001         -0.0603      0.0396     -1.52  0.12809
## Year2002         -0.0634      0.0374     -1.70  0.08987 .
## Year2003         -0.0746      0.0374     -2.00  0.04594 *
## Year2004         -0.1730      0.0381     -4.54  5.7e-06 ***
## Year2005         -0.0597      0.0358     -1.67  0.09557 .
## Year2006         -0.1134      0.0382     -2.97  0.00302 **
## Year2007         -0.0642      0.0353     -1.82  0.06850 .
## Year2008         -0.1042      0.0357     -2.92  0.00357 **
## Year2009         -0.1054      0.0357     -2.95  0.00318 **
## Year2010         -0.1237      0.0359     -3.44  0.00058 ***
## Year2011         -0.1213      0.0345     -3.52  0.00044 ***
## Year2012         -0.1556      0.0369     -4.22  2.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.419
## Multiple R-squared:  0.0696, Adjusted R-squared:  0.0655
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 461 weights are ~= 1. The remaining 4576 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.150  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          1.99e-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 1.056 1          1.027
## LastAuthorFemale  1.035 1          1.017
## Year              1.068 16          1.002

```

Residuals from first and last author

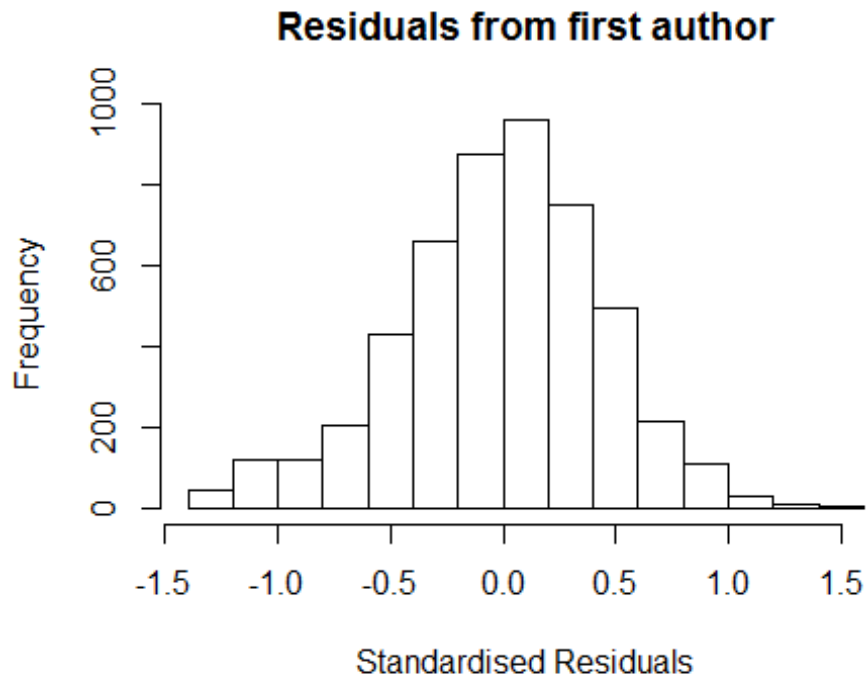


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3506 -0.2850  0.0119  0.2874  1.5557
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2018     0.0287   41.91 < 2e-16 ***
## FirstAuthorFemale1  0.0456     0.0131    3.47  0.00052 ***
## LastAuthorFemale1  0.0647     0.0144    4.51  6.6e-06 ***
## Year1997           0.0474     0.0427    1.11  0.26694
## Year1998           0.0385     0.0429    0.90  0.36994
## Year1999           0.0502     0.0389    1.29  0.19701
## Year2000           0.0224     0.0418    0.54  0.59219
## Year2001          -0.0425     0.0401   -1.06  0.28880
## Year2002          -0.0442     0.0392   -1.13  0.26003
## Year2003          -0.0426     0.0385   -1.10  0.26924
## Year2004          -0.1495     0.0399   -3.75  0.00018 ***
## Year2005          -0.0291     0.0368   -0.79  0.42911
```

```

## Year2006          -0.0857      0.0394   -2.18  0.02948 *
## Year2007          -0.0293      0.0365   -0.80  0.42292
## Year2008          -0.0664      0.0374   -1.78  0.07576 .
## Year2009          -0.0679      0.0372   -1.83  0.06794 .
## Year2010          -0.0858      0.0374   -2.29  0.02182 *
## Year2011          -0.0887      0.0360   -2.47  0.01360 *
## Year2012          -0.1116      0.0383   -2.91  0.00359 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.427
## Multiple R-squared:  0.02,   Adjusted R-squared:  0.0165
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 418 weights are ~= 1. The remaining 4619 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.157  0.867   0.951   0.898   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.99e-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 1.047 1      1.023
## Year              1.047 16      1.001

```



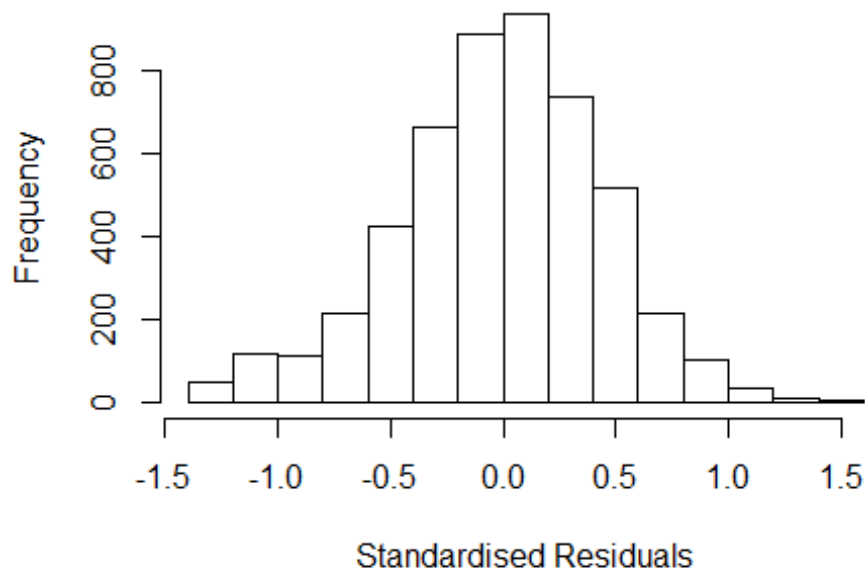
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3163 -0.2826 0.0129 0.2872 1.5419
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2132 0.0285 42.50 < 2e-16 ***
## FirstAuthorFemale1 0.0542 0.0131 4.14 3.5e-05 ***
## Year1997 0.0489 0.0425 1.15 0.25054
## Year1998 0.0394 0.0430 0.92 0.35894
## Year1999 0.0489 0.0389 1.26 0.20911
## Year2000 0.0253 0.0417 0.61 0.54410
## Year2001 -0.0401 0.0401 -1.00 0.31792
## Year2002 -0.0447 0.0392 -1.14 0.25435
## Year2003 -0.0397 0.0386 -1.03 0.30394
## Year2004 -0.1506 0.0399 -3.78 0.00016 ***
## Year2005 -0.0284 0.0369 -0.77 0.44133
## Year2006 -0.0782 0.0392 -2.00 0.04578 *
```

```

## Year2007          -0.0217      0.0366   -0.59  0.55278
## Year2008          -0.0635      0.0374   -1.70  0.08971 .
## Year2009          -0.0669      0.0372   -1.80  0.07182 .
## Year2010          -0.0828      0.0374   -2.21  0.02700 *
## Year2011          -0.0860      0.0359   -2.39  0.01677 *
## Year2012          -0.1077      0.0383   -2.81  0.00496 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.427
## Multiple R-squared:  0.0159, Adjusted R-squared:  0.0126
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 428 weights are ~= 1. The remaining 4609 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.165  0.869  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      1.99e-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 1.025 1          1.013
## Year              1.025 16          1.001

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.337 -0.283 0.012 0.287 1.541
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2133 0.0287 42.26 < 2e-16 ***
## LastAuthorFemale1 0.0721 0.0143 5.04 4.8e-07 ***
## Year1997 0.0473 0.0426 1.11 0.26641
## Year1998 0.0404 0.0432 0.94 0.34926
## Year1999 0.0514 0.0390 1.32 0.18809
## Year2000 0.0231 0.0420 0.55 0.58166
## Year2001 -0.0393 0.0401 -0.98 0.32628
## Year2002 -0.0393 0.0393 -1.00 0.31726
## Year2003 -0.0400 0.0387 -1.03 0.30080
## Year2004 -0.1430 0.0399 -3.58 0.00034 ***
## Year2005 -0.0237 0.0371 -0.64 0.52260
## Year2006 -0.0801 0.0394 -2.03 0.04220 *
```

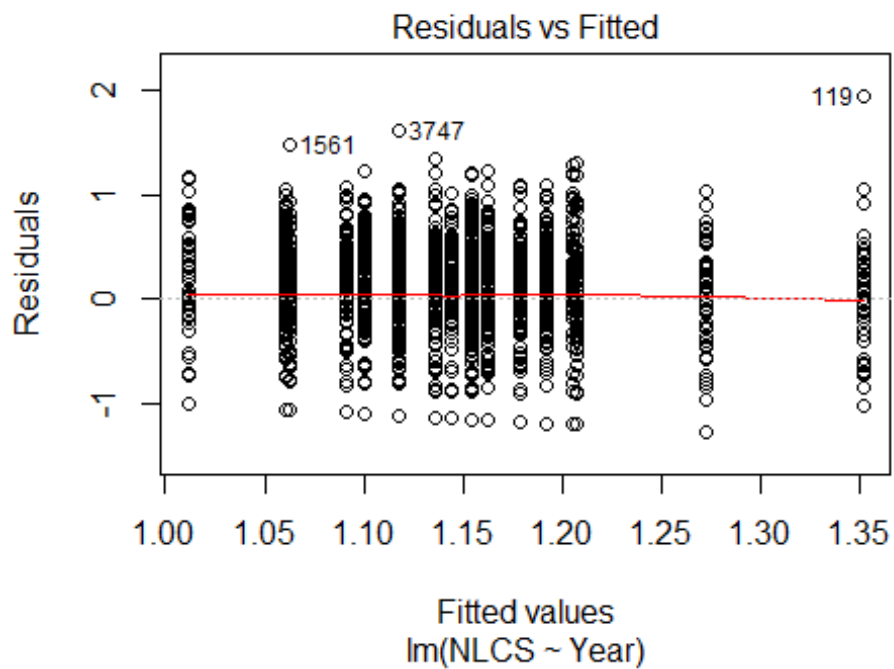
```

## Year2007          -0.0256      0.0365   -0.70   0.48280
## Year2008          -0.0593      0.0374   -1.59   0.11300
## Year2009          -0.0616      0.0373   -1.65   0.09813 .
## Year2010          -0.0763      0.0372   -2.05   0.04061 *
## Year2011          -0.0814      0.0359   -2.26   0.02361 *
## Year2012          -0.1018      0.0383   -2.65   0.00796 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.428
## Multiple R-squared:  0.0174, Adjusted R-squared:  0.0141
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 420 weights are ~= 1. The remaining 4617 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.167  0.867  0.951  0.898  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.99e-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: 5037"
## [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
## 135 169 137 140 177 173 197 170 167 179 234 245 254 277 298
## 2011 2012
## 292 283
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 51 71 61 70 91 96 118 117 101 124 139 147 163 190 201
## 2011 2012

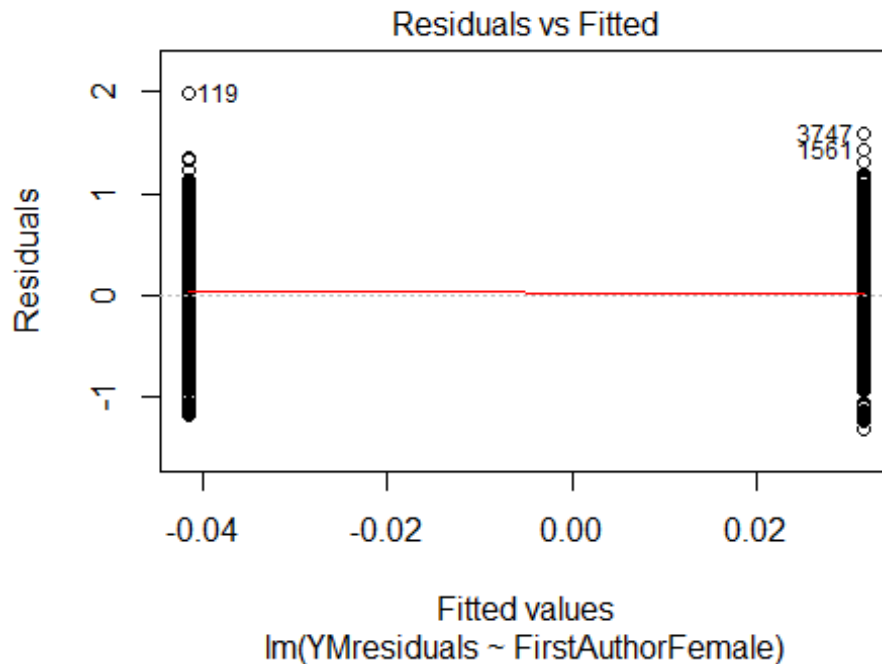
```



```
## 215 183
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 47 64 51 57 71 69 87 98 77 103 123 122 137 165 173
## 2011 2012
## 180 158
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 33, df = 16, p-value = 0.007
```

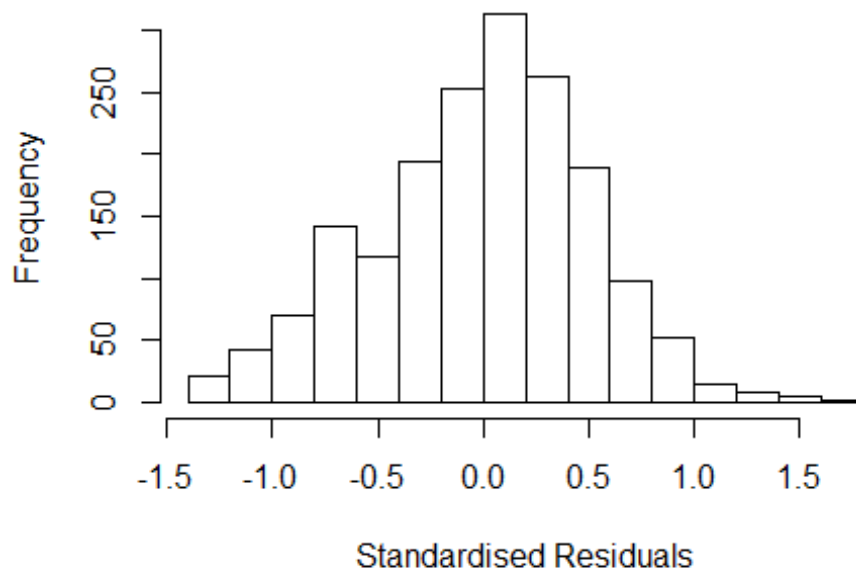


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.5, df = 1, p-value = 0.06
```



```
## [1] "Female first author team size 2018 geometric mean: 3.73380253781573"
## [1] "Male first author team size 2018 geometric mean: 3.5959119350478"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6900, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.70084985779008"
## [1] "Male last author team size 2018 geometric mean: 3.67150264857719"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7900, 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 1.095 1      1.047
## LastAuthorFemale  1.080 1      1.039
## UniqueAuthors    1.307 4      1.034
## Year              1.356 16     1.010
```

Residuals from first and last author and team size



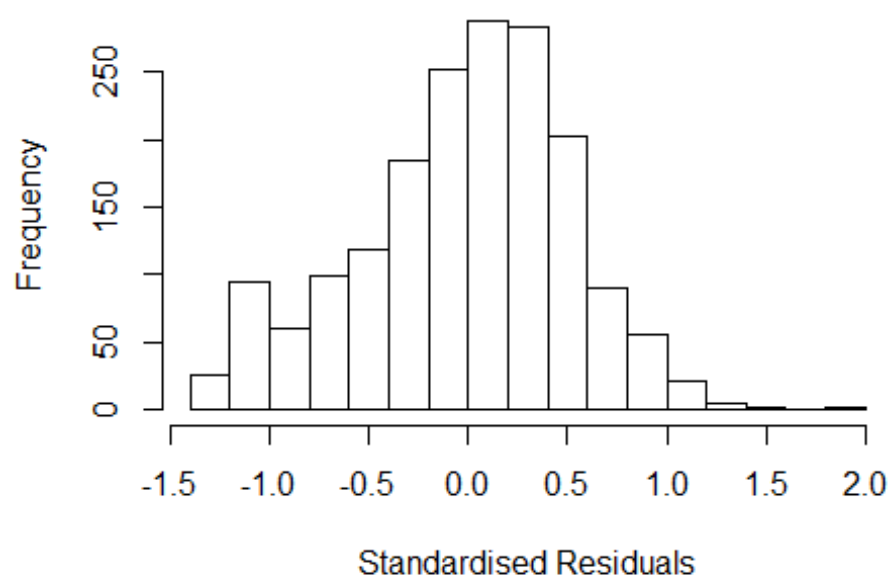
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3901 -0.3340 0.0331 0.3343 1.7484
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9967 0.0811 12.29 < 2e-16 ***
## FirstAuthorFemale1 0.0856 0.0265 3.23 0.00125 **
## LastAuthorFemale1 -0.0154 0.0260 -0.59 0.55308
## UniqueAuthors2 0.3622 0.0581 6.23 5.8e-10 ***
## UniqueAuthors3 0.3696 0.0530 6.98 4.2e-12 ***
## UniqueAuthors4 0.4122 0.0550 7.50 1.0e-13 ***
## UniqueAuthors5 0.5489 0.0487 11.26 < 2e-16 ***
## Year1997 -0.1783 0.1131 -1.58 0.11525
## Year1998 -0.0758 0.0995 -0.76 0.44598
## Year1999 -0.4229 0.1192 -3.55 0.00040 ***
```

```

## Year2000          -0.2579      0.1091    -2.36  0.01821 *
## Year2001          -0.1878      0.1083    -1.73  0.08312 .
## Year2002          -0.2270      0.0944    -2.40  0.01633 *
## Year2003          -0.2740      0.0880    -3.11  0.00188 **
## Year2004          -0.3642      0.0915    -3.98  7.2e-05 ***
## Year2005          -0.3019      0.0901    -3.35  0.00083 ***
## Year2006          -0.2871      0.0842    -3.41  0.00066 ***
## Year2007          -0.2603      0.0852    -3.06  0.00228 **
## Year2008          -0.3587      0.0848    -4.23  2.5e-05 ***
## Year2009          -0.2704      0.0810    -3.34  0.00086 ***
## Year2010          -0.3899      0.0820    -4.75  2.2e-06 ***
## Year2011          -0.3512      0.0820    -4.28  2.0e-05 ***
## Year2012          -0.2411      0.0835    -2.89  0.00392 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.132, Adjusted R-squared:  0.121
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 143 weights are ~= 1. The remaining 1639 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.173  0.854  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          5.61e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.082 1          1.040
## LastAuthorFemale  1.055 1          1.027
## Year              1.075 16          1.002

```

Residuals from first and last author



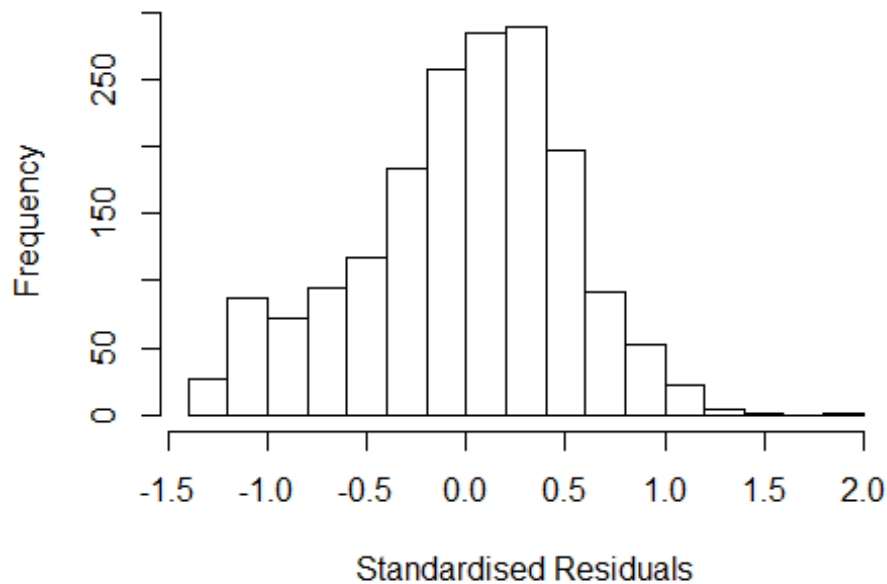
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3032 -0.3346 0.0367 0.3531 1.9864
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3076 0.0753 17.37 < 2e-16 ***
## FirstAuthorFemale1 0.1180 0.0281 4.20 2.8e-05 ***
## LastAuthorFemale1 -0.0429 0.0269 -1.60 0.11064
## Year1997 -0.1257 0.1193 -1.05 0.29225
## Year1998 -0.0301 0.1034 -0.29 0.77093
## Year1999 -0.4349 0.1298 -3.35 0.00082 ***
## Year2000 -0.2348 0.1204 -1.95 0.05138 .
## Year2001 -0.1380 0.1174 -1.18 0.23987
## Year2002 -0.1860 0.0990 -1.88 0.06040 .
## Year2003 -0.1915 0.0908 -2.11 0.03498 *
## Year2004 -0.3142 0.0954 -3.29 0.00101 **
## Year2005 -0.2160 0.0908 -2.38 0.01743 *
```

```

## Year2006          -0.2209      0.0875   -2.52  0.01169 *
## Year2007          -0.1477      0.0876   -1.69  0.09182 .
## Year2008          -0.2602      0.0861   -3.02  0.00255 **
## Year2009          -0.1873      0.0834   -2.25  0.02475 *
## Year2010          -0.3022      0.0849   -3.56  0.00038 ***
## Year2011          -0.2532      0.0847   -2.99  0.00285 **
## Year2012          -0.1223      0.0849   -1.44  0.14994
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.033, Adjusted R-squared:  0.0232
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 139 weights are ~= 1. The remaining 1643 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0811 0.8560 0.9480 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.053 1      1.026
## Year      1.053 16      1.002

```

Residuals from first author



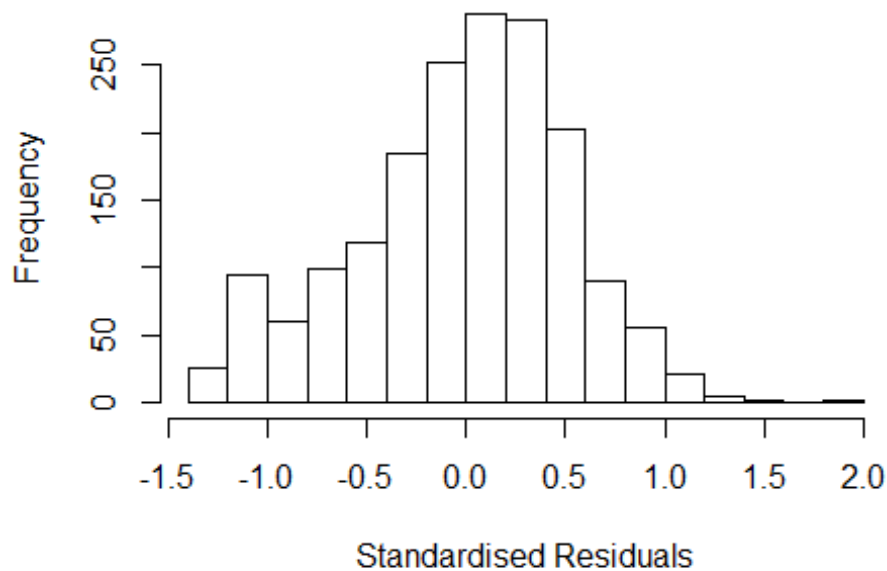
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2784 -0.3501 0.0378 0.3512 1.9969
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2971 0.0756 17.16 < 2e-16 ***
## FirstAuthorFemale1 0.1066 0.0279 3.83 0.00013 ***
## Year1997 -0.1253 0.1201 -1.04 0.29674
## Year1998 -0.0305 0.1037 -0.29 0.76887
## Year1999 -0.4351 0.1293 -3.37 0.00078 ***
## Year2000 -0.2367 0.1206 -1.96 0.04978 *
## Year2001 -0.1385 0.1182 -1.17 0.24151
## Year2002 -0.1872 0.0997 -1.88 0.06053 .
## Year2003 -0.1914 0.0912 -2.10 0.03600 *
## Year2004 -0.3108 0.0959 -3.24 0.00121 **
## Year2005 -0.2154 0.0915 -2.35 0.01867 *
## Year2006 -0.2207 0.0880 -2.51 0.01225 *
```

```

## Year2007          -0.1450      0.0879   -1.65   0.09915 .
## Year2008          -0.2604      0.0865   -3.01   0.00266 **
## Year2009          -0.1860      0.0838   -2.22   0.02666 *
## Year2010          -0.3036      0.0854   -3.55   0.00039 ***
## Year2011          -0.2546      0.0851   -2.99   0.00280 **
## Year2012          -0.1253      0.0852   -1.47   0.14188
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.502
## Multiple R-squared:  0.0317, Adjusted R-squared:  0.0224
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 148 weights are ~= 1. The remaining 1634 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0782 0.8580 0.9480 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      5.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.024 1          1.012
## Year            1.024 16          1.001

```


Residuals from last author



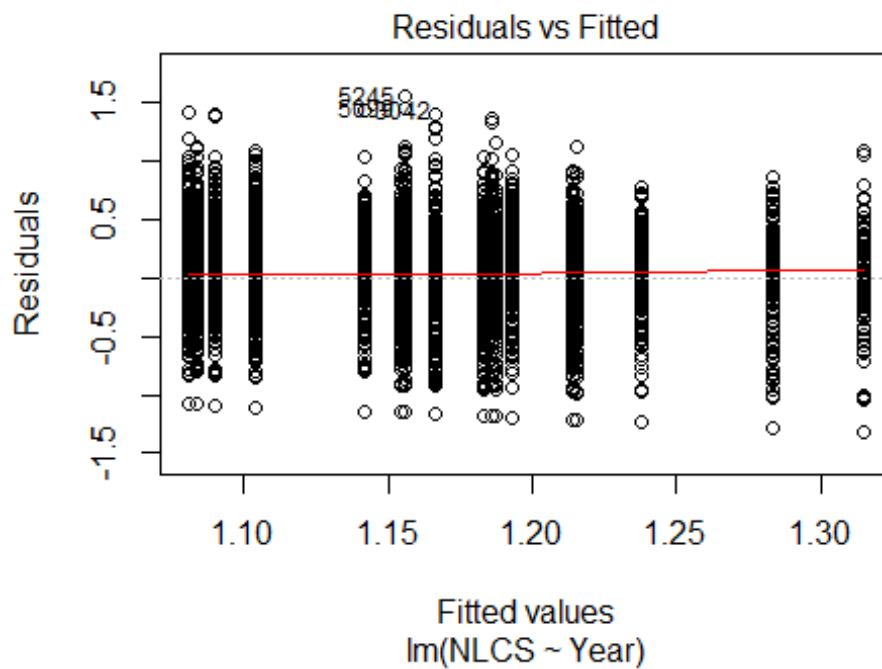
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2565 -0.3435 0.0383 0.3456 1.9493
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3447 0.0756 17.80 <2e-16 ***
## LastAuthorFemale1 -0.0125 0.0267 -0.47 0.6394
## Year1997 -0.1247 0.1229 -1.01 0.3104
## Year1998 -0.0317 0.1048 -0.30 0.7620
## Year1999 -0.4218 0.1299 -3.25 0.0012 **
## Year2000 -0.2291 0.1198 -1.91 0.0559 .
## Year2001 -0.1215 0.1203 -1.01 0.3128
## Year2002 -0.1756 0.0993 -1.77 0.0773 .
## Year2003 -0.1702 0.0919 -1.85 0.0642 .
## Year2004 -0.2940 0.0961 -3.06 0.0023 **
## Year2005 -0.2083 0.0923 -2.26 0.0242 *
## Year2006 -0.2049 0.0886 -2.31 0.0208 *
```

```

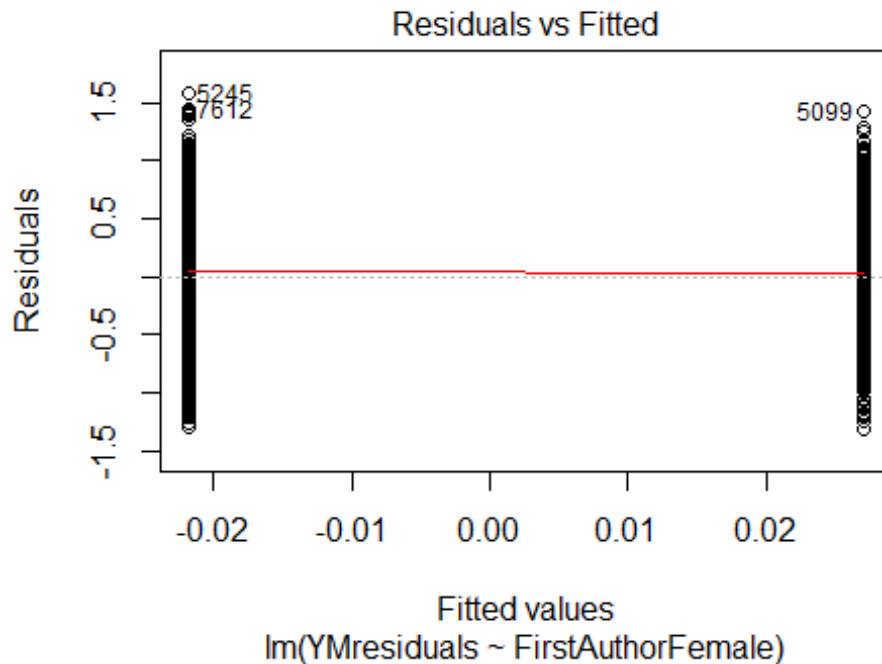
## Year2007          -0.1234      0.0886    -1.39    0.1640
## Year2008          -0.2388      0.0875    -2.73    0.0064 **
## Year2009          -0.1529      0.0841    -1.82    0.0691 .
## Year2010          -0.2738      0.0854    -3.21    0.0014 **
## Year2011          -0.2235      0.0854    -2.62    0.0090 **
## Year2012          -0.0882      0.0857    -1.03    0.3032
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.502
## Multiple R-squared:  0.0228, Adjusted R-squared:  0.0134
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 139 weights are ~= 1. The remaining 1643 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.098  0.852  0.949   0.895  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1782"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
## 269 319 379 345 363 346 349 275 316 365 440 511 513 536 658
## 2011 2012
## 634 686
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 130 165 205 165 185 168 218 186 218 254 302 368 390 405 462
## 2011 2012

```

```
## 453 449
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 111 140 179 133 158 145 186 157 177 228 261 330 350 350 400
## 2011 2012
## 397 398
## [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
```

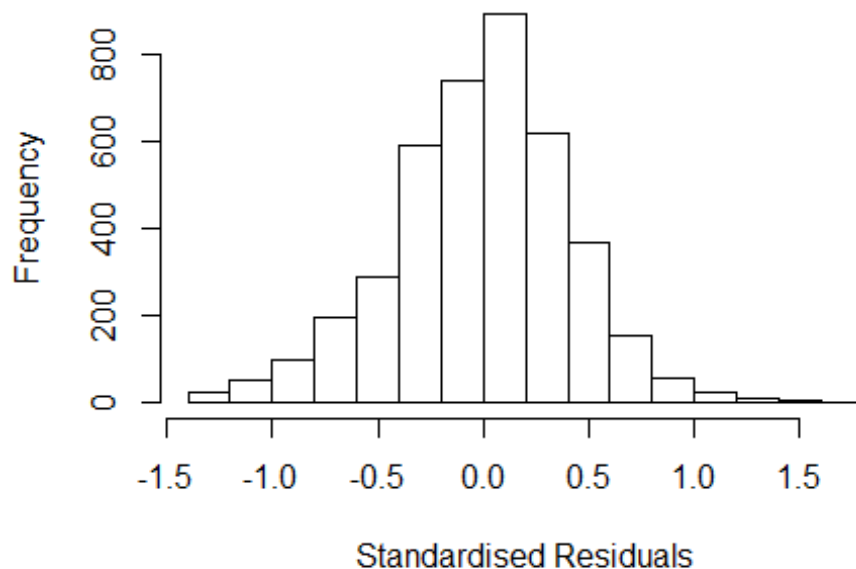


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 37, df = 1, p-value = 1e-09
```



```
## [1] "Female first author team size 2018 geometric mean: 6.04098891718205"
## [1] "Male first author team size 2018 geometric mean: 6.06272627378335"
##
## 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: 6.11998802232574"
## [1] "Male last author team size 2018 geometric mean: 6.01603946132621"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 19000, 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 1.054 1      1.027
## LastAuthorFemale  1.035 1      1.017
## UniqueAuthors    1.144 4      1.017
## Year             1.179 16      1.005
```

Residuals from first and last author and team size



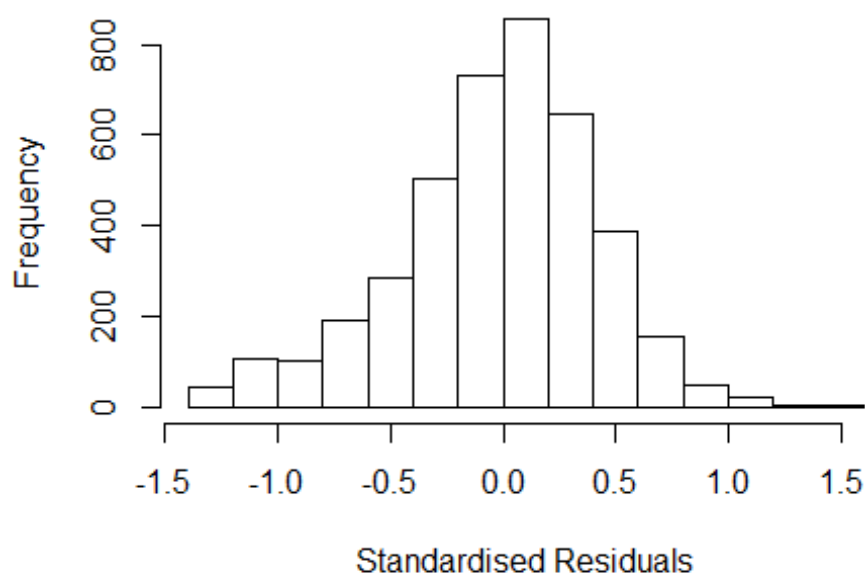
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3909 -0.2566  0.0144  0.2602  1.6562
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9106    0.0597   15.26 < 2e-16 ***
## FirstAuthorFemale1 0.0209    0.0131    1.60  0.11016
## LastAuthorFemale1 0.0356    0.0141    2.52  0.01162 *
## UniqueAuthors2    0.3383    0.0405    8.36 < 2e-16 ***
## UniqueAuthors3    0.4169    0.0394   10.59 < 2e-16 ***
## UniqueAuthors4    0.4883    0.0375   13.03 < 2e-16 ***
## UniqueAuthors5    0.6030    0.0347   17.36 < 2e-16 ***
## Year1997         -0.0711    0.0594   -1.20  0.23121
## Year1998         -0.1227    0.0583   -2.10  0.03547 *
## Year1999         -0.1557    0.0593   -2.63  0.00866 **
```

```

## Year2000          -0.1746      0.0568   -3.07  0.00213 **
## Year2001          -0.1742      0.0611   -2.85  0.00435 **
## Year2002          -0.1890      0.0552   -3.42  0.00063 ***
## Year2003          -0.2699      0.0556   -4.85  1.3e-06 ***
## Year2004          -0.2728      0.0544   -5.02  5.4e-07 ***
## Year2005          -0.2523      0.0535   -4.72  2.4e-06 ***
## Year2006          -0.2179      0.0528   -4.12  3.8e-05 ***
## Year2007          -0.2521      0.0522   -4.83  1.4e-06 ***
## Year2008          -0.2466      0.0516   -4.78  1.8e-06 ***
## Year2009          -0.3046      0.0521   -5.84  5.6e-09 ***
## Year2010          -0.3473      0.0517   -6.72  2.1e-11 ***
## Year2011          -0.3106      0.0516   -6.02  1.9e-09 ***
## Year2012          -0.3470      0.0514   -6.75  1.7e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.385
## Multiple R-squared:  0.175, Adjusted R-squared:  0.171
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 348 weights are ~= 1. The remaining 3752 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0251 0.8640 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          2.44e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.027 1 1.013
## LastAuthorFemale 1.020 1 1.010
## Year 1.047 16 1.001

```

Residuals from first and last author



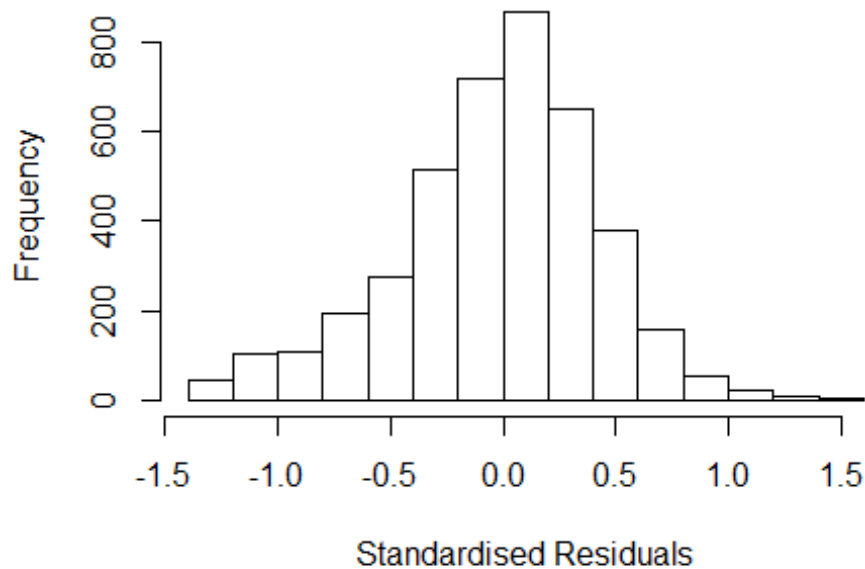
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3648 -0.2710  0.0202  0.2618  1.5586
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3647    0.0461   29.62 < 2e-16 ***
## FirstAuthorFemale1  0.0353    0.0137    2.57  0.01018 *
## LastAuthorFemale1  0.0275    0.0146    1.88  0.06025 .
## Year1997          -0.0627    0.0607   -1.03  0.30145
## Year1998          -0.1178    0.0586   -2.01  0.04447 *
## Year1999          -0.1338    0.0574   -2.33  0.01989 *
## Year2000          -0.1516    0.0555   -2.73  0.00633 **
## Year2001          -0.1467    0.0604   -2.43  0.01514 *
## Year2002          -0.1461    0.0541   -2.70  0.00693 **
## Year2003          -0.2173    0.0552   -3.94  8.3e-05 ***
## Year2004          -0.2278    0.0536   -4.25  2.2e-05 ***
## Year2005          -0.1874    0.0532   -3.52  0.00043 ***
```

```

## Year2006          -0.1456      0.0517   -2.82  0.00486 **
## Year2007          -0.2123      0.0523   -4.06  5.1e-05 ***
## Year2008          -0.2027      0.0505   -4.01  6.1e-05 ***
## Year2009          -0.2631      0.0515   -5.11  3.3e-07 ***
## Year2010          -0.2940      0.0519   -5.66  1.6e-08 ***
## Year2011          -0.2739      0.0510   -5.37  8.3e-08 ***
## Year2012          -0.2853      0.0509   -5.61  2.2e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0293, Adjusted R-squared:  0.025
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 340 weights are ~= 1. The remaining 3760 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0955 0.8550 0.9500 0.8880 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.44e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.027 1      1.013
## Year      1.027 16      1.001

```


Residuals from first author



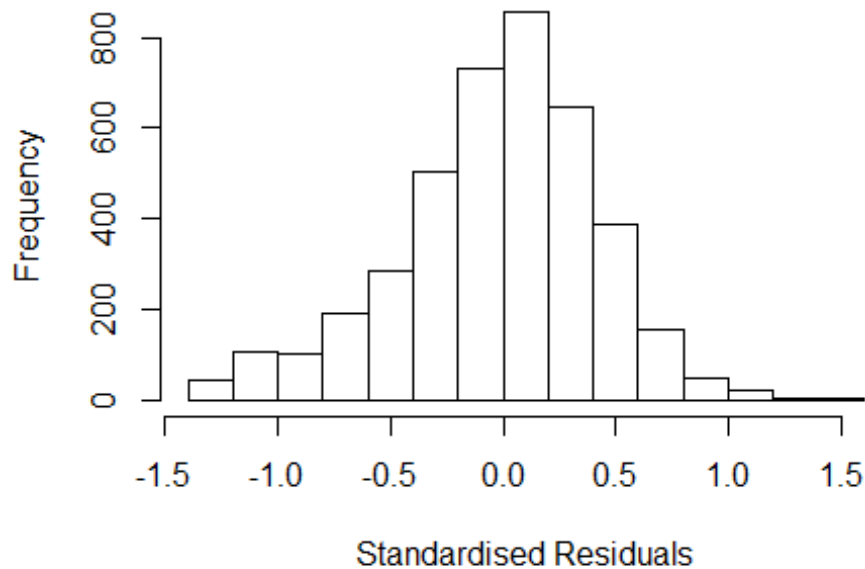
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.373 -0.267 0.020 0.264 1.550
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3733 0.0454 30.26 < 2e-16 ***
## FirstAuthorFemale1 0.0389 0.0139 2.80 0.00507 **
## Year1997 -0.0676 0.0604 -1.12 0.26261
## Year1998 -0.1202 0.0584 -2.06 0.03967 *
## Year1999 -0.1335 0.0573 -2.33 0.01976 *
## Year2000 -0.1541 0.0552 -2.79 0.00523 **
## Year2001 -0.1493 0.0600 -2.49 0.01280 *
## Year2002 -0.1487 0.0537 -2.77 0.00567 **
## Year2003 -0.2204 0.0549 -4.01 6.1e-05 ***
## Year2004 -0.2320 0.0533 -4.35 1.4e-05 ***
## Year2005 -0.1898 0.0529 -3.59 0.00034 ***
## Year2006 -0.1464 0.0514 -2.85 0.00444 **
```

```

## Year2007          -0.2128      0.0521   -4.09  4.4e-05 ***
## Year2008          -0.2044      0.0502   -4.07  4.8e-05 ***
## Year2009          -0.2650      0.0512   -5.18  2.4e-07 ***
## Year2010          -0.2956      0.0517   -5.72  1.1e-08 ***
## Year2011          -0.2764      0.0507   -5.45  5.5e-08 ***
## Year2012          -0.2860      0.0506   -5.65  1.7e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0284, Adjusted R-squared:  0.0243
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 345 weights are ~= 1. The remaining 3755 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0997 0.8550 0.9510 0.8880 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.44e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.021 1      1.011
## Year      1.021 16      1.001

```

Residuals from last author



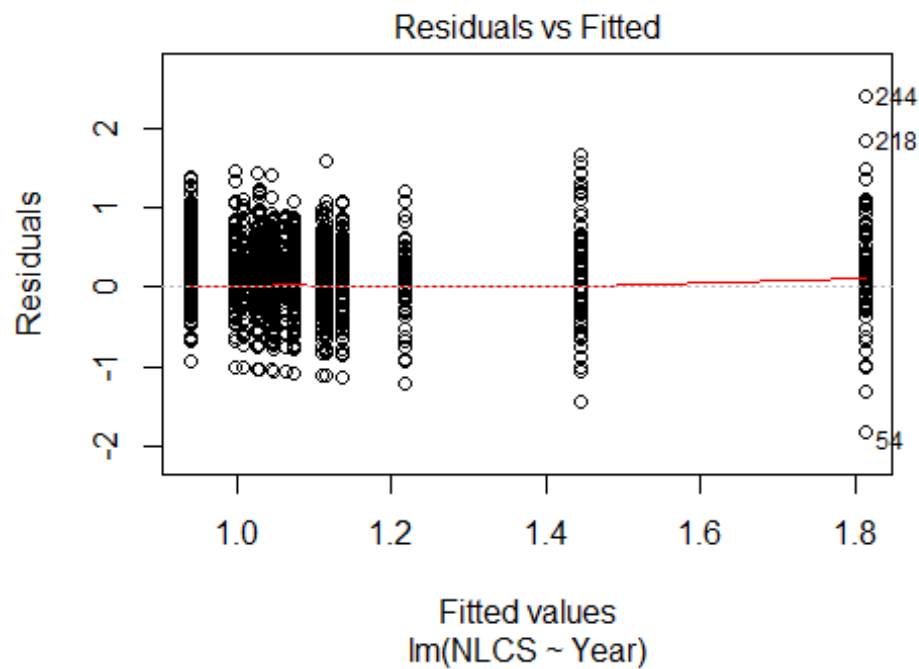
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3731 -0.2730 0.0208 0.2650 1.5430
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3731 0.0457 30.04 < 2e-16 ***
## LastAuthorFemale1 0.0327 0.0147 2.22 0.02617 *
## Year1997 -0.0603 0.0605 -1.00 0.31912
## Year1998 -0.1143 0.0584 -1.96 0.05037 .
## Year1999 -0.1304 0.0571 -2.28 0.02249 *
## Year2000 -0.1480 0.0553 -2.68 0.00744 **
## Year2001 -0.1399 0.0600 -2.33 0.01982 *
## Year2002 -0.1372 0.0538 -2.55 0.01078 *
## Year2003 -0.2105 0.0547 -3.85 0.00012 ***
## Year2004 -0.2204 0.0533 -4.13 3.6e-05 ***
## Year2005 -0.1792 0.0528 -3.39 0.00070 ***
## Year2006 -0.1382 0.0515 -2.69 0.00726 **
```

```

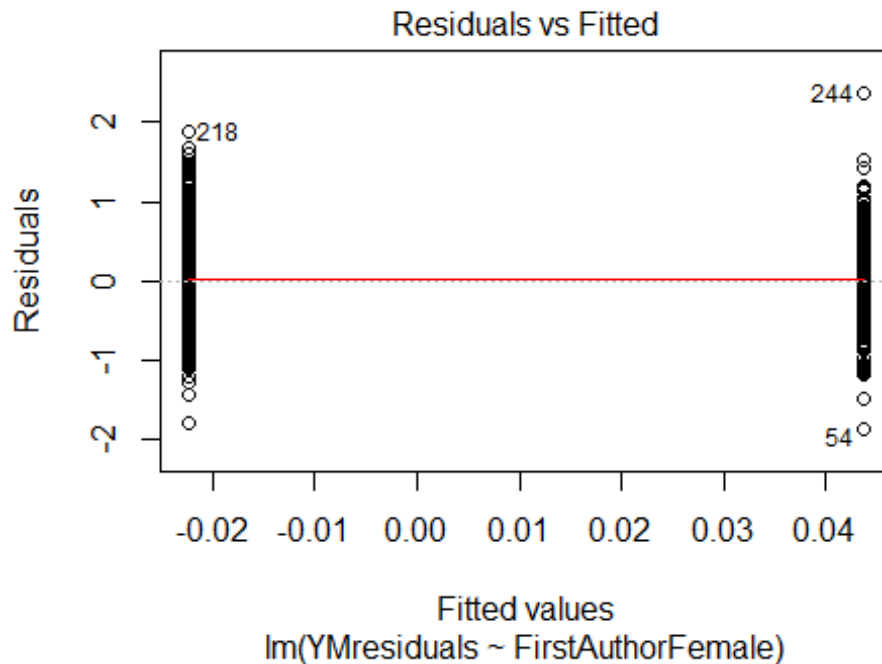
## Year2007          -0.2052      0.0520    -3.95  8.1e-05 ***
## Year2008          -0.1952      0.0502    -3.89  0.00010 ***
## Year2009          -0.2553      0.0512    -4.99  6.3e-07 ***
## Year2010          -0.2869      0.0516    -5.56  2.8e-08 ***
## Year2011          -0.2674      0.0508    -5.27  1.5e-07 ***
## Year2012          -0.2777      0.0506    -5.49  4.2e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0276, Adjusted R-squared:  0.0235
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 331 weights are ~= 1. The remaining 3769 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.104  0.856  0.951  0.888  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.44e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4100"
## [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
## 285 240 106 155 177 142 182 150 127 134 143 209 198 208 180
## 2011 2012
## 186 201
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 90 88 71 111 105 77 139 119 100 100 107 160 142 155 136
## 2011 2012

```

```
## 145 159
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 76 82 60 101 89 65 123 103 89 79 96 141 129 138 122
## 2011 2012
## 129 143
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 140, df = 16, p-value <2e-16
```

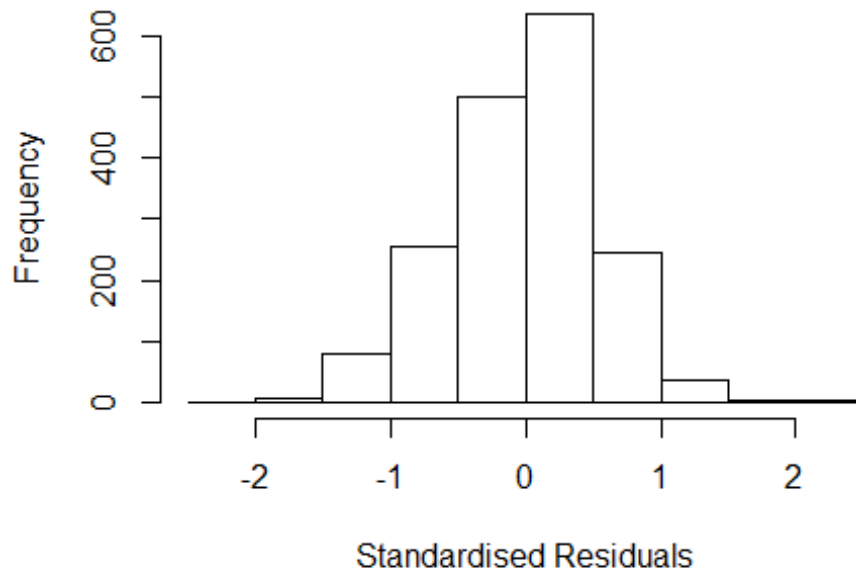


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 16, df = 1, p-value = 6e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 3.80590291023628"
## [1] "Male first author team size 2018 geometric mean: 3.42174955180653"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.58704732795888"
## [1] "Male last author team size 2018 geometric mean: 3.59073949006092"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1800, 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 1.065 1          1.032
## LastAuthorFemale  1.072 1          1.036
## UniqueAuthors    1.290 4          1.032
## Year              1.327 16         1.009
```

Residuals from first and last author and team size



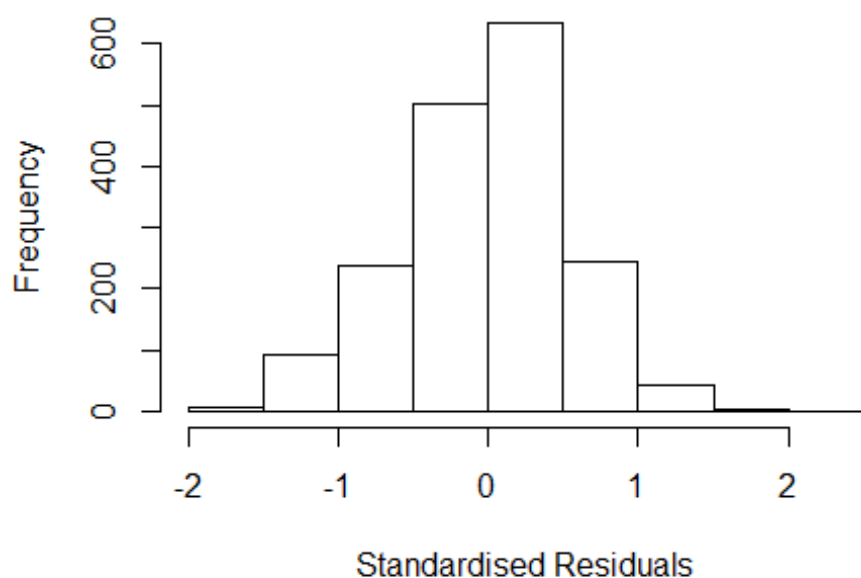
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -2.1222 -0.3436 0.0242 0.3486 2.0875
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.65094 0.12222 13.51 < 2e-16 ***
## FirstAuthorFemale1 0.06896 0.02709 2.55 0.01100 *
## LastAuthorFemale1 0.00628 0.02895 0.22 0.82823
## UniqueAuthors2 0.23274 0.06451 3.61 0.00032 ***
## UniqueAuthors3 0.20093 0.06492 3.10 0.00200 **
## UniqueAuthors4 0.29744 0.06529 4.56 5.6e-06 ***
## UniqueAuthors5 0.40229 0.06442 6.24 5.3e-10 ***
## Year1997 -0.49154 0.14965 -3.28 0.00104 **
## Year1998 -0.74294 0.12293 -6.04 1.8e-09 ***
## Year1999 -0.79610 0.11629 -6.85 1.0e-11 ***
```

```

## Year2000      -0.86617    0.13711    -6.32    3.4e-10 ***
## Year2001      -0.79898    0.12458    -6.41    1.8e-10 ***
## Year2002      -1.00668    0.12350    -8.15    6.8e-16 ***
## Year2003      -1.00525    0.12300    -8.17    5.7e-16 ***
## Year2004      -0.90854    0.11865    -7.66    3.1e-14 ***
## Year2005      -0.93276    0.11695    -7.98    2.7e-15 ***
## Year2006      -0.80588    0.11394    -7.07    2.2e-12 ***
## Year2007      -0.88811    0.11429    -7.77    1.3e-14 ***
## Year2008      -0.89605    0.11497    -7.79    1.1e-14 ***
## Year2009      -0.84245    0.11246    -7.49    1.1e-13 ***
## Year2010      -0.79036    0.11385    -6.94    5.4e-12 ***
## Year2011      -0.93180    0.11432    -8.15    6.8e-16 ***
## Year2012      -0.93526    0.11545    -8.10    1.0e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.51
## Multiple R-squared:  0.145, Adjusted R-squared:  0.134
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 1597 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0441 0.8530 0.9470 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          5.67e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1 1.021
## LastAuthorFemale 1.073 1 1.036
## Year 1.080 16 1.002

```


Residuals from first and last author



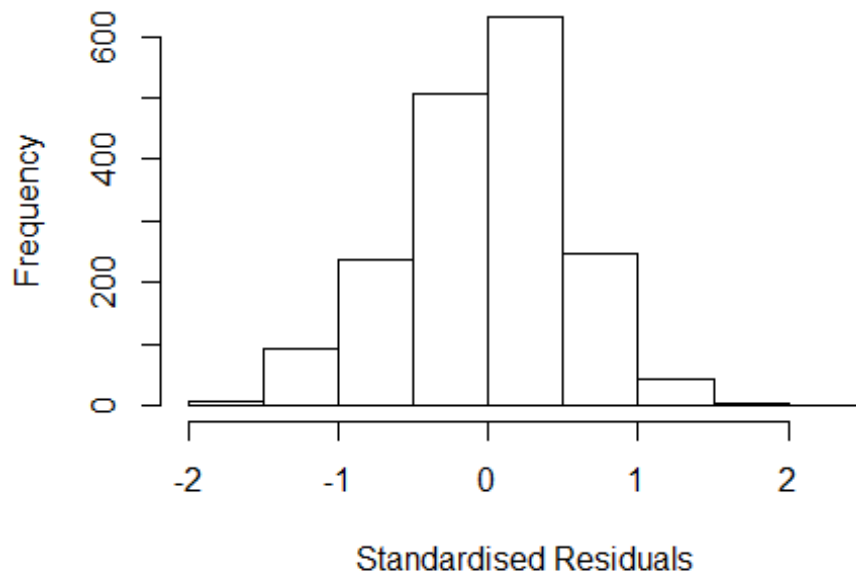
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9534 -0.3451  0.0357  0.3520  2.2479
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.8720     0.1056   17.73  < 2e-16 ***
## FirstAuthorFemale1  0.0815     0.0273    2.99   0.0028 **
## LastAuthorFemale1  0.0147     0.0294    0.50   0.6179
## Year1997          -0.4677     0.1493   -3.13   0.0018 **
## Year1998          -0.7031     0.1229   -5.72  1.2e-08 ***
## Year1999          -0.8001     0.1153   -6.94  5.5e-12 ***
## Year2000          -0.8509     0.1359   -6.26  4.8e-10 ***
## Year2001          -0.7748     0.1260   -6.15  9.8e-10 ***
## Year2002          -0.9962     0.1241   -8.03  1.8e-15 ***
## Year2003          -1.0089     0.1232   -8.19  4.9e-16 ***
## Year2004          -0.8904     0.1185   -7.51  9.2e-14 ***
## Year2005          -0.8956     0.1162   -7.71  2.1e-14 ***
```

```

## Year2006          -0.7685      0.1117    -6.88  8.4e-12 ***
## Year2007          -0.8473      0.1128    -7.51  9.1e-14 ***
## Year2008          -0.8444      0.1138    -7.42  1.9e-13 ***
## Year2009          -0.7880      0.1112    -7.09  2.0e-12 ***
## Year2010          -0.7429      0.1131    -6.57  6.7e-11 ***
## Year2011          -0.8778      0.1136    -7.73  1.9e-14 ***
## Year2012          -0.8814      0.1138    -7.75  1.6e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.114, Adjusted R-squared:  0.105
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 139 weights are ~= 1. The remaining 1626 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0173 0.8550 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      5.67e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.025 1      1.012
## Year      1.025 16      1.001

```

Residuals from first author



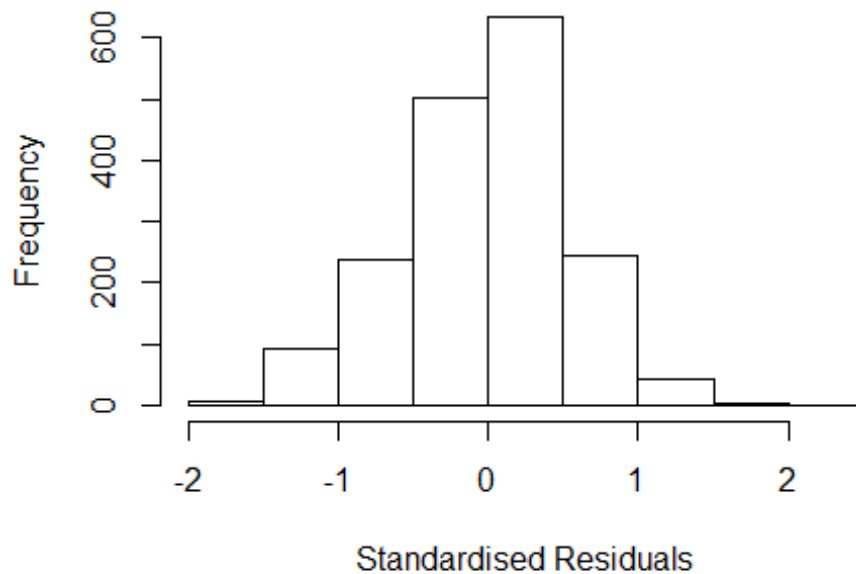
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.9580 -0.3450 0.0331 0.3541 2.2580
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.8748 0.1054 17.79 < 2e-16 ***
## FirstAuthorFemale1 0.0831 0.0270 3.07 0.0021 **
## Year1997 -0.4679 0.1494 -3.13 0.0018 **
## Year1998 -0.7043 0.1228 -5.73 1.2e-08 ***
## Year1999 -0.7996 0.1153 -6.93 5.8e-12 ***
## Year2000 -0.8506 0.1359 -6.26 4.9e-10 ***
## Year2001 -0.7743 0.1262 -6.14 1.0e-09 ***
## Year2002 -0.9946 0.1241 -8.02 2.0e-15 ***
## Year2003 -1.0085 0.1232 -8.18 5.2e-16 ***
## Year2004 -0.8895 0.1186 -7.50 9.9e-14 ***
## Year2005 -0.8959 0.1162 -7.71 2.1e-14 ***
## Year2006 -0.7670 0.1118 -6.86 9.7e-12 ***
```

```

## Year2007          -0.8467      0.1128   -7.51  9.7e-14 ***
## Year2008          -0.8438      0.1139   -7.41  2.0e-13 ***
## Year2009          -0.7866      0.1113   -7.07  2.2e-12 ***
## Year2010          -0.7429      0.1132   -6.57  6.8e-11 ***
## Year2011          -0.8763      0.1136   -7.71  2.0e-14 ***
## Year2012          -0.8812      0.1138   -7.74  1.6e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.114, Adjusted R-squared:  0.106
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1631 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0152 0.8530 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      5.67e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.057 1      1.028
## Year      1.057 16      1.002

```

Residuals from last author



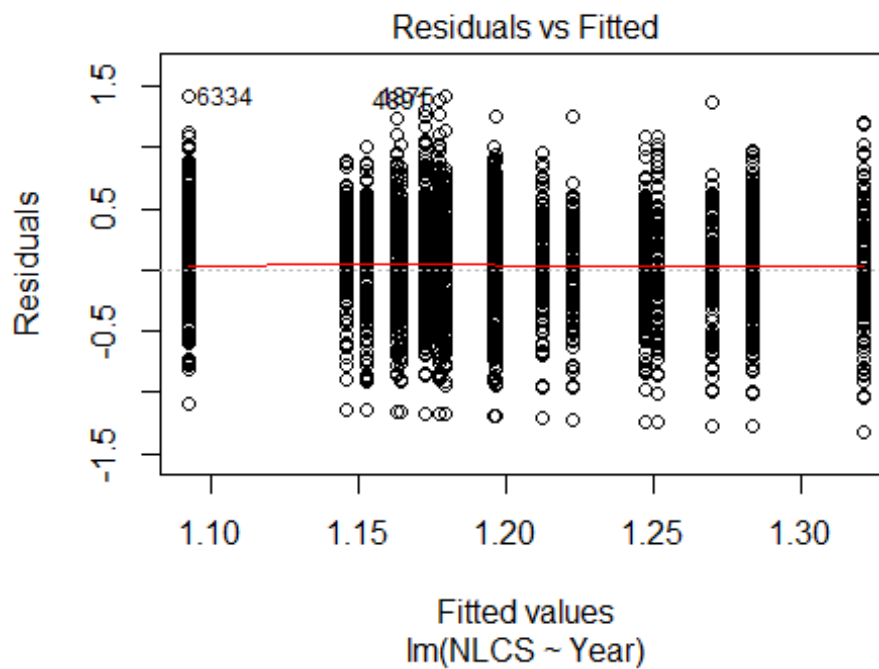
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8959 -0.3410 0.0291 0.3566 2.2948
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.8959 0.1035 18.32 < 2e-16 ***
## LastAuthorFemale1 0.0253 0.0293 0.87 0.387
## Year1997 -0.4638 0.1499 -3.09 0.002 **
## Year1998 -0.7040 0.1213 -5.80 7.7e-09 ***
## Year1999 -0.8013 0.1140 -7.03 2.9e-12 ***
## Year2000 -0.8483 0.1341 -6.32 3.2e-10 ***
## Year2001 -0.7841 0.1245 -6.30 3.8e-10 ***
## Year2002 -0.9965 0.1230 -8.10 9.8e-16 ***
## Year2003 -1.0094 0.1222 -8.26 2.9e-16 ***
## Year2004 -0.8888 0.1179 -7.54 7.4e-14 ***
## Year2005 -0.8981 0.1150 -7.81 9.7e-15 ***
## Year2006 -0.7679 0.1107 -6.93 5.7e-12 ***
```

```

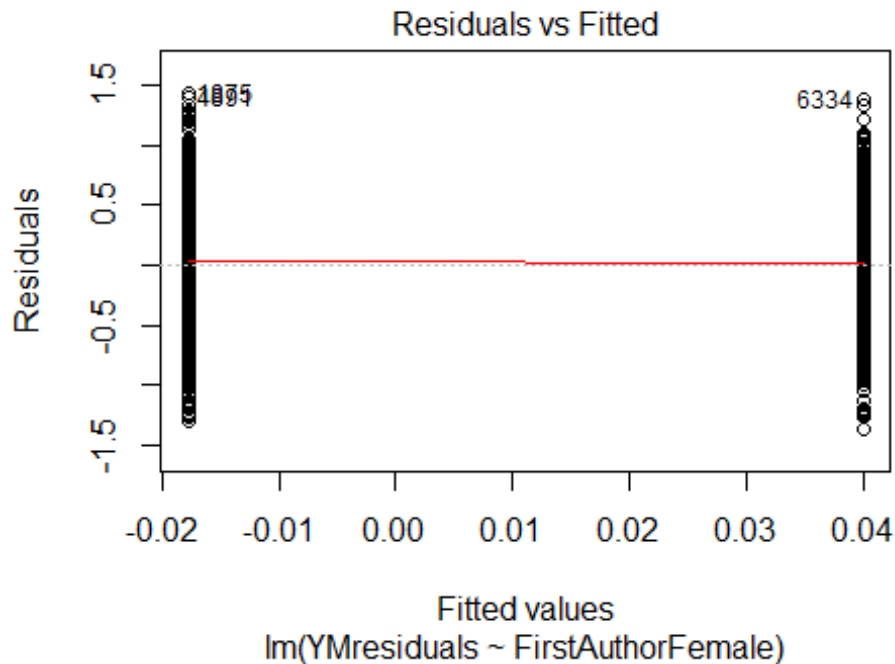
## Year2007          -0.8407      0.1118    -7.52  8.7e-14 ***
## Year2008          -0.8409      0.1128    -7.45  1.4e-13 ***
## Year2009          -0.7803      0.1102    -7.08  2.1e-12 ***
## Year2010          -0.7370      0.1125    -6.55  7.5e-11 ***
## Year2011          -0.8721      0.1127    -7.74  1.7e-14 ***
## Year2012          -0.8735      0.1126    -7.75  1.5e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.111, Adjusted R-squared:  0.102
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 175 weights are ~= 1. The remaining 1590 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0094 0.8520 0.9470 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      5.67e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1765"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
## 242 244 222 235 266 285 272 228 257 294 279 336 362 376 433
## 2011 2012
## 444 449
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 137 138 116 147 156 133 204 158 188 214 203 255 267 270 324
## 2011 2012

```

```
## 353 335
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 113 122 107 129 138 116 181 141 160 189 188 231 240 242 299
## 2011 2012
## 315 310
## [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
```

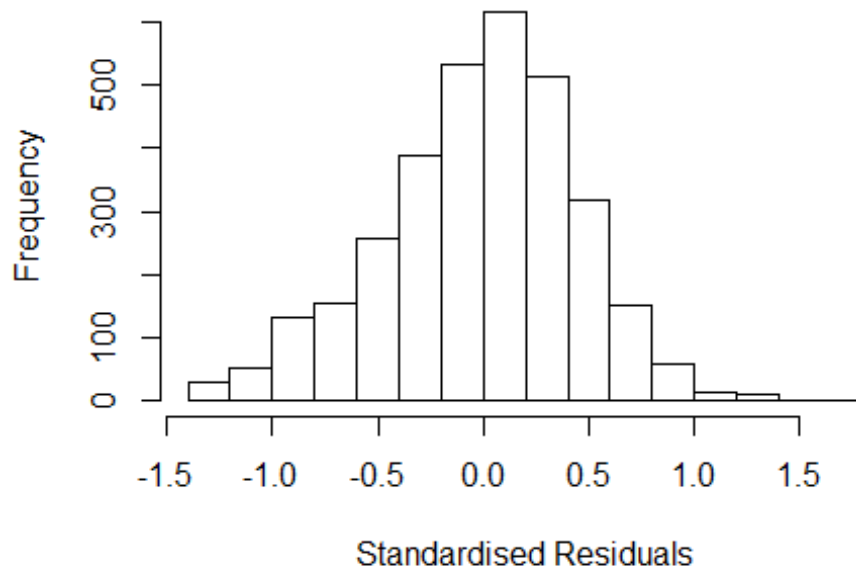


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 16, df = 1, p-value = 8e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 3.87322142104452"
## [1] "Male first author team size 2018 geometric mean: 3.92869285101258"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.53472961483042"
## [1] "Male last author team size 2018 geometric mean: 4.05194089107436"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.056 1      1.028
## LastAuthorFemale  1.050 1      1.025
## UniqueAuthors    1.135 4      1.016
## Year             1.184 16      1.005
```


Residuals from first and last author and team size



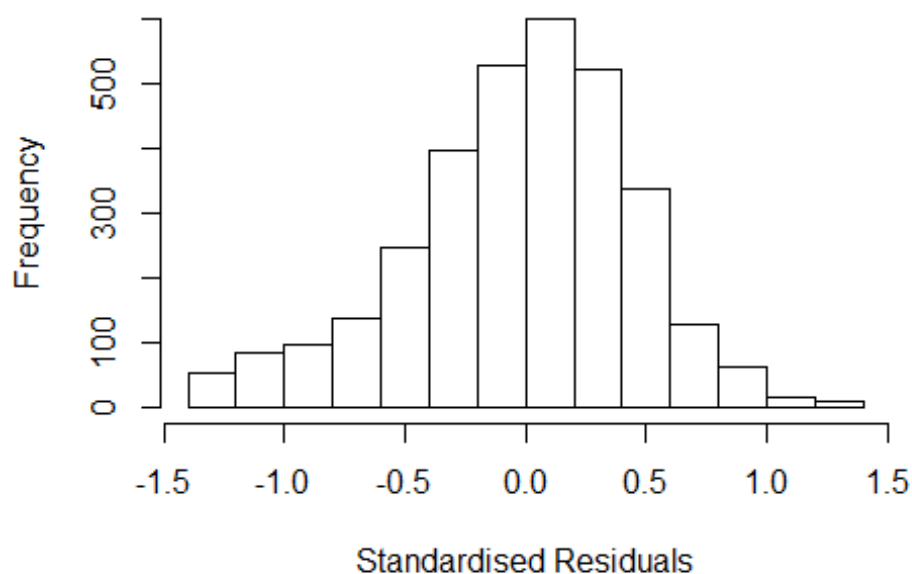
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3713 -0.2839 0.0215 0.2943 1.6381
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9721 0.0545 17.83 < 2e-16 ***
## FirstAuthorFemale1 0.0437 0.0172 2.55 0.01091 *
## LastAuthorFemale1 0.0256 0.0204 1.26 0.20899
## UniqueAuthors2 0.2858 0.0455 6.28 3.9e-10 ***
## UniqueAuthors3 0.3615 0.0448 8.06 1.0e-15 ***
## UniqueAuthors4 0.4036 0.0458 8.82 < 2e-16 ***
## UniqueAuthors5 0.4577 0.0448 10.22 < 2e-16 ***
## Year1997 0.0376 0.0602 0.62 0.53216
## Year1998 -0.0283 0.0647 -0.44 0.66227
## Year1999 -0.0856 0.0522 -1.64 0.10104
```

```

## Year2000          -0.1288      0.0600    -2.15   0.03176 *
## Year2001          -0.0930      0.0555    -1.67   0.09415 .
## Year2002          -0.1033      0.0527    -1.96   0.05014 .
## Year2003          -0.1769      0.0546    -3.24   0.00121 **
## Year2004          -0.1553      0.0548    -2.83   0.00464 **
## Year2005          -0.1452      0.0505    -2.87   0.00409 **
## Year2006          -0.0741      0.0494    -1.50   0.13337
## Year2007          -0.0708      0.0485    -1.46   0.14421
## Year2008          -0.1402      0.0500    -2.81   0.00506 **
## Year2009          -0.1592      0.0504    -3.16   0.00159 **
## Year2010          -0.1632      0.0485    -3.37   0.00077 ***
## Year2011          -0.1772      0.0472    -3.76   0.00017 ***
## Year2012          -0.2351      0.0499    -4.72   2.5e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.428
## Multiple R-squared:  0.0874, Adjusted R-squared:  0.0812
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 253 weights are ~= 1. The remaining 2968 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110  0.864  0.950  0.897  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.10e-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 1.050 1          1.025
## LastAuthorFemale  1.043 1          1.021
## Year              1.055 16          1.002

```

Residuals from first and last author



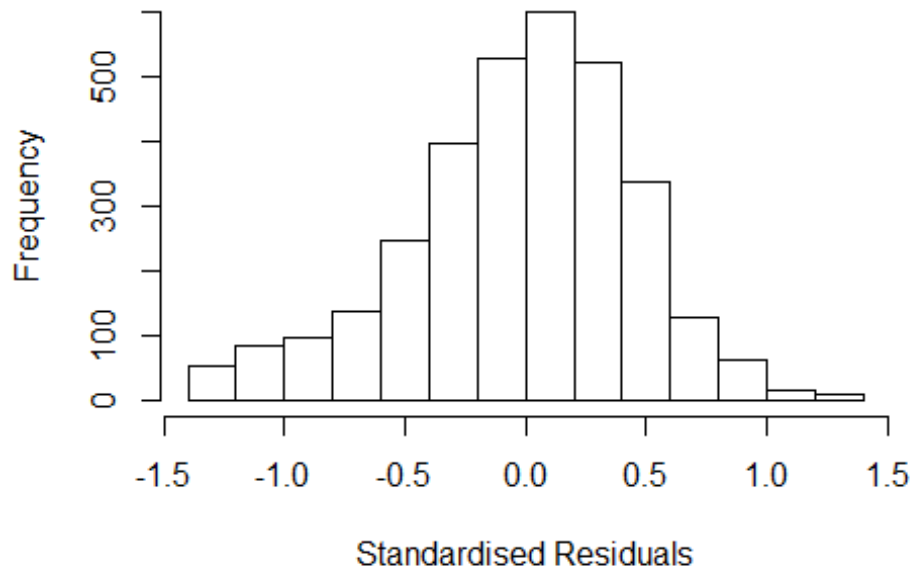
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3634 -0.2974 0.0221 0.2946 1.3821
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.267161 0.040788 31.07 <2e-16 ***
## FirstAuthorFemale1 0.047998 0.017290 2.78 0.0055 **
## LastAuthorFemale1 -0.000457 0.020135 -0.02 0.9819
## Year1997 0.048720 0.063004 0.77 0.4394
## Year1998 -0.001739 0.062953 -0.03 0.9780
## Year1999 -0.034243 0.053654 -0.64 0.5234
## Year2000 -0.105957 0.059373 -1.78 0.0744 .
## Year2001 -0.024783 0.056233 -0.44 0.6594
## Year2002 -0.057285 0.053923 -1.06 0.2882
## Year2003 -0.091696 0.054914 -1.67 0.0951 .
## Year2004 -0.086758 0.056674 -1.53 0.1259
## Year2005 -0.065054 0.051178 -1.27 0.2038
```

```

## Year2006          -0.011293    0.051046   -0.22    0.8249
## Year2007          -0.000795    0.048897   -0.02    0.9870
## Year2008          -0.058822    0.050950   -1.15    0.2484
## Year2009          -0.096625    0.052515   -1.84    0.0659 .
## Year2010          -0.081207    0.049118   -1.65    0.0984 .
## Year2011          -0.097821    0.047835   -2.04    0.0409 *
## Year2012          -0.164780    0.050365   -3.27    0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.0145, Adjusted R-squared:  0.00895
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 263 weights are ~= 1. The remaining 2958 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.284  0.862  0.950  0.891  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.10e-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 1.034 1      1.017
## Year      1.034 16      1.001

```

Residuals from first author



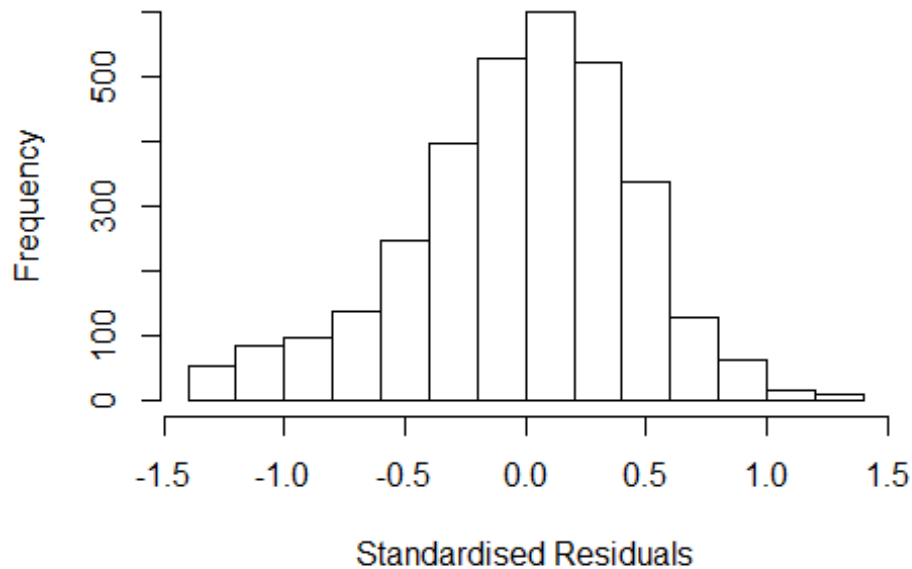
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3637 -0.2973 0.0221 0.2946 1.3822
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.267088 0.040668 31.16 <2e-16 ***
## FirstAuthorFemale1 0.047920 0.017188 2.79 0.0053 **
## Year1997 0.048708 0.062991 0.77 0.4394
## Year1998 -0.001722 0.062927 -0.03 0.9782
## Year1999 -0.034216 0.053655 -0.64 0.5237
## Year2000 -0.105957 0.059371 -1.78 0.0744 .
## Year2001 -0.024766 0.056216 -0.44 0.6596
## Year2002 -0.057284 0.053924 -1.06 0.2882
## Year2003 -0.091661 0.054885 -1.67 0.0950 .
## Year2004 -0.086755 0.056675 -1.53 0.1259
## Year2005 -0.065058 0.051176 -1.27 0.2037
## Year2006 -0.011292 0.051048 -0.22 0.8250
```

```

## Year2007          -0.000781    0.048897   -0.02    0.9873
## Year2008          -0.058811    0.050959   -1.15    0.2485
## Year2009          -0.096630    0.052511   -1.84    0.0658 .
## Year2010          -0.081209    0.049116   -1.65    0.0983 .
## Year2011          -0.097827    0.047822   -2.05    0.0409 *
## Year2012          -0.164780    0.050366   -3.27    0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.0145, Adjusted R-squared:  0.00926
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 262 weights are ~= 1. The remaining 2959 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.284  0.862  0.950  0.891  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.10e-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 1.027 1      1.014
## Year              1.027 16      1.001

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.335 -0.296 0.020 0.295 1.392
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28e+00 4.04e-02 31.61 <2e-16 ***
## LastAuthorFemale1 1.07e-02 2.00e-02 0.54 0.5926
## Year1997 4.81e-02 6.28e-02 0.77 0.4443
## Year1998 2.18e-05 6.30e-02 0.00 0.9997
## Year1999 -3.49e-02 5.36e-02 -0.65 0.5155
## Year2000 -1.02e-01 5.91e-02 -1.72 0.0853 .
## Year2001 -2.41e-02 5.60e-02 -0.43 0.6673
## Year2002 -5.49e-02 5.39e-02 -1.02 0.3084
## Year2003 -8.61e-02 5.49e-02 -1.57 0.1172
## Year2004 -8.29e-02 5.65e-02 -1.47 0.1426
## Year2005 -6.15e-02 5.09e-02 -1.21 0.2271
## Year2006 -3.36e-03 5.08e-02 -0.07 0.9474
```

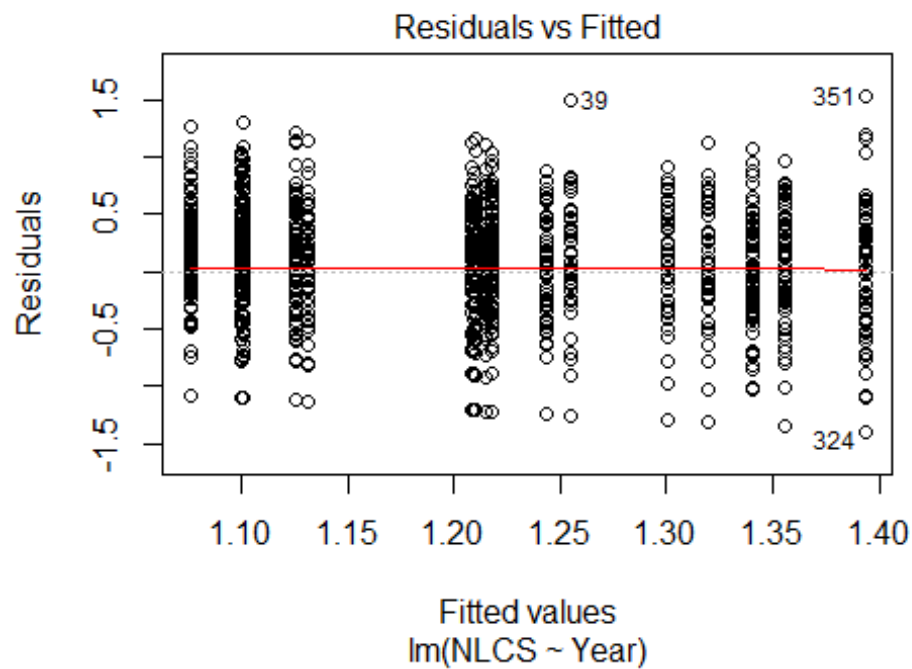
```

## Year2007          5.54e-03  4.87e-02  0.11  0.9096
## Year2008          -5.23e-02  5.09e-02  -1.03  0.3039
## Year2009          -9.13e-02  5.22e-02  -1.75  0.0803 .
## Year2010          -7.56e-02  4.90e-02  -1.54  0.1230
## Year2011          -9.19e-02  4.77e-02  -1.93  0.0541 .
## Year2012          -1.61e-01  5.02e-02  -3.20  0.0014 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.0122, Adjusted R-squared:  0.00698
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 259 weights are ~= 1. The remaining 2962 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.277  0.861  0.950  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      3.10e-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: 3221"
## [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
##   72  101   79   84   83   92   84   71   55   73   78   96  164  168  171
## 2011 2012
##  174  161
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   44   65   52   57   46   59   64   54   42   57   66   79  137  129  133
## 2011 2012

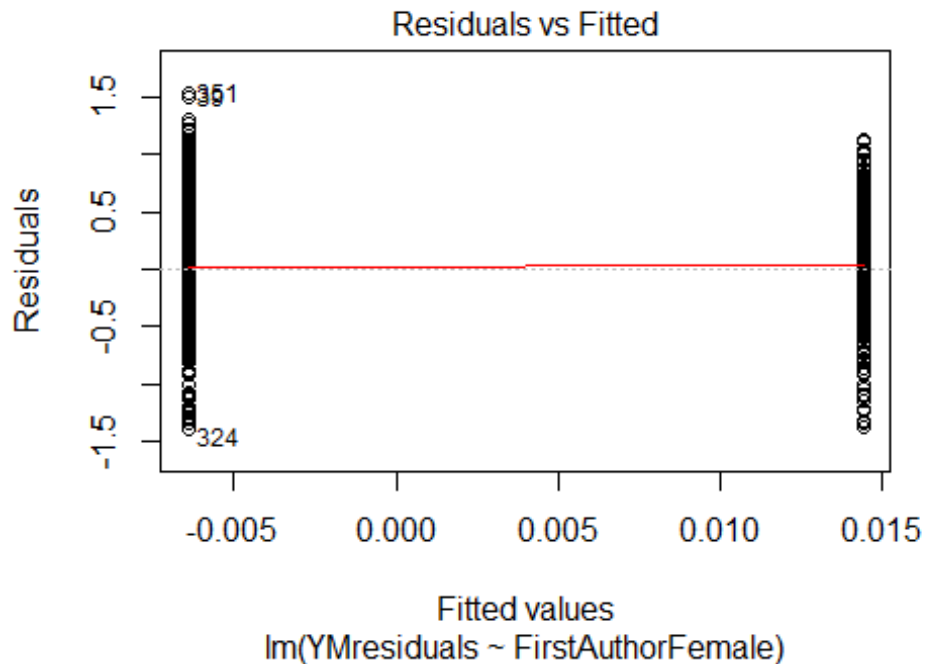
```



```
## 127 123
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 40 63 47 48 44 53 56 45 37 54 59 70 117 116 117
## 2011 2012
## 108 108
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.06
```

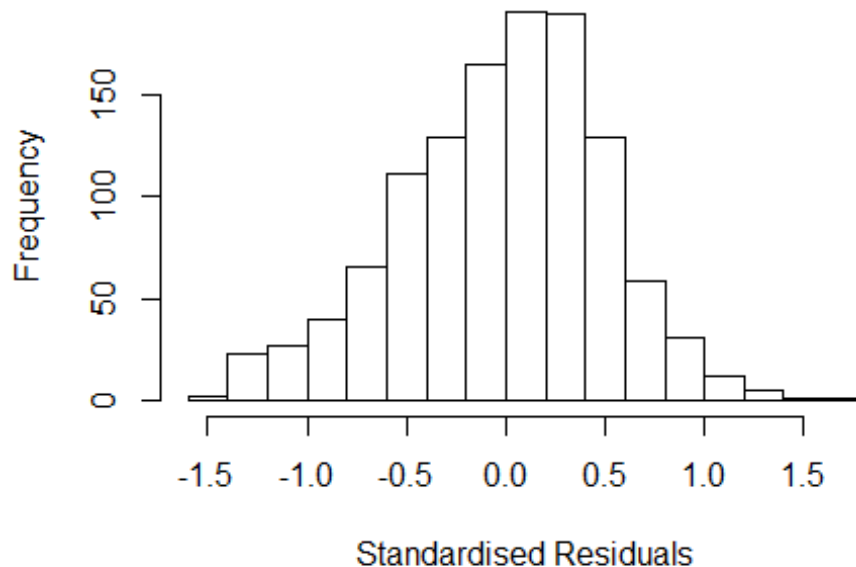


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.35, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 4.41715821439273"
## [1] "Male first author team size 2018 geometric mean: 4.39278034783948"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.07599538831547"
## [1] "Male last author team size 2018 geometric mean: 4.52994307650437"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 930, 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  1.164 1      1.079
## UniqueAuthors     1.448 4      1.047
## Year              1.524 16     1.013
```

Residuals from first and last author and team size



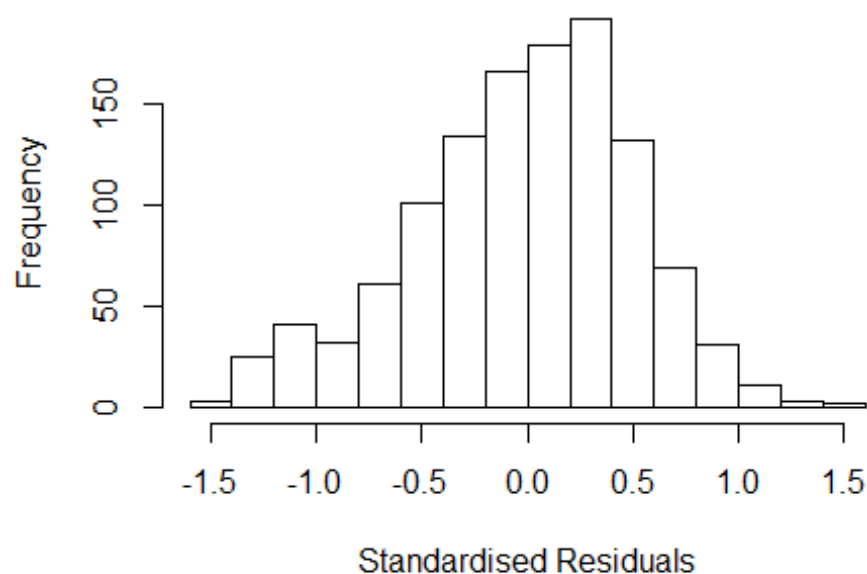
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4424 -0.3430  0.0386  0.3263  1.6241
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0203     0.1231   8.29 3.1e-16 ***
## FirstAuthorFemale1  0.0402     0.0343   1.17  0.24088
## LastAuthorFemale1  0.1083     0.0394   2.75  0.00605 **
## UniqueAuthors2     0.2095     0.0983   2.13  0.03325 *
## UniqueAuthors3     0.2545     0.0975   2.61  0.00916 **
## UniqueAuthors4     0.3429     0.0980   3.50  0.00049 ***
## UniqueAuthors5     0.4685     0.0948   4.94  8.9e-07 ***
## Year1997          0.0632     0.1209   0.52  0.60117
## Year1998          0.0191     0.1325   0.14  0.88557
## Year1999          0.0201     0.1274   0.16  0.87452
```

```

## Year2000          -0.0255      0.1277   -0.20  0.84146
## Year2001          -0.1258      0.1142   -1.10  0.27068
## Year2002          -0.1206      0.1133   -1.06  0.28753
## Year2003          -0.0549      0.1246   -0.44  0.65989
## Year2004          -0.1772      0.1376   -1.29  0.19828
## Year2005          -0.1818      0.1263   -1.44  0.15016
## Year2006          -0.1021      0.1279   -0.80  0.42489
## Year2007          -0.0339      0.1166   -0.29  0.77109
## Year2008          -0.2532      0.1109   -2.28  0.02264 *
## Year2009          -0.1716      0.1080   -1.59  0.11245
## Year2010          -0.2689      0.1079   -2.49  0.01281 *
## Year2011          -0.2385      0.1149   -2.08  0.03818 *
## Year2012          -0.2761      0.1150   -2.40  0.01650 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.09,   Adjusted R-squared:  0.0727
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 96 weights are ~= 1. The remaining 1086 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.246  0.869  0.948  0.900  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.46e-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 1.119 1          1.058
## LastAuthorFemale  1.130 1          1.063
## Year              1.085 16          1.003

```

Residuals from first and last author



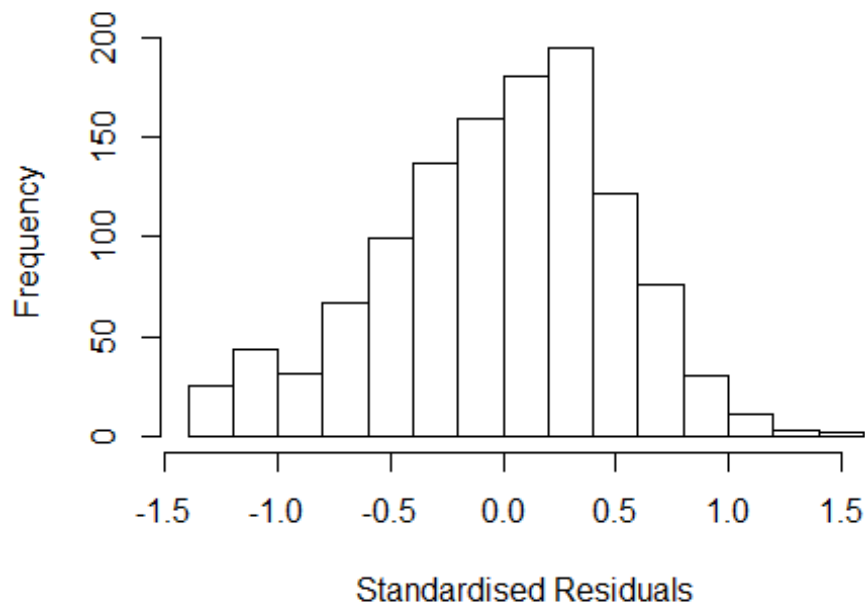
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4512 -0.3482  0.0318  0.3347  1.5636
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.25732    0.09607   13.09  <2e-16 ***
## FirstAuthorFemale1 0.01636    0.03423    0.48   0.633
## LastAuthorFemale1 0.08356    0.04018    2.08   0.038 *
## Year1997         0.10802    0.11714    0.92   0.357
## Year1998         0.09397    0.12999    0.72   0.470
## Year1999         0.09806    0.13096    0.75   0.454
## Year2000         0.04978    0.12716    0.39   0.696
## Year2001        -0.03155    0.11328   -0.28   0.781
## Year2002        -0.04062    0.11620   -0.35   0.727
## Year2003         0.01227    0.12266    0.10   0.920
## Year2004        -0.07350    0.13927   -0.53   0.598
## Year2005        -0.11346    0.12561   -0.90   0.367
```

```

## Year2006      -0.00816    0.12470   -0.07    0.948
## Year2007      0.07910    0.11129    0.71    0.477
## Year2008     -0.14674    0.10817   -1.36    0.175
## Year2009     -0.03240    0.10414   -0.31    0.756
## Year2010     -0.13657    0.10392   -1.31    0.189
## Year2011     -0.12891    0.11042   -1.17    0.243
## Year2012     -0.14212    0.11140   -1.28    0.202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.504
## Multiple R-squared:  0.0333, Adjusted R-squared:  0.0183
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 90 weights are ~= 1. The remaining 1092 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.315  0.875  0.952  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      8.46e-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 1.037 1         1.018
## Year              1.037 16         1.001

```

Residuals from first author



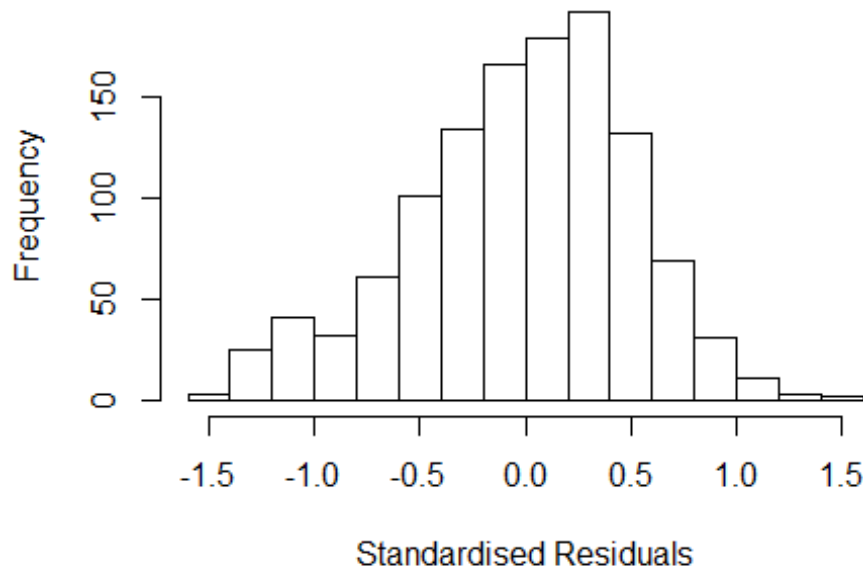
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3939 -0.3507  0.0354  0.3340  1.5499
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2737     0.0965   13.20  <2e-16 ***
## FirstAuthorFemale1  0.0359     0.0331    1.08    0.28
## Year1997          0.0997     0.1176    0.85    0.40
## Year1998          0.0843     0.1312    0.64    0.52
## Year1999          0.0954     0.1319    0.72    0.47
## Year2000          0.0447     0.1281    0.35    0.73
## Year2001         -0.0387     0.1135   -0.34    0.73
## Year2002         -0.0470     0.1170   -0.40    0.69
## Year2003          0.0110     0.1233    0.09    0.93
## Year2004         -0.0722     0.1402   -0.51    0.61
## Year2005         -0.1159     0.1266   -0.92    0.36
## Year2006         -0.0138     0.1261   -0.11    0.91
```

```

## Year2007          0.0768      0.1123      0.68      0.49
## Year2008          -0.1550      0.1087     -1.43      0.15
## Year2009          -0.0406      0.1049     -0.39      0.70
## Year2010          -0.1408      0.1047     -1.34      0.18
## Year2011          -0.1384      0.1111     -1.25      0.21
## Year2012          -0.1488      0.1125     -1.32      0.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.503
## Multiple R-squared:  0.0296, Adjusted R-squared:  0.0154
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 1090 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.322  0.870   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      8.46e-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 1.05 1      1.025
## Year              1.05 16      1.002

```


Residuals from last author



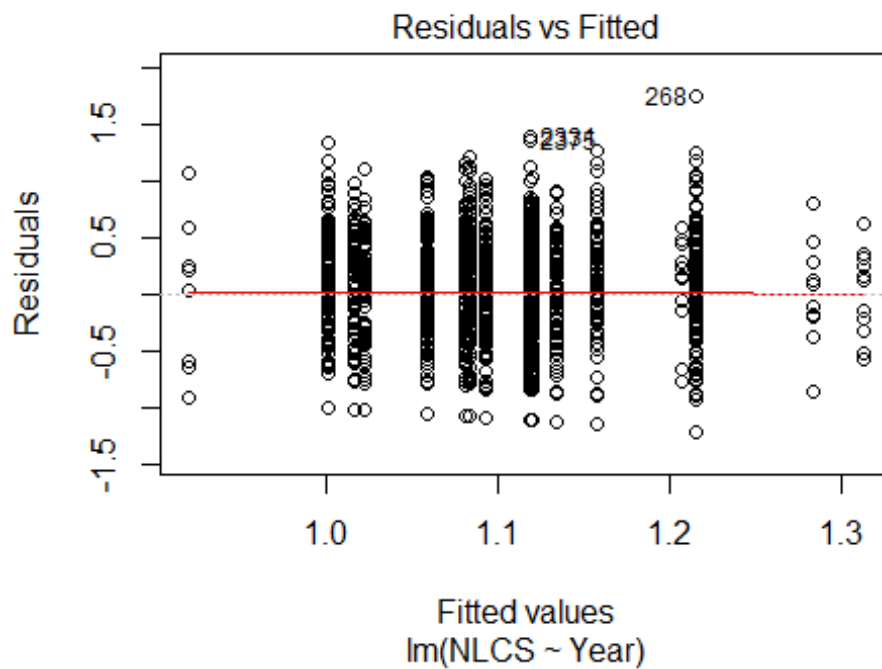
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4435 -0.3431  0.0357  0.3352  1.5616
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.26037    0.09618   13.10  <2e-16 ***
## LastAuthorFemale1 0.08874    0.03873    2.29   0.022 *
## Year1997        0.10956    0.11719    0.93   0.350
## Year1998        0.09442    0.13028    0.72   0.469
## Year1999        0.09705    0.13101    0.74   0.459
## Year2000        0.05050    0.12711    0.40   0.691
## Year2001       -0.03114    0.11359   -0.27   0.784
## Year2002       -0.03876    0.11630   -0.33   0.739
## Year2003        0.01171    0.12269    0.10   0.924
## Year2004       -0.07233    0.13991   -0.52   0.605
## Year2005       -0.11120    0.12567   -0.88   0.376
## Year2006       -0.00665    0.12495   -0.05   0.958
```

```

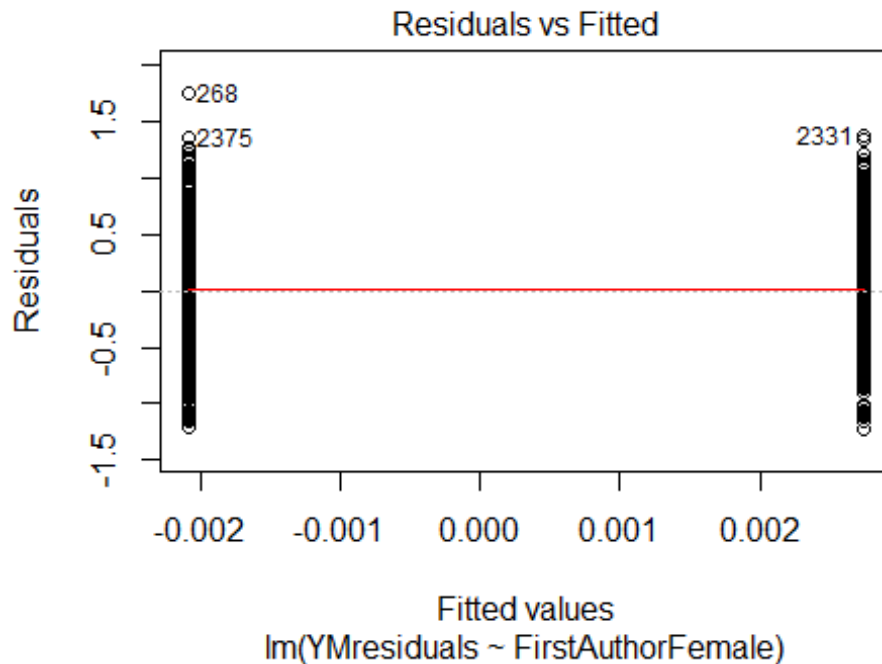
## Year2007      0.08143      0.11150      0.73      0.465
## Year2008     -0.14606      0.10844     -1.35      0.178
## Year2009     -0.03105      0.10430     -0.30      0.766
## Year2010     -0.13508      0.10412     -1.30      0.195
## Year2011     -0.12738      0.11058     -1.15      0.250
## Year2012     -0.14024      0.11127     -1.26      0.208
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.504
## Multiple R-squared:  0.033, Adjusted R-squared:  0.0189
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 94 weights are ~= 1. The remaining 1088 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.316  0.873  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      8.46e-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: 1182"
## [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
## 230 197 12 19 16 180 13 129 133 143 186 218 207 231 243
## 2011 2012
## 233 242
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 132 105 8 12 10 98 11 94 89 98 130 155 162 174 182
## 2011 2012

```

```
## 168 183
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 121 97 5 12 9 81 11 83 77 87 115 141 137 151 158
## 2011 2012
## 136 162
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.07
```

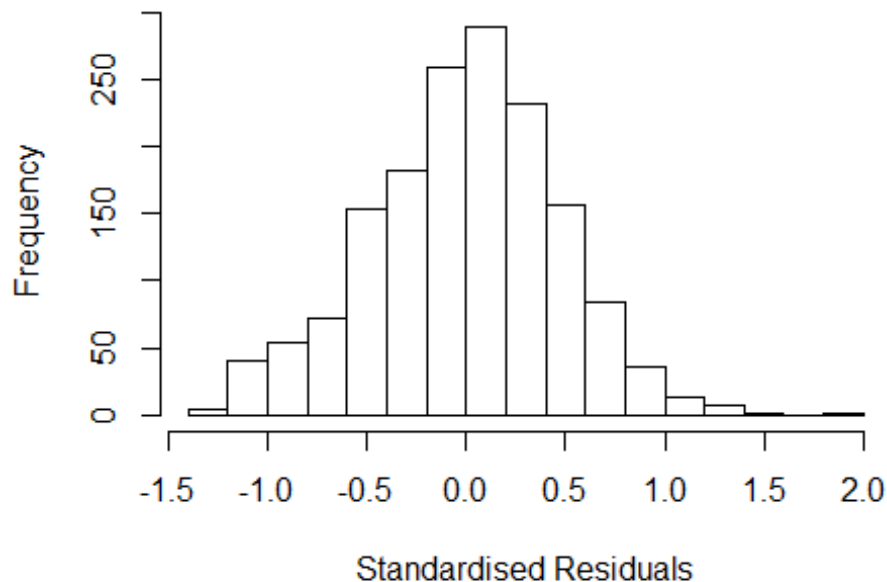


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.7, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.44772153851732"
## [1] "Male first author team size 2018 geometric mean: 2.84342931512182"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3700, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.25209737591862"
## [1] "Male last author team size 2018 geometric mean: 3.01082129013667"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3400, 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 1.082 1      1.040
## LastAuthorFemale  1.078 1      1.038
## UniqueAuthors     1.231 4      1.026
## Year              1.305 16     1.008
```

Residuals from first and last author and team size



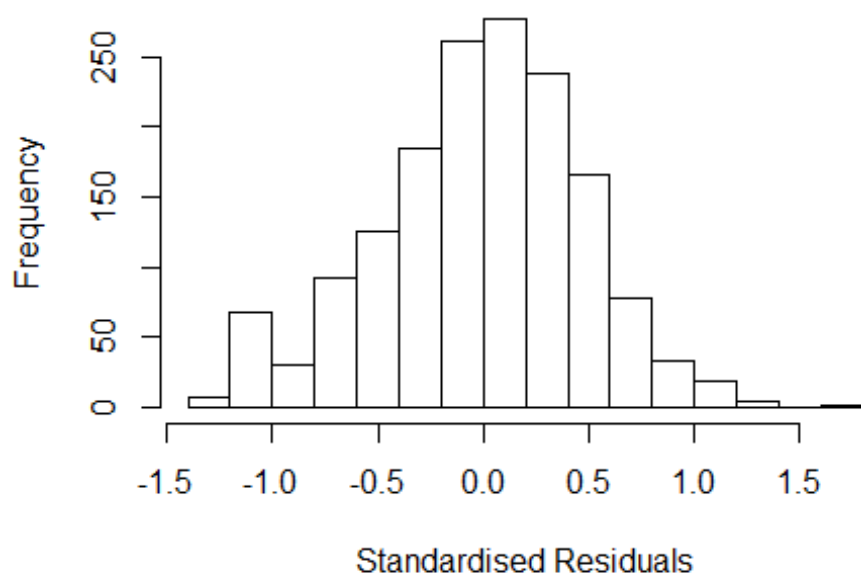
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3254 -0.2985  0.0232  0.3041  1.9354
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.030569   0.061122   16.86 < 2e-16 ***
## FirstAuthorFemale1 -0.002957   0.024860   -0.12  0.905
## LastAuthorFemale1  0.023088   0.025781    0.90  0.371
## UniqueAuthors2     0.184017   0.046078    3.99 6.8e-05 ***
## UniqueAuthors3     0.192804   0.044866    4.30 1.8e-05 ***
## UniqueAuthors4     0.193985   0.046890    4.14 3.7e-05 ***
## UniqueAuthors5     0.297762   0.041863    7.11 1.7e-12 ***
## Year1997          -0.058381   0.065303   -0.89  0.371
## Year1998          -0.158018   0.460308   -0.34  0.731
## Year1999           0.040430   0.137137    0.29  0.768
```

```

## Year2000      0.029668    0.160982    0.18    0.854
## Year2001      0.000352    0.078557    0.00    0.996
## Year2002      0.071265    0.131634    0.54    0.588
## Year2003     -0.181769    0.074906   -2.43    0.015 *
## Year2004     -0.182083    0.075677   -2.41    0.016 *
## Year2005     -0.100668    0.066976   -1.50    0.133
## Year2006     -0.123137    0.065315   -1.89    0.060 .
## Year2007     -0.158817    0.062087   -2.56    0.011 *
## Year2008     -0.160340    0.064864   -2.47    0.014 *
## Year2009     -0.104410    0.063133   -1.65    0.098 .
## Year2010     -0.093327    0.060592   -1.54    0.124
## Year2011     -0.127474    0.069681   -1.83    0.068 .
## Year2012     -0.189791    0.063845   -2.97    0.003 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.0572, Adjusted R-squared:  0.0439
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1445 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0331 0.8690 0.9510 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          6.32e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "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 1.049 1 1.024
## Year 1.076 16 1.002

```

Residuals from first and last author



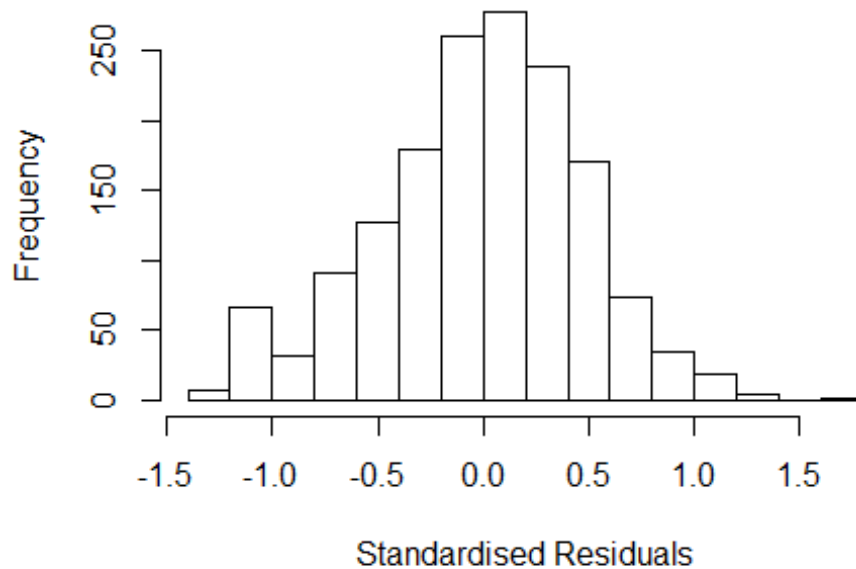
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2335 -0.3111  0.0266  0.3186  1.7600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20599    0.05105   23.62  <2e-16 ***
## FirstAuthorFemale1  0.01433    0.02524    0.57  0.5702
## LastAuthorFemale1  0.01319    0.02594    0.51  0.6111
## Year1997        -0.07735    0.06543   -1.18  0.2373
## Year1998        -0.25278    0.45871   -0.55  0.5817
## Year1999         0.02467    0.15319    0.16  0.8721
## Year2000         0.02850    0.16509    0.17  0.8630
## Year2001        -0.00499    0.07795   -0.06  0.9490
## Year2002         0.09970    0.13016    0.77  0.4438
## Year2003        -0.17160    0.07794   -2.20  0.0278 *
## Year2004        -0.18271    0.07715   -2.37  0.0180 *
## Year2005        -0.09680    0.06905   -1.40  0.1611
```

```

## Year2006      -0.11350    0.06664   -1.70    0.0887 .
## Year2007      -0.14558    0.06351   -2.29    0.0220 *
## Year2008      -0.13259    0.06559   -2.02    0.0434 *
## Year2009      -0.08488    0.06344   -1.34    0.1811
## Year2010      -0.06647    0.06130   -1.08    0.2784
## Year2011      -0.10480    0.06833   -1.53    0.1253
## Year2012      -0.18742    0.06616   -2.83    0.0047 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.46
## Multiple R-squared:  0.0161, Adjusted R-squared:  0.00474
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 139 weights are ~= 1. The remaining 1444 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110  0.863   0.948   0.899   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.32e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50          2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



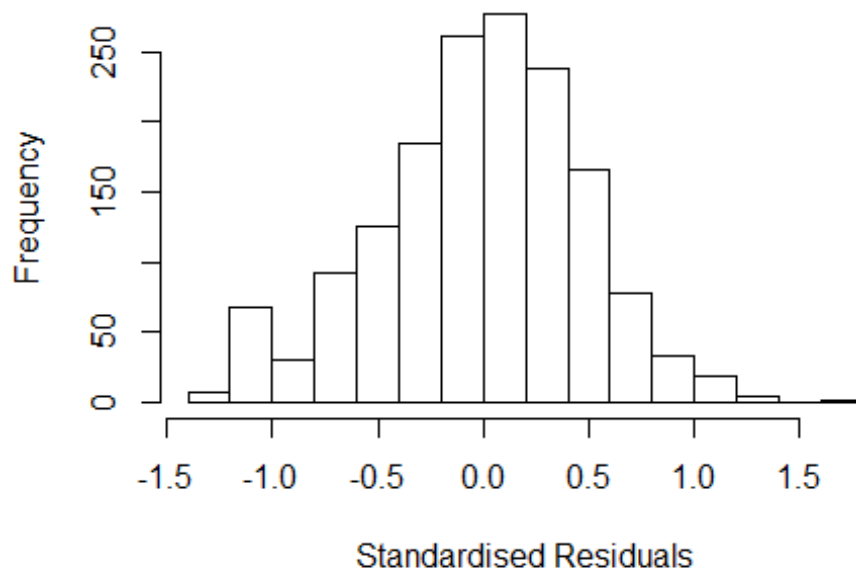
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2261 -0.3111 0.0276 0.3153 1.7566
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20941 0.05031 24.04 <2e-16 ***
## FirstAuthorFemale1 0.01664 0.02502 0.67 0.5061
## Year1997 -0.07749 0.06554 -1.18 0.2373
## Year1998 -0.25178 0.45089 -0.56 0.5766
## Year1999 0.02295 0.15225 0.15 0.8802
## Year2000 0.02944 0.16506 0.18 0.8585
## Year2001 -0.00456 0.07789 -0.06 0.9533
## Year2002 0.09993 0.12965 0.77 0.4410
## Year2003 -0.17198 0.07803 -2.20 0.0277 *
## Year2004 -0.18362 0.07704 -2.38 0.0173 *
## Year2005 -0.09704 0.06911 -1.40 0.1605
## Year2006 -0.11447 0.06666 -1.72 0.0861 .
```

```

## Year2007          -0.14541    0.06360   -2.29    0.0224 *
## Year2008          -0.13261    0.06561   -2.02    0.0434 *
## Year2009          -0.08399    0.06349   -1.32    0.1861
## Year2010          -0.06550    0.06141   -1.07    0.2863
## Year2011          -0.10312    0.06840   -1.51    0.1319
## Year2012          -0.18838    0.06613   -2.85    0.0044 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.459
## Multiple R-squared:  0.0159, Adjusted R-squared:  0.00523
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1447 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.111  0.863  0.948  0.899  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.32e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



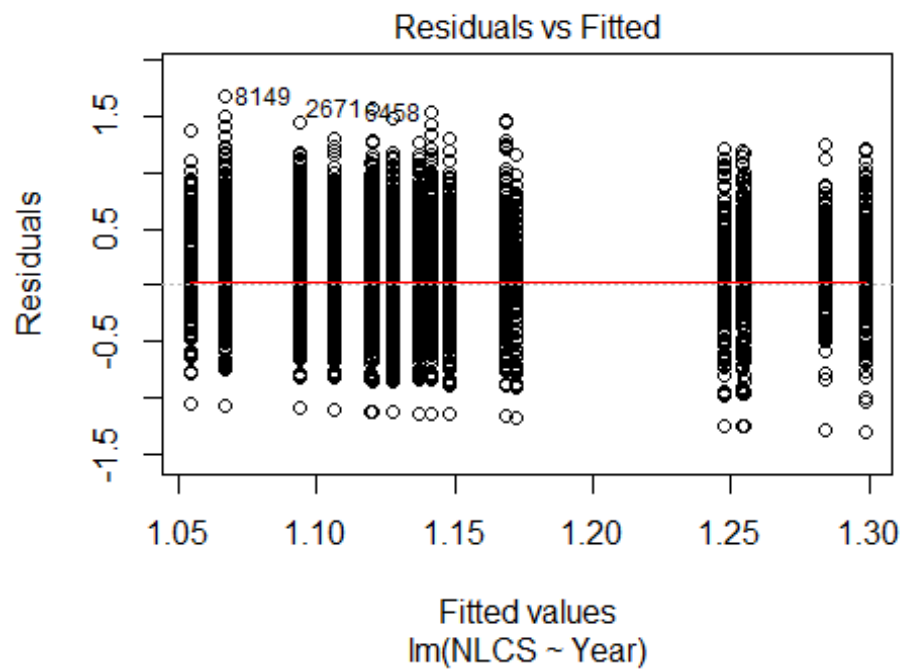
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2253 -0.3122  0.0287  0.3198  1.7566
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20937    0.05089   23.77  <2e-16 ***
## LastAuthorFemale1 0.01594    0.02571    0.62  0.5354
## Year1997       -0.07519    0.06515   -1.15  0.2486
## Year1998       -0.25443    0.45607   -0.56  0.5770
## Year1999        0.02309    0.15298    0.15  0.8800
## Year2000        0.03189    0.16365    0.19  0.8455
## Year2001       -0.00407    0.07810   -0.05  0.9585
## Year2002        0.10079    0.13007    0.77  0.4385
## Year2003       -0.16995    0.07777   -2.19  0.0290 *
## Year2004       -0.18073    0.07701   -2.35  0.0191 *
## Year2005       -0.09426    0.06868   -1.37  0.1701
## Year2006       -0.11149    0.06633   -1.68  0.0930 .
```

```

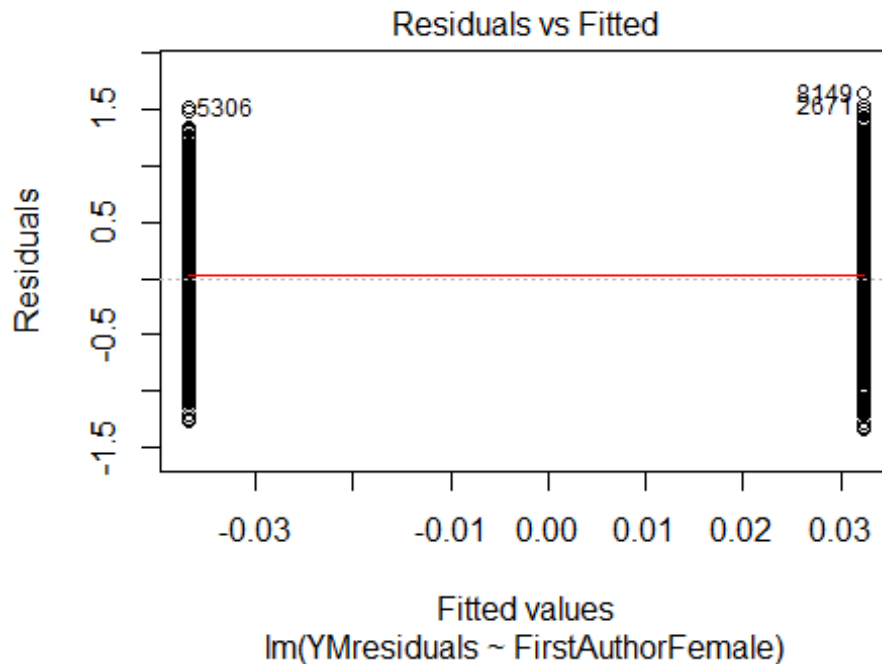
## Year2007          -0.14393      0.06320    -2.28    0.0229 *
## Year2008          -0.13058      0.06538    -2.00    0.0460 *
## Year2009          -0.08209      0.06299    -1.30    0.1927
## Year2010          -0.06410      0.06099    -1.05    0.2935
## Year2011          -0.10292      0.06823    -1.51    0.1317
## Year2012          -0.18405      0.06561    -2.81    0.0051 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.459
## Multiple R-squared:  0.0158, Adjusted R-squared:  0.00511
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 142 weights are ~= 1. The remaining 1441 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.111  0.864  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      6.32e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1583"
## [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
## 317 337 336 290 335 367 351 275 293 341 460 443 526 574 501
## 2011 2012
## 593 548
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 165 176 155 160 184 197 272 212 224 272 361 341 414 448 385
## 2011 2012

```

```
## 470 435
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 148 164 138 133 157 168 240 176 196 232 310 292 360 393 342
## 2011 2012
## 415 382
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 35, df = 16, p-value = 0.005
```

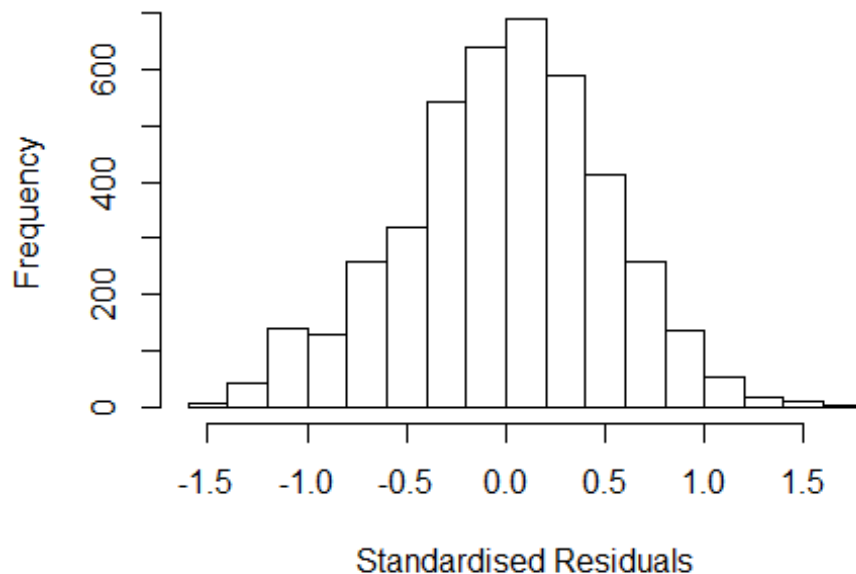


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.1, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 4.08452453140749"
## [1] "Male first author team size 2018 geometric mean: 3.9985703792937"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.77820331994016"
## [1] "Male last author team size 2018 geometric mean: 4.33675110982758"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26000, 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 1.109 1      1.053
## LastAuthorFemale  1.097 1      1.047
## UniqueAuthors     1.133 4      1.016
## Year              1.175 16     1.005
```

Residuals from first and last author and team size



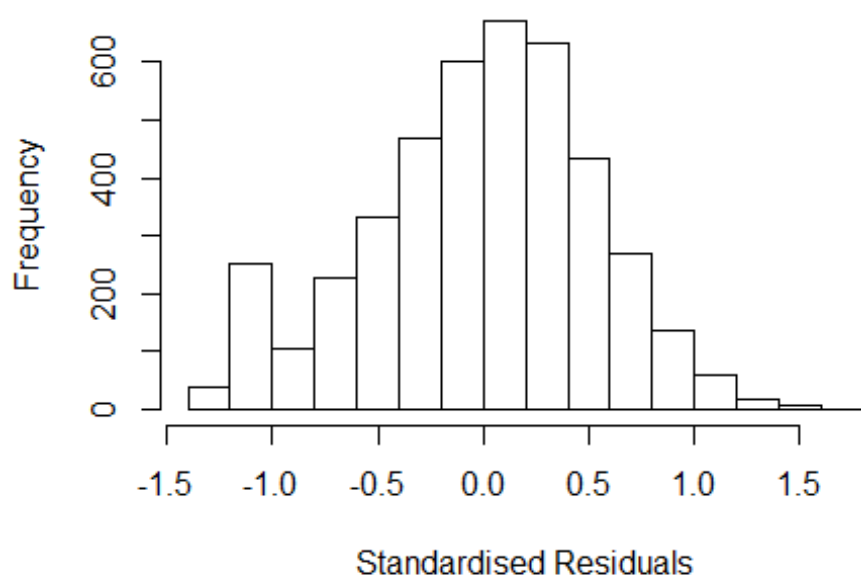
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4980 -0.3335  0.0159  0.3474  1.6883
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9350    0.0521   17.93 < 2e-16 ***
## FirstAuthorFemale1 0.0489    0.0171    2.85  0.0044 **
## LastAuthorFemale1  0.0233    0.0175    1.33  0.1846
## UniqueAuthors2     0.2901    0.0402    7.22 6.2e-13 ***
## UniqueAuthors3     0.3187    0.0382    8.34 < 2e-16 ***
## UniqueAuthors4     0.4146    0.0374   11.08 < 2e-16 ***
## UniqueAuthors5     0.5256    0.0349   15.06 < 2e-16 ***
## Year1997           0.0142    0.0628    0.23  0.8213
## Year1998          -0.0180    0.0572   -0.31  0.7536
## Year1999          -0.0498    0.0643   -0.78  0.4384
```

```

## Year2000          -0.1164      0.0588   -1.98    0.0479 *
## Year2001          -0.1461      0.0589   -2.48    0.0131 *
## Year2002          -0.1773      0.0573   -3.10    0.0020 **
## Year2003          -0.2547      0.0590   -4.31    1.6e-05 ***
## Year2004          -0.1396      0.0581   -2.40    0.0164 *
## Year2005          -0.2451      0.0526   -4.66    3.3e-06 ***
## Year2006          -0.2132      0.0513   -4.15    3.3e-05 ***
## Year2007          -0.2729      0.0557   -4.90    9.7e-07 ***
## Year2008          -0.2229      0.0499   -4.47    8.0e-06 ***
## Year2009          -0.2336      0.0499   -4.68    3.0e-06 ***
## Year2010          -0.2135      0.0508   -4.20    2.7e-05 ***
## Year2011          -0.2041      0.0501   -4.07    4.8e-05 ***
## Year2012          -0.3208      0.0522   -6.14    9.0e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.503
## Multiple R-squared:  0.0993, Adjusted R-squared:  0.0946
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 342 weights are ~= 1. The remaining 3904 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.236  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      2.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.083 1      1.041
## LastAuthorFemale  1.079 1      1.039
## Year              1.070 16      1.002

```


Residuals from first and last author



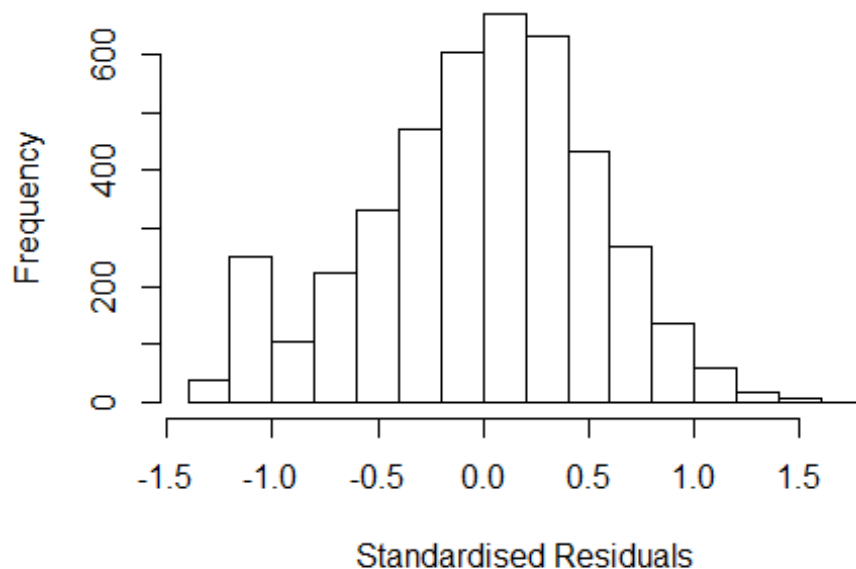
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3594 -0.3461  0.0266  0.3529  1.6453
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.262208   0.044445  28.40  < 2e-16 ***
## FirstAuthorFemale1 0.067085   0.017794   3.77  0.00017 ***
## LastAuthorFemale1 0.000282   0.018105   0.02  0.98756
## Year1997         0.030086   0.063198   0.48  0.63405
## Year1998         0.008372   0.058925   0.14  0.88702
## Year1999         0.002683   0.066393   0.04  0.96776
## Year2000        -0.084275   0.059682  -1.41  0.15800
## Year2001        -0.117286   0.060633  -1.93  0.05314 .
## Year2002        -0.147188   0.059862  -2.46  0.01398 *
## Year2003        -0.219462   0.064052  -3.43  0.00062 ***
## Year2004        -0.087687   0.060035  -1.46  0.14420
## Year2005        -0.185406   0.055443  -3.34  0.00083 ***
```

```

## Year2006      -0.143389    0.052638    -2.72    0.00648 **
## Year2007      -0.214832    0.058027    -3.70    0.00022 ***
## Year2008      -0.146139    0.051470    -2.84    0.00454 **
## Year2009      -0.170048    0.052350    -3.25    0.00117 **
## Year2010      -0.152713    0.052798    -2.89    0.00384 **
## Year2011      -0.141154    0.051520    -2.74    0.00617 **
## Year2012      -0.232873    0.054203    -4.30    1.8e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.0196, Adjusted R-squared:  0.0154
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 366 weights are ~= 1. The remaining 3880 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.287  0.860  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      2.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.039 1      1.019
## Year      1.039 16      1.001

```

Residuals from first author



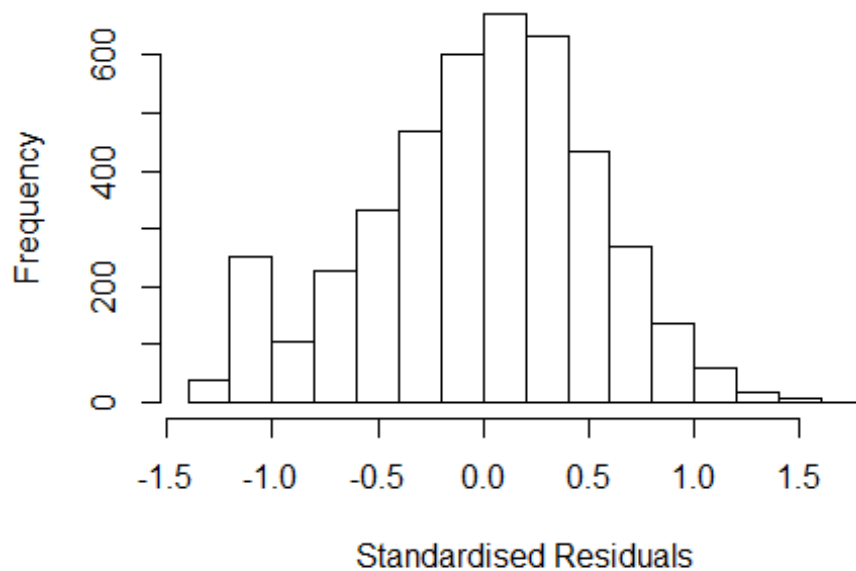
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3595 -0.3461 0.0265 0.3529 1.6454
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26228 0.04437 28.45 < 2e-16 ***
## FirstAuthorFemale1 0.06714 0.01743 3.85 0.00012 ***
## Year1997 0.03008 0.06320 0.48 0.63415
## Year1998 0.00840 0.05885 0.14 0.88656
## Year1999 0.00269 0.06640 0.04 0.96771
## Year2000 -0.08428 0.05968 -1.41 0.15798
## Year2001 -0.11725 0.06061 -1.93 0.05312 .
## Year2002 -0.14718 0.05986 -2.46 0.01399 *
## Year2003 -0.21944 0.06405 -3.43 0.00062 ***
## Year2004 -0.08770 0.06003 -1.46 0.14409
## Year2005 -0.18537 0.05536 -3.35 0.00082 ***
## Year2006 -0.14337 0.05257 -2.73 0.00641 **
```

```

## Year2007          -0.21481    0.05804   -3.70  0.00022 ***
## Year2008          -0.14612    0.05145   -2.84  0.00453 **
## Year2009          -0.17002    0.05227   -3.25  0.00115 **
## Year2010          -0.15269    0.05268   -2.90  0.00377 **
## Year2011          -0.14113    0.05149   -2.74  0.00616 **
## Year2012          -0.23284    0.05404   -4.31  1.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.0196, Adjusted R-squared:  0.0157
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 366 weights are ~= 1. The remaining 3880 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.286  0.860  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      2.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.033 1          1.017
## Year            1.033 16          1.001

```

Residuals from last author



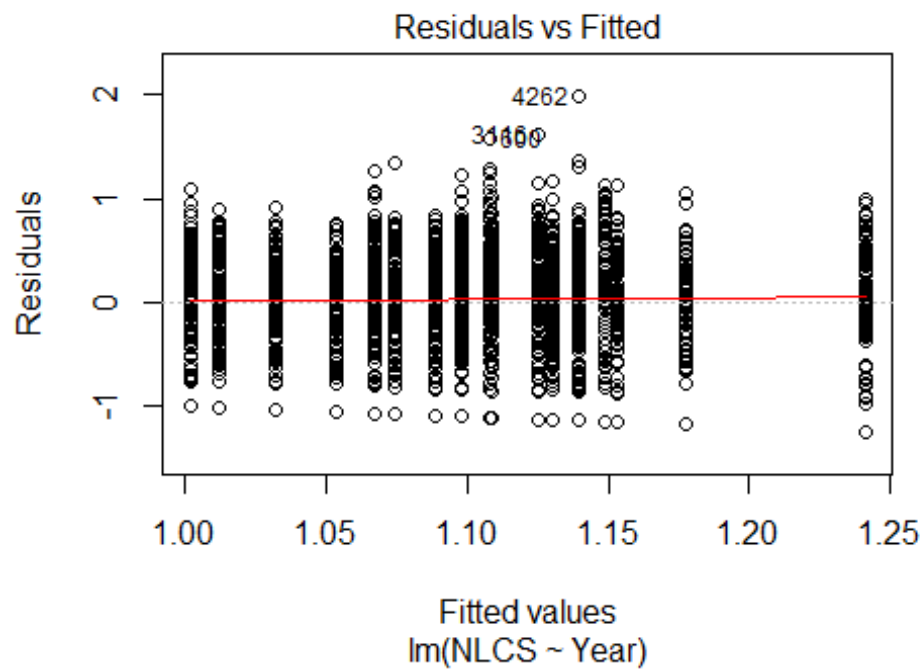
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3343 -0.3479 0.0276 0.3499 1.6611
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28374 0.04402 29.17 < 2e-16 ***
## LastAuthorFemale1 0.01483 0.01773 0.84 0.40291
## Year1997 0.03574 0.06306 0.57 0.57089
## Year1998 0.00888 0.05938 0.15 0.88120
## Year1999 0.01067 0.06642 0.16 0.87237
## Year2000 -0.07657 0.05950 -1.29 0.19826
## Year2001 -0.11494 0.06080 -1.89 0.05876 .
## Year2002 -0.14185 0.05980 -2.37 0.01773 *
## Year2003 -0.21308 0.06423 -3.32 0.00092 ***
## Year2004 -0.08140 0.05994 -1.36 0.17456
## Year2005 -0.17659 0.05526 -3.20 0.00140 **
## Year2006 -0.13086 0.05239 -2.50 0.01254 *
```

```

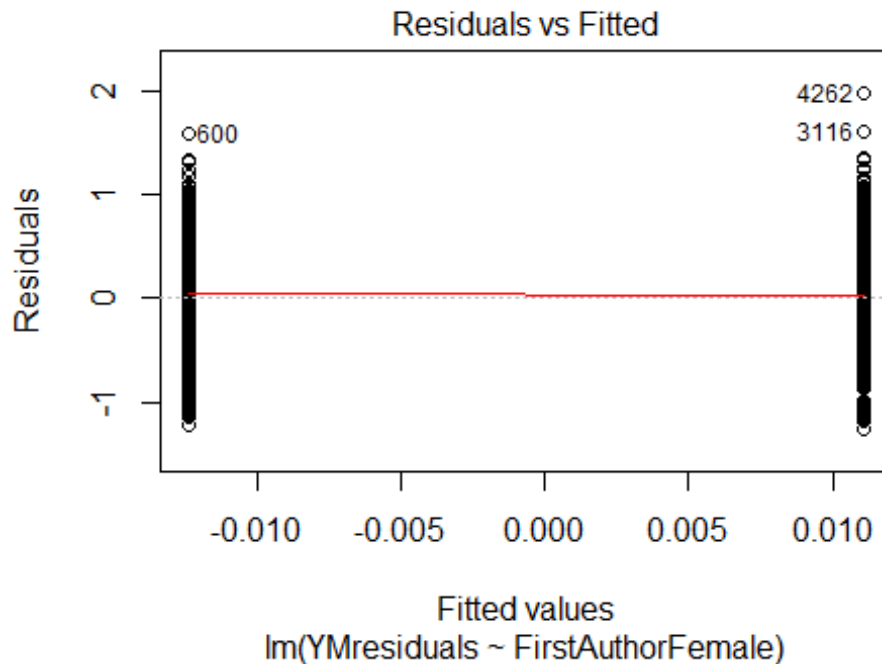
## Year2007          -0.20685      0.05785    -3.58  0.00035 ***
## Year2008          -0.13313      0.05128    -2.60  0.00945 **
## Year2009          -0.15862      0.05219    -3.04  0.00239 **
## Year2010          -0.14222      0.05270    -2.70  0.00699 **
## Year2011          -0.12917      0.05128    -2.52  0.01180 *
## Year2012          -0.21765      0.05387    -4.04  5.4e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.016, Adjusted R-squared:  0.012
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 343 weights are ~= 1. The remaining 3903 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.277  0.861  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      2.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4246"
## [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
## 156 181 155 186 174 210 215 165 151 190 197 222 253 272 266
## 2011 2012
## 313 320
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 98 105 96 101 87 84 165 120 113 141 142 156 182 192 190
## 2011 2012

```

```
## 239 231
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 81 91 83 94 80 77 149 104 99 122 121 131 165 172 172
## 2011 2012
## 218 212
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 65, df = 16, p-value = 7e-08
```

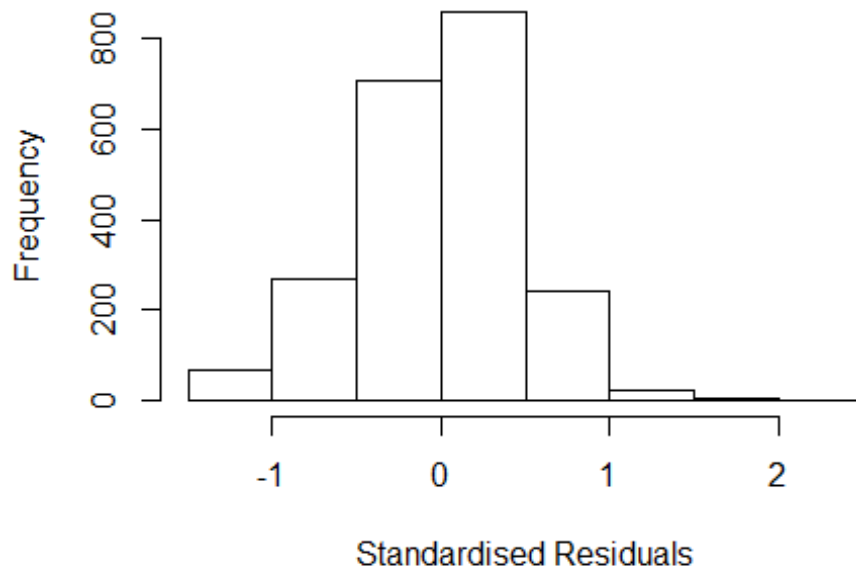


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.023, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 5.08782949936274"
## [1] "Male first author team size 2018 geometric mean: 4.16632434105281"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5400, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.67830712901884"
## [1] "Male last author team size 2018 geometric mean: 4.60802217851475"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4500, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.087 1      1.043
## LastAuthorFemale  1.099 1      1.048
## UniqueAuthors     1.178 4      1.021
## Year               1.273 16     1.008
```


Residuals from first and last author and team size



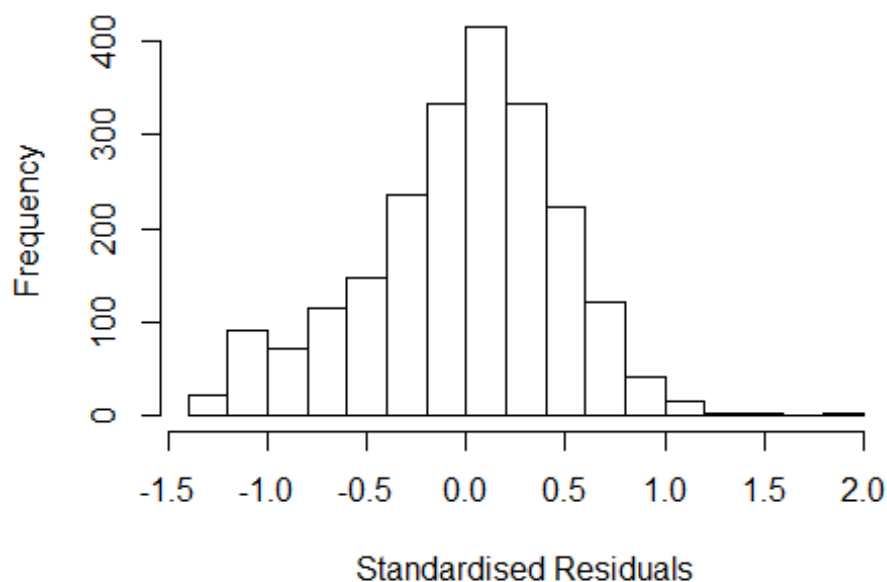
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3684 -0.2945 0.0188 0.3019 2.3467
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8635 0.0971 8.89 < 2e-16 ***
## FirstAuthorFemale1 -0.0041 0.0212 -0.19 0.846
## LastAuthorFemale1 -0.1106 0.0250 -4.43 9.8e-06 ***
## UniqueAuthors2 0.3056 0.0515 5.94 3.4e-09 ***
## UniqueAuthors3 0.3548 0.0507 7.01 3.3e-12 ***
## UniqueAuthors4 0.4055 0.0483 8.39 < 2e-16 ***
## UniqueAuthors5 0.5049 0.0452 11.17 < 2e-16 ***
## Year1997 0.0578 0.1058 0.55 0.585
## Year1998 -0.0329 0.1142 -0.29 0.773
## Year1999 0.0111 0.1023 0.11 0.914
```

```

## Year2000          -0.0131      0.1002   -0.13    0.896
## Year2001          -0.0811      0.1038   -0.78    0.435
## Year2002          -0.1174      0.0952   -1.23    0.218
## Year2003          -0.1628      0.0983   -1.66    0.098 .
## Year2004          -0.1781      0.0996   -1.79    0.074 .
## Year2005          -0.2211      0.0955   -2.32    0.021 *
## Year2006          -0.1495      0.0956   -1.56    0.118
## Year2007          -0.1861      0.0966   -1.93    0.054 .
## Year2008          -0.0242      0.0979   -0.25    0.805
## Year2009          -0.0968      0.0950   -1.02    0.309
## Year2010          -0.1544      0.0957   -1.61    0.107
## Year2011          -0.0842      0.0937   -0.90    0.369
## Year2012          -0.0687      0.0937   -0.73    0.464
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.443
## Multiple R-squared:  0.123, Adjusted R-squared:  0.114
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 1891 is an outlier with |weight| = 0 ( < 4.6e-05);
## 191 weights are ~= 1. The remaining 1979 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0263 0.8570 0.9480 0.8970 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.61e-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 1.061 1          1.030
## LastAuthorFemale 1.043 1          1.021
## Year 1.089 16          1.003

```

Residuals from first and last author



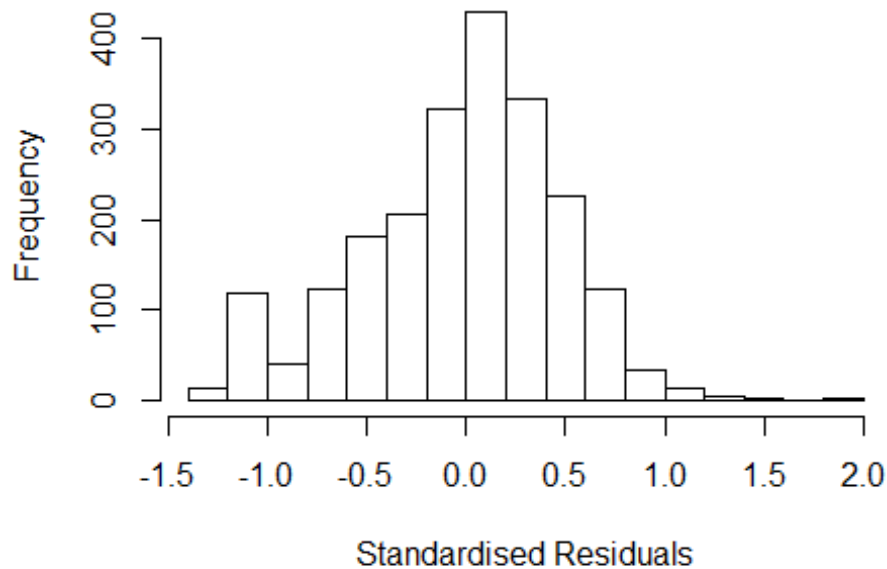
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.321 -0.320 0.035 0.310 1.956
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2314 0.0953 12.92 < 2e-16 ***
## FirstAuthorFemale1 0.0190 0.0218 0.87 0.384
## LastAuthorFemale1 -0.1400 0.0264 -5.31 1.2e-07 ***
## Year1997 0.0707 0.1099 0.64 0.520
## Year1998 -0.0382 0.1199 -0.32 0.750
## Year1999 -0.0059 0.1079 -0.05 0.956
## Year2000 -0.0364 0.1048 -0.35 0.728
## Year2001 -0.0814 0.1107 -0.74 0.462
## Year2002 -0.1143 0.1011 -1.13 0.258
## Year2003 -0.1238 0.1043 -1.19 0.235
## Year2004 -0.1427 0.1053 -1.36 0.175
## Year2005 -0.1882 0.1029 -1.83 0.068 .
```

```

## Year2006          -0.1506      0.1027    -1.47     0.143
## Year2007          -0.1738      0.1038    -1.68     0.094 .
## Year2008          -0.0303      0.1045    -0.29     0.772
## Year2009          -0.0848      0.1017    -0.83     0.404
## Year2010          -0.1287      0.1021    -1.26     0.207
## Year2011          -0.0610      0.1003    -0.61     0.543
## Year2012          -0.0354      0.1001    -0.35     0.724
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.454
## Multiple R-squared:  0.033, Adjusted R-squared:  0.025
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 1986 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0244 0.8530 0.9470 0.8930 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.61e-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 1.052 1      1.026
## Year              1.052 16      1.002

```

Residuals from first author



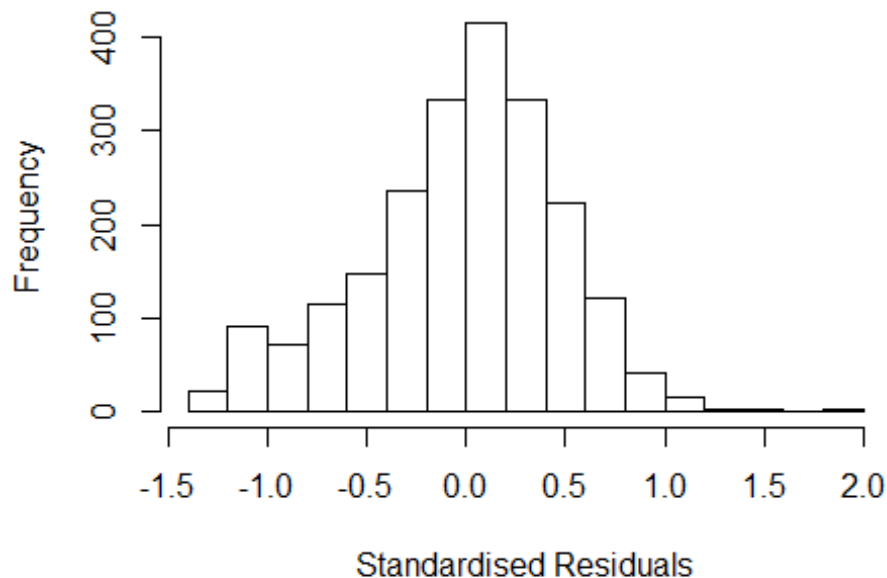
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2791 -0.3207  0.0396  0.3078  1.9891
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.199604   0.096141  12.48   <2e-16 ***
## FirstAuthorFemale1 -0.000492   0.021993  -0.02    0.982
## Year1997         0.079473   0.111220   0.71    0.475
## Year1998        -0.038980   0.121068  -0.32    0.748
## Year1999         0.012074   0.109073   0.11    0.912
## Year2000        -0.015901   0.106273  -0.15    0.881
## Year2001        -0.066665   0.110890  -0.60    0.548
## Year2002        -0.104380   0.102084  -1.02    0.307
## Year2003        -0.120325   0.105359  -1.14    0.254
## Year2004        -0.136564   0.106092  -1.29    0.198
## Year2005        -0.184726   0.103998  -1.78    0.076 .
## Year2006        -0.148968   0.103847  -1.43    0.152
```

```

## Year2007          -0.170785    0.104864   -1.63    0.104
## Year2008          -0.033450    0.106147   -0.32    0.753
## Year2009          -0.089193    0.102440   -0.87    0.384
## Year2010          -0.132522    0.103042   -1.29    0.199
## Year2011          -0.062718    0.101277   -0.62    0.536
## Year2012          -0.039497    0.101133   -0.39    0.696
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.456
## Multiple R-squared:  0.0173, Adjusted R-squared:  0.00952
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 186 weights are ~= 1. The remaining 1985 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0177 0.8570 0.9470 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      4.61e-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 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3076 -0.3210 0.0386 0.3116 1.9476
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23819 0.09518 13.01 < 2e-16 ***
## LastAuthorFemale1 -0.13664 0.02634 -5.19 2.3e-07 ***
## Year1997 0.06940 0.11014 0.63 0.529
## Year1998 -0.03891 0.12022 -0.32 0.746
## Year1999 -0.00516 0.10827 -0.05 0.962
## Year2000 -0.03592 0.10523 -0.34 0.733
## Year2001 -0.08056 0.11081 -0.73 0.467
## Year2002 -0.11307 0.10142 -1.11 0.265
## Year2003 -0.12271 0.10462 -1.17 0.241
## Year2004 -0.13976 0.10548 -1.32 0.185
## Year2005 -0.18866 0.10325 -1.83 0.068 .
## Year2006 -0.14917 0.10306 -1.45 0.148
```

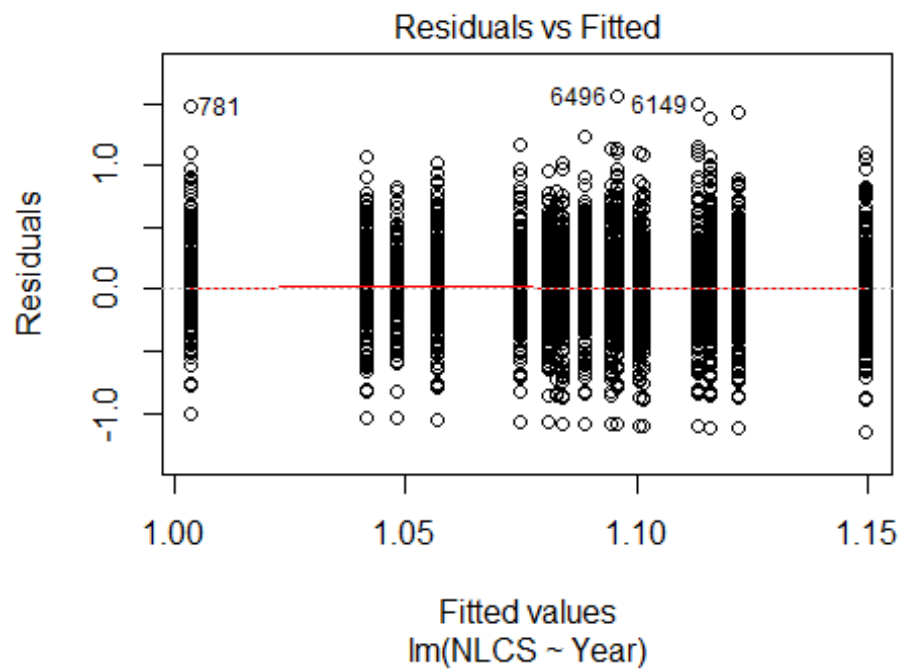
```

## Year2007          -0.17248      0.10408    -1.66      0.098 .
## Year2008          -0.02826      0.10471    -0.27      0.787
## Year2009          -0.08366      0.10199    -0.82      0.412
## Year2010          -0.12530      0.10225    -1.23      0.221
## Year2011          -0.05976      0.10062    -0.59      0.553
## Year2012          -0.03299      0.10044    -0.33      0.743
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.454
## Multiple R-squared:  0.0327, Adjusted R-squared:  0.0251
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 187 weights are ~= 1. The remaining 1984 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0263 0.8530 0.9480 0.8930 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.61e-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: 2171"
## [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
## 397 362 324 324 327 352 261 232 228 204 278 321 334 383 357
## 2011 2012
## 371 363
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 215 221 174 199 168 139 172 137 165 146 196 238 246 286 267
## 2011 2012

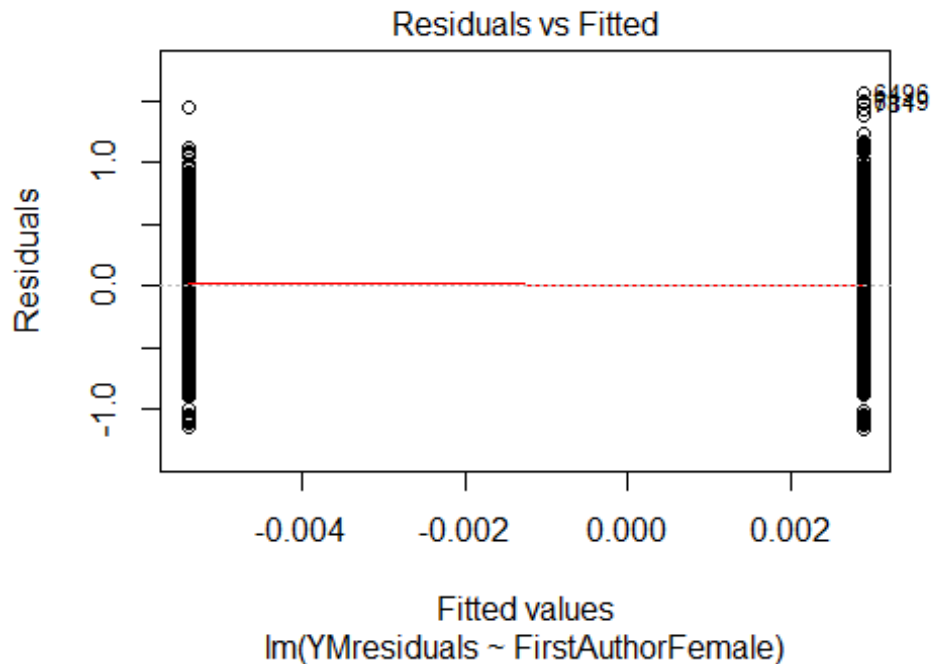
```



```
## 294 278
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 186 202 140 176 140 120 151 114 137 133 168 209 211 240 246
## 2011 2012
## 245 242
## [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
```

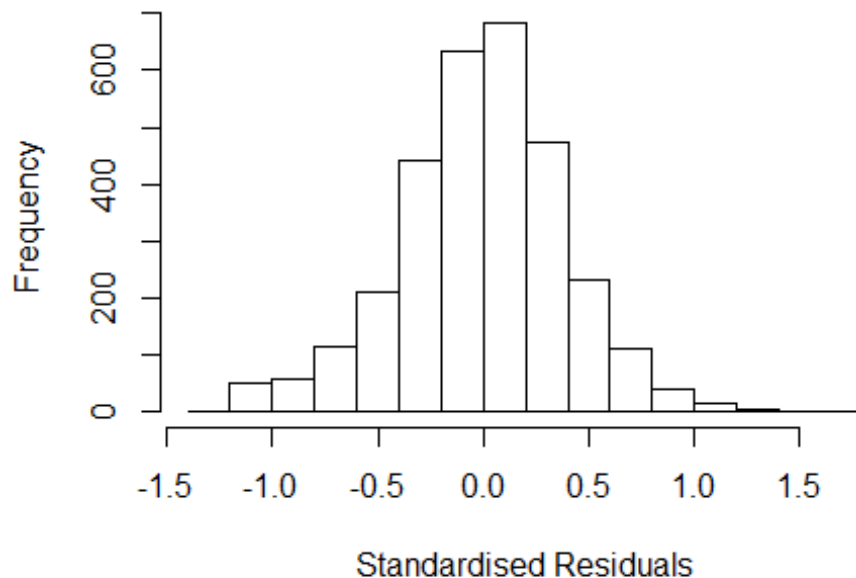


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.7, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 4.38395567096293"
## [1] "Male first author team size 2018 geometric mean: 4.77096878637523"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6300, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.61712020962621"
## [1] "Male last author team size 2018 geometric mean: 4.58545816989577"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4900, 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  1.027 1      1.013
## UniqueAuthors    1.193 4      1.022
## Year              1.265 16      1.007
```

Residuals from first and last author and team size



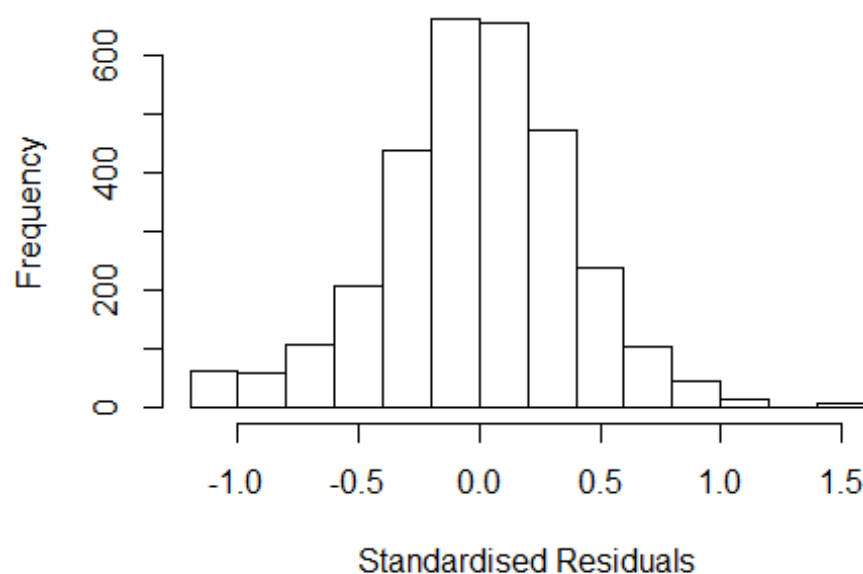
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22211 -0.23772  0.00711  0.23719  1.76142
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.94816    0.05493   17.26 < 2e-16 ***
## FirstAuthorFemale1  0.00372    0.01461    0.25  0.7990
## LastAuthorFemale1 -0.02109    0.01735   -1.22  0.2243
## UniqueAuthors2     0.13848    0.05047    2.74  0.0061 **
## UniqueAuthors3     0.16047    0.04987    3.22  0.0013 **
## UniqueAuthors4     0.15706    0.05011    3.13  0.0017 **
## UniqueAuthors5     0.27418    0.04940    5.55 3.1e-08 ***
## Year1997         -0.10315    0.04386   -2.35  0.0187 *
## Year1998         -0.02368    0.04538   -0.52  0.6018
## Year1999         -0.06813    0.04050   -1.68  0.0926 .
```

```

## Year2000      -0.06123      0.04252      -1.44      0.1499
## Year2001      -0.04404      0.04462      -0.99      0.3237
## Year2002      -0.02123      0.04195      -0.51      0.6129
## Year2003      -0.05764      0.04138      -1.39      0.1638
## Year2004      -0.08885      0.04238      -2.10      0.0361 *
## Year2005      -0.10216      0.04047      -2.52      0.0117 *
## Year2006      -0.02881      0.04229      -0.68      0.4958
## Year2007      -0.03581      0.03912      -0.92      0.3601
## Year2008       0.01714      0.04003       0.43      0.6685
## Year2009      -0.05608      0.04107      -1.37      0.1723
## Year2010      -0.05855      0.04030      -1.45      0.1464
## Year2011      -0.02439      0.03915      -0.62      0.5334
## Year2012      -0.05457      0.04040      -1.35      0.1768
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.354
## Multiple R-squared:  0.0436, Adjusted R-squared:  0.0367
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2893 is an outlier with |weight| = 0 ( < 3.3e-05);
## 256 weights are ~= 1. The remaining 2803 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0227 0.8640 0.9510 0.8920 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           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 1.065 1 1.032
## LastAuthorFemale 1.014 1 1.007
## Year 1.071 16 1.002

```

Residuals from first and last author



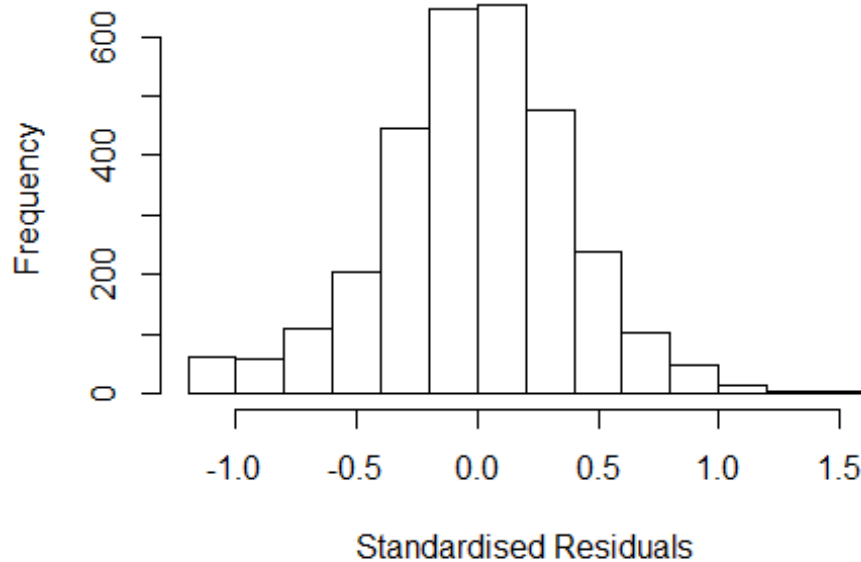
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.180649 -0.240016  0.000343  0.237273  1.556727
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.10874    0.03038   36.50  <2e-16 ***
## FirstAuthorFemale1  0.00340    0.01476    0.23   0.818
## LastAuthorFemale1 -0.03167    0.01739   -1.82   0.069 .
## Year1997          -0.10227    0.04421   -2.31   0.021 *
## Year1998          -0.01072    0.04535   -0.24   0.813
## Year1999          -0.04750    0.04084   -1.16   0.245
## Year2000          -0.02782    0.04278   -0.65   0.516
## Year2001          -0.00610    0.04420   -0.14   0.890
## Year2002           0.00807    0.04205    0.19   0.848
## Year2003          -0.01359    0.04160   -0.33   0.744
## Year2004          -0.03685    0.04288   -0.86   0.390
## Year2005          -0.06791    0.04054   -1.67   0.094 .
```

```

## Year2006          0.00845    0.04318    0.20    0.845
## Year2007          0.01470    0.03944    0.37    0.709
## Year2008          0.07191    0.04001    1.80    0.072 .
## Year2009         -0.00695    0.04136   -0.17    0.867
## Year2010         -0.01866    0.04058   -0.46    0.646
## Year2011          0.02187    0.03917    0.56    0.577
## Year2012         -0.01047    0.04018   -0.26    0.794
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.0112, Adjusted R-squared:  0.0053
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 256 weights are ~= 1. The remaining 2804 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0197 0.8650 0.9520 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.062 1      1.030
## Year              1.062 16      1.002

```

Residuals from first author



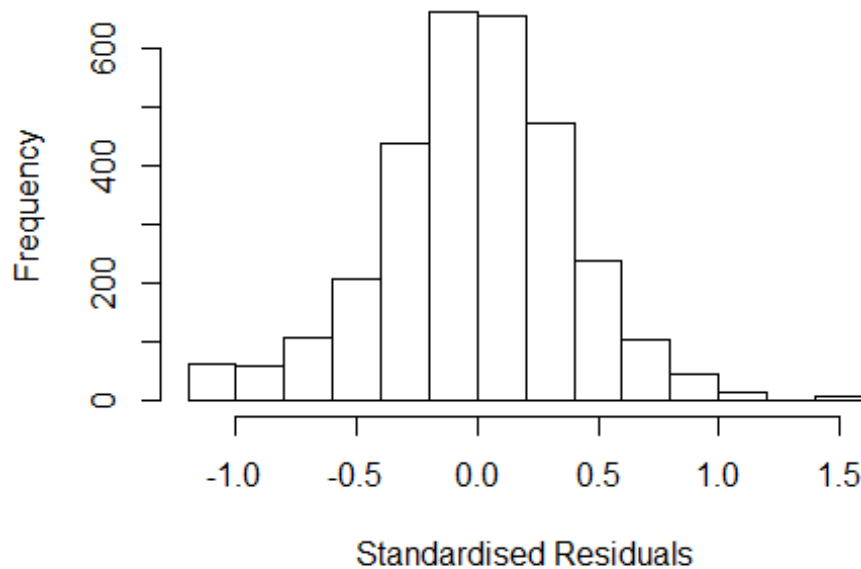
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17588 -0.23716 0.00261 0.23749 1.56222
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.104315 0.030349 36.39 <2e-16 ***
## FirstAuthorFemale1 0.000957 0.014747 0.06 0.948
## Year1997 -0.102051 0.044270 -2.31 0.021 *
## Year1998 -0.011117 0.045385 -0.24 0.807
## Year1999 -0.047156 0.040972 -1.15 0.250
## Year2000 -0.029757 0.042713 -0.70 0.486
## Year2001 -0.006737 0.044370 -0.15 0.879
## Year2002 0.006676 0.042056 0.16 0.874
## Year2003 -0.014930 0.041727 -0.36 0.721
## Year2004 -0.038021 0.043001 -0.88 0.377
## Year2005 -0.067982 0.040632 -1.67 0.094 .
## Year2006 0.008051 0.043213 0.19 0.852
```

```

## Year2007          0.012214    0.039550    0.31    0.757
## Year2008          0.070608    0.040137    1.76    0.079 .
## Year2009         -0.007794    0.041441   -0.19    0.851
## Year2010         -0.018886    0.040727   -0.46    0.643
## Year2011          0.019443    0.039214    0.50    0.620
## Year2012         -0.011533    0.040350   -0.29    0.775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.359
## Multiple R-squared:  0.0101, Adjusted R-squared:  0.00453
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 246 weights are ~= 1. The remaining 2814 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0188 0.8670 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      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 1.011 1          1.005
## Year            1.011 16          1.000

```


Residuals from last author



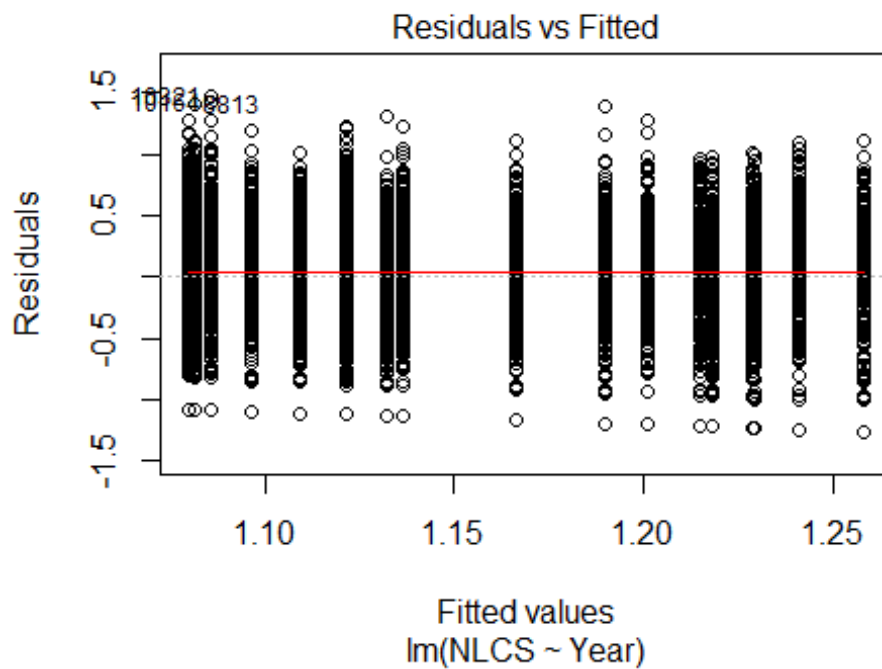
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.181905 -0.240259 0.000486 0.238144 1.555346
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10954 0.03026 36.66 <2e-16 ***
## LastAuthorFemale1 -0.03130 0.01737 -1.80 0.072 .
## Year1997 -0.10215 0.04420 -2.31 0.021 *
## Year1998 -0.01052 0.04528 -0.23 0.816
## Year1999 -0.04721 0.04079 -1.16 0.247
## Year2000 -0.02787 0.04277 -0.65 0.515
## Year2001 -0.00601 0.04419 -0.14 0.892
## Year2002 0.00838 0.04202 0.20 0.842
## Year2003 -0.01352 0.04161 -0.32 0.745
## Year2004 -0.03652 0.04278 -0.85 0.393
## Year2005 -0.06765 0.04052 -1.67 0.095 .
## Year2006 0.00901 0.04299 0.21 0.834
```

```

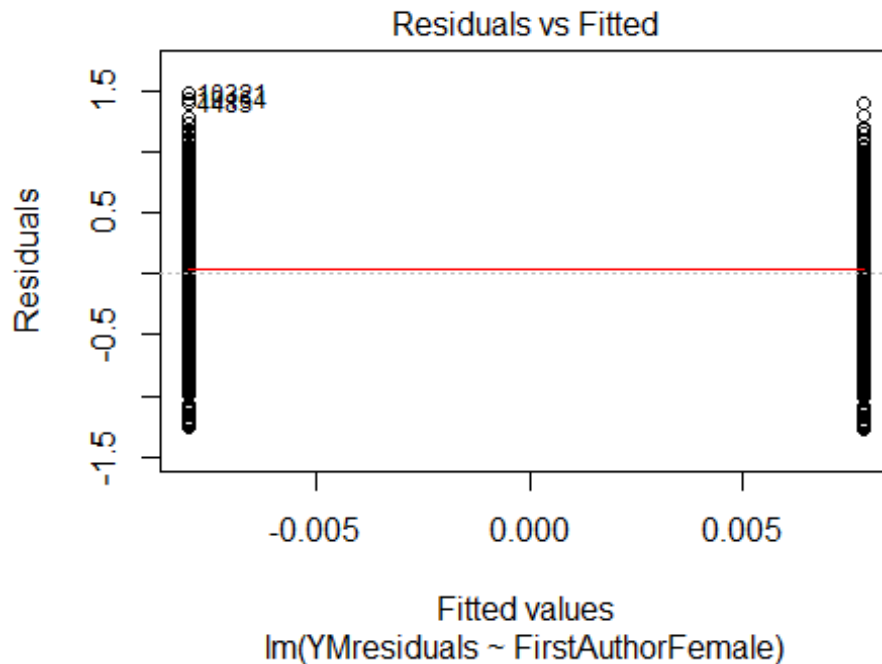
## Year2007      0.01517      0.03930      0.39      0.700
## Year2008      0.07237      0.03991      1.81      0.070 .
## Year2009     -0.00666      0.04135     -0.16      0.872
## Year2010     -0.01803      0.04029     -0.45      0.654
## Year2011      0.02241      0.03896      0.58      0.565
## Year2012     -0.00988      0.04006     -0.25      0.805
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.0111, Adjusted R-squared:  0.00561
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 252 weights are ~= 1. The remaining 2808 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0202 0.8650 0.9520 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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: 3060"
## [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
## 422 357 403 390 463 518 441 374 380 459 572 642 634 705 725
## 2011 2012
## 735 862
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 302 262 267 274 274 285 319 304 296 383 468 526 529 593 612
## 2011 2012

```

```
## 595 721
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 263 231 238 246 242 247 289 256 252 334 412 489 471 526 542
## 2011 2012
## 528 651
## [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 = 9e-05
```

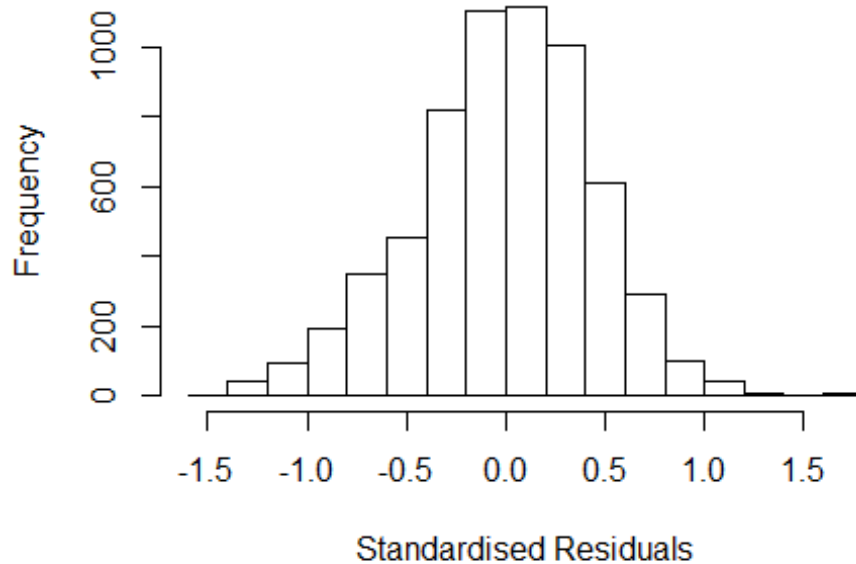


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 14, df = 1, p-value = 2e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.12430445723455"
## [1] "Male first author team size 2018 geometric mean: 4.30835458148887"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 49000, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.05518089489588"
## [1] "Male last author team size 2018 geometric mean: 4.31382161329188"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 54000, 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 1.043 1          1.021
## LastAuthorFemale  1.033 1          1.016
## UniqueAuthors    1.106 4          1.013
## Year             1.112 16          1.003
```

Residuals from first and last author and team size



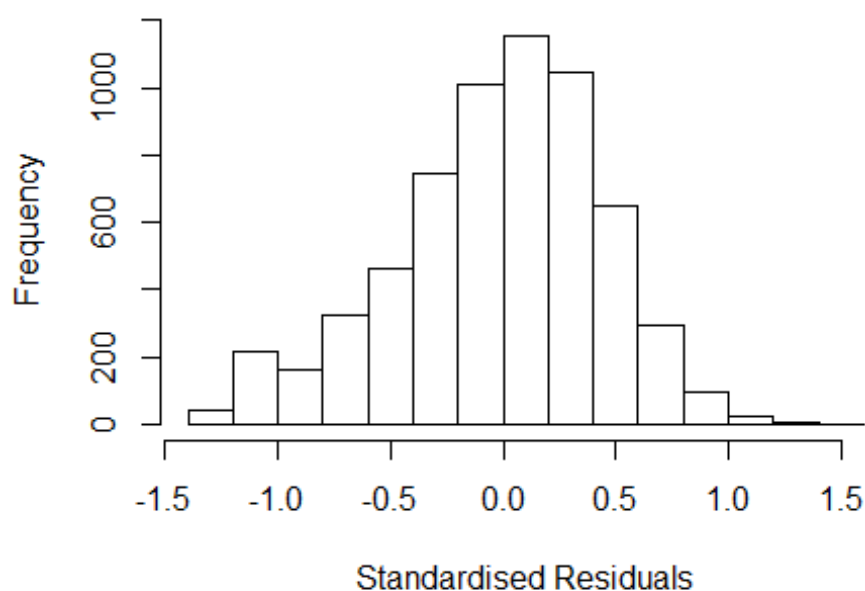
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4376 -0.2805 0.0131 0.2915 1.7891
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9820 0.0365 26.89 < 2e-16 ***
## FirstAuthorFemale1 0.0132 0.0116 1.14 0.2535
## LastAuthorFemale1 -0.0361 0.0116 -3.11 0.0019 **
## UniqueAuthors2 0.2945 0.0243 12.12 < 2e-16 ***
## UniqueAuthors3 0.3425 0.0236 14.51 < 2e-16 ***
## UniqueAuthors4 0.3712 0.0241 15.41 < 2e-16 ***
## UniqueAuthors5 0.4334 0.0225 19.26 < 2e-16 ***
## Year1997 0.0221 0.0453 0.49 0.6251
## Year1998 -0.0214 0.0426 -0.50 0.6151
## Year1999 -0.0536 0.0428 -1.25 0.2100
```

```

## Year2000          -0.0406      0.0415   -0.98    0.3282
## Year2001          -0.0428      0.0425   -1.01    0.3140
## Year2002          -0.0567      0.0412   -1.38    0.1685
## Year2003          -0.1088      0.0433   -2.51    0.0120 *
## Year2004          -0.1279      0.0431   -2.96    0.0030 **
## Year2005          -0.1585      0.0390   -4.06    4.9e-05 ***
## Year2006          -0.1528      0.0389   -3.93    8.5e-05 ***
## Year2007          -0.1756      0.0368   -4.77    1.9e-06 ***
## Year2008          -0.2018      0.0372   -5.43    5.8e-08 ***
## Year2009          -0.1525      0.0381   -4.00    6.3e-05 ***
## Year2010          -0.2056      0.0379   -5.42    6.2e-08 ***
## Year2011          -0.2181      0.0369   -5.91    3.7e-09 ***
## Year2012          -0.2204      0.0372   -5.92    3.4e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.106, Adjusted R-squared:  0.103
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 483 weights are ~= 1. The remaining 5734 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0372 0.8680 0.9500 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          1.61e-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 1.024 1 1.012
## LastAuthorFemale 1.023 1 1.011
## Year 1.036 16 1.001

```

Residuals from first and last author



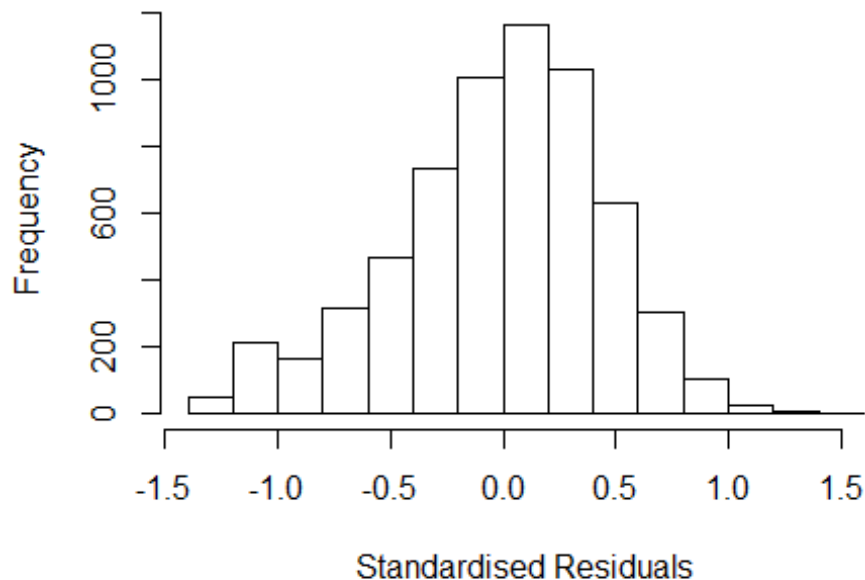
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2899 -0.2934  0.0279  0.2933  1.4586
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2581    0.0342   36.77 < 2e-16 ***
## FirstAuthorFemale1  0.0318    0.0120    2.66  0.00781 **
## LastAuthorFemale1 -0.0405    0.0120   -3.38  0.00072 ***
## Year1997          0.0146    0.0464    0.31  0.75325
## Year1998         -0.0385    0.0451   -0.85  0.39348
## Year1999         -0.0257    0.0444   -0.58  0.56320
## Year2000         -0.0289    0.0425   -0.68  0.49725
## Year2001         -0.0113    0.0436   -0.26  0.79640
## Year2002         -0.0159    0.0420   -0.38  0.70562
## Year2003         -0.0769    0.0447   -1.72  0.08542 .
## Year2004         -0.0819    0.0437   -1.87  0.06095 .
## Year2005         -0.1130    0.0404   -2.79  0.00522 **
```

```

## Year2006          -0.1194      0.0408    -2.93   0.00342 **
## Year2007          -0.1324      0.0387    -3.42   0.00062 ***
## Year2008          -0.1527      0.0391    -3.91   9.3e-05 ***
## Year2009          -0.1159      0.0397    -2.92   0.00351 **
## Year2010          -0.1566      0.0397    -3.95   7.9e-05 ***
## Year2011          -0.1638      0.0387    -4.23   2.3e-05 ***
## Year2012          -0.1625      0.0389    -4.18   3.0e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.0188, Adjusted R-squared:  0.0159
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 498 weights are ~= 1. The remaining 5719 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.232  0.862  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      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 1.017 1          1.009
## Year              1.017 16          1.001

```


Residuals from first author



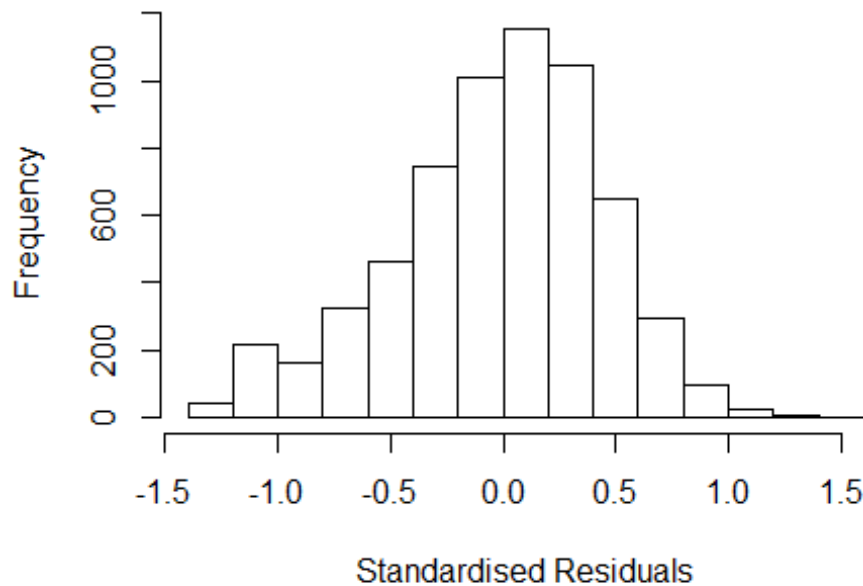
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2804 -0.2946 0.0258 0.2967 1.4712
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2428 0.0339 36.68 < 2e-16 ***
## FirstAuthorFemale1 0.0245 0.0120 2.04 0.04126 *
## Year1997 0.0132 0.0464 0.28 0.77665
## Year1998 -0.0363 0.0452 -0.80 0.42192
## Year1999 -0.0216 0.0444 -0.49 0.62756
## Year2000 -0.0267 0.0426 -0.63 0.53047
## Year2001 -0.0105 0.0437 -0.24 0.80984
## Year2002 -0.0118 0.0421 -0.28 0.77906
## Year2003 -0.0743 0.0447 -1.66 0.09621 .
## Year2004 -0.0798 0.0438 -1.82 0.06864 .
## Year2005 -0.1118 0.0405 -2.76 0.00580 **
## Year2006 -0.1156 0.0409 -2.83 0.00470 **
```

```

## Year2007          -0.1297      0.0388   -3.34  0.00084 ***
## Year2008          -0.1502      0.0392   -3.83  0.00013 ***
## Year2009          -0.1132      0.0398   -2.84  0.00451 **
## Year2010          -0.1557      0.0398   -3.91  9.2e-05 ***
## Year2011          -0.1610      0.0388   -4.15  3.4e-05 ***
## Year2012          -0.1613      0.0390   -4.14  3.6e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.017, Adjusted R-squared:  0.0143
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 500 weights are ~= 1. The remaining 5717 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.225  0.862  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      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 1.017 1          1.008
## Year            1.017 16          1.001

```

Residuals from last author



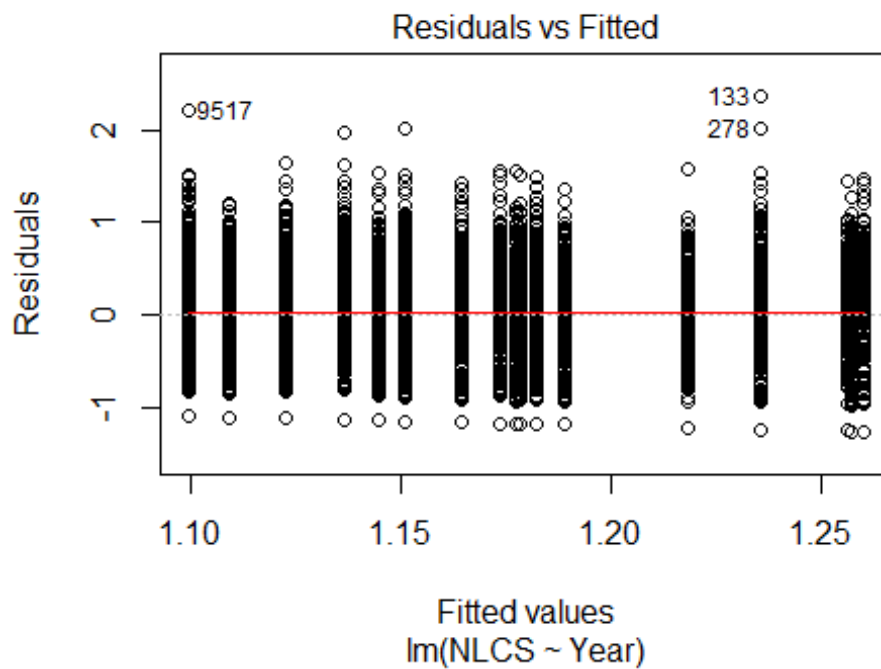
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2846 -0.2918  0.0272  0.2946  1.4422
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2686     0.0339   37.40 < 2e-16 ***
## LastAuthorFemale1 -0.0346     0.0120   -2.88  0.00394 **
## Year1997          0.0160     0.0464    0.35  0.72988
## Year1998         -0.0361     0.0451   -0.80  0.42300
## Year1999         -0.0240     0.0444   -0.54  0.58929
## Year2000         -0.0273     0.0426   -0.64  0.52144
## Year2001         -0.0100     0.0437   -0.23  0.81879
## Year2002         -0.0124     0.0421   -0.29  0.76871
## Year2003         -0.0738     0.0447   -1.65  0.09878 .
## Year2004         -0.0791     0.0438   -1.81  0.07098 .
## Year2005         -0.1104     0.0405   -2.73  0.00642 **
## Year2006         -0.1155     0.0408   -2.83  0.00460 **
```

```

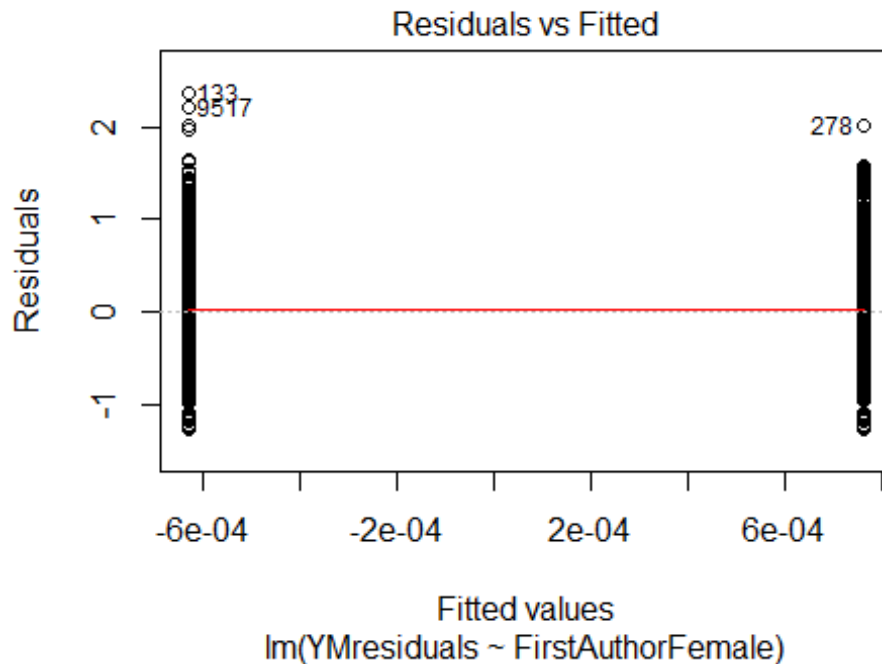
## Year2007          -0.1277      0.0387   -3.30  0.00097 ***
## Year2008          -0.1481      0.0391   -3.79  0.00015 ***
## Year2009          -0.1111      0.0397   -2.80  0.00516 **
## Year2010          -0.1512      0.0396   -3.81  0.00014 ***
## Year2011          -0.1578      0.0387   -4.08  4.6e-05 ***
## Year2012          -0.1578      0.0389   -4.05  5.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.0176, Adjusted R-squared:  0.015
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 515 weights are ~= 1. The remaining 5702 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.243  0.862  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      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: 6217"
## [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
## 408 377 387 352 408 415 466 454 485 575 650 685 790 899 986
## 2011 2012
## 1081 1079
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 241 238 242 263 178 178 338 356 375 462 497 535 597 679 771
## 2011 2012

```

```
## 809 841
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 224 216 220 236 161 161 302 314 330 403 439 482 522 601 692
## 2011 2012
## 743 772
## [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
```

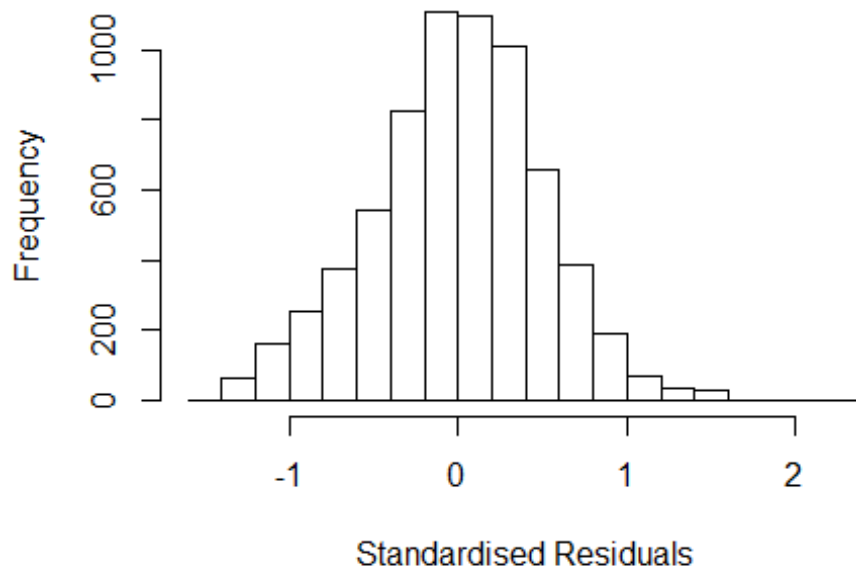


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.5, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 4.06976701774829"
## [1] "Male first author team size 2018 geometric mean: 3.74162590357697"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 94000, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.82812196551593"
## [1] "Male last author team size 2018 geometric mean: 4.08274570045757"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 92000, 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 1.121 1      1.059
## LastAuthorFemale  1.115 1      1.056
## UniqueAuthors     1.083 4      1.010
## Year               1.103 16     1.003
```

Residuals from first and last author and team size



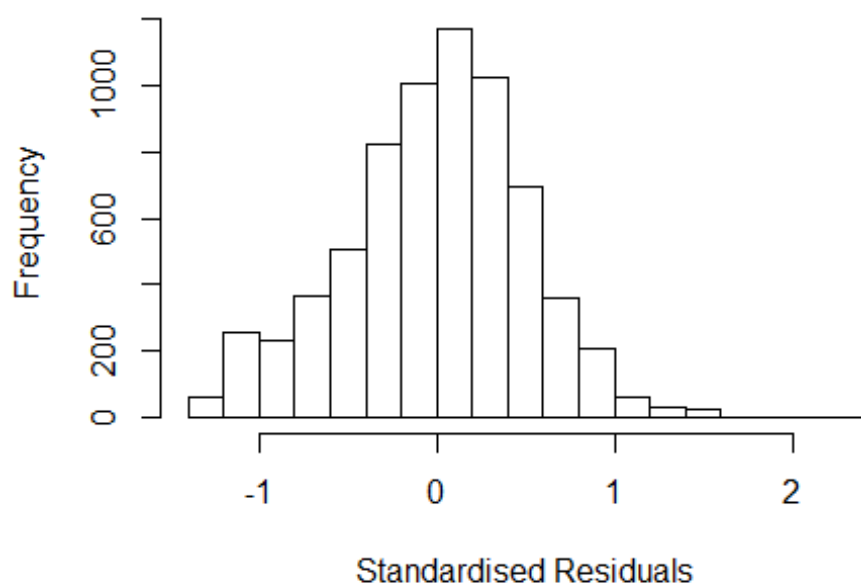
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4058 -0.3227  0.0152  0.3306  2.3237
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.08964    0.04803   22.69 < 2e-16 ***
## FirstAuthorFemale1 -0.00198    0.01323   -0.15  0.88099
## LastAuthorFemale1 -0.01655    0.01321   -1.25  0.21045
## UniqueAuthors2     0.14737    0.02494    5.91 3.6e-09 ***
## UniqueAuthors3     0.20615    0.02399    8.59 < 2e-16 ***
## UniqueAuthors4     0.23089    0.02448    9.43 < 2e-16 ***
## UniqueAuthors5     0.28889    0.02225   12.99 < 2e-16 ***
## Year1997           0.02924    0.06203    0.47  0.63746
## Year1998          -0.05938    0.06190   -0.96  0.33747
## Year1999           0.02246    0.05710    0.39  0.69412
```

```

## Year2000      0.04443      0.06155      0.72  0.47034
## Year2001     -0.01711      0.05974     -0.29  0.77461
## Year2002     -0.11256      0.05337     -2.11  0.03499 *
## Year2003     -0.07897      0.05275     -1.50  0.13443
## Year2004     -0.07455      0.05180     -1.44  0.15019
## Year2005     -0.09583      0.05139     -1.86  0.06228 .
## Year2006     -0.15835      0.05120     -3.09  0.00199 **
## Year2007     -0.09580      0.04981     -1.92  0.05451 .
## Year2008     -0.12877      0.04948     -2.60  0.00927 **
## Year2009     -0.12284      0.04969     -2.47  0.01345 *
## Year2010     -0.18205      0.04906     -3.71  0.00021 ***
## Year2011     -0.14624      0.04893     -2.99  0.00281 **
## Year2012     -0.13010      0.04874     -2.67  0.00762 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.482
## Multiple R-squared:  0.0449, Adjusted R-squared:  0.0418
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 65 is an outlier with |weight| = 0 ( < 1.5e-05);
## 607 weights are ~= 1. The remaining 6210 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.082  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           1.47e-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 1.114 1 1.056
## LastAuthorFemale 1.101 1 1.049
## Year 1.034 16 1.001

```


Residuals from first and last author



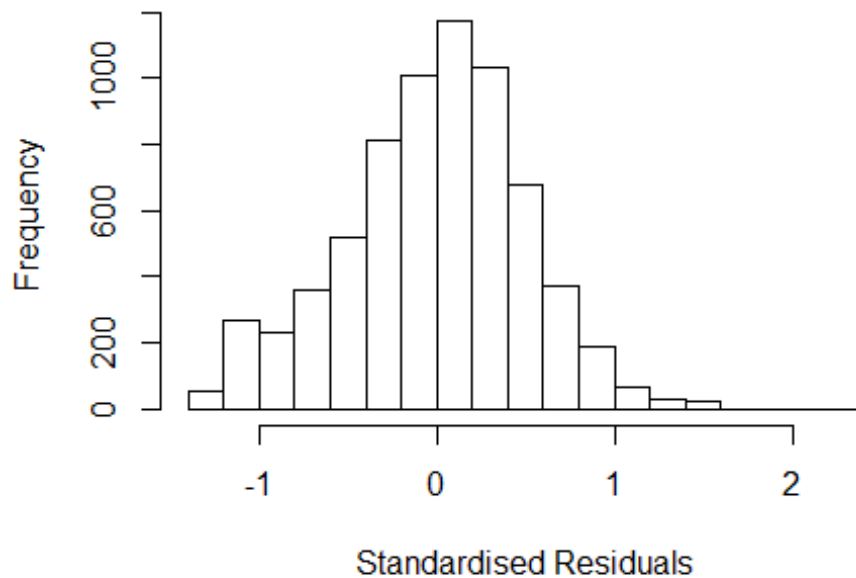
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3103 -0.3294 0.0266 0.3278 2.3846
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23727 0.04659 26.56 <2e-16 ***
## FirstAuthorFemale1 0.01208 0.01341 0.90 0.368
## LastAuthorFemale1 -0.03290 0.01333 -2.47 0.014 *
## Year1997 0.03829 0.06304 0.61 0.544
## Year1998 -0.03540 0.06278 -0.56 0.573
## Year1999 0.06095 0.05795 1.05 0.293
## Year2000 0.05980 0.06051 0.99 0.323
## Year2001 0.00846 0.06097 0.14 0.890
## Year2002 -0.07133 0.05439 -1.31 0.190
## Year2003 -0.03881 0.05396 -0.72 0.472
## Year2004 -0.02073 0.05265 -0.39 0.694
## Year2005 -0.04888 0.05231 -0.93 0.350
```

```

## Year2006          -0.11589    0.05253   -2.21    0.027 *
## Year2007          -0.05069    0.05106   -0.99    0.321
## Year2008          -0.08479    0.05054   -1.68    0.093 .
## Year2009          -0.06863    0.05084   -1.35    0.177
## Year2010          -0.12441    0.05013   -2.48    0.013 *
## Year2011          -0.08738    0.05003   -1.75    0.081 .
## Year2012          -0.07363    0.04986   -1.48    0.140
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.00989,    Adjusted R-squared:  0.00727
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 65 is an outlier with |weight| = 0 ( < 1.5e-05);
## 549 weights are ~= 1. The remaining 6268 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0567 0.8620 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          1.47e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1          1.012
## Year              1.024 16          1.001

```

Residuals from first author



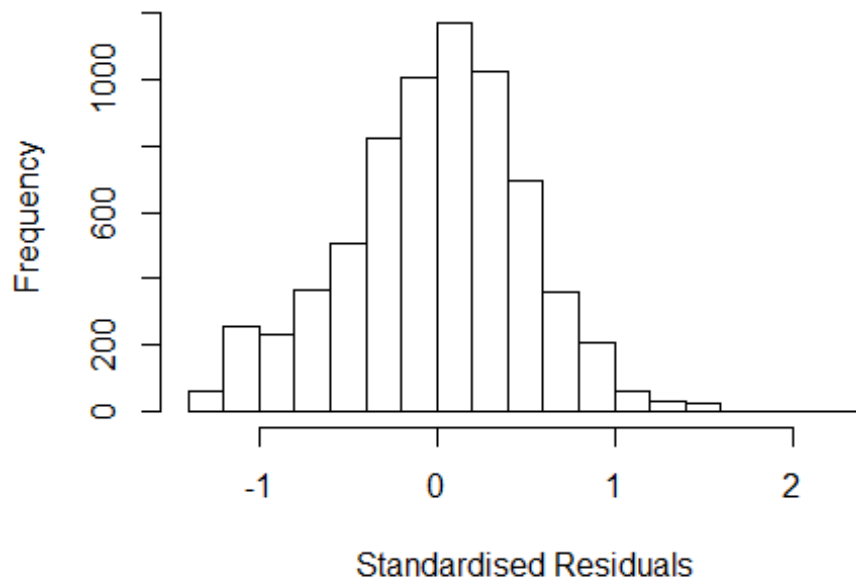
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2918 -0.3276  0.0275  0.3273  2.3706
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22903    0.04634   26.52  <2e-16 ***
## FirstAuthorFemale1 0.00141    0.01286    0.11   0.913
## Year1997         0.03940    0.06300    0.63   0.532
## Year1998        -0.03606    0.06277   -0.57   0.566
## Year1999         0.06136    0.05778    1.06   0.288
## Year2000         0.05887    0.06040    0.97   0.330
## Year2001         0.00863    0.06079    0.14   0.887
## Year2002        -0.07137    0.05430   -1.31   0.189
## Year2003        -0.03760    0.05383   -0.70   0.485
## Year2004        -0.02029    0.05257   -0.39   0.699
## Year2005        -0.05130    0.05216   -0.98   0.325
## Year2006        -0.11613    0.05247   -2.21   0.027 *
```

```

## Year2007          -0.05110    0.05094   -1.00    0.316
## Year2008          -0.08592    0.05043   -1.70    0.088 .
## Year2009          -0.06948    0.05075   -1.37    0.171
## Year2010          -0.12459    0.04999   -2.49    0.013 *
## Year2011          -0.08820    0.04990   -1.77    0.077 .
## Year2012          -0.07379    0.04974   -1.48    0.138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.00898,    Adjusted R-squared:  0.0065
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 65 is an outlier with |weight| = 0 ( < 1.5e-05);
## 525 weights are ~= 1. The remaining 6292 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0495 0.8630 0.9510 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      1.47e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.012 1          1.006
## Year          1.012 16          1.000

```

Residuals from last author



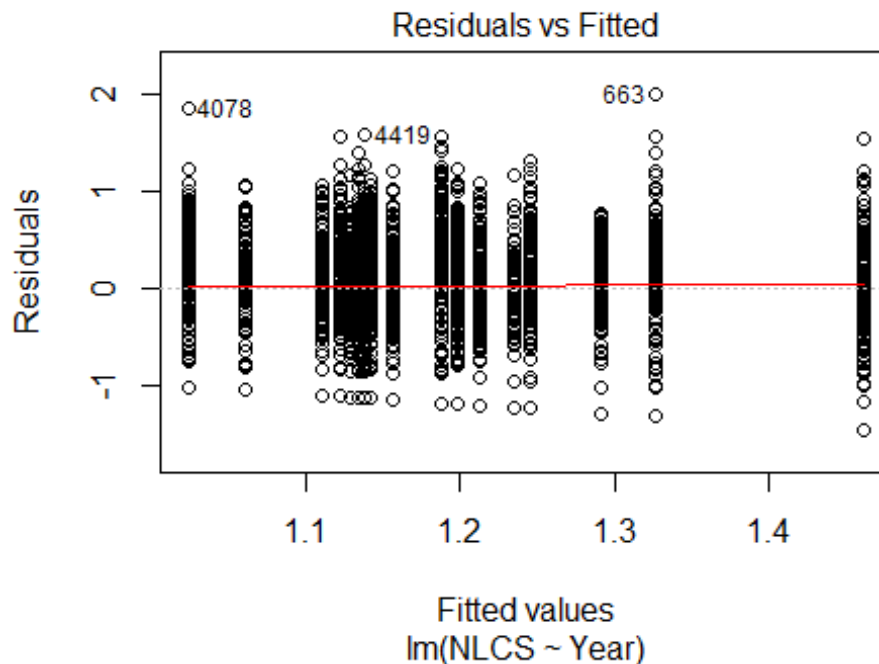
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3023 -0.3295 0.0272 0.3278 2.3888
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2412 0.0464 26.77 <2e-16 ***
## LastAuthorFemale1 -0.0290 0.0128 -2.27 0.023 *
## Year1997 0.0386 0.0631 0.61 0.541
## Year1998 -0.0355 0.0628 -0.57 0.572
## Year1999 0.0610 0.0580 1.05 0.292
## Year2000 0.0601 0.0605 0.99 0.321
## Year2001 0.0088 0.0610 0.14 0.885
## Year2002 -0.0701 0.0544 -1.29 0.197
## Year2003 -0.0383 0.0540 -0.71 0.478
## Year2004 -0.0202 0.0527 -0.38 0.701
## Year2005 -0.0485 0.0523 -0.93 0.354
## Year2006 -0.1146 0.0525 -2.18 0.029 *
```

```

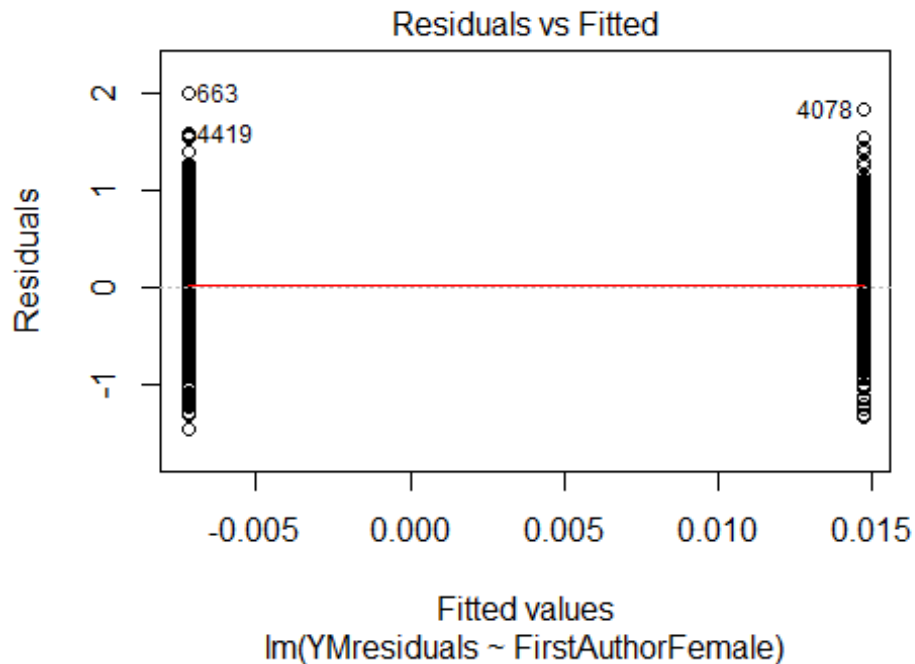
## Year2007          -0.0500      0.0511   -0.98    0.327
## Year2008          -0.0835      0.0505   -1.65    0.099 .
## Year2009          -0.0673      0.0508   -1.32    0.186
## Year2010          -0.1228      0.0501   -2.45    0.014 *
## Year2011          -0.0860      0.0500   -1.72    0.086 .
## Year2012          -0.0717      0.0498   -1.44    0.150
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.00976,    Adjusted R-squared:  0.00728
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 65 is an outlier with |weight| = 0 ( < 1.5e-05);
## 548 weights are ~= 1. The remaining 6269 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0543 0.8630 0.9510 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          1.47e-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: 6818"
## [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
## 319 235 258 246 242 170 239 182 198 194 217 266 274 292 271
## 2011 2012
## 295 283
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 192 131 131 147 140 97 165 118 146 133 149 193 206 190 195

```

```
## 2011 2012
## 233 228
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 147 114 106 126 122 88 131 95 112 104 120 161 171 167 182
## 2011 2012
## 205 188
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 150, df = 16, p-value <2e-16
```

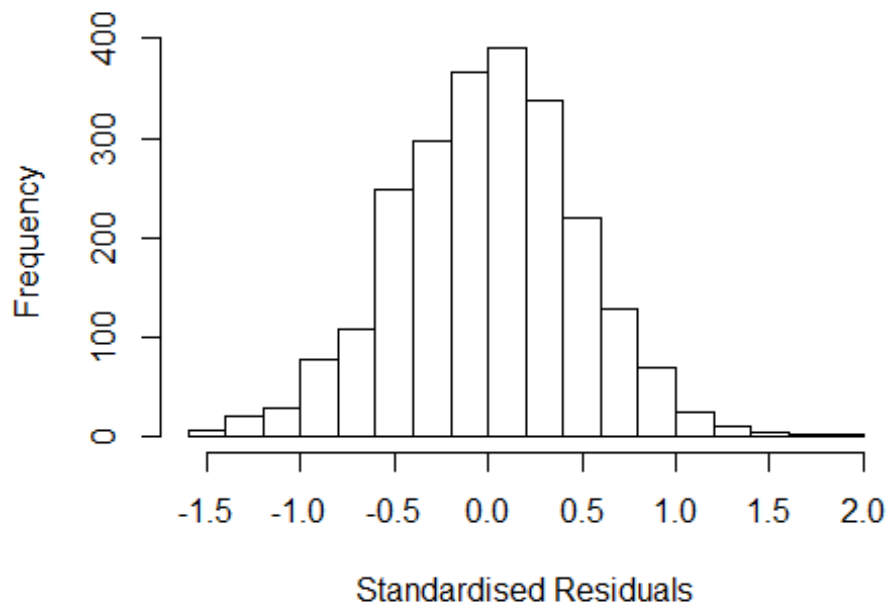


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.6, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 6.58733271697746"
## [1] "Male first author team size 2018 geometric mean: 5.27943487974901"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6000, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.49701824273253"
## [1] "Male last author team size 2018 geometric mean: 5.60626946384653"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5100, 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 1.077 1 1.038
## LastAuthorFemale 1.086 1 1.042
## UniqueAuthors 1.202 4 1.023
## Year 1.311 16 1.009
```


Residuals from first and last author and team size



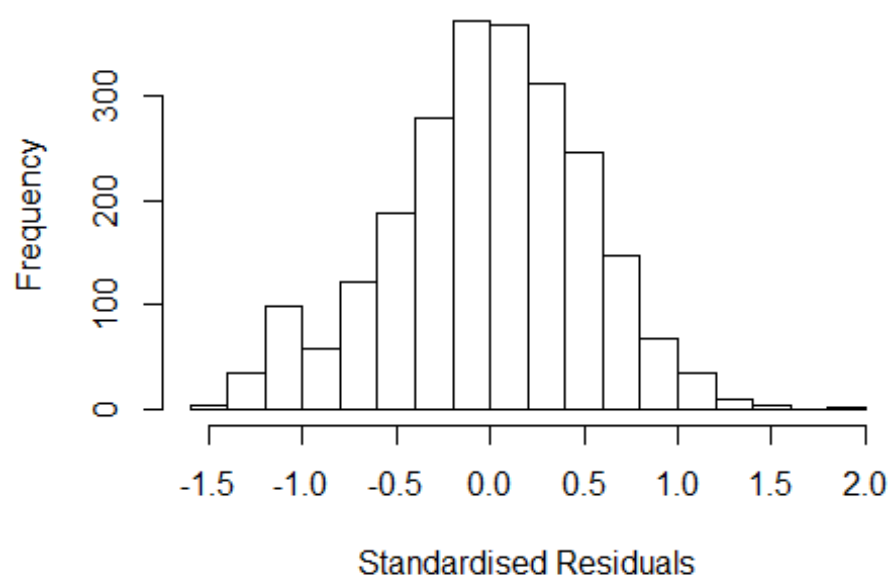
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5709 -0.3239 0.0101 0.3106 1.8414
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.83236 0.06901 12.06 < 2e-16 ***
## FirstAuthorFemale1 0.00312 0.02150 0.14 0.88479
## LastAuthorFemale1 0.04583 0.02592 1.77 0.07722 .
## UniqueAuthors2 0.40693 0.05260 7.74 1.5e-14 ***
## UniqueAuthors3 0.45963 0.04889 9.40 < 2e-16 ***
## UniqueAuthors4 0.54844 0.04577 11.98 < 2e-16 ***
## UniqueAuthors5 0.70536 0.04176 16.89 < 2e-16 ***
## Year1997 -0.01261 0.07832 -0.16 0.87205
## Year1998 0.19648 0.07717 2.55 0.01096 *
## Year1999 0.14757 0.07416 1.99 0.04672 *
```

```

## Year2000      -0.05187    0.06840   -0.76  0.44838
## Year2001      -0.15324    0.07126   -2.15  0.03163 *
## Year2002      -0.21585    0.07304   -2.96  0.00316 **
## Year2003      -0.27200    0.07046   -3.86  0.00012 ***
## Year2004      -0.30477    0.07324   -4.16  3.3e-05 ***
## Year2005      -0.35904    0.07120   -5.04  4.9e-07 ***
## Year2006      -0.25594    0.06874   -3.72  0.00020 ***
## Year2007      -0.29403    0.06663   -4.41  1.1e-05 ***
## Year2008      -0.21554    0.06714   -3.21  0.00134 **
## Year2009      -0.27700    0.06898   -4.02  6.1e-05 ***
## Year2010      -0.38240    0.07099   -5.39  7.9e-08 ***
## Year2011      -0.27231    0.06836   -3.98  7.0e-05 ***
## Year2012      -0.28628    0.06758   -4.24  2.4e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.474
## Multiple R-squared:  0.203, Adjusted R-squared:  0.195
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 209 weights are ~= 1. The remaining 2130 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0983 0.8710 0.9520 0.9030 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.28e-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 1.045 1 1.022
## LastAuthorFemale 1.062 1 1.030
## Year 1.084 16 1.003

```

Residuals from first and last author



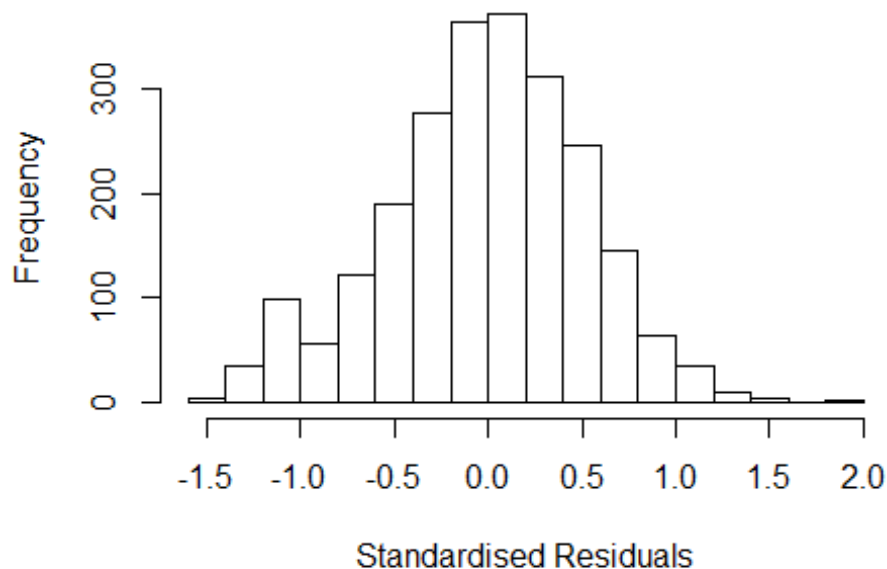
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4596 -0.3308 0.0123 0.3521 1.8663
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3137 0.0639 20.56 < 2e-16 ***
## FirstAuthorFemale1 0.0231 0.0231 1.00 0.31698
## LastAuthorFemale1 0.0222 0.0285 0.78 0.43557
## Year1997 -0.0421 0.0878 -0.48 0.63109
## Year1998 0.1208 0.0804 1.50 0.13290
## Year1999 0.1458 0.0831 1.76 0.07936 .
## Year2000 -0.0130 0.0747 -0.17 0.86155
## Year2001 -0.0823 0.0769 -1.07 0.28485
## Year2002 -0.1009 0.0790 -1.28 0.20191
## Year2003 -0.1833 0.0772 -2.38 0.01761 *
## Year2004 -0.2290 0.0792 -2.89 0.00388 **
## Year2005 -0.2635 0.0778 -3.39 0.00072 ***
```

```

## Year2006          -0.1685      0.0757   -2.23  0.02609 *
## Year2007          -0.2300      0.0740   -3.11  0.00190 **
## Year2008          -0.1188      0.0727   -1.63  0.10253
## Year2009          -0.1921      0.0774   -2.48  0.01311 *
## Year2010          -0.2944      0.0817   -3.60  0.00032 ***
## Year2011          -0.1875      0.0764   -2.45  0.01420 *
## Year2012          -0.1866      0.0753   -2.48  0.01327 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.0496, Adjusted R-squared:  0.0422
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 205 weights are ~= 1. The remaining 2134 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.135  0.862  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      4.28e-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 1.035 1          1.017
## Year              1.035 16          1.001

```

Residuals from first author



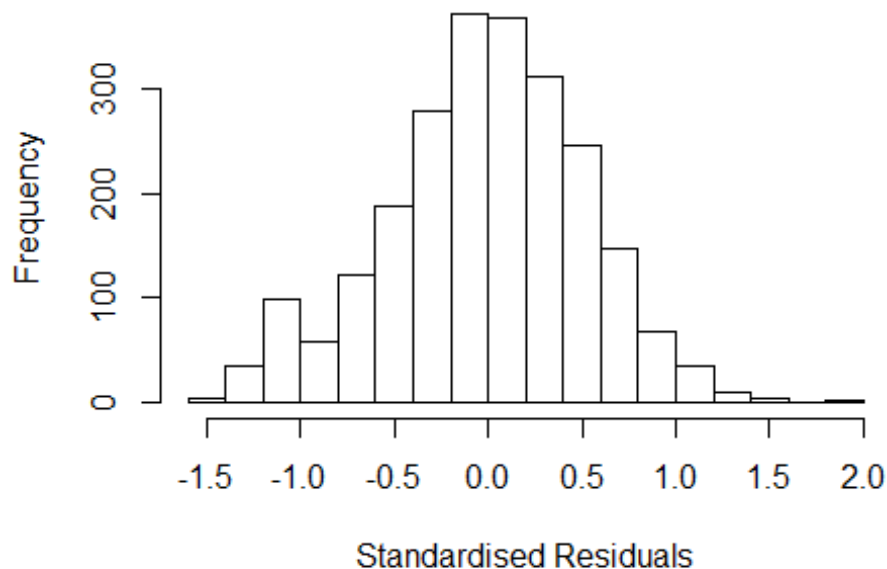
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4630 -0.3294 0.0148 0.3535 1.8864
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3145 0.0639 20.58 < 2e-16 ***
## FirstAuthorFemale1 0.0260 0.0230 1.13 0.25874
## Year1997 -0.0407 0.0877 -0.46 0.64261
## Year1998 0.1221 0.0803 1.52 0.12858
## Year1999 0.1485 0.0828 1.79 0.07279 .
## Year2000 -0.0119 0.0746 -0.16 0.87320
## Year2001 -0.0809 0.0769 -1.05 0.29258
## Year2002 -0.0988 0.0788 -1.26 0.20957
## Year2003 -0.1823 0.0772 -2.36 0.01825 *
## Year2004 -0.2274 0.0791 -2.88 0.00406 **
## Year2005 -0.2597 0.0774 -3.36 0.00080 ***
## Year2006 -0.1646 0.0753 -2.19 0.02896 *
```

```

## Year2007          -0.2270      0.0737   -3.08  0.00209 **
## Year2008          -0.1162      0.0724   -1.60  0.10865
## Year2009          -0.1883      0.0769   -2.45  0.01445 *
## Year2010          -0.2906      0.0813   -3.57  0.00036 ***
## Year2011          -0.1847      0.0761   -2.43  0.01526 *
## Year2012          -0.1833      0.0749   -2.45  0.01454 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.0492, Adjusted R-squared:  0.0423
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 206 weights are ~= 1. The remaining 2133 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.125  0.863  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      4.28e-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 1.051 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
## -1.4668 -0.3333  0.0111  0.3520  1.8568
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3199     0.0635  20.77 < 2e-16 ***
## LastAuthorFemale1  0.0266     0.0283   0.94  0.34703
## Year1997        -0.0421     0.0878  -0.48  0.63191
## Year1998         0.1197     0.0804   1.49  0.13648
## Year1999         0.1469     0.0831   1.77  0.07711 .
## Year2000        -0.0138     0.0747  -0.18  0.85343
## Year2001        -0.0837     0.0768  -1.09  0.27570
## Year2002        -0.1007     0.0789  -1.28  0.20222
## Year2003        -0.1824     0.0772  -2.36  0.01818 *
## Year2004        -0.2285     0.0792  -2.89  0.00393 **
## Year2005        -0.2632     0.0778  -3.38  0.00073 ***
## Year2006        -0.1672     0.0757  -2.21  0.02731 *
```

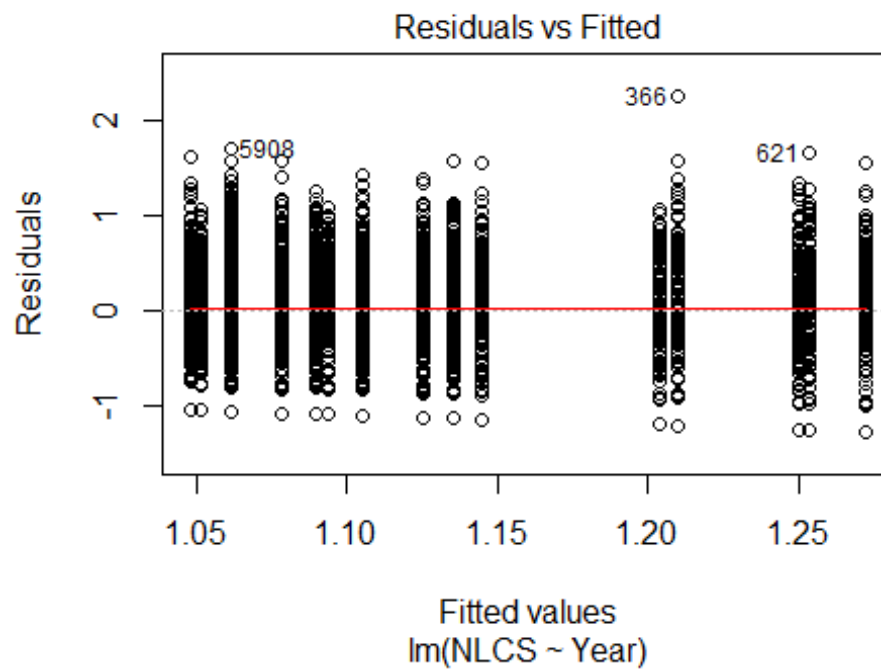
```

## Year2007          -0.2276      0.0739   -3.08   0.00209 **
## Year2008          -0.1162      0.0727   -1.60   0.10998
## Year2009          -0.1906      0.0773   -2.47   0.01368 *
## Year2010          -0.2934      0.0818   -3.59   0.00034 ***
## Year2011          -0.1838      0.0762   -2.41   0.01599 *
## Year2012          -0.1840      0.0752   -2.45   0.01455 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.0492, Adjusted R-squared:  0.0422
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 205 weights are ~= 1. The remaining 2134 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.139  0.863  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      4.28e-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: 2339"
## [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
## 325 312 347 329 338 331 308 309 337 605 447 592 572 560 638
## 2011 2012
## 595 737
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 163 157 187 180 168 133 200 215 226 277 286 344 391 379 424
## 2011 2012

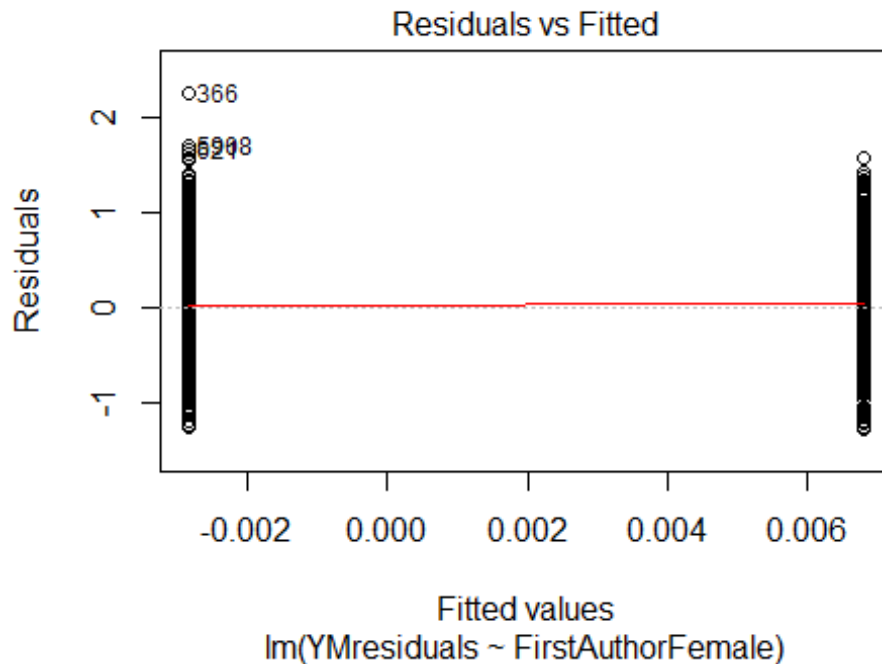
```



```
## 396 437
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 131 132 157 153 137 103 170 168 190 233 237 287 336 307 348
## 2011 2012
## 348 361
## [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
```

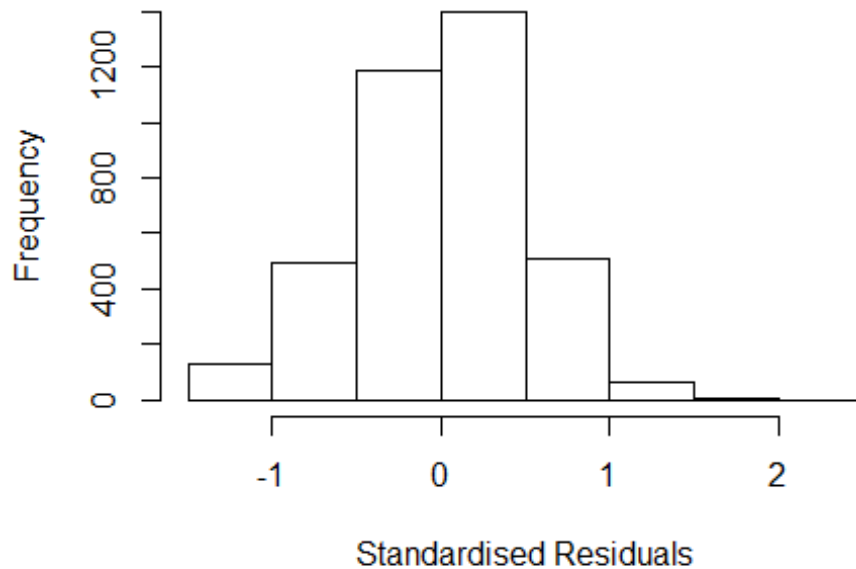


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.61, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 5.17916620520078"
## [1] "Male first author team size 2018 geometric mean: 4.81889859157762"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.79345591113183"
## [1] "Male last author team size 2018 geometric mean: 5.01979580293001"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 13000, 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 1.064 1      1.032
## LastAuthorFemale  1.062 1      1.030
## UniqueAuthors    1.125 4      1.015
## Year             1.167 16     1.005
```

Residuals from first and last author and team size



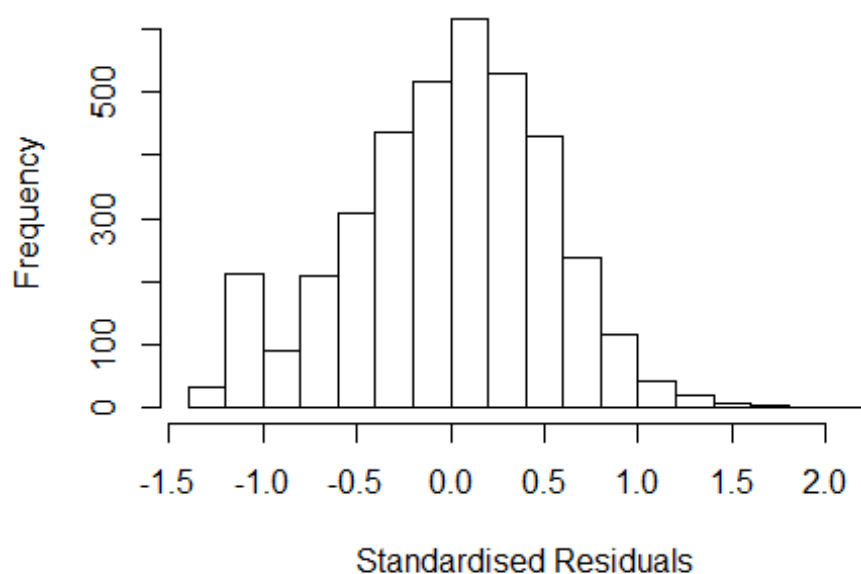
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4673 -0.3399 0.0243 0.3402 2.3322
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00410 0.06247 16.07 < 2e-16 ***
## FirstAuthorFemale1 0.00107 0.01915 0.06 0.95544
## LastAuthorFemale1 -0.00347 0.02205 -0.16 0.87504
## UniqueAuthors2 0.11467 0.04716 2.43 0.01509 *
## UniqueAuthors3 0.19411 0.04555 4.26 2.1e-05 ***
## UniqueAuthors4 0.27840 0.04568 6.10 1.2e-09 ***
## UniqueAuthors5 0.46217 0.04209 10.98 < 2e-16 ***
## Year1997 -0.04207 0.07727 -0.54 0.58614
## Year1998 -0.00834 0.06762 -0.12 0.90186
## Year1999 -0.14028 0.06779 -2.07 0.03859 *
```

```

## Year2000      -0.07892      0.06771      -1.17      0.24387
## Year2001      -0.11792      0.07170      -1.64      0.10011
## Year2002      -0.14638      0.06469      -2.26      0.02369 *
## Year2003      -0.17491      0.06919      -2.53      0.01151 *
## Year2004      -0.23070      0.06201      -3.72      0.00020 ***
## Year2005      -0.15254      0.06178      -2.47      0.01360 *
## Year2006      -0.16436      0.06068      -2.71      0.00679 **
## Year2007      -0.14167      0.06090      -2.33      0.02005 *
## Year2008      -0.23257      0.05926      -3.92      8.9e-05 ***
## Year2009      -0.23084      0.06109      -3.78      0.00016 ***
## Year2010      -0.21886      0.05848      -3.74      0.00018 ***
## Year2011      -0.24434      0.05880      -4.16      3.3e-05 ***
## Year2012      -0.28036      0.05896      -4.75      2.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.499
## Multiple R-squared:  0.0972, Adjusted R-squared:  0.0919
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 323 weights are ~= 1. The remaining 3475 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000  0.867  0.950  0.903  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.63e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.047 1      1.023
## LastAuthorFemale  1.045 1      1.022
## Year              1.067 16      1.002

```

Residuals from first and last author



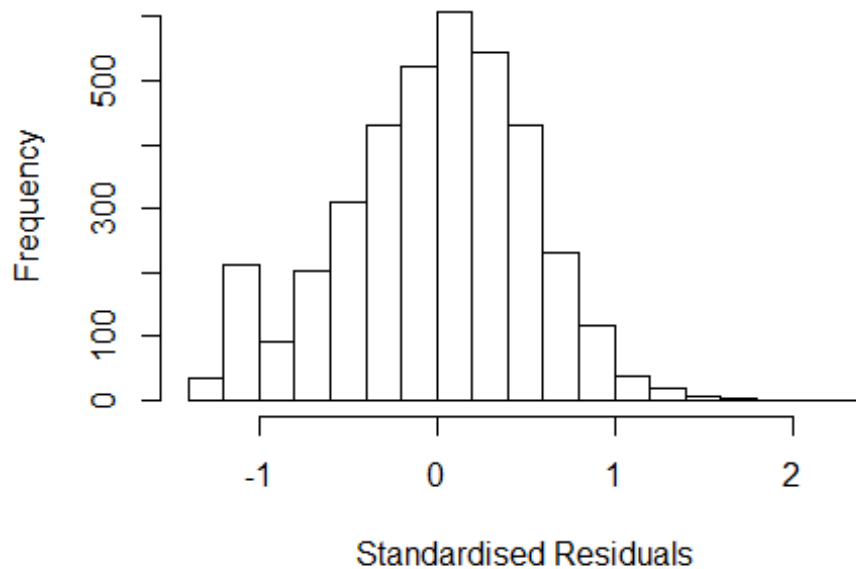
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3032 -0.3465 0.0272 0.3547 2.1970
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2540 0.0559 22.42 <2e-16 ***
## FirstAuthorFemale1 0.0274 0.0199 1.38 0.1676
## LastAuthorFemale1 -0.0170 0.0230 -0.74 0.4609
## Year1997 -0.0102 0.0795 -0.13 0.8978
## Year1998 0.0217 0.0724 0.30 0.7639
## Year1999 -0.1447 0.0707 -2.05 0.0408 *
## Year2000 -0.0480 0.0722 -0.67 0.5060
## Year2001 -0.0645 0.0770 -0.84 0.4021
## Year2002 -0.1143 0.0678 -1.69 0.0920 .
## Year2003 -0.1627 0.0721 -2.26 0.0241 *
## Year2004 -0.2053 0.0657 -3.13 0.0018 **
## Year2005 -0.0985 0.0653 -1.51 0.1317
```

```

## Year2006          -0.1119      0.0642   -1.74   0.0815 .
## Year2007          -0.0934      0.0652   -1.43   0.1519
## Year2008          -0.1935      0.0636   -3.04   0.0024 **
## Year2009          -0.1795      0.0650   -2.76   0.0058 **
## Year2010          -0.1488      0.0621   -2.40   0.0166 *
## Year2011          -0.1697      0.0623   -2.72   0.0065 **
## Year2012          -0.1985      0.0627   -3.16   0.0016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.0144, Adjusted R-squared:  0.00973
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 342 weights are ~= 1. The remaining 3456 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0319 0.8670 0.9480 0.9020 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.63e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1      1.016
## Year      1.032 16      1.001

```

Residuals from first author



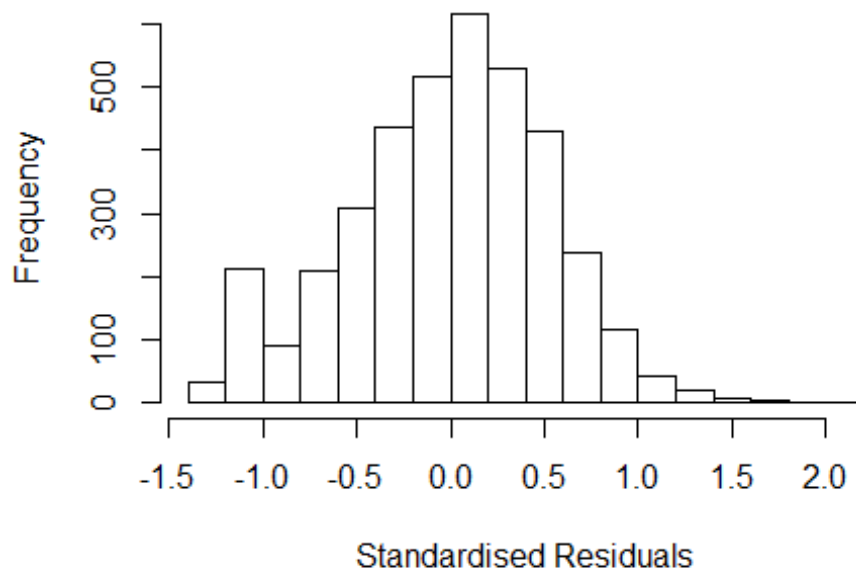
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2986 -0.3462  0.0258  0.3543  2.2012
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2498    0.0553   22.60  <2e-16 ***
## FirstAuthorFemale1  0.0250    0.0198    1.26   0.2065
## Year1997        -0.0085    0.0794   -0.11   0.9147
## Year1998         0.0238    0.0722    0.33   0.7413
## Year1999        -0.1436    0.0706   -2.04   0.0419 *
## Year2000        -0.0473    0.0722   -0.65   0.5125
## Year2001        -0.0615    0.0767   -0.80   0.4223
## Year2002        -0.1118    0.0675   -1.65   0.0980 .
## Year2003        -0.1605    0.0718   -2.23   0.0255 *
## Year2004        -0.2039    0.0655   -3.11   0.0019 **
## Year2005        -0.0964    0.0651   -1.48   0.1388
## Year2006        -0.1099    0.0639   -1.72   0.0855 .
```

```

## Year2007          -0.0911      0.0649   -1.40    0.1607
## Year2008          -0.1919      0.0634   -3.03    0.0025 **
## Year2009          -0.1772      0.0647   -2.74    0.0062 **
## Year2010          -0.1469      0.0618   -2.37    0.0176 *
## Year2011          -0.1684      0.0622   -2.71    0.0068 **
## Year2012          -0.1978      0.0627   -3.16    0.0016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.0143, Adjusted R-squared:  0.00983
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 342 weights are ~= 1. The remaining 3456 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0306 0.8670 0.9480 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.63e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



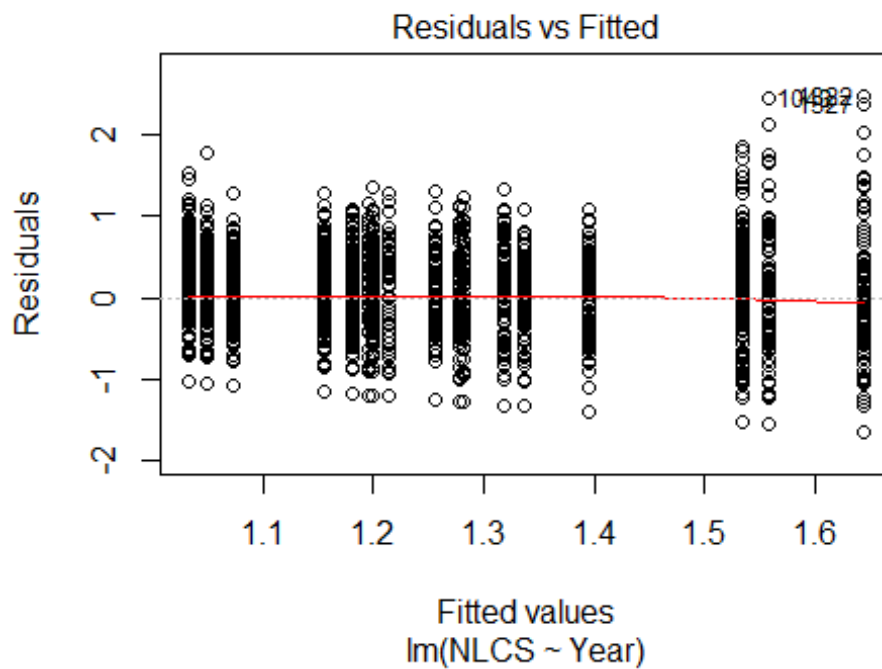
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2800 -0.3500 0.0324 0.3565 2.1933
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2577 0.0559 22.49 <2e-16 ***
## LastAuthorFemale1 -0.0117 0.0229 -0.51 0.6101
## Year1997 -0.0110 0.0795 -0.14 0.8905
## Year1998 0.0223 0.0724 0.31 0.7584
## Year1999 -0.1435 0.0707 -2.03 0.0424 *
## Year2000 -0.0463 0.0723 -0.64 0.5214
## Year2001 -0.0612 0.0769 -0.80 0.4267
## Year2002 -0.1121 0.0678 -1.65 0.0984 .
## Year2003 -0.1600 0.0721 -2.22 0.0266 *
## Year2004 -0.2019 0.0656 -3.08 0.0021 **
## Year2005 -0.0930 0.0651 -1.43 0.1531
## Year2006 -0.1078 0.0641 -1.68 0.0929 .
```

```

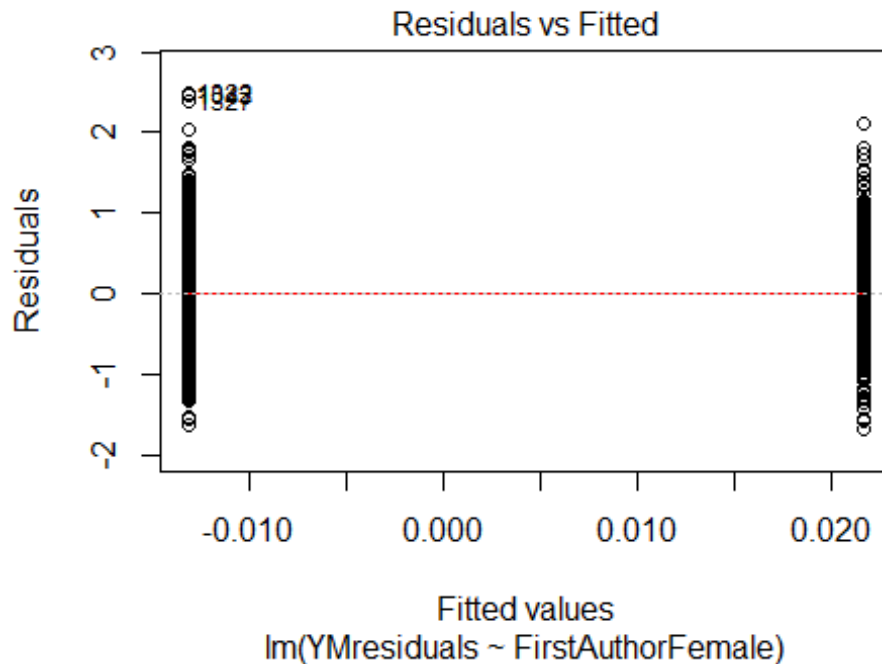
## Year2007          -0.0893      0.0652    -1.37    0.1705
## Year2008          -0.1899      0.0636    -2.99    0.0028 **
## Year2009          -0.1743      0.0648    -2.69    0.0072 **
## Year2010          -0.1449      0.0620    -2.34    0.0195 *
## Year2011          -0.1640      0.0621    -2.64    0.0083 **
## Year2012          -0.1948      0.0626    -3.11    0.0019 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.518
## Multiple R-squared:  0.0139, Adjusted R-squared:  0.00943
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 350 weights are ~= 1. The remaining 3448 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.033  0.867  0.948  0.901  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.63e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3798"
## [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
## 122  96  82  93  111  136  134  117  127  140  176  178  225  265  296
## 2011 2012
## 273  317
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 72  66  63  68  69  80  105  90  98  118  146  151  179  214  250
## 2011 2012

```

```
## 230 259
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 61 57 60 62 65 70 93 81 89 107 134 130 154 196 220
## 2011 2012
## 204 228
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 120, df = 16, p-value <2e-16
```

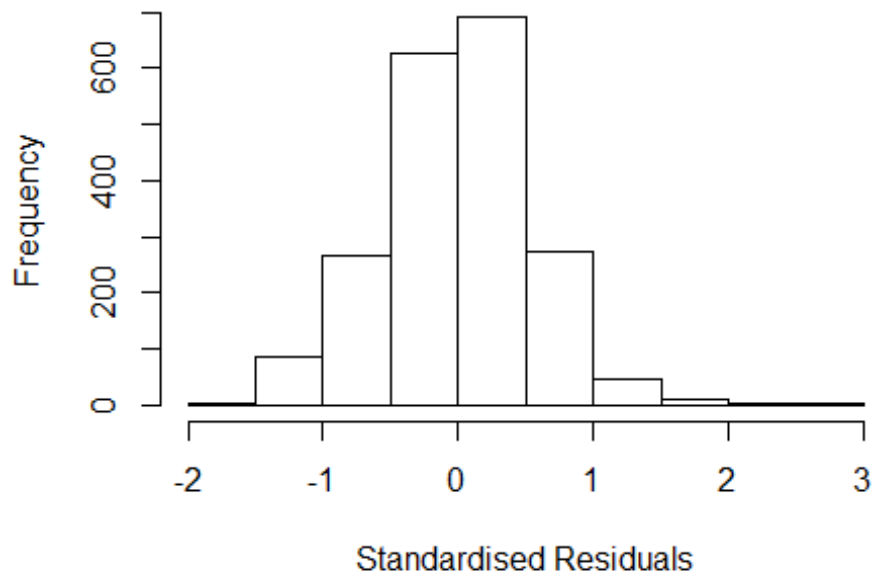


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.9, df = 1, p-value = 0.05
```



```
## [1] "Female first author team size 2018 geometric mean: 3.77960265947802"
## [1] "Male first author team size 2018 geometric mean: 4.15519870258701"
##
## 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] "Female last author team size 2018 geometric mean: 3.79377246953243"
## [1] "Male last author team size 2018 geometric mean: 4.04677875531432"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7700, 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 1.111 1      1.054
## LastAuthorFemale  1.121 1      1.059
## UniqueAuthors     1.172 4      1.020
## Year              1.235 16     1.007
```

Residuals from first and last author and team size



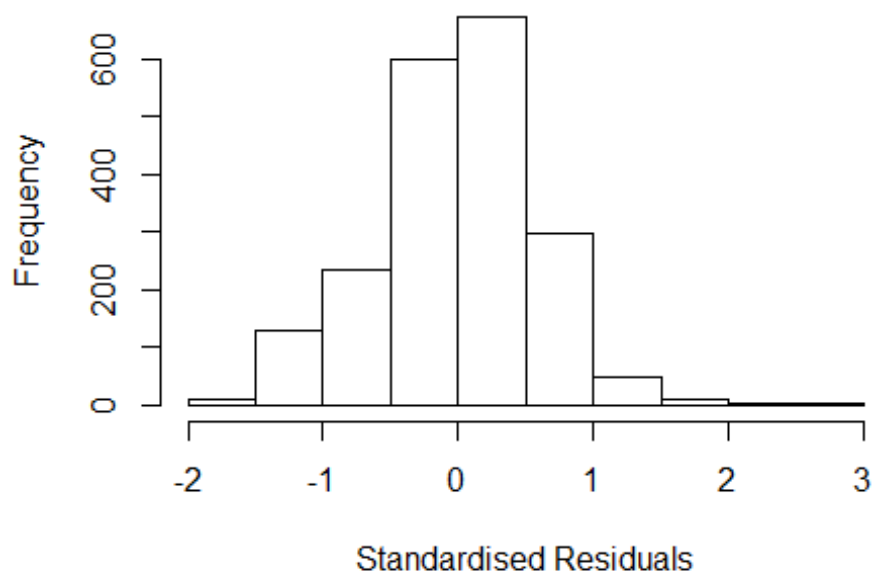
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1043  9744281230 4.000 2004      2742      1      2.966
## 1322 16344376061 4.116 2005      2742      1      2.599
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7464 -0.3507  0.0113  0.3446  2.9659
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91196    0.09464   9.64 < 2e-16 ***
## FirstAuthorFemale1 -0.01927    0.02643  -0.73  0.4661
## LastAuthorFemale1 -0.04401    0.02587  -1.70  0.0891 .
## UniqueAuthors2    0.34539    0.05570   6.20 6.8e-10 ***
## UniqueAuthors3    0.42330    0.05137   8.24 3.1e-16 ***
## UniqueAuthors4    0.48401    0.05301   9.13 < 2e-16 ***
## UniqueAuthors5    0.63309    0.05028  12.59 < 2e-16 ***
## Year1997         -0.04215    0.11618  -0.36  0.7168
## Year1998         -0.06665    0.12002  -0.56  0.5787
```

```

## Year1999      0.00166      0.10114      0.02      0.9869
## Year2000      0.02272      0.10619      0.21      0.8306
## Year2001      0.13101      0.09908      1.32      0.1862
## Year2002     -0.00536      0.10122     -0.05      0.9578
## Year2003      0.05508      0.09376      0.59      0.5569
## Year2004      0.18539      0.11334      1.64      0.1021
## Year2005      0.24536      0.10282      2.39      0.0171 *
## Year2006      0.23131      0.09797      2.36      0.0183 *
## Year2007     -0.14161      0.08976     -1.58      0.1148
## Year2008     -0.16414      0.08617     -1.90      0.0569 .
## Year2009     -0.21314      0.08580     -2.48      0.0131 *
## Year2010     -0.16100      0.08489     -1.90      0.0580 .
## Year2011     -0.28217      0.08708     -3.24      0.0012 **
## Year2012     -0.25892      0.08756     -2.96      0.0031 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.18,   Adjusted R-squared:  0.171
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(551,714) are outliers with |weight| = 0 ( < 5e-05);
## 160 weights are ~= 1. The remaining 1849 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0137 0.8660 0.9530 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.087 1      1.042
## LastAuthorFemale 1.098 1      1.048
## Year      1.078 16      1.002

```

Residuals from first and last author



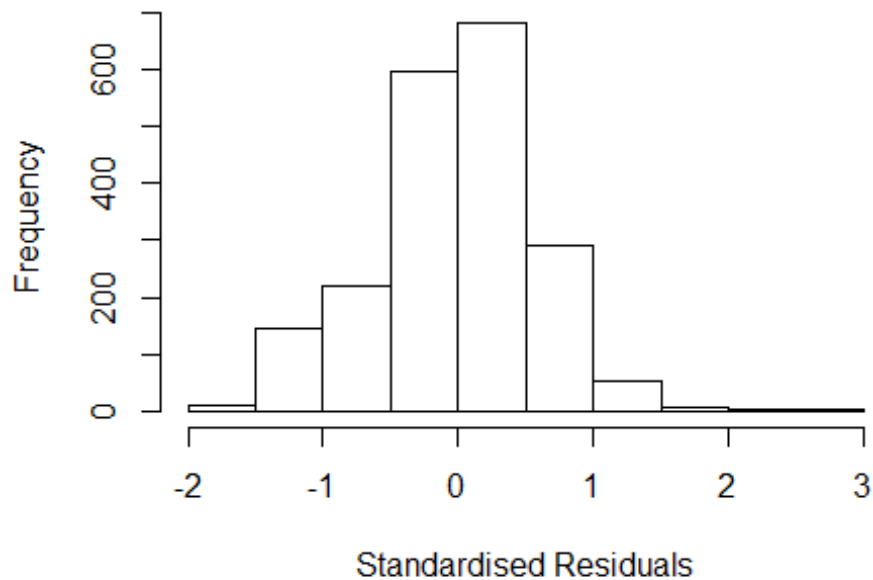
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1043  9744281230 4.000 2004      2742      1      2.557
## 1322 16344376061 4.116 2005      2742      1      2.612
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6137 -0.3692  0.0162  0.3758  2.6124
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3068    0.0838   15.59  <2e-16 ***
## FirstAuthorFemale1 -0.0111    0.0275   -0.40   0.6879
## LastAuthorFemale1  -0.0826    0.0272   -3.04   0.0024 **
## Year1997          -0.0744    0.1224   -0.61   0.5433
## Year1998          -0.0468    0.1221   -0.38   0.7015
## Year1999           0.0270    0.1099    0.25   0.8063
## Year2000           0.0103    0.1091    0.09   0.9246
## Year2001           0.1646    0.1048    1.57   0.1164
## Year2002           0.0404    0.1056    0.38   0.7020
## Year2003           0.1031    0.1001    1.03   0.3027
## Year2004           0.2302    0.1192    1.93   0.0537 .
```

```

## Year2005          0.2904      0.1097      2.65      0.0082 **
## Year2006          0.3069      0.1032      2.97      0.0030 **
## Year2007         -0.0651      0.0948     -0.69      0.4927
## Year2008         -0.0909      0.0903     -1.01      0.3141
## Year2009         -0.1530      0.0902     -1.70      0.0900 .
## Year2010         -0.0934      0.0895     -1.04      0.2969
## Year2011         -0.2349      0.0915     -2.57      0.0103 *
## Year2012         -0.2130      0.0927     -2.30      0.0217 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.0854, Adjusted R-squared:  0.0772
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(551,714) are outliers with |weight| = 0 ( < 5e-05);
## 170 weights are ~= 1. The remaining 1839 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0064 0.8630 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          4.97e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.031 1          1.015
## Year              1.031 16          1.001

```


Residuals from first author



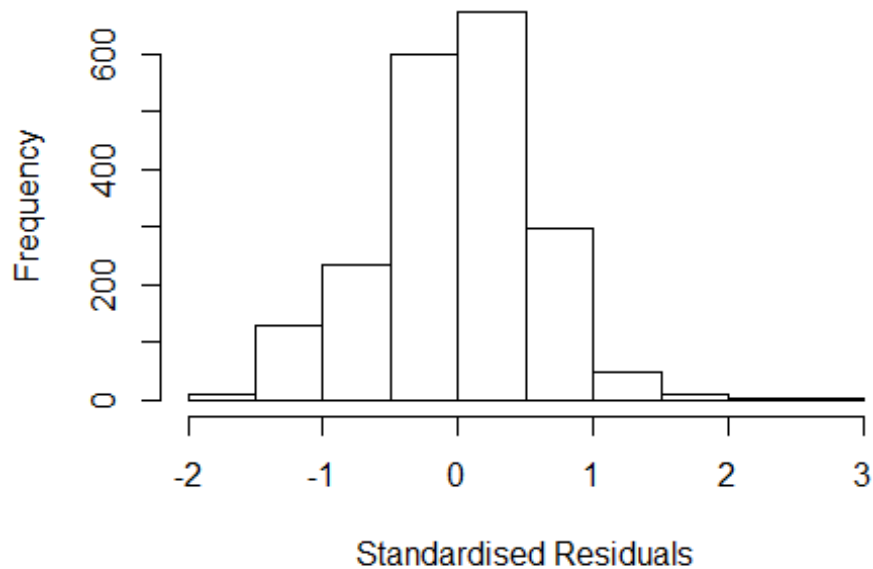
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1043  9744281230 4.000 2004      2742      1      2.557
## 1322 16344376061 4.116 2005      2742      1      2.612
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5840 -0.3587  0.0201  0.3772  2.5852
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.28865    0.08510   15.14  <2e-16 ***
## FirstAuthorFemale1 -0.03467    0.02681   -1.29   0.1961
## Year1997       -0.07528    0.12230   -0.62   0.5383
## Year1998       -0.04822    0.12160   -0.40   0.6917
## Year1999        0.01143    0.11113    0.10   0.9181
## Year2000        0.00535    0.11025    0.05   0.9613
## Year2001        0.15968    0.10466    1.53   0.1272
## Year2002        0.03304    0.10683    0.31   0.7571
## Year2003        0.09847    0.10083    0.98   0.3289
## Year2004        0.23761    0.12007    1.98   0.0480 *
## Year2005        0.27678    0.11005    2.52   0.0120 *
```

```

## Year2006          0.29533    0.10419    2.83    0.0046 **
## Year2007          -0.07562    0.09588   -0.79    0.4304
## Year2008          -0.09486    0.09196   -1.03    0.3024
## Year2009          -0.16230    0.09149   -1.77    0.0762 .
## Year2010          -0.10377    0.09077   -1.14    0.2531
## Year2011          -0.23939    0.09267   -2.58    0.0099 **
## Year2012          -0.21817    0.09432   -2.31    0.0208 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.543
## Multiple R-squared:  0.0814, Adjusted R-squared:  0.0735
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 714 is an outlier with |weight| = 0 ( < 5e-05);
## 175 weights are ~= 1. The remaining 1835 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0008 0.8620 0.9490 0.8950 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.042 1          1.021
## Year            1.042 16          1.001

```

Residuals from last author



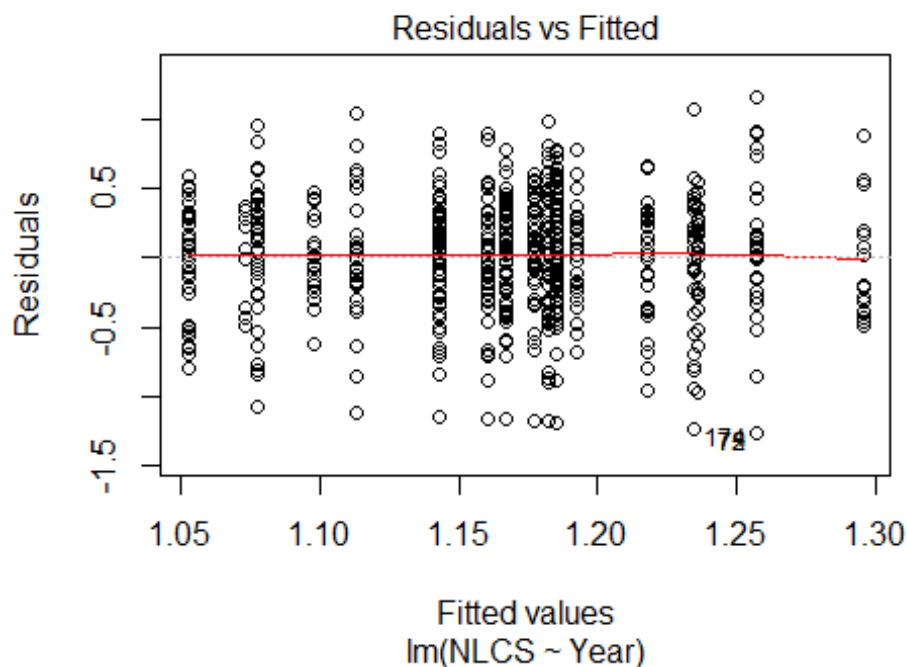
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1043  9744281230 4.000 2004      2742      1      2.557
## 1322 16344376061 4.116 2005      2742      1      2.612
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6079 -0.3689  0.0195  0.3767  2.6081
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3013     0.0828   15.71  <2e-16 ***
## LastAuthorFemale1 -0.0856     0.0265   -3.23  0.0013 **
## Year1997         -0.0737     0.1224   -0.60  0.5470
## Year1998         -0.0459     0.1221   -0.38  0.7072
## Year1999          0.0282     0.1099    0.26  0.7978
## Year2000          0.0119     0.1092    0.11  0.9135
## Year2001          0.1647     0.1048    1.57  0.1165
## Year2002          0.0410     0.1056    0.39  0.6975
## Year2003          0.1033     0.1001    1.03  0.3024
## Year2004          0.2291     0.1193    1.92  0.0549 .
## Year2005          0.2922     0.1094    2.67  0.0076 **
```

```

## Year2006          0.3066      0.1033      2.97      0.0030 **
## Year2007          -0.0650      0.0949     -0.69      0.4931
## Year2008          -0.0909      0.0903     -1.01      0.3143
## Year2009          -0.1537      0.0902     -1.70      0.0886 .
## Year2010          -0.0938      0.0895     -1.05      0.2949
## Year2011          -0.2349      0.0915     -2.57      0.0103 *
## Year2012          -0.2127      0.0928     -2.29      0.0220 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.543
## Multiple R-squared:  0.0854, Adjusted R-squared:  0.0776
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(551,714) are outliers with |weight| = 0 ( < 5e-05);
## 171 weights are ~ = 1. The remaining 1838 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0071 0.8620 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          4.97e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2011"
## [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
##   59   53   58   41   57   63   60   58   68   55   70   84   76  105  105
## 2011 2012
##   107   89
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

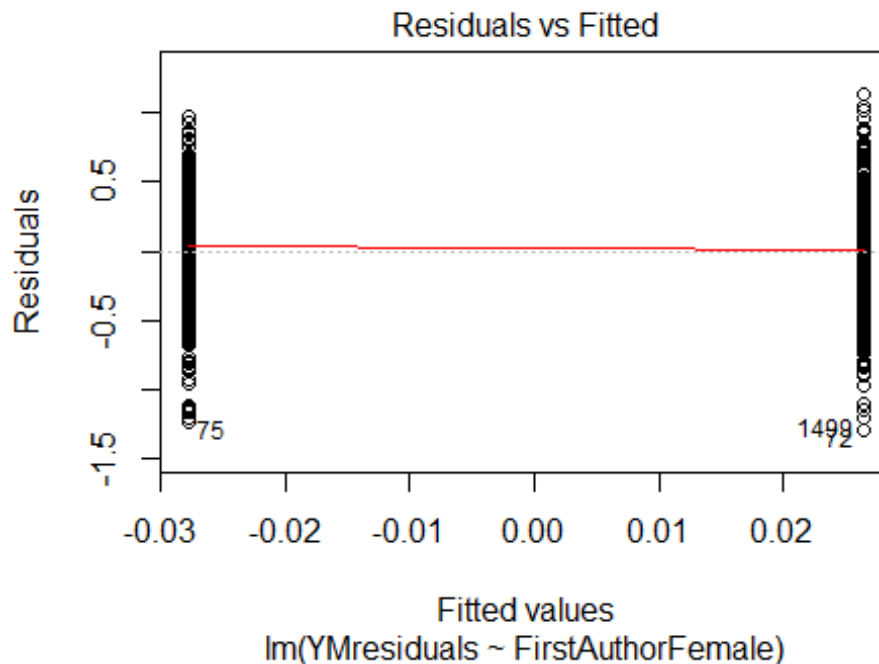
```
##      9   28   31   17   23   18   29   33   39   32   39   52   48   70   74
## 2011 2012
##    71   54
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   24   24   12   14    8   18   23   30   23   33   39   34   63   57
## 2011 2012
##    54   46
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```



```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.49, df = 1, p-value = 0.5
## [1] "Female first author team size 2018 geometric mean: 4.63183139782037"
## [1] "Male first author team size 2018 geometric mean: 4.58081722307047"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 470, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.82543119058968"
## [1] "Male last author team size 2018 geometric mean: 4.47610187331959"

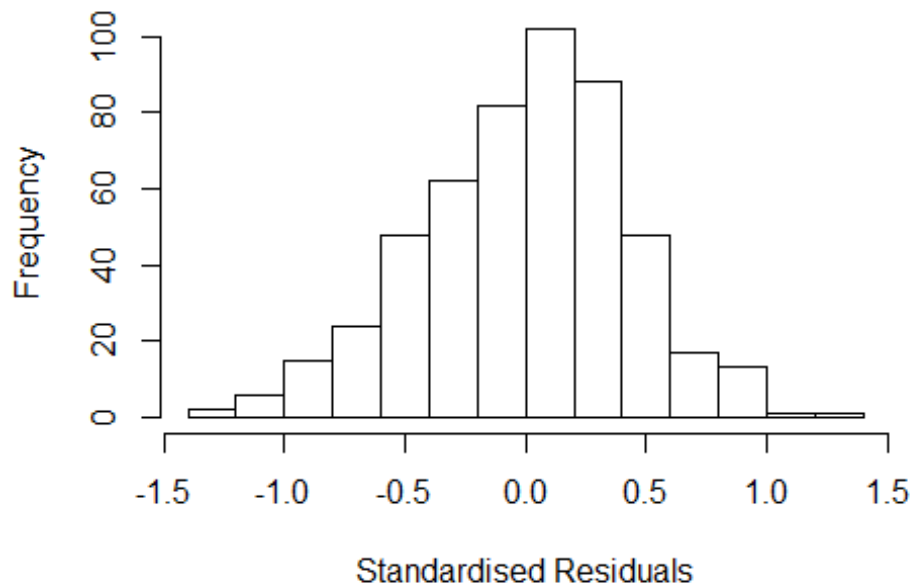
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 510, p-value = 0.7
## 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	1.270	1	1.127
LastAuthorFemale	1.277	1	1.130
UniqueAuthors	2.317	4	1.111
Year	3.361	16	1.039

Residuals from first and last author and team size



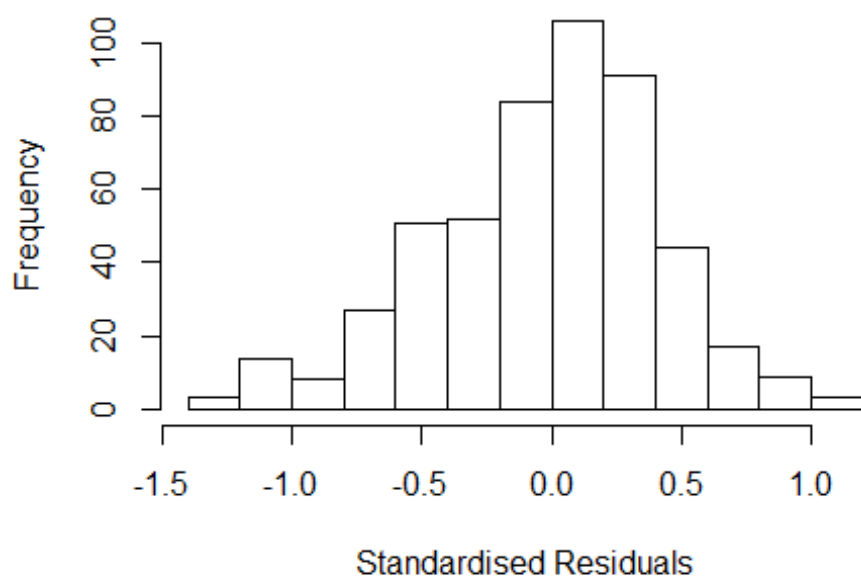
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3012 -0.2908  0.0273  0.2796  1.2592
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.89635    0.15386   5.83    1e-08 ***
## FirstAuthorFemale1 0.09073    0.04177   2.17    0.0303 *
## LastAuthorFemale1  0.02308    0.04605   0.50    0.6164
## UniqueAuthors2    0.22526    0.10743   2.10    0.0365 *
## UniqueAuthors3    0.25239    0.10147   2.49    0.0132 *
## UniqueAuthors4    0.24858    0.10443   2.38    0.0177 *
## UniqueAuthors5    0.29440    0.09967   2.95    0.0033 **
## Year1997          0.00443    0.17593   0.03    0.9799
## Year1998          0.17788    0.15403   1.15    0.2487
## Year1999          0.16343    0.19831   0.82    0.4103
```

```

## Year2000      0.12311      0.15740      0.78      0.4345
## Year2001      0.00393      0.16865      0.02      0.9814
## Year2002      0.01047      0.16143      0.06      0.9483
## Year2003     -0.07614      0.18000     -0.42      0.6725
## Year2004     -0.15635      0.14816     -1.06      0.2918
## Year2005      0.04060      0.15643      0.26      0.7953
## Year2006     -0.09702      0.15885     -0.61      0.5417
## Year2007      0.03869      0.15330      0.25      0.8009
## Year2008     -0.14266      0.14883     -0.96      0.3383
## Year2009     -0.00841      0.14059     -0.06      0.9523
## Year2010     -0.03885      0.14720     -0.26      0.7920
## Year2011     -0.05124      0.14796     -0.35      0.7293
## Year2012      0.03009      0.15251      0.20      0.8437
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.0842, Adjusted R-squared:  0.0427
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 470 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.290  0.863  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      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 1.313 1      1.146
## LastAuthorFemale  1.174 1      1.083
## Year              1.518 16      1.013

```


Residuals from first and last author



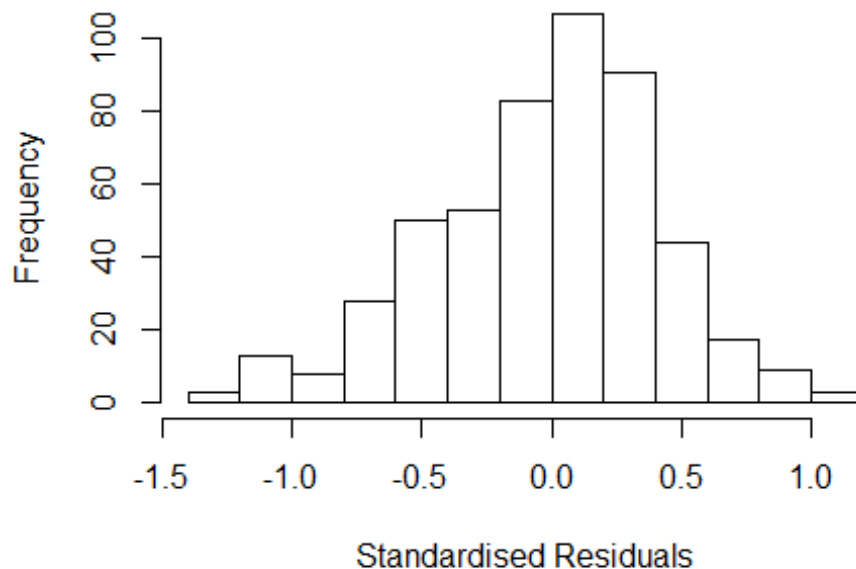
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2880 -0.2843 0.0372 0.2711 1.1799
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09735 0.12442 8.82 <2e-16 ***
## FirstAuthorFemale1 0.10020 0.04330 2.31 0.021 *
## LastAuthorFemale1 -0.00342 0.04573 -0.07 0.940
## Year1997 0.03760 0.16723 0.22 0.822
## Year1998 0.22267 0.14982 1.49 0.138
## Year1999 0.16519 0.18772 0.88 0.379
## Year2000 0.16441 0.15647 1.05 0.294
## Year2001 0.08086 0.15989 0.51 0.613
## Year2002 0.04566 0.15611 0.29 0.770
## Year2003 -0.01936 0.16274 -0.12 0.905
## Year2004 -0.12883 0.14471 -0.89 0.374
## Year2005 0.06750 0.14943 0.45 0.652
```

```

## Year2006      -0.10367    0.15452   -0.67    0.503
## Year2007      0.09390    0.14588    0.64    0.520
## Year2008     -0.08578    0.14105   -0.61    0.543
## Year2009      0.02848    0.13485    0.21    0.833
## Year2010      0.01177    0.13971    0.08    0.933
## Year2011      0.00593    0.13941    0.04    0.966
## Year2012      0.07873    0.14563    0.54    0.589
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.0477, Adjusted R-squared:  0.0128
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 468 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.290  0.858   0.946   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      1.96e-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 1.292 1         1.137
## Year              1.292 16         1.008

```

Residuals from first author



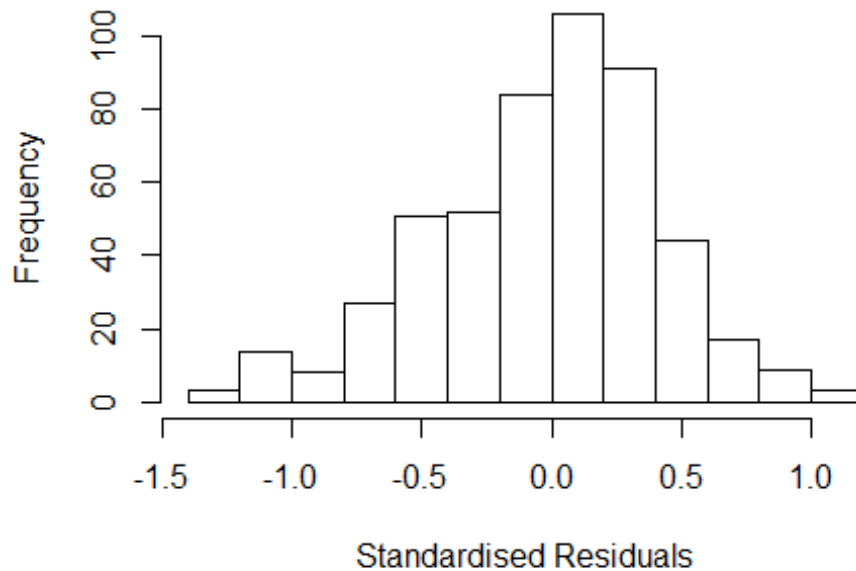
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2894 -0.2820 0.0376 0.2713 1.1807
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09616 0.12388 8.85 <2e-16 ***
## FirstAuthorFemale1 0.10010 0.04341 2.31 0.022 *
## Year1997 0.03799 0.16737 0.23 0.821
## Year1998 0.22176 0.14990 1.48 0.140
## Year1999 0.16548 0.18692 0.89 0.376
## Year2000 0.16368 0.15665 1.04 0.297
## Year2001 0.08124 0.15994 0.51 0.612
## Year2002 0.04566 0.15614 0.29 0.770
## Year2003 -0.01947 0.16263 -0.12 0.905
## Year2004 -0.12837 0.14392 -0.89 0.373
## Year2005 0.06786 0.14912 0.46 0.649
## Year2006 -0.10401 0.15434 -0.67 0.501
```

```

## Year2007          0.09317    0.14490    0.64    0.521
## Year2008         -0.08617    0.14102   -0.61    0.541
## Year2009          0.02843    0.13476    0.21    0.833
## Year2010          0.01145    0.13965    0.08    0.935
## Year2011          0.00517    0.13933    0.04    0.970
## Year2012          0.07774    0.14542    0.53    0.593
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.0475, Adjusted R-squared:  0.0145
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 468 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.300  0.861  0.948  0.895  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.163 1      1.079
## Year              1.163 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
## -1.2710 -0.2964 0.0388 0.2802 1.2390
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1070 0.1299 8.52 <2e-16 ***
## LastAuthorFemale1 0.0180 0.0454 0.40 0.69
## Year1997 0.0691 0.1760 0.39 0.69
## Year1998 0.2472 0.1551 1.59 0.11
## Year1999 0.1949 0.1842 1.06 0.29
## Year2000 0.1694 0.1627 1.04 0.30
## Year2001 0.0784 0.1625 0.48 0.63
## Year2002 0.0854 0.1593 0.54 0.59
## Year2003 0.0163 0.1698 0.10 0.92
## Year2004 -0.1003 0.1502 -0.67 0.50
## Year2005 0.0795 0.1539 0.52 0.61
## Year2006 -0.0635 0.1582 -0.40 0.69
```

```

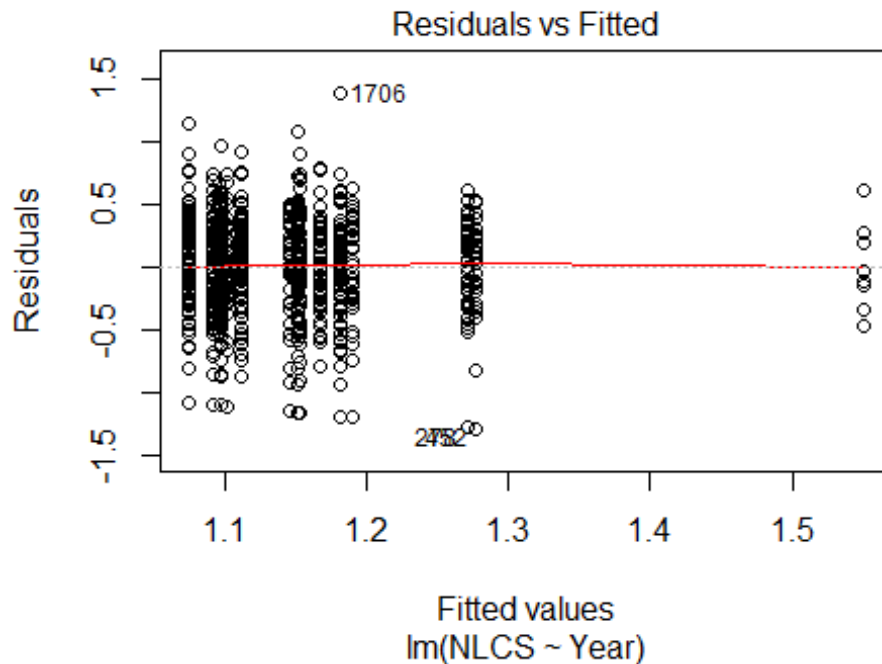
## Year2007          0.1460      0.1474      0.99      0.32
## Year2008          -0.0346     0.1458     -0.24     0.81
## Year2009          0.0645     0.1402     0.46     0.65
## Year2010          0.0608     0.1422     0.43     0.67
## Year2011          0.0563     0.1420     0.40     0.69
## Year2012          0.1315     0.1499     0.88     0.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.0364, Adjusted R-squared:  0.00303
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 465 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.304  0.862  0.947  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      1.96e-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 2744"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 2007
## 1
##
## 2007
## 1
##
## 2007
## 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: 1"
## [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
## 70 66 72 74 83 64 70 58 60 84 88 85 85 132 113
## 2011 2012
## 133 129
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 44 41 38 52 34 8 47 45 44 51 59 64 62 86 75
## 2011 2012
## 95 93
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 41 38 30 44 29 8 45 40 40 39 50 55 59 74 64
## 2011 2012
## 82 83
## [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

```



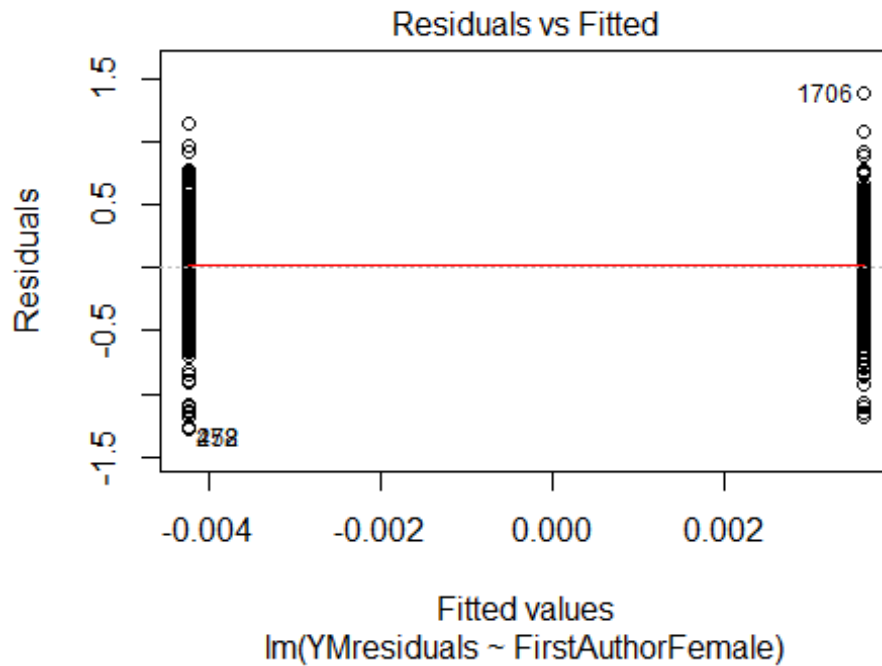
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.7, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 5.91932913704855"
## [1] "Male first author team size 2018 geometric mean: 5.22170306627186"

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

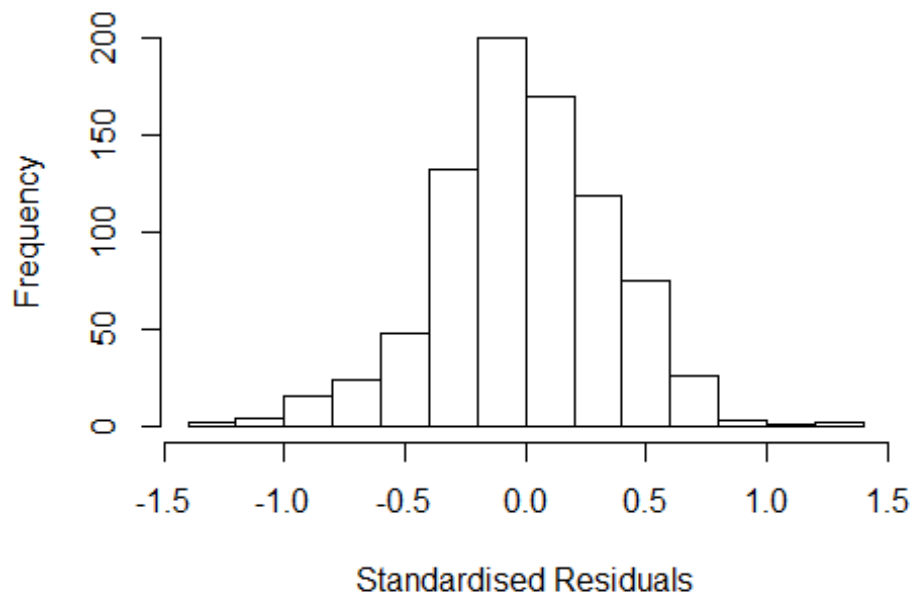
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 680, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.12149207180142"
## [1] "Male last author team size 2018 geometric mean: 6.08833894215941"

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

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 470, 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 1.114 1          1.056
## LastAuthorFemale  1.082 1          1.040
## UniqueAuthors    1.663 4          1.066
## Year              1.724 16         1.017
```

Residuals from first and last author and team size



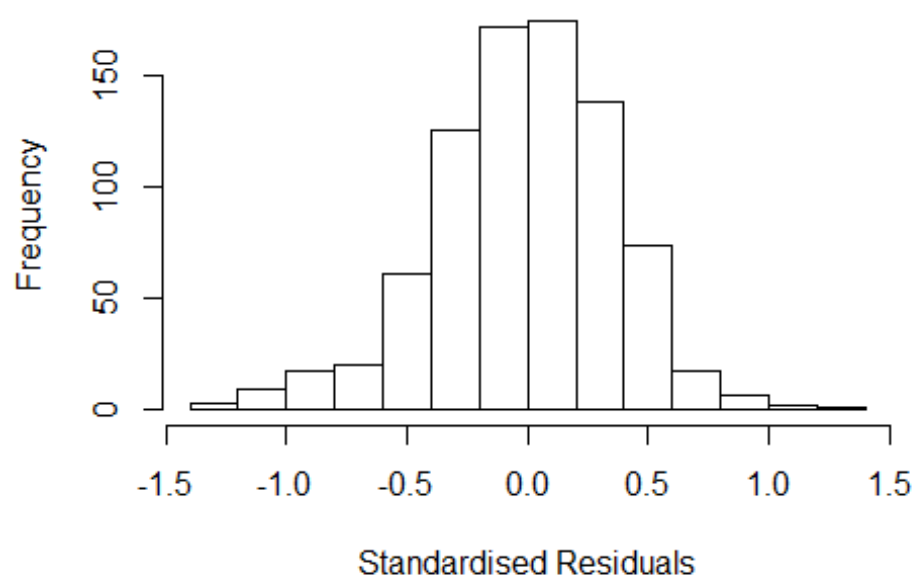
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2392 -0.2142 -0.0182 0.2259 1.3633
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00986 0.09828 10.28 < 2e-16 ***
## FirstAuthorFemale1 -0.02955 0.02587 -1.14 0.2538
## LastAuthorFemale1 0.07556 0.02637 2.87 0.0043 **
## UniqueAuthors2 0.06466 0.09385 0.69 0.4910
## UniqueAuthors3 0.14972 0.08747 1.71 0.0873 .
## UniqueAuthors4 0.24525 0.08216 2.99 0.0029 **
## UniqueAuthors5 0.32571 0.07979 4.08 4.9e-05 ***
## Year1997 -0.13645 0.09650 -1.41 0.1577
## Year1998 0.00441 0.09947 0.04 0.9646
## Year1999 0.01395 0.08560 0.16 0.8706
```

```

## Year2000          0.06757    0.08972    0.75    0.4516
## Year2001          0.21318    0.13226    1.61    0.1074
## Year2002         -0.17283    0.09119   -1.90    0.0584 .
## Year2003         -0.15934    0.08736   -1.82    0.0685 .
## Year2004         -0.08193    0.09543   -0.86    0.3908
## Year2005         -0.11800    0.08906   -1.32    0.1856
## Year2006         -0.05344    0.09198   -0.58    0.5614
## Year2007         -0.20441    0.08939   -2.29    0.0225 *
## Year2008         -0.10893    0.08585   -1.27    0.2049
## Year2009         -0.14409    0.08443   -1.71    0.0883 .
## Year2010         -0.18147    0.08725   -2.08    0.0378 *
## Year2011         -0.09641    0.08415   -1.15    0.2523
## Year2012         -0.22340    0.08562   -2.61    0.0092 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.128, Adjusted R-squared:  0.104
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 759 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0749 0.8700 0.9530 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          1.22e-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 1.084 1 1.041
## LastAuthorFemale 1.055 1 1.027
## Year 1.131 16 1.004

```

Residuals from first and last author



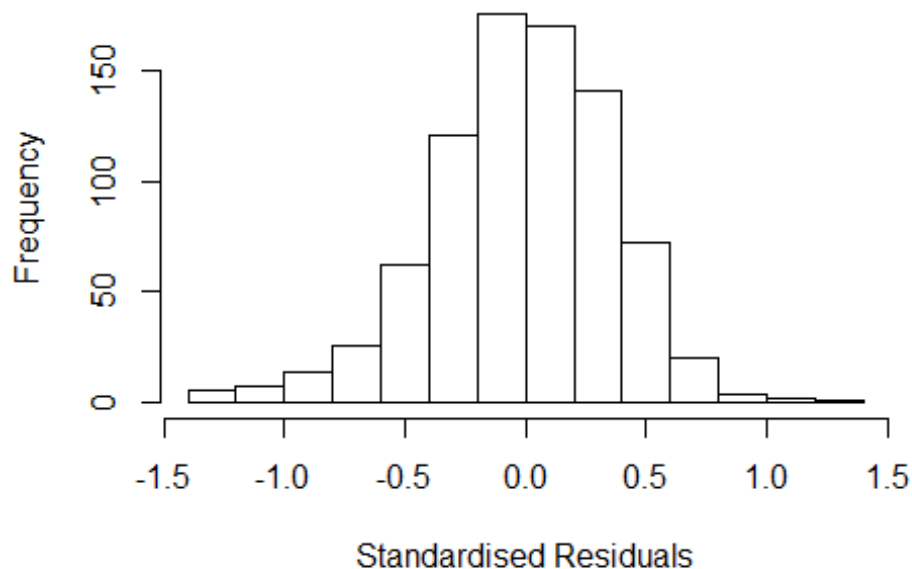
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31709 -0.24457 0.00602 0.23814 1.34902
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20078 0.07656 15.68 <2e-16 ***
## FirstAuthorFemale1 -0.01524 0.02641 -0.58 0.5641
## LastAuthorFemale1 0.07130 0.02704 2.64 0.0085 **
## Year1997 -0.11709 0.09747 -1.20 0.2300
## Year1998 0.03873 0.10538 0.37 0.7133
## Year1999 0.04501 0.08940 0.50 0.6148
## Year2000 0.11416 0.09145 1.25 0.2123
## Year2001 0.31118 0.13768 2.26 0.0241 *
## Year2002 -0.09247 0.09339 -0.99 0.3224
## Year2003 -0.12889 0.09035 -1.43 0.1541
## Year2004 -0.03942 0.09713 -0.41 0.6850
## Year2005 -0.05909 0.09353 -0.63 0.5277
```

```

## Year2006          -0.00221    0.09018   -0.02    0.9804
## Year2007          -0.12653    0.09000   -1.41    0.1601
## Year2008          -0.07768    0.08972   -0.87    0.3869
## Year2009          -0.11375    0.08707   -1.31    0.1918
## Year2010          -0.14393    0.09006   -1.60    0.1104
## Year2011          -0.03786    0.08562   -0.44    0.6585
## Year2012          -0.15780    0.08831   -1.79    0.0744 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.0535, Adjusted R-squared:  0.0323
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 755 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.121  0.877   0.952   0.901   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.22e-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 1.08 1          1.039
## Year              1.08 16          1.002

```

Residuals from first author



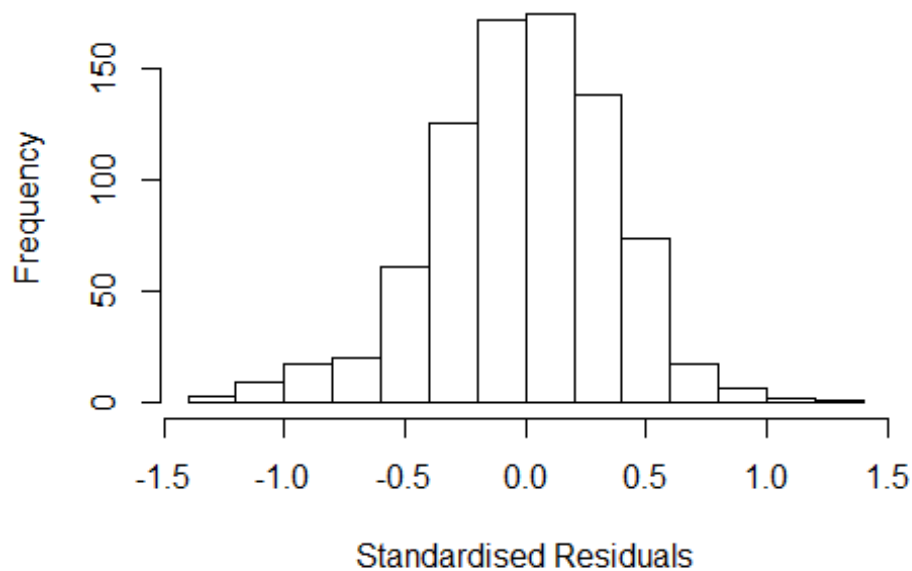
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.331924 -0.235396 -0.000374 0.240018 1.385519
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.216512 0.077590 15.68 <2e-16 ***
## FirstAuthorFemale1 -0.005528 0.026576 -0.21 0.835
## Year1997 -0.111486 0.099397 -1.12 0.262
## Year1998 0.042517 0.105225 0.40 0.686
## Year1999 0.040139 0.090573 0.44 0.658
## Year2000 0.115412 0.091006 1.27 0.205
## Year2001 0.323477 0.143900 2.25 0.025 *
## Year2002 -0.093020 0.094368 -0.99 0.325
## Year2003 -0.123129 0.090651 -1.36 0.175
## Year2004 -0.038920 0.099439 -0.39 0.696
## Year2005 -0.050063 0.094766 -0.53 0.597
## Year2006 -0.000138 0.091565 0.00 0.999
```

```

## Year2007          -0.118787    0.092172    -1.29    0.198
## Year2008          -0.083855    0.091411    -0.92    0.359
## Year2009          -0.110882    0.088104    -1.26    0.209
## Year2010          -0.132647    0.090806    -1.46    0.144
## Year2011          -0.028503    0.086405    -0.33    0.742
## Year2012          -0.150002    0.089871    -1.67    0.095 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.357
## Multiple R-squared:  0.0458, Adjusted R-squared:  0.0256
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 760 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0985 0.8770 0.9510 0.9010 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.22e-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 1.045 1      1.022
## Year      1.045 16      1.001

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30875 -0.24009 0.00344 0.23705 1.34475
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19595 0.07633 15.67 <2e-16 ***
## LastAuthorFemale1 0.06909 0.02700 2.56 0.011 *
## Year1997 -0.11782 0.09786 -1.20 0.229
## Year1998 0.03661 0.10541 0.35 0.728
## Year1999 0.04256 0.08965 0.47 0.635
## Year2000 0.11280 0.09181 1.23 0.220
## Year2001 0.31267 0.13872 2.25 0.024 *
## Year2002 -0.09365 0.09374 -1.00 0.318
## Year2003 -0.12934 0.09062 -1.43 0.154
## Year2004 -0.04430 0.09709 -0.46 0.648
## Year2005 -0.06181 0.09381 -0.66 0.510
## Year2006 -0.00689 0.09035 -0.08 0.939
```

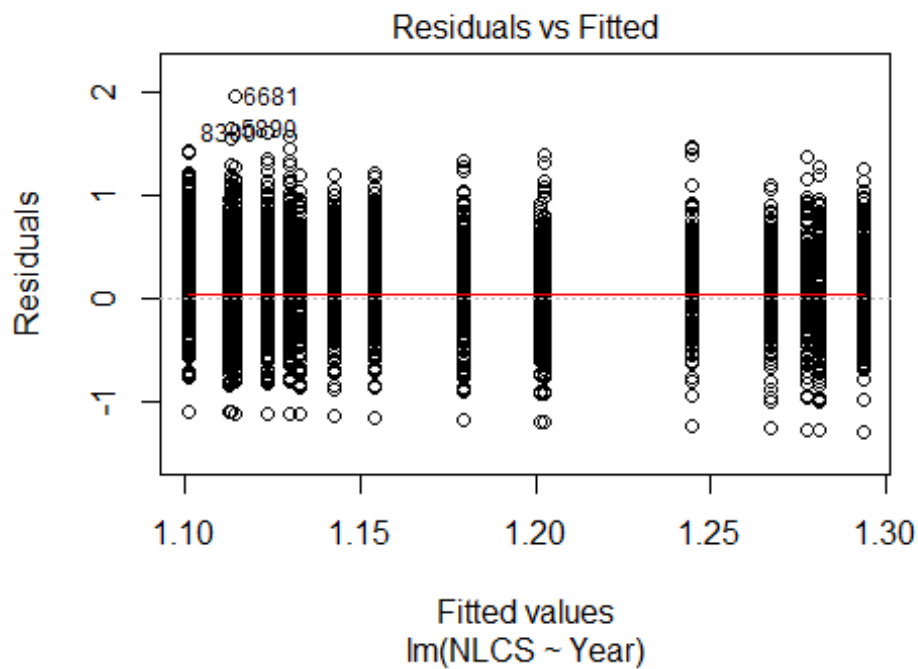


```

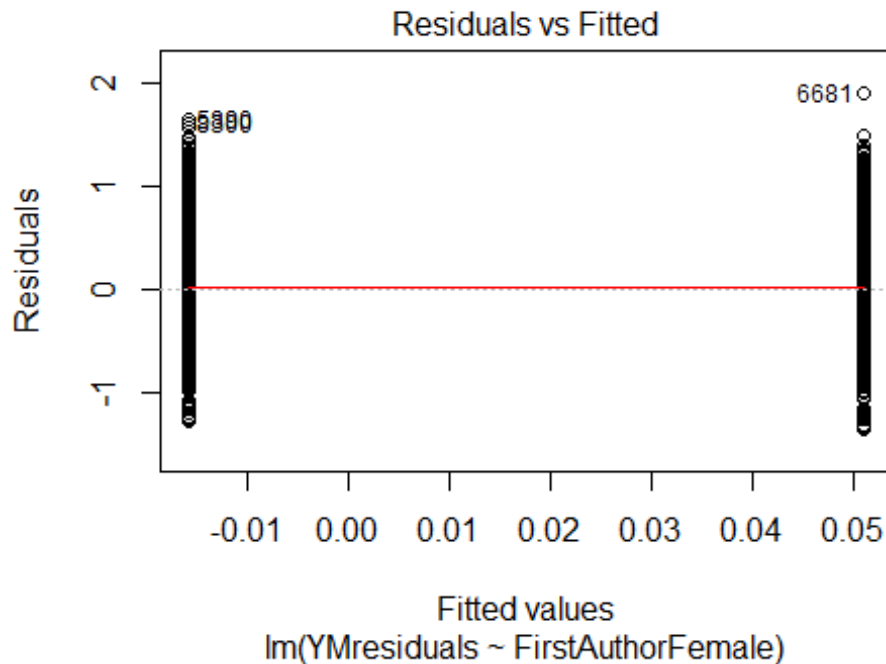
## Year2007          -0.12870      0.09008    -1.43      0.153
## Year2008          -0.08187      0.08954    -0.91      0.361
## Year2009          -0.11775      0.08718    -1.35      0.177
## Year2010          -0.14530      0.09045    -1.61      0.109
## Year2011          -0.04179      0.08574    -0.49      0.626
## Year2012          -0.16178      0.08820    -1.83      0.067 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.357
## Multiple R-squared:  0.0531, Adjusted R-squared:  0.033
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 757 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.124  0.877  0.950  0.901  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.22e-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: 821"
## [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
## 386 390 409 455 404 480 385 308 320 384 436 611 691 685 640
## 2011 2012
## 582 631
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 212 195 200 251 232 228 281 227 226 286 326 455 545 519 489
## 2011 2012

```

```
## 436 484
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 174 151 176 209 197 195 236 191 192 238 279 410 473 455 429
## 2011 2012
## 382 407
## [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
```

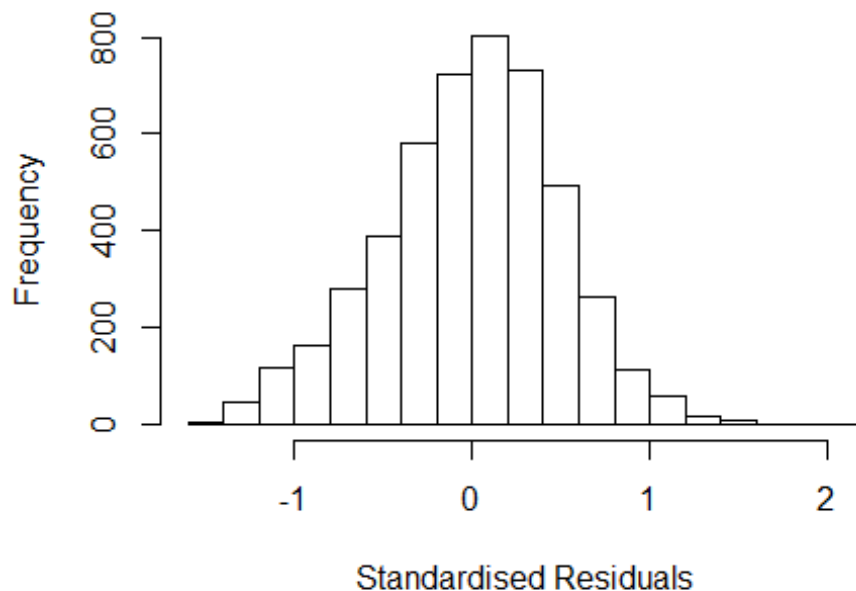


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.1, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.74306075687368"
## [1] "Male first author team size 2018 geometric mean: 4.59423834282178"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 34000, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.54133846712606"
## [1] "Male last author team size 2018 geometric mean: 4.69294003524867"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 28000, 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 1.053 1          1.026
## LastAuthorFemale  1.042 1          1.021
## UniqueAuthors    1.141 4          1.017
## Year              1.166 16         1.005
```

Residuals from first and last author and team size



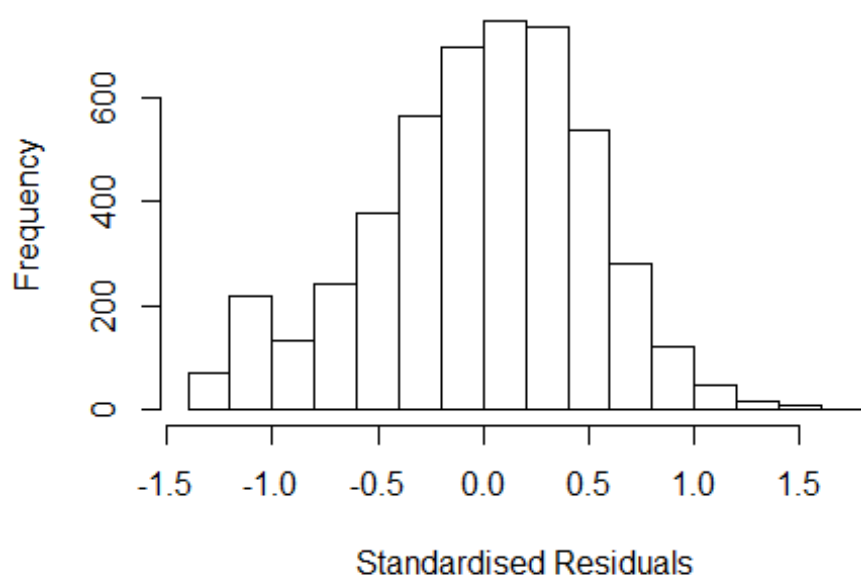
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5035 -0.3280 0.0213 0.3379 2.0288
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94577 0.04723 20.02 < 2e-16 ***
## FirstAuthorFemale1 0.05552 0.01703 3.26 0.00112 **
## LastAuthorFemale1 0.00765 0.01970 0.39 0.69797
## UniqueAuthors2 0.24976 0.03999 6.25 4.6e-10 ***
## UniqueAuthors3 0.29160 0.03883 7.51 7.1e-14 ***
## UniqueAuthors4 0.39086 0.03814 10.25 < 2e-16 ***
## UniqueAuthors5 0.53649 0.03612 14.85 < 2e-16 ***
## Year1997 -0.01973 0.06036 -0.33 0.74381
## Year1998 -0.04153 0.05271 -0.79 0.43073
## Year1999 0.01362 0.04928 0.28 0.78220
```

```

## Year2000      -0.02830    0.04728   -0.60  0.54949
## Year2001      -0.10551    0.04664   -2.26  0.02372 *
## Year2002      -0.13910    0.04703   -2.96  0.00311 **
## Year2003      -0.16565    0.04868   -3.40  0.00067 ***
## Year2004      -0.22033    0.04902   -4.49  7.1e-06 ***
## Year2005      -0.16879    0.04719   -3.58  0.00035 ***
## Year2006      -0.12726    0.04883   -2.61  0.00919 **
## Year2007      -0.19486    0.04416   -4.41  1.0e-05 ***
## Year2008      -0.22356    0.04293   -5.21  2.0e-07 ***
## Year2009      -0.23470    0.04343   -5.40  6.8e-08 ***
## Year2010      -0.21910    0.04368   -5.02  5.5e-07 ***
## Year2011      -0.24319    0.04458   -5.46  5.1e-08 ***
## Year2012      -0.25607    0.04531   -5.65  1.7e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.106, Adjusted R-squared:  0.102
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 377 weights are ~= 1. The remaining 4417 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.041  0.869  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          2.09e-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 1.040 1          1.020
## LastAuthorFemale 1.028 1          1.014
## Year 1.033 16          1.001

```

Residuals from first and last author

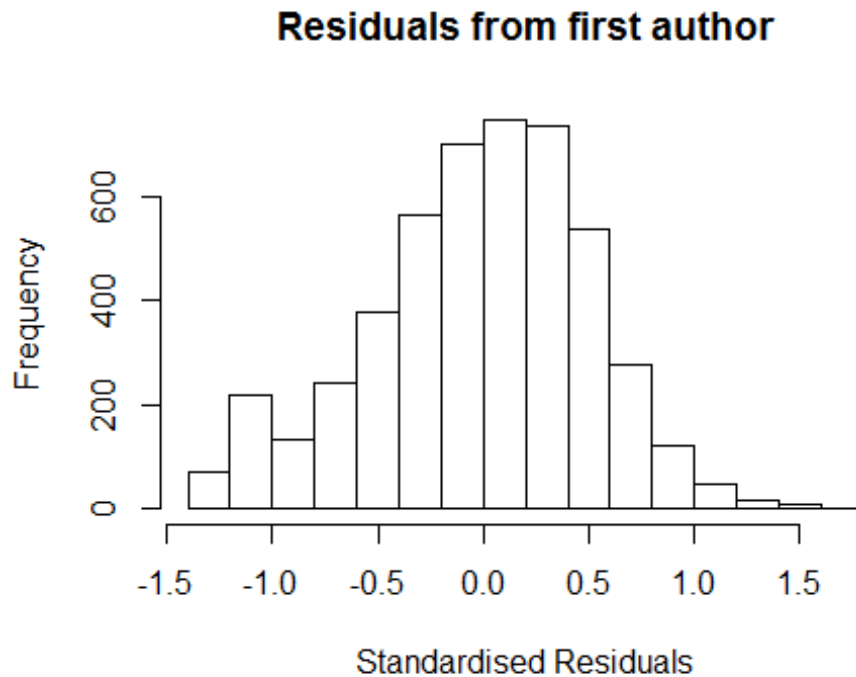


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3426 -0.3382 0.0259 0.3440 1.6303
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26578 0.03849 32.88 < 2e-16 ***
## FirstAuthorFemale1 0.06762 0.01749 3.87 0.00011 ***
## LastAuthorFemale1 -0.00232 0.02035 -0.11 0.90922
## Year1997 0.00916 0.06452 0.14 0.88705
## Year1998 -0.01813 0.05602 -0.32 0.74624
## Year1999 0.00494 0.05090 0.10 0.92265
## Year2000 0.00933 0.04962 0.19 0.85083
## Year2001 -0.06129 0.05063 -1.21 0.22608
## Year2002 -0.08986 0.05076 -1.77 0.07677 .
## Year2003 -0.13242 0.05190 -2.55 0.01075 *
## Year2004 -0.16307 0.05338 -3.05 0.00226 **
## Year2005 -0.09154 0.04989 -1.83 0.06659 .
```

```

## Year2006      -0.04659    0.05075   -0.92  0.35860
## Year2007      -0.11490    0.04679   -2.46  0.01410 *
## Year2008      -0.14510    0.04549   -3.19  0.00143 **
## Year2009      -0.15986    0.04614   -3.46  0.00054 ***
## Year2010      -0.13264    0.04590   -2.89  0.00387 **
## Year2011      -0.15512    0.04733   -3.28  0.00105 **
## Year2012      -0.16474    0.04825   -3.41  0.00064 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.0156, Adjusted R-squared:  0.0119
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 396 weights are ~= 1. The remaining 4398 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.262  0.868  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      2.09e-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      1.012
## Year      1.023 16      1.001

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3421 -0.3379 0.0253 0.3440 1.6307
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26557 0.03840 32.95 < 2e-16 ***
## FirstAuthorFemale1 0.06728 0.01737 3.87 0.00011 ***
## Year1997 0.00909 0.06451 0.14 0.88799
## Year1998 -0.01816 0.05602 -0.32 0.74578
## Year1999 0.00490 0.05091 0.10 0.92332
## Year2000 0.00928 0.04962 0.19 0.85160
## Year2001 -0.06141 0.05063 -1.21 0.22522
## Year2002 -0.08989 0.05077 -1.77 0.07671 .
## Year2003 -0.13242 0.05190 -2.55 0.01076 *
## Year2004 -0.16315 0.05338 -3.06 0.00225 **
## Year2005 -0.09171 0.04988 -1.84 0.06603 .
## Year2006 -0.04665 0.05077 -0.92 0.35820
```

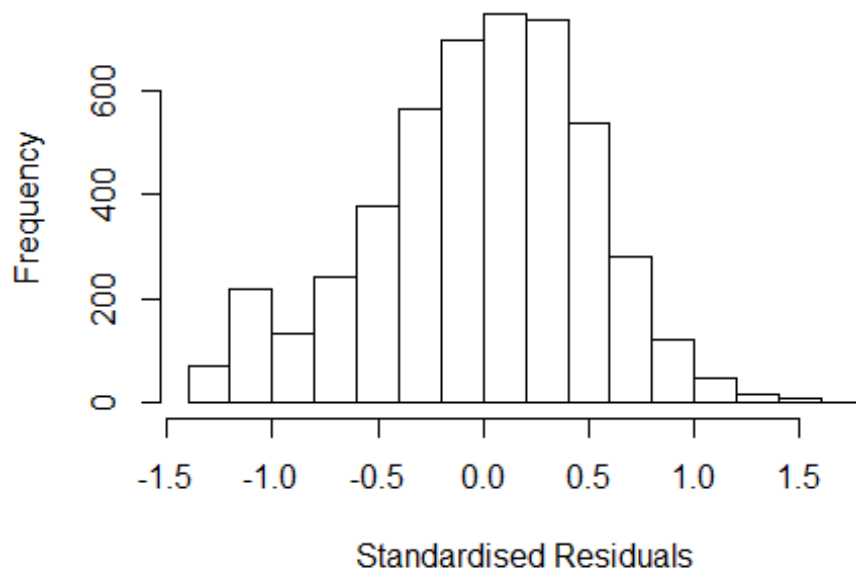


```

## Year2007          -0.11496    0.04681    -2.46    0.01409 *
## Year2008          -0.14525    0.04549    -3.19    0.00142 **
## Year2009          -0.15990    0.04614    -3.47    0.00053 ***
## Year2010          -0.13270    0.04590    -2.89    0.00386 **
## Year2011          -0.15523    0.04732    -3.28    0.00104 **
## Year2012          -0.16481    0.04826    -3.42    0.00064 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.0156, Adjusted R-squared:  0.0121
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 395 weights are ~= 1. The remaining 4399 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.261  0.868  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      2.09e-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 1.011 1          1.006
## Year            1.011 16          1.000

```

Residuals from last author



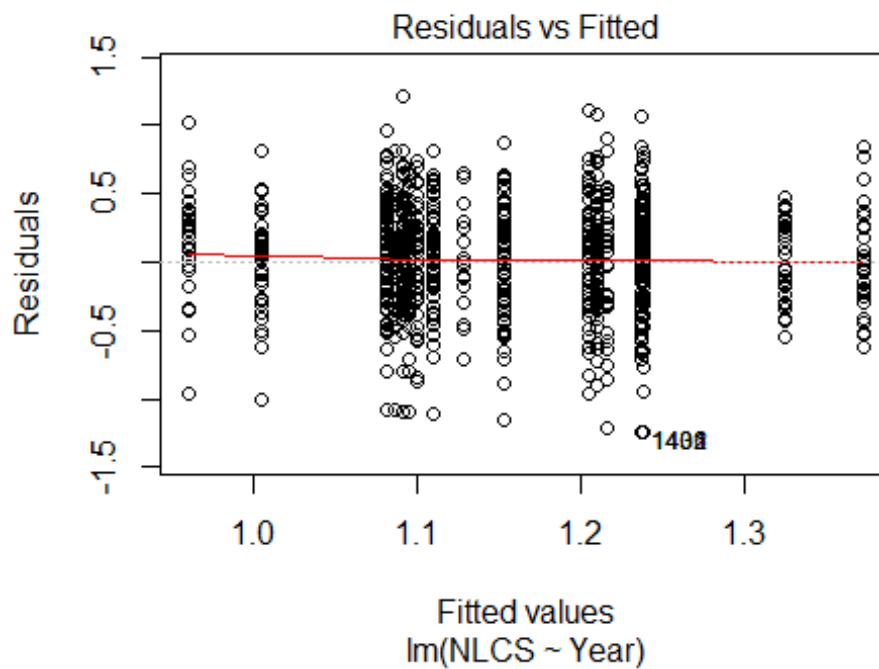
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3003 -0.3393 0.0252 0.3442 1.6166
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27779 0.03855 33.14 < 2e-16 ***
## LastAuthorFemale1 0.01099 0.02012 0.55 0.58480
## Year1997 0.00378 0.06449 0.06 0.95331
## Year1998 -0.02199 0.05624 -0.39 0.69579
## Year1999 0.00463 0.05117 0.09 0.92791
## Year2000 0.01152 0.04983 0.23 0.81715
## Year2001 -0.06354 0.05085 -1.25 0.21158
## Year2002 -0.09100 0.05093 -1.79 0.07404 .
## Year2003 -0.13185 0.05203 -2.53 0.01130 *
## Year2004 -0.16154 0.05365 -3.01 0.00262 **
## Year2005 -0.08944 0.05009 -1.79 0.07421 .
## Year2006 -0.04610 0.05094 -0.90 0.36554
```

```

## Year2007          -0.11145      0.04700      -2.37   0.01777 *
## Year2008          -0.14339      0.04577      -3.13   0.00174 **
## Year2009          -0.15582      0.04640      -3.36   0.00079 ***
## Year2010          -0.12772      0.04612      -2.77   0.00564 **
## Year2011          -0.14635      0.04744      -3.08   0.00205 **
## Year2012          -0.16027      0.04844      -3.31   0.00094 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.0126, Adjusted R-squared:  0.00905
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 365 weights are ~= 1. The remaining 4429 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.271  0.866  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      2.09e-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: 4794"
## [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
##   79   65   59   68   65   99   87   65   70   98   95  118  141  135  108
## 2011 2012
##  118  107
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   28   22   32   34   36   55   47   45   40   62   57   74   74   78   73
## 2011 2012

```

```
## 75 66
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 24 19 28 25 29 49 39 36 32 56 49 67 64 66 59
## 2011 2012
## 64 55
## [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
```



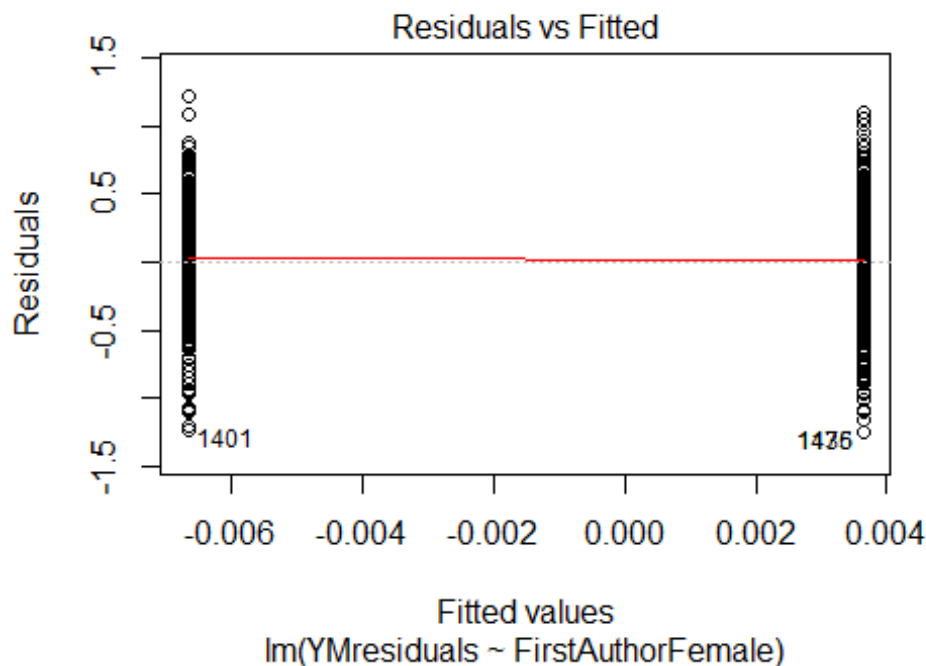
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 6.67442627456796"
## [1] "Male first author team size 2018 geometric mean: 6.36468438084797"

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

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 470, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.89180787860675"
## [1] "Male last author team size 2018 geometric mean: 6.37073198407779"

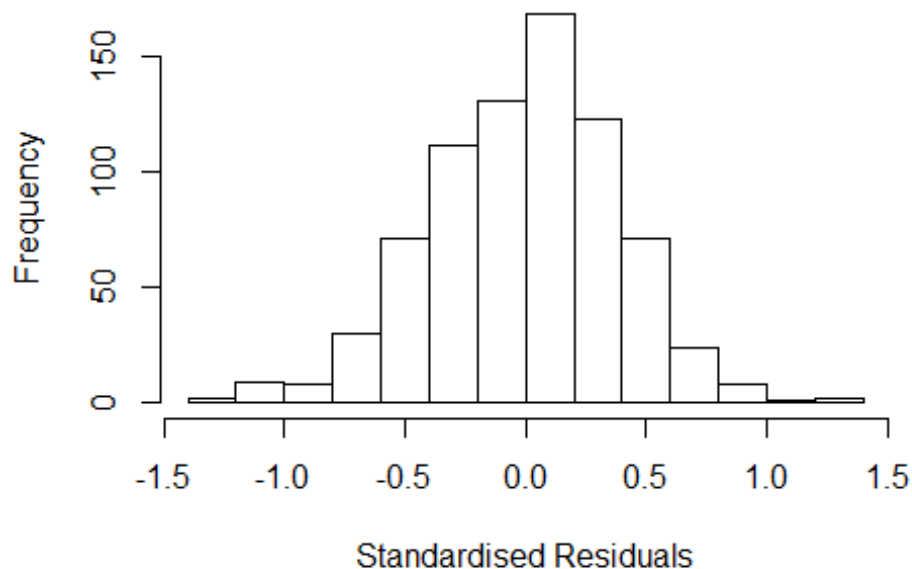
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 400, p-value = 0.9
## 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	1.137	1	1.066
LastAuthorFemale	1.136	1	1.066
UniqueAuthors	1.682	4	1.067
Year	1.741	16	1.017

Residuals from first and last author and team size



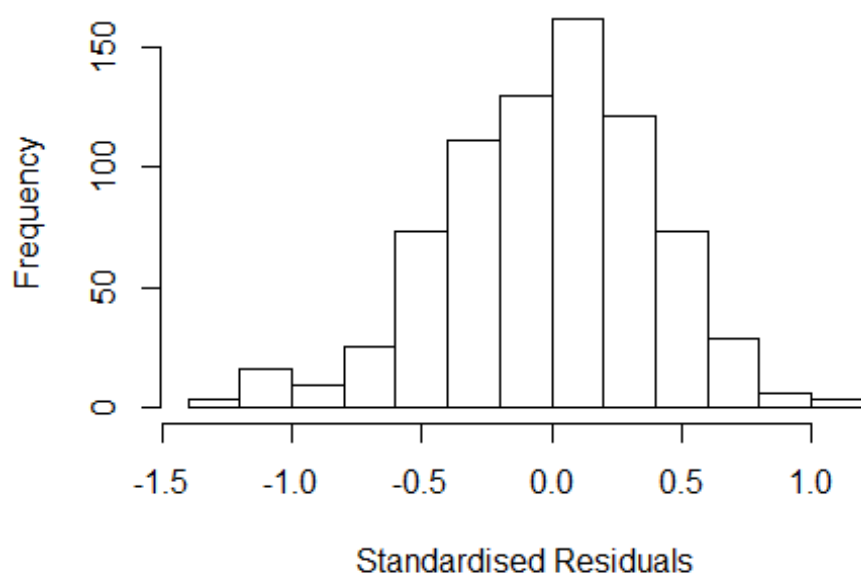
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3281 -0.2673  0.0217  0.2584  1.3945
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0735     0.1057  10.16 < 2e-16 ***
## FirstAuthorFemale1  0.0120     0.0319   0.38  0.70743
## LastAuthorFemale1 -0.0167     0.0370  -0.45  0.65193
## UniqueAuthors2     0.2171     0.1088   2.00  0.04638 *
## UniqueAuthors3     0.1501     0.1001   1.50  0.13404
## UniqueAuthors4     0.1786     0.0941   1.90  0.05795 .
## UniqueAuthors5     0.3248     0.0882   3.68  0.00025 ***
## Year1997          -0.2379     0.1062  -2.24  0.02532 *
## Year1998          -0.2908     0.1174  -2.48  0.01351 *
## Year1999          -0.1264     0.1333  -0.95  0.34328
```

```

## Year2000          0.0240      0.0978      0.25  0.80602
## Year2001         -0.1589      0.1007     -1.58  0.11504
## Year2002         -0.2095      0.0979     -2.14  0.03267 *
## Year2003         -0.3166      0.0971     -3.26  0.00116 **
## Year2004         -0.2018      0.1012     -1.99  0.04657 *
## Year2005         -0.2170      0.0898     -2.42  0.01588 *
## Year2006         -0.1829      0.0966     -1.89  0.05879 .
## Year2007         -0.1719      0.0899     -1.91  0.05634 .
## Year2008         -0.1180      0.0926     -1.27  0.20283
## Year2009         -0.0822      0.0943     -0.87  0.38361
## Year2010         -0.1575      0.0919     -1.71  0.08714 .
## Year2011         -0.2547      0.0975     -2.61  0.00920 **
## Year2012         -0.2434      0.0982     -2.48  0.01341 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.0944, Adjusted R-squared:  0.0674
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 696 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.171  0.875  0.950  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      1.31e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.073 1      1.036
## LastAuthorFemale  1.107 1      1.052
## 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
## -1.2943 -0.2721  0.0116  0.2637  1.1790
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3060     0.0776   16.82  <2e-16 ***
## FirstAuthorFemale1  0.0207     0.0319    0.65   0.5163
## LastAuthorFemale1 -0.0216     0.0385   -0.56   0.5743
## Year1997          -0.2323     0.1051   -2.21   0.0274 *
## Year1998          -0.2661     0.1167   -2.28   0.0229 *
## Year1999          -0.1015     0.1320   -0.77   0.4421
## Year2000           0.0346     0.0976    0.35   0.7234
## Year2001          -0.1588     0.0994   -1.60   0.1106
## Year2002          -0.1750     0.0999   -1.75   0.0801 .
## Year2003          -0.2963     0.0972   -3.05   0.0024 **
## Year2004          -0.1915     0.1005   -1.91   0.0569 .
## Year2005          -0.2083     0.0906   -2.30   0.0217 *
```

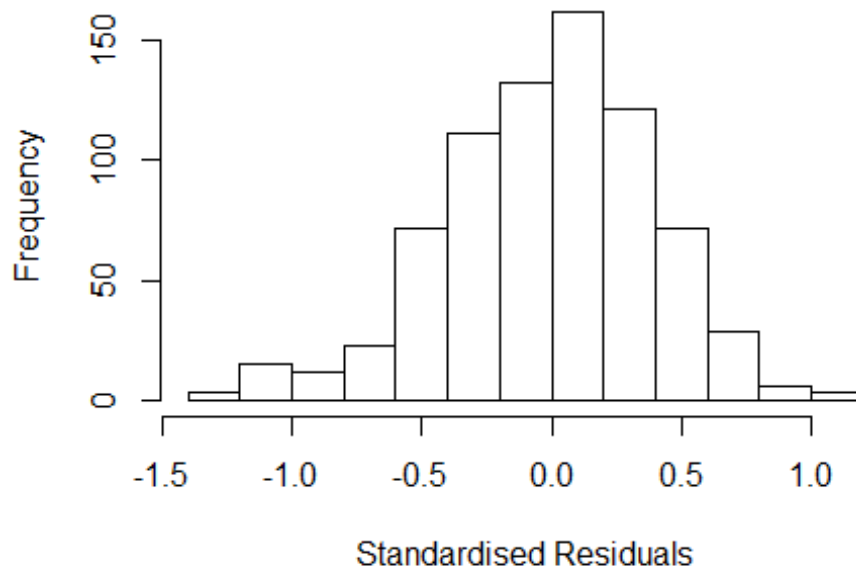


```

## Year2006          -0.1649      0.0942    -1.75    0.0804 .
## Year2007          -0.1282      0.0911    -1.41    0.1596
## Year2008          -0.0798      0.0925    -0.86    0.3890
## Year2009          -0.0324      0.0936    -0.35    0.7293
## Year2010          -0.1150      0.0920    -1.25    0.2119
## Year2011          -0.2047      0.0973    -2.10    0.0358 *
## Year2012          -0.2037      0.0961    -2.12    0.0344 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.0397, Adjusted R-squared:  0.0164
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 697 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.259  0.877   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      1.31e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.046 1          1.023
## Year              1.046 16          1.001

```

Residuals from first author



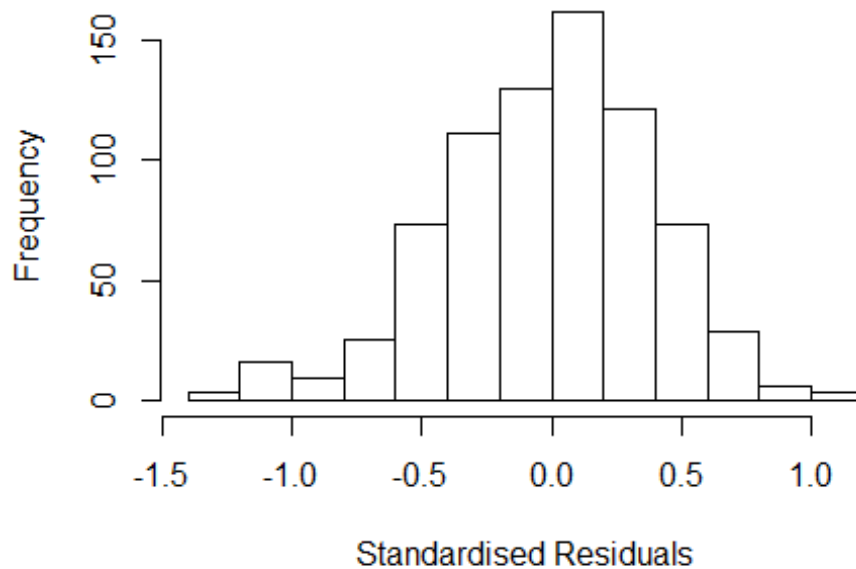
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2871 -0.2769 0.0111 0.2641 1.1860
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3042 0.0768 16.98 <2e-16 ***
## FirstAuthorFemale1 0.0172 0.0315 0.55 0.5855
## Year1997 -0.2312 0.1046 -2.21 0.0273 *
## Year1998 -0.2645 0.1159 -2.28 0.0228 *
## Year1999 -0.1032 0.1309 -0.79 0.4307
## Year2000 0.0366 0.0969 0.38 0.7059
## Year2001 -0.1587 0.0988 -1.61 0.1087
## Year2002 -0.1754 0.0992 -1.77 0.0775 .
## Year2003 -0.2945 0.0964 -3.06 0.0023 **
## Year2004 -0.1921 0.0999 -1.92 0.0548 .
## Year2005 -0.2103 0.0896 -2.35 0.0192 *
## Year2006 -0.1674 0.0932 -1.80 0.0730 .
```

```

## Year2007          -0.1281      0.0904    -1.42    0.1569
## Year2008          -0.0829      0.0918    -0.90    0.3672
## Year2009          -0.0343      0.0927    -0.37    0.7116
## Year2010          -0.1192      0.0908    -1.31    0.1898
## Year2011          -0.2076      0.0970    -2.14    0.0326 *
## Year2012          -0.2054      0.0956    -2.15    0.0320 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.0392, Adjusted R-squared:  0.0173
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 699 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.263  0.878  0.948  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      1.31e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.076 1          1.037
## Year            1.076 16          1.002

```

Residuals from last author



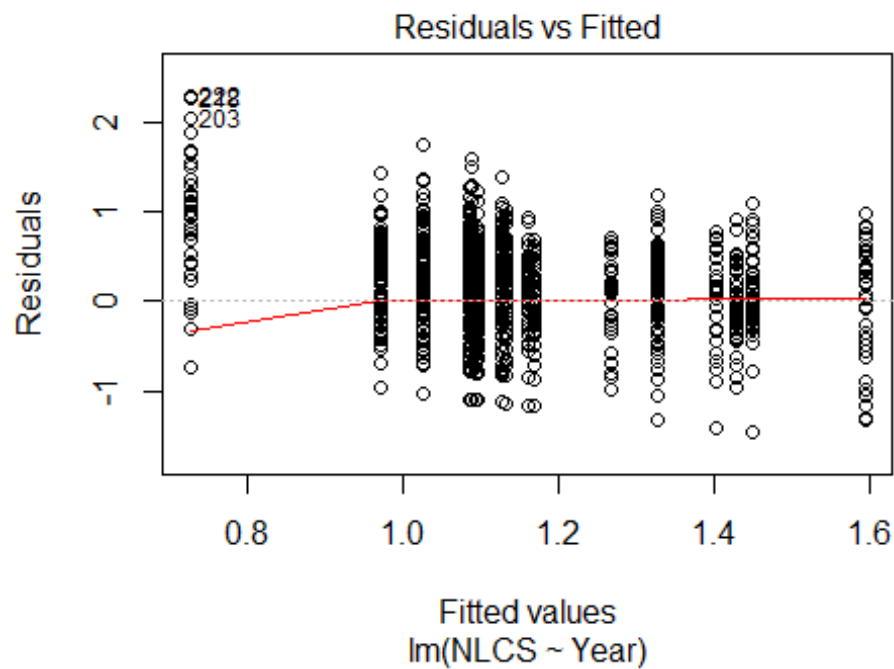
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2778 -0.2718 0.0178 0.2633 1.1933
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3116 0.0767 17.09 <2e-16 ***
## LastAuthorFemale1 -0.0169 0.0379 -0.45 0.6556
## Year1997 -0.2296 0.1049 -2.19 0.0290 *
## Year1998 -0.2659 0.1171 -2.27 0.0235 *
## Year1999 -0.0998 0.1313 -0.76 0.4472
## Year2000 0.0348 0.0980 0.36 0.7225
## Year2001 -0.1581 0.0995 -1.59 0.1126
## Year2002 -0.1766 0.0996 -1.77 0.0765 .
## Year2003 -0.2955 0.0973 -3.04 0.0025 **
## Year2004 -0.1911 0.1006 -1.90 0.0579 .
## Year2005 -0.2090 0.0906 -2.31 0.0212 *
## Year2006 -0.1640 0.0943 -1.74 0.0824 .
```

```

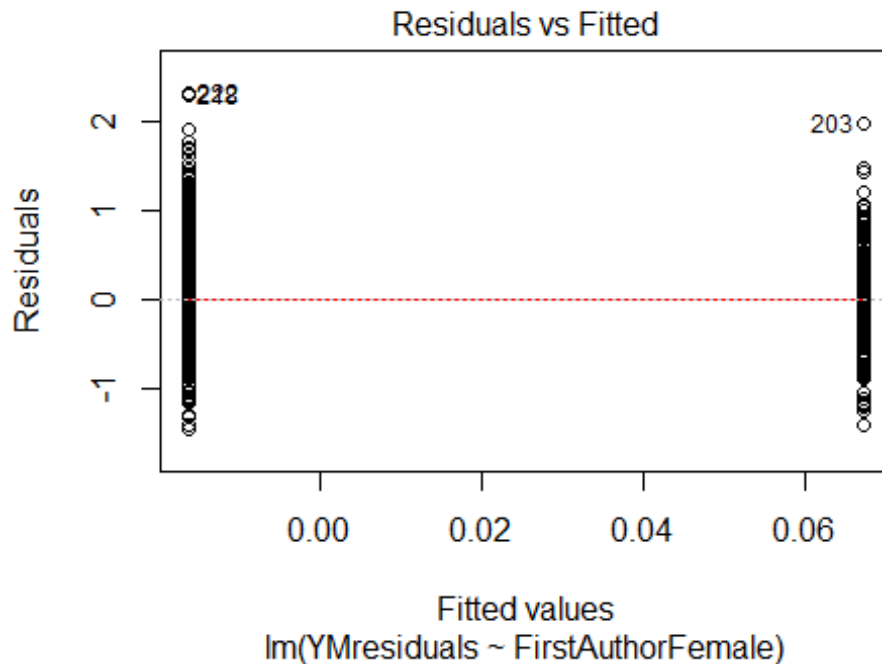
## Year2007          -0.1275      0.0909    -1.40    0.1610
## Year2008          -0.0780      0.0926    -0.84    0.3994
## Year2009          -0.0338      0.0937    -0.36    0.7182
## Year2010          -0.1160      0.0919    -1.26    0.2073
## Year2011          -0.2039      0.0972    -2.10    0.0362 *
## Year2012          -0.2029      0.0961    -2.11    0.0350 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.039, Adjusted R-squared:  0.017
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 700 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.279  0.881  0.948  0.906  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.31e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 761"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
##   81  136   90   69  108   86  116   78  107   84  123  149  147  160  201
## 2011 2012
##   203  179
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   35   80   51   26   43   35   74   45   71   61   88  109  116  130  162
## 2011 2012

```

```
## 168 149
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 31 73 45 25 39 32 70 40 64 54 80 100 110 124 147
## 2011 2012
## 143 136
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 100, df = 16, p-value = 4e-14
```

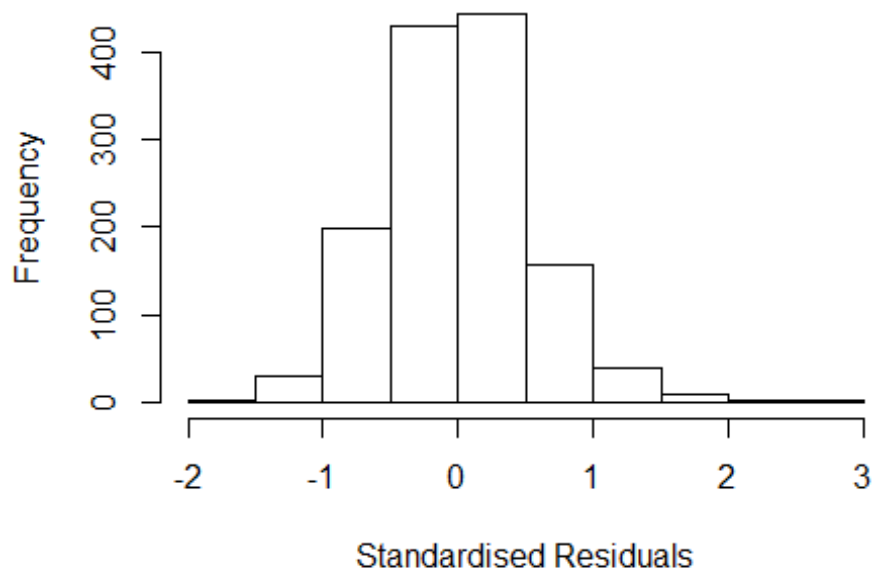


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.8, df = 1, p-value = 0.09
```



```
## [1] "Female first author team size 2018 geometric mean: 5.3264628618059"
## [1] "Male first author team size 2018 geometric mean: 6.05476558684364"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.75684471223238"
## [1] "Male last author team size 2018 geometric mean: 6.10772549169103"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.007
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.140 1          1.068
## LastAuthorFemale  1.111 1          1.054
## UniqueAuthors     1.387 4          1.042
## Year              1.425 16         1.011
```

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
## 218 0031022767 2.995 1997      2748      1      2.683
## 222 0031033534 3.011 1997      2748      1      2.508
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.56148 -0.32593 -0.00205  0.34577  2.68314
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1502     0.1517   7.58 6.4e-14 ***
## FirstAuthorFemale1  0.0341     0.0375   0.91 0.36328
## LastAuthorFemale1  0.0340     0.0410   0.83 0.40632
## UniqueAuthors2     0.3440     0.0702   4.90 1.1e-06 ***
## UniqueAuthors3     0.4523     0.0629   7.19 1.1e-12 ***
## UniqueAuthors4     0.5231     0.0602   8.69 < 2e-16 ***
## UniqueAuthors5     0.7142     0.0516  13.83 < 2e-16 ***
## Year1997          -1.3614     0.1993  -6.83 1.3e-11 ***
## Year1998           -0.1632     0.1613  -1.01 0.31177
```

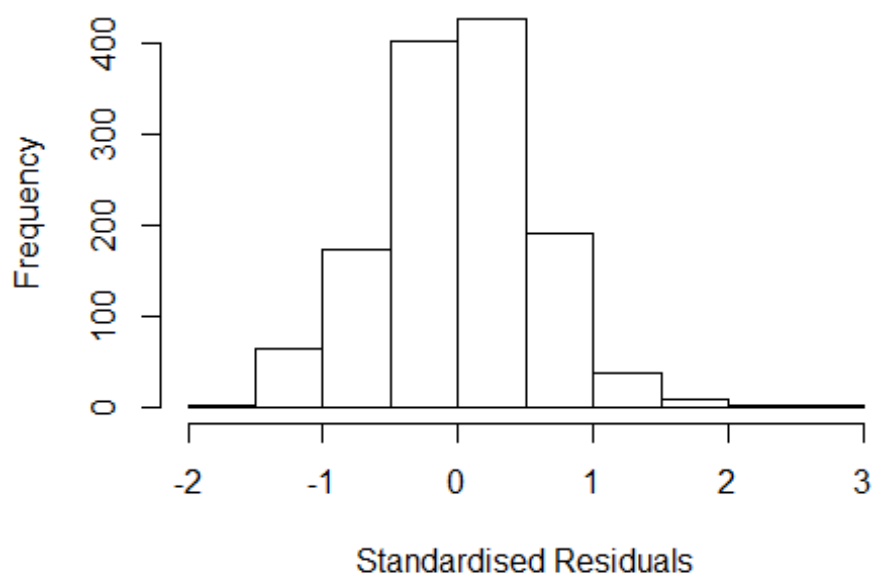


```

## Year1999          -0.2653      0.1760    -1.51    0.13190
## Year2000          -0.3352      0.1573    -2.13    0.03326 *
## Year2001          -0.4277      0.1714    -2.50    0.01267 *
## Year2002          -0.3235      0.1565    -2.07    0.03898 *
## Year2003          -0.5355      0.1591    -3.37    0.00079 ***
## Year2004          -0.6200      0.1569    -3.95    8.2e-05 ***
## Year2005          -0.5572      0.1557    -3.58    0.00036 ***
## Year2006          -0.5500      0.1600    -3.44    0.00060 ***
## Year2007          -0.6334      0.1550    -4.09    4.6e-05 ***
## Year2008          -0.5665      0.1530    -3.70    0.00022 ***
## Year2009          -0.5613      0.1556    -3.61    0.00032 ***
## Year2010          -0.6597      0.1531    -4.31    1.8e-05 ***
## Year2011          -0.6042      0.1536    -3.93    8.8e-05 ***
## Year2012          -0.5495      0.1528    -3.60    0.00034 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.29,   Adjusted R-squared:  0.277
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 4 observations c(33,94,100,101) are outliers with |weight| = 0 ( < 7.6e-
05);
## 106 weights are ~= 1. The remaining 1203 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0717 0.8650 0.9480 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          7.62e-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 1.104 1          1.051
## LastAuthorFemale 1.089 1          1.044
## Year 1.112 16          1.003

```

Residuals from first and last author



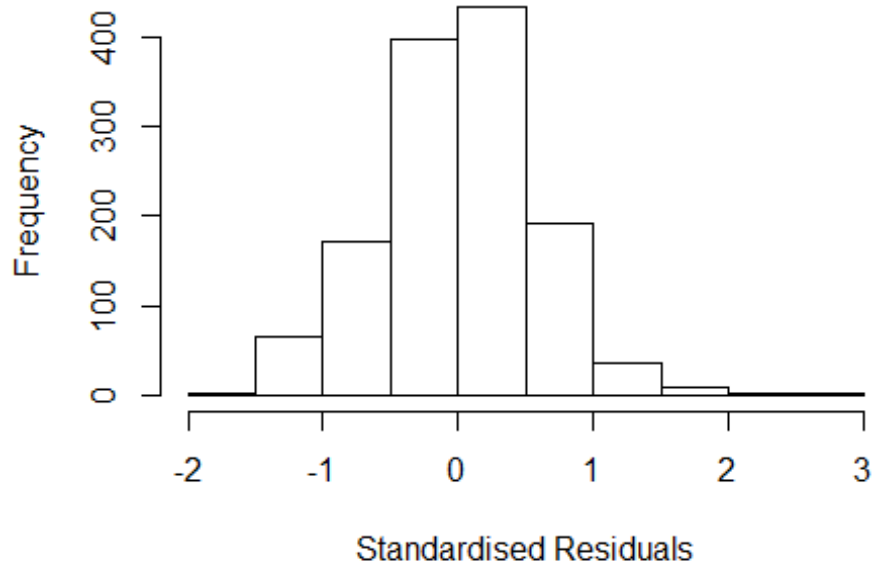
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 218 0031022767 2.995 1997    2748      1    2.609
## 222 0031033534 3.011 1997    2748      1    2.625
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5419 -0.3819  0.0171  0.3923  2.6249
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.6975     0.1312   12.94 < 2e-16 ***
## FirstAuthorFemale1  0.0827     0.0384    2.15  0.0317 *
## LastAuthorFemale1  0.0328     0.0419    0.78  0.4340
## Year1997        -1.3114     0.1835   -7.14 1.5e-12 ***
## Year1998         -0.2172     0.1447   -1.50  0.1335
## Year1999         -0.3127     0.1703   -1.84  0.0664 .
## Year2000         -0.2983     0.1462   -2.04  0.0415 *
## Year2001         -0.4656     0.1563   -2.98  0.0030 **
## Year2002         -0.3311     0.1433   -2.31  0.0210 *
## Year2003         -0.5358     0.1478   -3.63  0.0003 ***
## Year2004         -0.6269     0.1432   -4.38 1.3e-05 ***
```

```

## Year2005          -0.5627      0.1404    -4.01    6.5e-05 ***
## Year2006          -0.5692      0.1463    -3.89     0.0001 ***
## Year2007          -0.7266      0.1480    -4.91    1.0e-06 ***
## Year2008          -0.5984      0.1416    -4.23    2.6e-05 ***
## Year2009          -0.6236      0.1447    -4.31    1.8e-05 ***
## Year2010          -0.7282      0.1417    -5.14    3.2e-07 ***
## Year2011          -0.6335      0.1409    -4.50    7.6e-06 ***
## Year2012          -0.5690      0.1400    -4.06    5.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.539
## Multiple R-squared:  0.139, Adjusted R-squared:  0.127
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 2 observations c(100,101) are outliers with |weight| = 0 ( < 7.6e-05);
## 90 weights are ~= 1. The remaining 1221 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0291 0.8650 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          7.62e-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 1.059 1          1.029
## Year              1.059 16          1.002

```

Residuals from first author



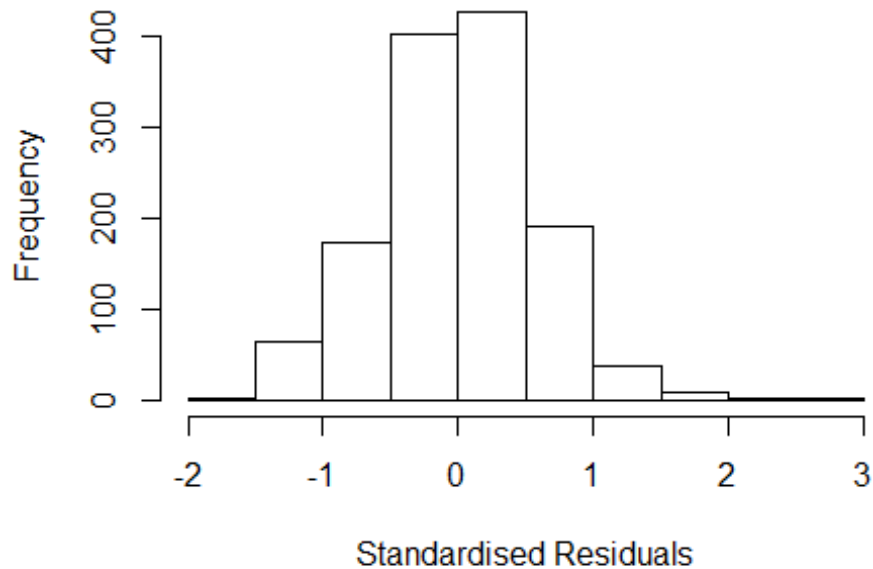
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 218 0031022767 2.995 1997      2748      1      2.609
## 222 0031033534 3.011 1997      2748      1      2.625
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5203 -0.3857  0.0218  0.3990  2.6220
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.7015     0.1311   12.97 < 2e-16 ***
## FirstAuthorFemale1  0.0898     0.0378    2.37  0.01770 *
## Year1997         -1.3125     0.1837   -7.15  1.5e-12 ***
## Year1998         -0.2187     0.1446   -1.51  0.13068
## Year1999         -0.3112     0.1713   -1.82  0.06940 .
## Year2000         -0.2989     0.1460   -2.05  0.04084 *
## Year2001         -0.4691     0.1564   -3.00  0.00276 **
## Year2002         -0.3313     0.1432   -2.31  0.02087 *
## Year2003         -0.5343     0.1478   -3.61  0.00031 ***
## Year2004         -0.6269     0.1430   -4.38  1.3e-05 ***
## Year2005         -0.5618     0.1405   -4.00  6.7e-05 ***
```

```

## Year2006          -0.5676      0.1464    -3.88  0.00011 ***
## Year2007          -0.7293      0.1479    -4.93  9.2e-07 ***
## Year2008          -0.5992      0.1417    -4.23  2.5e-05 ***
## Year2009          -0.6238      0.1449    -4.31  1.8e-05 ***
## Year2010          -0.7298      0.1418    -5.14  3.1e-07 ***
## Year2011          -0.6354      0.1409    -4.51  7.1e-06 ***
## Year2012          -0.5689      0.1401    -4.06  5.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.139, Adjusted R-squared:  0.127
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 2 observations c(100,101) are outliers with |weight| = 0 ( < 7.6e-05);
## 97 weights are ~= 1. The remaining 1214 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0313 0.8640 0.9490 0.8980 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.62e-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 1.044 1          1.022
## Year          1.044 16          1.001

```

Residuals from last author



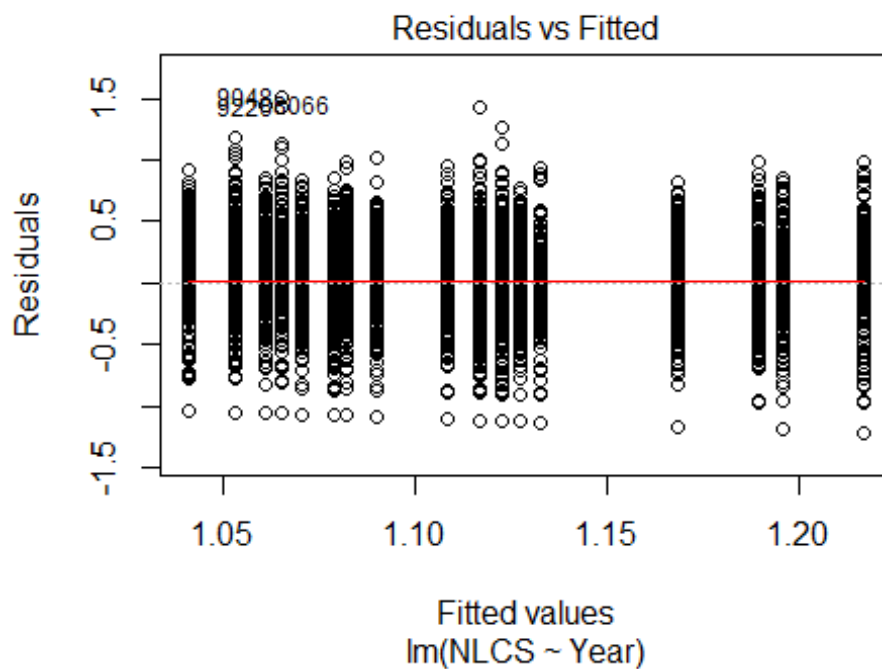
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 218 0031022767 2.995 1997      2748      1      2.609
## 222 0031033534 3.011 1997      2748      1      2.625
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4941 -0.3941  0.0147  0.4114  2.6169
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.7024     0.1328   12.82 < 2e-16 ***
## LastAuthorFemale1  0.0559     0.0415    1.35  0.17782
## Year1997         -1.3083     0.1838   -7.12  1.8e-12 ***
## Year1998         -0.2082     0.1463   -1.42  0.15496
## Year1999         -0.3030     0.1696   -1.79  0.07431 .
## Year2000         -0.2990     0.1480   -2.02  0.04354 *
## Year2001         -0.4630     0.1578   -2.93  0.00340 **
## Year2002         -0.3251     0.1449   -2.24  0.02504 *
## Year2003         -0.5339     0.1490   -3.58  0.00035 ***
## Year2004         -0.6165     0.1446   -4.26  2.2e-05 ***
## Year2005         -0.5576     0.1419   -3.93  9.0e-05 ***
```

```

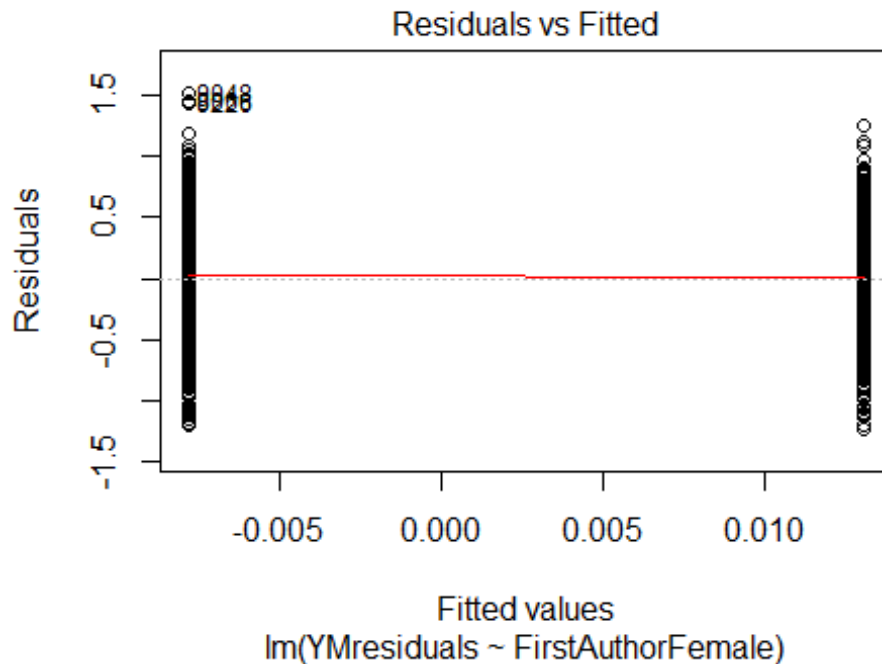
## Year2006      -0.5638      0.1481      -3.81      0.00015 ***
## Year2007      -0.7107      0.1492      -4.76      2.1e-06 ***
## Year2008      -0.5891      0.1432      -4.11      4.2e-05 ***
## Year2009      -0.6115      0.1460      -4.19      3.0e-05 ***
## Year2010      -0.7106      0.1428      -4.98      7.3e-07 ***
## Year2011      -0.6253      0.1424      -4.39      1.2e-05 ***
## Year2012      -0.5616      0.1416      -3.97      7.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.539
## Multiple R-squared:  0.137, Adjusted R-squared:  0.125
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 2 observations c(100,101) are outliers with |weight| = 0 ( < 7.6e-05);
## 85 weights are ~= 1. The remaining 1226 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0138 0.8660 0.9480 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          7.62e-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: 1313"
## [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
## 437 485 445 526 436 414 422 424 487 508 525 558 540 459 485
## 2011 2012
## 448 438
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 232 260 264 281 219 178 276 286 324 324 355 395 355 324 329
## 2011 2012
## 318 318
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 199 219 221 247 191 162 230 233 260 271 301 317 299 276 291
## 2011 2012
## 263 275
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 41, df = 16, p-value = 6e-04
```

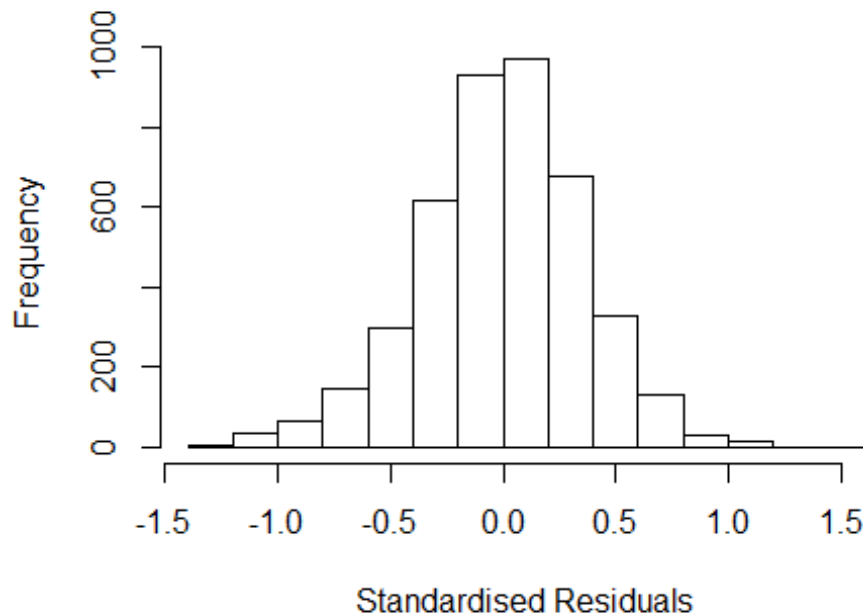


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 8e-04
```

```
## [1] "Female first author team size 2018 geometric mean: 3.86640301369104"
## [1] "Male first author team size 2018 geometric mean: 3.67247262669842"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14000, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.62094273504381"
## [1] "Male last author team size 2018 geometric mean: 3.84741577504649"
##
## 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 1.038 1      1.019
## LastAuthorFemale  1.025 1      1.012
## UniqueAuthors     1.131 4      1.016
## Year              1.151 16     1.004
```

Residuals from first and last author and team size



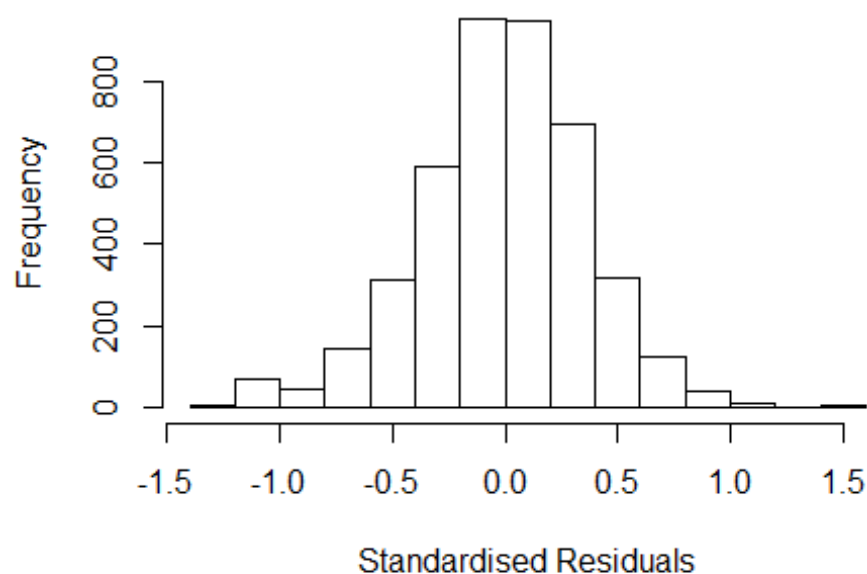
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34890 -0.22644 0.00789 0.22836 1.43765
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02586 0.04551 22.54 < 2e-16 ***
## FirstAuthorFemale1 -0.00557 0.01133 -0.49 0.62338
## LastAuthorFemale1 0.01333 0.01400 0.95 0.34121
## UniqueAuthors2 0.20261 0.03843 5.27 1.4e-07 ***
## UniqueAuthors3 0.21575 0.03856 5.59 2.3e-08 ***
## UniqueAuthors4 0.23967 0.03936 6.09 1.2e-09 ***
## UniqueAuthors5 0.32304 0.03921 8.24 2.3e-16 ***
## Year1997 -0.02547 0.03776 -0.67 0.50001
## Year1998 -0.04409 0.03842 -1.15 0.25113
## Year1999 -0.04791 0.03625 -1.32 0.18637
```

```

## Year2000      -0.12002    0.03957   -3.03  0.00243 **
## Year2001      -0.10896    0.03936   -2.77  0.00566 **
## Year2002      -0.13213    0.03753   -3.52  0.00043 ***
## Year2003      -0.12964    0.03488   -3.72  0.00020 ***
## Year2004      -0.16539    0.03470   -4.77  1.9e-06 ***
## Year2005      -0.14337    0.03482   -4.12  3.9e-05 ***
## Year2006      -0.15779    0.03372   -4.68  3.0e-06 ***
## Year2007      -0.13879    0.03428   -4.05  5.3e-05 ***
## Year2008      -0.16069    0.03460   -4.64  3.5e-06 ***
## Year2009      -0.20179    0.03442   -5.86  4.9e-09 ***
## Year2010      -0.20160    0.03703   -5.44  5.5e-08 ***
## Year2011      -0.21255    0.03702   -5.74  1.0e-08 ***
## Year2012      -0.21432    0.03732   -5.74  1.0e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.338
## Multiple R-squared:  0.0635, Adjusted R-squared:  0.0587
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 366 weights are ~= 1. The remaining 3889 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0312 0.8640 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          2.35e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.033 1 1.016
## LastAuthorFemale 1.018 1 1.009
## Year 1.049 16 1.001

```

Residuals from first and last author



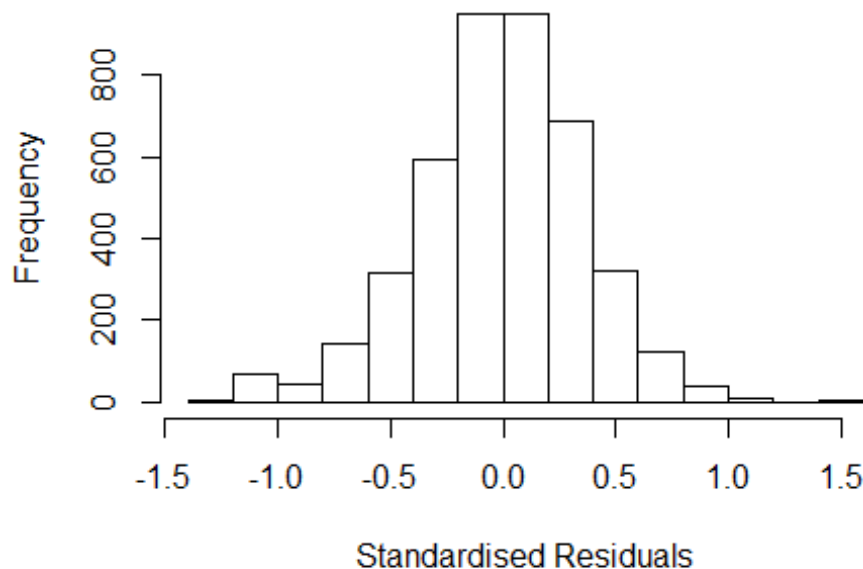
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24025 -0.22625 0.00187 0.23190 1.52737
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23036 0.02764 44.51 < 2e-16 ***
## FirstAuthorFemale1 0.00989 0.01141 0.87 0.38615
## LastAuthorFemale1 0.01644 0.01404 1.17 0.24152
## Year1997 -0.02197 0.03679 -0.60 0.55037
## Year1998 -0.04267 0.03784 -1.13 0.25946
## Year1999 -0.03862 0.03573 -1.08 0.27977
## Year2000 -0.10656 0.03912 -2.72 0.00648 **
## Year2001 -0.08977 0.03902 -2.30 0.02146 *
## Year2002 -0.11363 0.03687 -3.08 0.00207 **
## Year2003 -0.10374 0.03450 -3.01 0.00265 **
## Year2004 -0.14363 0.03473 -4.14 3.6e-05 ***
## Year2005 -0.11642 0.03456 -3.37 0.00076 ***
```

```

## Year2006      -0.13854    0.03309   -4.19  2.9e-05 ***
## Year2007      -0.11923    0.03386   -3.52  0.00043 ***
## Year2008      -0.13950    0.03410   -4.09  4.4e-05 ***
## Year2009      -0.18575    0.03385   -5.49  4.3e-08 ***
## Year2010      -0.17450    0.03663   -4.76  2.0e-06 ***
## Year2011      -0.18373    0.03684   -4.99  6.4e-07 ***
## Year2012      -0.18426    0.03646   -5.05  4.5e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.0236, Adjusted R-squared:  0.0194
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 352 weights are ~= 1. The remaining 3903 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0073 0.8640 0.9520 0.8930 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.35e-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 1.032 1      1.016
## Year      1.032 16      1.001

```

Residuals from first author



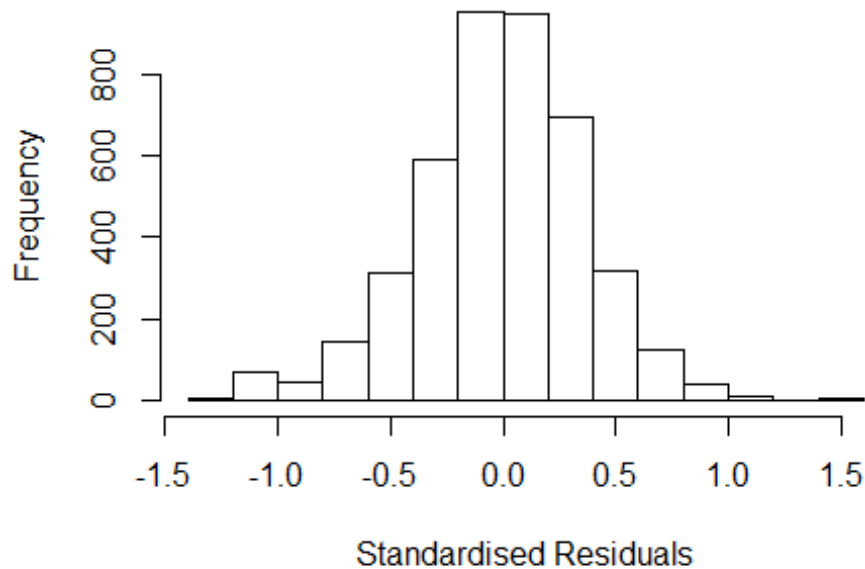
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24431 -0.22769 0.00132 0.23086 1.52445
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2334 0.0274 44.97 < 2e-16 ***
## FirstAuthorFemale1 0.0110 0.0114 0.96 0.33704
## Year1997 -0.0223 0.0367 -0.61 0.54284
## Year1998 -0.0441 0.0378 -1.17 0.24306
## Year1999 -0.0396 0.0356 -1.11 0.26708
## Year2000 -0.1065 0.0391 -2.73 0.00645 **
## Year2001 -0.0909 0.0390 -2.33 0.01985 *
## Year2002 -0.1140 0.0368 -3.10 0.00197 **
## Year2003 -0.1040 0.0345 -3.02 0.00256 **
## Year2004 -0.1435 0.0347 -4.14 3.5e-05 ***
## Year2005 -0.1161 0.0345 -3.37 0.00077 ***
## Year2006 -0.1380 0.0330 -4.18 3.0e-05 ***
```

```

## Year2007          -0.1187      0.0338   -3.51  0.00045 ***
## Year2008          -0.1392      0.0340   -4.09  4.4e-05 ***
## Year2009          -0.1862      0.0338   -5.51  3.8e-08 ***
## Year2010          -0.1745      0.0366   -4.77  1.9e-06 ***
## Year2011          -0.1838      0.0368   -4.99  6.3e-07 ***
## Year2012          -0.1841      0.0364   -5.05  4.5e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.0232, Adjusted R-squared:  0.0193
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 362 weights are ~= 1. The remaining 3893 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0082 0.8640 0.9510 0.8930 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.35e-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 1.017 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
## -1.23318 -0.22874  0.00135  0.23122  1.52326
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2332     0.0274   44.95 < 2e-16 ***
## LastAuthorFemale1  0.0174     0.0140    1.24  0.21577
## Year1997        -0.0219     0.0368   -0.60  0.55100
## Year1998        -0.0424     0.0378   -1.12  0.26225
## Year1999        -0.0383     0.0357   -1.07  0.28292
## Year2000        -0.1060     0.0391   -2.71  0.00672 **
## Year2001        -0.0888     0.0389   -2.28  0.02259 *
## Year2002        -0.1127     0.0368   -3.06  0.00220 **
## Year2003        -0.1035     0.0345   -3.00  0.00271 **
## Year2004        -0.1424     0.0346   -4.11  4.0e-05 ***
## Year2005        -0.1154     0.0345   -3.35  0.00083 ***
## Year2006        -0.1383     0.0331   -4.18  2.9e-05 ***
```

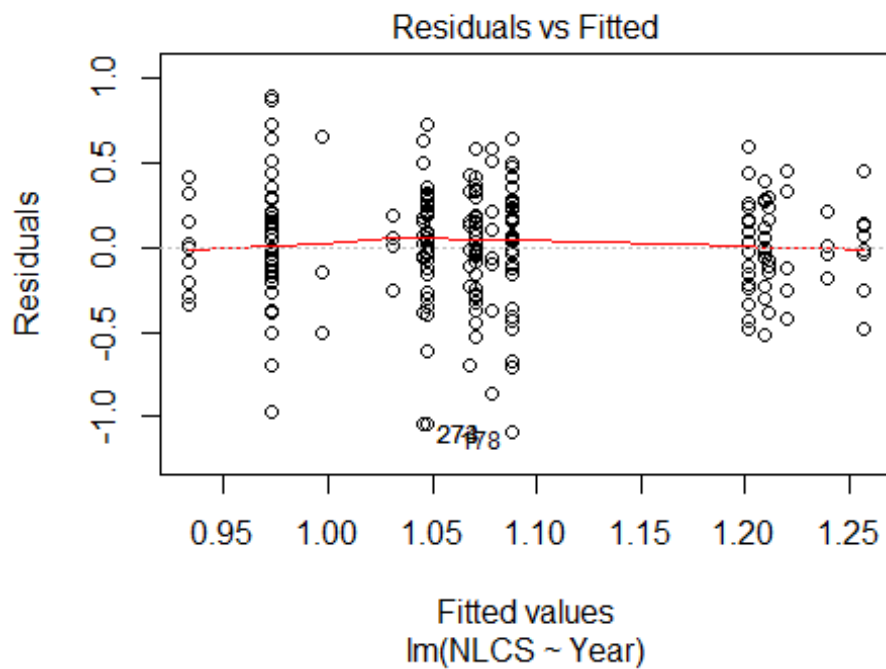


```

## Year2007          -0.1181      0.0338   -3.49  0.00048 ***
## Year2008          -0.1385      0.0340   -4.07  4.8e-05 ***
## Year2009          -0.1845      0.0338   -5.46  4.9e-08 ***
## Year2010          -0.1733      0.0366   -4.74  2.3e-06 ***
## Year2011          -0.1824      0.0368   -4.96  7.4e-07 ***
## Year2012          -0.1828      0.0363   -5.03  5.2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.0234, Adjusted R-squared:  0.0195
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 347 weights are ~= 1. The remaining 3908 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0084 0.8640 0.9510 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.35e-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: 4255"
## [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
##   10   7   10   2   16   14   9   19   13   14   14   12   11   41   36
## 2011 2012
##   36   46
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    3    8    0    8    5    4   16   10    7    8   10   11   34   29
## 2011 2012

```

```
## 34 42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1 2 7 0 8 5 3 14 10 6 6 10 11 30 23
## 2011 2012
## 31 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 = 19, df = 15, p-value = 0.2
```



```
##
## 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: 4.43361269112587"
## [1] "Male first author team size 2018 geometric mean: 4.56488704369751"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 320, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.7230771051061"
## [1] "Male last author team size 2018 geometric mean: 4.75122372337201"

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

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

## 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

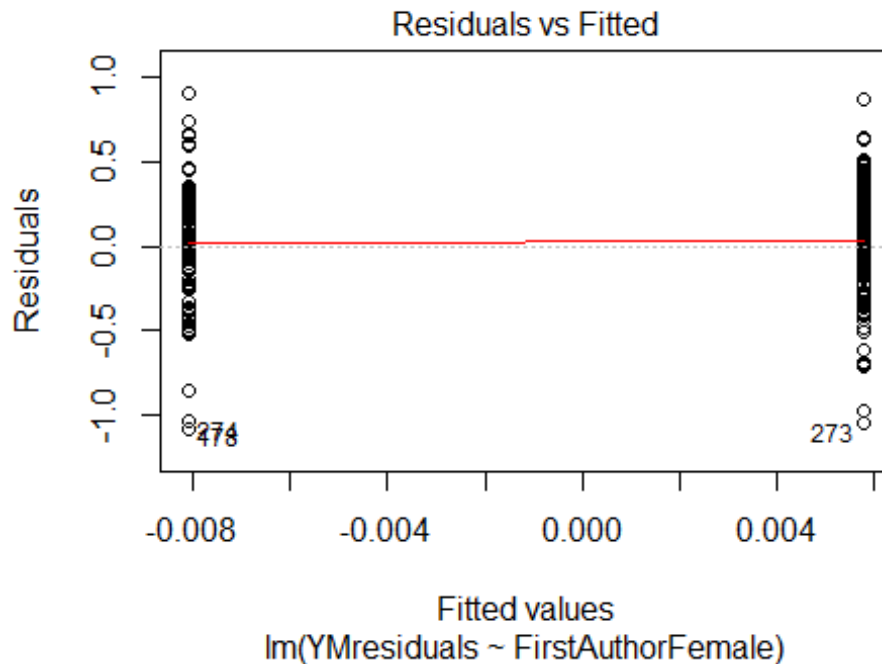
## [1] "Regression 3: First author gender, Year as factors"

## 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"

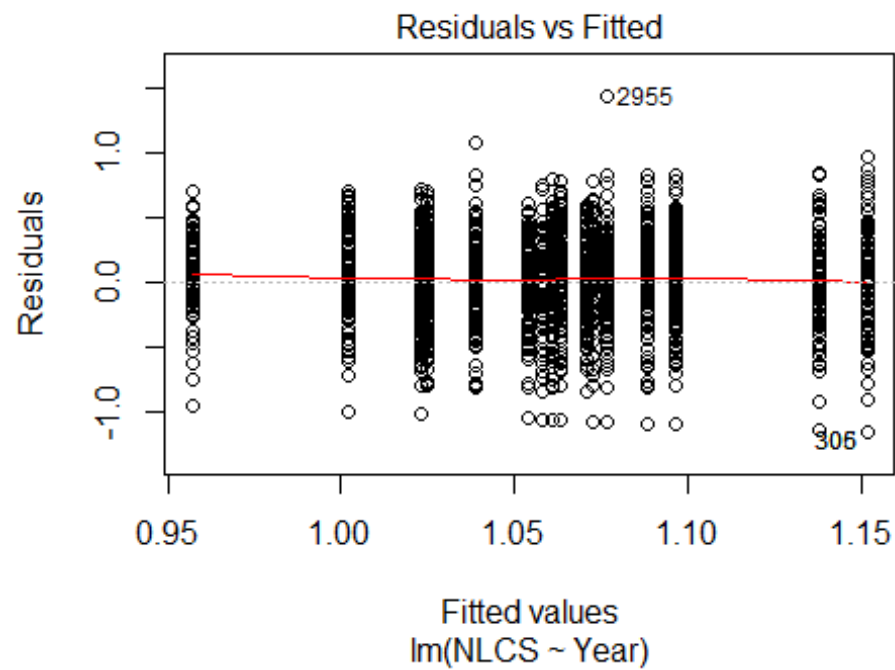
## Warning in lmrob.fit(x, y, control, init = init, mf = mf): M-step did NOT
## converge. Returning unconverged SM-estimate

```

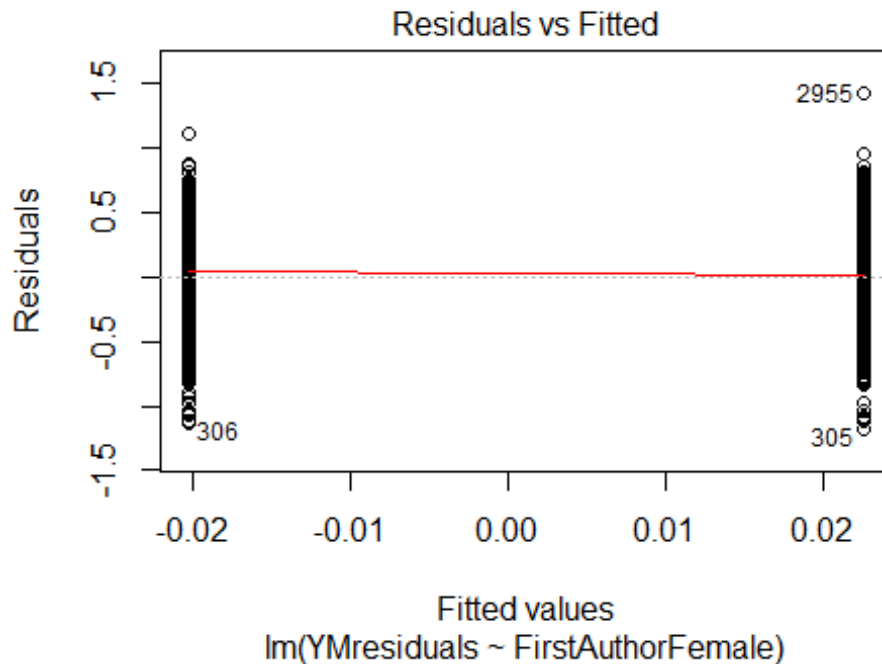


```
## [1] "Sample size for the above analysis: 205"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2802"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 154 143 154 124 148 139 133 107 147 164 178 161 193 168 190
## 2011 2012
## 186 188
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 114 111 103 80 94 90 102 67 105 134 136 128 152 134 152
## 2011 2012
## 145 146
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 103 101 97 75 82 83 90 57 97 127 120 117 136 119 131
## 2011 2012
## 126 124
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
```

```
##
## data: NLCS by Year
## Bartlett's K-squared = 29, df = 16, p-value = 0.02
```

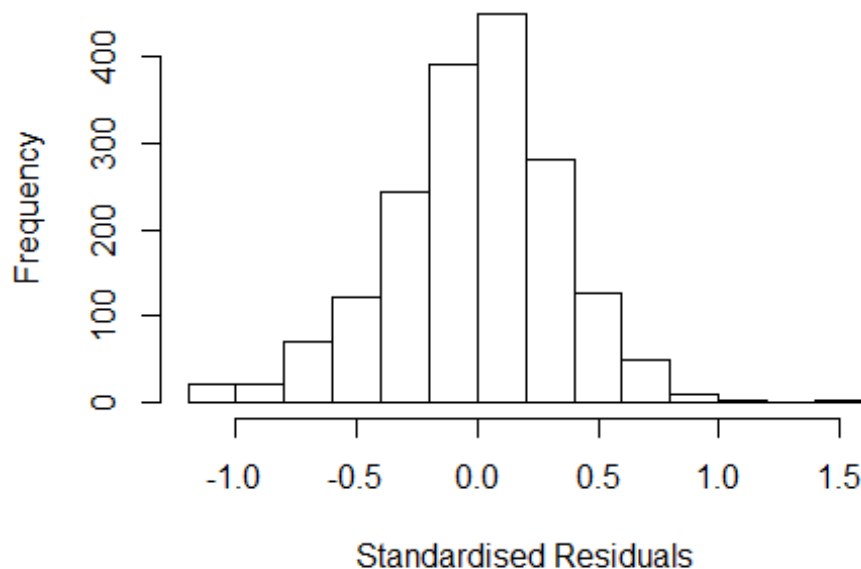


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7, df = 1, p-value = 0.008
```



```
## [1] "Female first author team size 2018 geometric mean: 3.97987327790085"
## [1] "Male first author team size 2018 geometric mean: 3.5418128024423"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.03620434460444"
## [1] "Male last author team size 2018 geometric mean: 3.59605241147151"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2800, 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 1.067 1      1.033
## LastAuthorFemale  1.127 1      1.061
## UniqueAuthors    1.327 4      1.036
## Year             1.481 16      1.012
```

Residuals from first and last author and team size



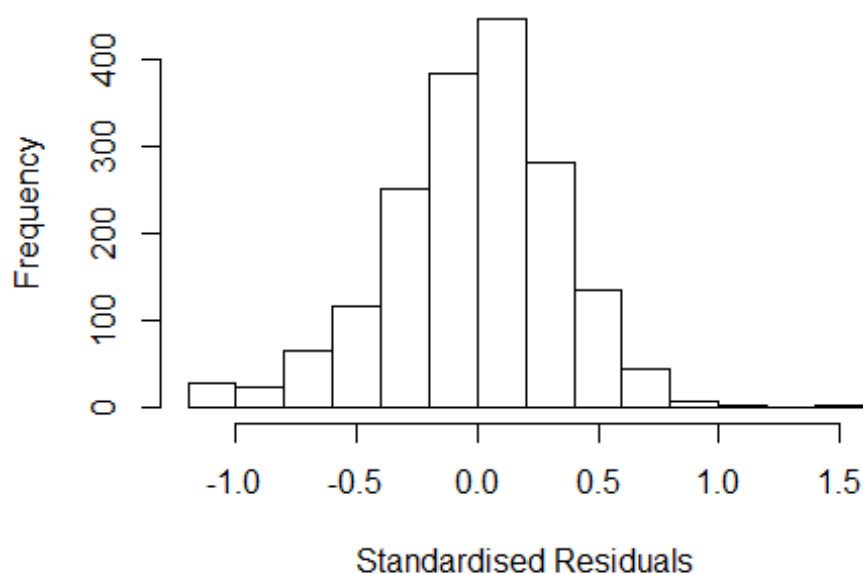
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1756 -0.2177 0.0108 0.2105 1.4930
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0299 0.0516 19.94 <2e-16 ***
## FirstAuthorFemale1 0.0243 0.0169 1.44 0.1514
## LastAuthorFemale1 0.0238 0.0179 1.33 0.1822
## UniqueAuthors2 0.0738 0.0442 1.67 0.0948 .
## UniqueAuthors3 0.0796 0.0440 1.81 0.0705 .
## UniqueAuthors4 0.0781 0.0463 1.69 0.0915 .
## UniqueAuthors5 0.1460 0.0471 3.10 0.0019 **
## Year1997 0.0476 0.0521 0.91 0.3610
## Year1998 0.0256 0.0524 0.49 0.6256
## Year1999 -0.0326 0.0542 -0.60 0.5481
```

```

## Year2000          -0.0306      0.0491   -0.62   0.5333
## Year2001          -0.0724      0.0617   -1.17   0.2404
## Year2002          -0.0330      0.0484   -0.68   0.4951
## Year2003          -0.1381      0.0624   -2.21   0.0270 *
## Year2004          -0.0439      0.0488   -0.90   0.3682
## Year2005          -0.0916      0.0470   -1.95   0.0517 .
## Year2006          -0.0610      0.0456   -1.34   0.1804
## Year2007          -0.0916      0.0486   -1.88   0.0598 .
## Year2008          -0.1005      0.0474   -2.12   0.0342 *
## Year2009          -0.0390      0.0494   -0.79   0.4294
## Year2010          -0.0531      0.0468   -1.13   0.2566
## Year2011          -0.0590      0.0454   -1.30   0.1940
## Year2012          -0.0481      0.0522   -0.92   0.3567
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.324
## Multiple R-squared:  0.025, Adjusted R-squared:  0.0128
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 148 weights are ~= 1. The remaining 1637 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0012 0.8630 0.9500 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.60e-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 1.058 1          1.029
## LastAuthorFemale 1.093 1          1.045
## Year          1.133 16          1.004

```


Residuals from first and last author



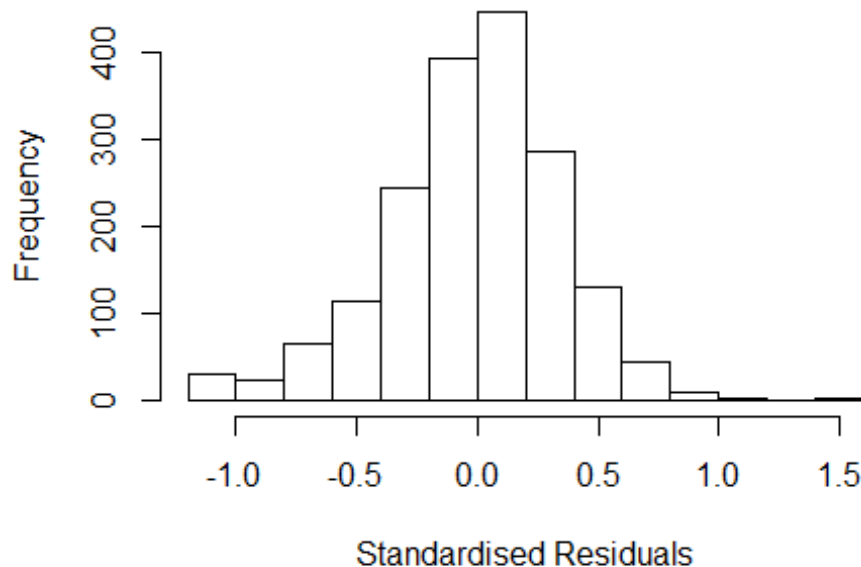
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1718 -0.2216  0.0102  0.2131  1.4024
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0966    0.0368   29.79  <2e-16 ***
## FirstAuthorFemale1  0.0284    0.0171    1.66   0.097 .
## LastAuthorFemale1  0.0220    0.0178    1.24   0.216
## Year1997          0.0468    0.0519    0.90   0.367
## Year1998          0.0286    0.0523    0.55   0.584
## Year1999         -0.0285    0.0545   -0.52   0.601
## Year2000         -0.0226    0.0491   -0.46   0.645
## Year2001         -0.0629    0.0607   -1.04   0.300
## Year2002         -0.0169    0.0487   -0.35   0.729
## Year2003         -0.1257    0.0622   -2.02   0.043 *
## Year2004         -0.0312    0.0492   -0.64   0.526
## Year2005         -0.0789    0.0470   -1.68   0.093 .
```

```

## Year2006          -0.0432      0.0458   -0.94    0.346
## Year2007          -0.0705      0.0479   -1.47    0.142
## Year2008          -0.0829      0.0469   -1.77    0.077 .
## Year2009          -0.0207      0.0489   -0.42    0.672
## Year2010          -0.0338      0.0462   -0.73    0.464
## Year2011          -0.0374      0.0450   -0.83    0.407
## Year2012          -0.0174      0.0510   -0.34    0.733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.325
## Multiple R-squared:  0.0141, Adjusted R-squared:  0.00405
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1636 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0224 0.8630 0.9490 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.60e-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 1.049 1      1.024
## Year              1.049 16      1.002

```

Residuals from first author



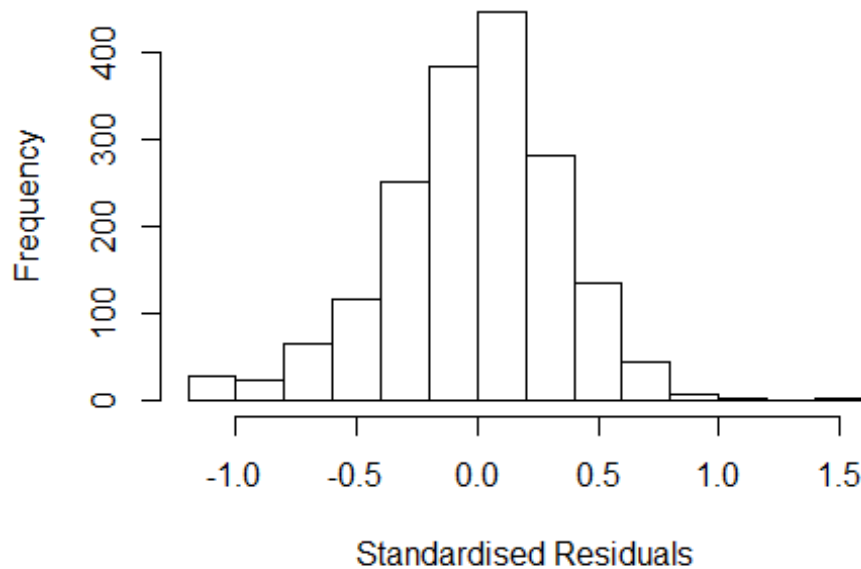
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1785 -0.2221 0.0096 0.2165 1.4154
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1013 0.0368 29.97 <2e-16 ***
## FirstAuthorFemale1 0.0311 0.0170 1.83 0.068 .
## Year1997 0.0461 0.0519 0.89 0.375
## Year1998 0.0296 0.0523 0.57 0.571
## Year1999 -0.0274 0.0546 -0.50 0.616
## Year2000 -0.0195 0.0490 -0.40 0.690
## Year2001 -0.0629 0.0605 -1.04 0.299
## Year2002 -0.0164 0.0487 -0.34 0.736
## Year2003 -0.1245 0.0624 -1.99 0.046 *
## Year2004 -0.0333 0.0493 -0.68 0.499
## Year2005 -0.0768 0.0469 -1.64 0.102
## Year2006 -0.0420 0.0459 -0.92 0.360
```

```

## Year2007          -0.0681      0.0479   -1.42    0.155
## Year2008          -0.0802      0.0467   -1.72    0.086 .
## Year2009          -0.0173      0.0487   -0.35    0.723
## Year2010          -0.0315      0.0461   -0.68    0.495
## Year2011          -0.0358      0.0449   -0.80    0.425
## Year2012          -0.0152      0.0510   -0.30    0.766
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.324
## Multiple R-squared:  0.0131, Adjusted R-squared:  0.00361
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1644 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0177 0.8660 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      5.60e-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 1.082 1          1.040
## Year            1.082 16          1.002

```

Residuals from last author



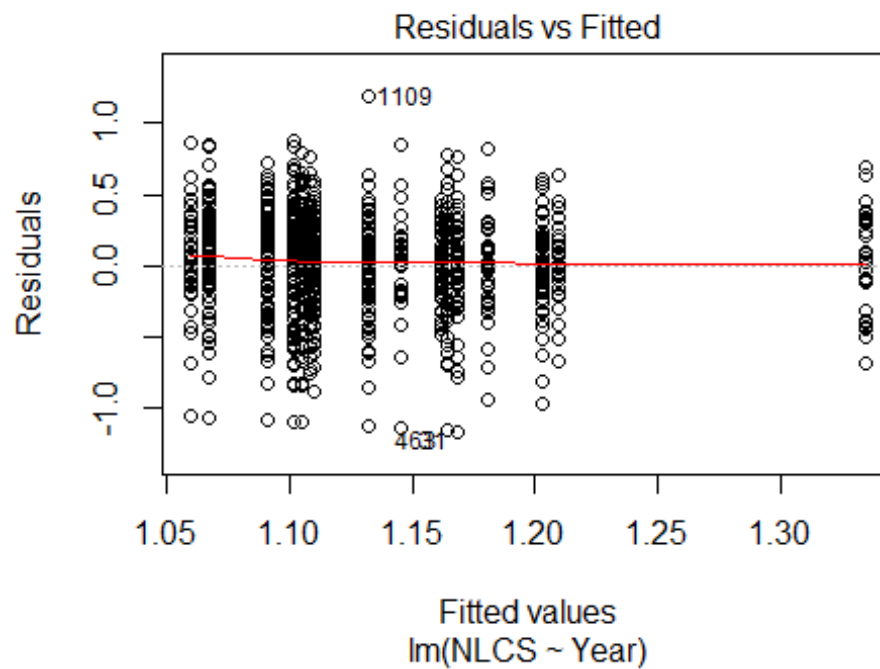
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1519 -0.2270  0.0108  0.2128  1.4129
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1066    0.0361  30.62  <2e-16 ***
## LastAuthorFemale1  0.0260    0.0177   1.47   0.142
## Year1997          0.0453    0.0518   0.87   0.382
## Year1998          0.0278    0.0519   0.54   0.592
## Year1999         -0.0318    0.0543  -0.59   0.558
## Year2000         -0.0196    0.0489  -0.40   0.688
## Year2001         -0.0645    0.0606  -1.06   0.287
## Year2002         -0.0135    0.0483  -0.28   0.779
## Year2003         -0.1207    0.0615  -1.96   0.050 *
## Year2004         -0.0272    0.0490  -0.55   0.579
## Year2005         -0.0754    0.0469  -1.61   0.108
## Year2006         -0.0392    0.0457  -0.86   0.391
```

```

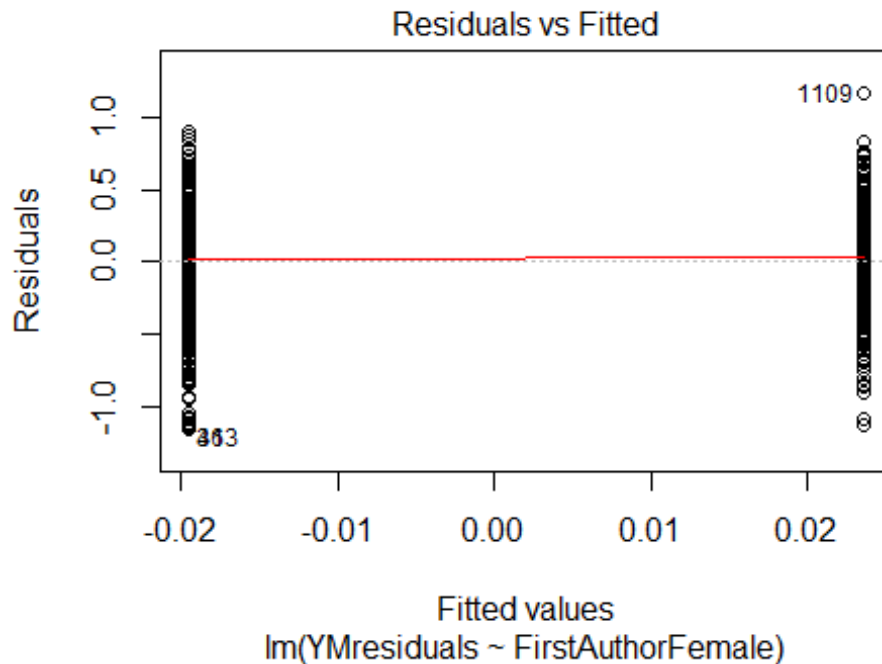
## Year2007          -0.0654      0.0480    -1.36      0.173
## Year2008          -0.0779      0.0470    -1.66      0.098 .
## Year2009          -0.0182      0.0488    -0.37      0.709
## Year2010          -0.0301      0.0462    -0.65      0.515
## Year2011          -0.0334      0.0449    -0.74      0.457
## Year2012          -0.0152      0.0512    -0.30      0.766
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.326
## Multiple R-squared:  0.0124, Adjusted R-squared:  0.00286
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1644 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0209 0.8660 0.9520 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.60e-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: 1785"
## [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
##   64   48   51   66   46   52   55   55   53   67   76   81   76   95   99
## 2011 2012
##  103  122
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   43   32   28   45   29   32   39   44   43   62   57   69   61   80   87
## 2011 2012

```

```
## 90 108
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 36 24 25 42 26 28 32 31 40 54 46 63 56 76 75
## 2011 2012
## 78 98
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 29, df = 16, p-value = 0.02
```

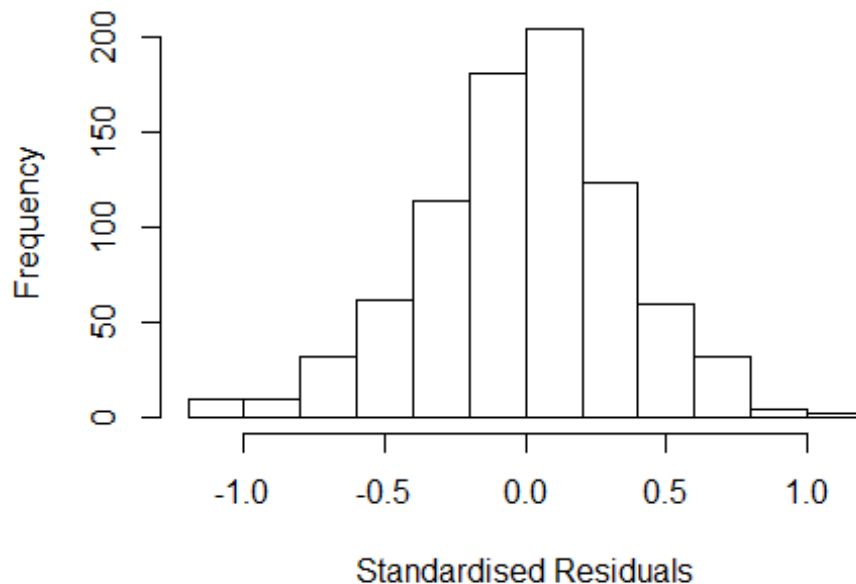


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.071, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 4.89042205222888"
## [1] "Male first author team size 2018 geometric mean: 5.14518288970775"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 930, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.84231391710808"
## [1] "Male last author team size 2018 geometric mean: 5.06547422162712"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1000, 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 1.197  1      1.094
## LastAuthorFemale  1.094  1      1.046
## UniqueAuthors    1.570  4      1.058
## Year             1.807 16      1.019
```


Residuals from first and last author and team size



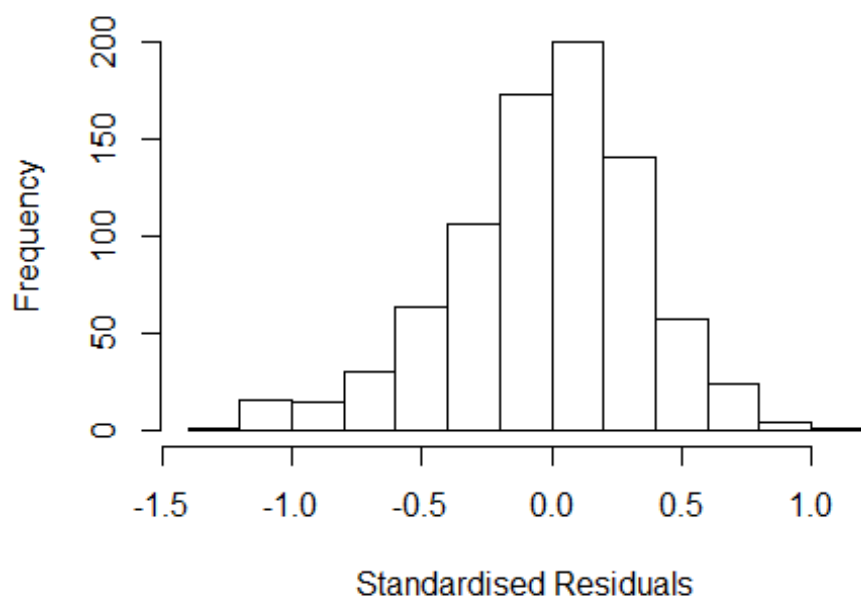
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1991 -0.2253 0.0105 0.2150 1.1227
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.88156 0.12589 7.00 5.3e-12 ***
## FirstAuthorFemale1 0.06344 0.02616 2.43 0.01552 *
## LastAuthorFemale1 -0.02022 0.02944 -0.69 0.49241
## UniqueAuthors2 0.29257 0.11925 2.45 0.01436 *
## UniqueAuthors3 0.33353 0.11604 2.87 0.00416 **
## UniqueAuthors4 0.39142 0.11694 3.35 0.00085 ***
## UniqueAuthors5 0.38935 0.11403 3.41 0.00067 ***
## Year1997 0.14044 0.11455 1.23 0.22056
## Year1998 -0.07535 0.09962 -0.76 0.44966
## Year1999 -0.08550 0.07623 -1.12 0.26235
```

```

## Year2000      -0.17563      0.10385      -1.69      0.09119 .
## Year2001      -0.00723      0.09196      -0.08      0.93739
## Year2002      -0.09421      0.09120      -1.03      0.30188
## Year2003      -0.10096      0.10702      -0.94      0.34577
## Year2004      -0.16958      0.08442      -2.01      0.04488 *
## Year2005      -0.13288      0.07832      -1.70      0.09014 .
## Year2006      -0.03435      0.07999      -0.43      0.66767
## Year2007      -0.13257      0.07461      -1.78      0.07595 .
## Year2008      -0.14385      0.08342      -1.72      0.08501 .
## Year2009      -0.11794      0.07255      -1.63      0.10441
## Year2010      -0.12379      0.07998      -1.55      0.12207
## Year2011      -0.13734      0.07481      -1.84      0.06673 .
## Year2012      -0.16800      0.07589      -2.21      0.02713 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.334
## Multiple R-squared:  0.0935, Adjusted R-squared:  0.0688
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 762 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.170  0.859  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      1.20e-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 1.149 1      1.072
## LastAuthorFemale  1.076 1      1.037
## Year              1.219 16      1.006

```

Residuals from first and last author



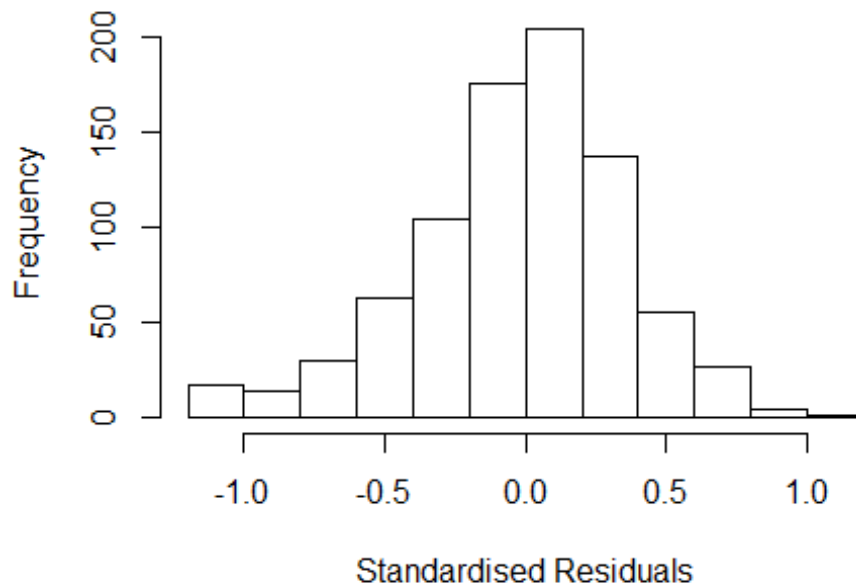
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2016 -0.2287 0.0119 0.2164 1.1476
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2016 0.0639 18.81 <2e-16 ***
## FirstAuthorFemale1 0.0738 0.0261 2.83 0.0048 **
## LastAuthorFemale1 -0.0191 0.0296 -0.64 0.5199
## Year1997 0.1689 0.1008 1.68 0.0942 .
## Year1998 -0.0845 0.0923 -0.92 0.3599
## Year1999 -0.0652 0.0737 -0.89 0.3761
## Year2000 -0.1467 0.0960 -1.53 0.1266
## Year2001 -0.0115 0.0840 -0.14 0.8916
## Year2002 -0.0668 0.0870 -0.77 0.4429
## Year2003 -0.0705 0.1078 -0.65 0.5131
## Year2004 -0.1289 0.0839 -1.54 0.1248
## Year2005 -0.1172 0.0792 -1.48 0.1395
```

```

## Year2006          -0.0236      0.0780   -0.30   0.7623
## Year2007          -0.1151      0.0745   -1.55   0.1227
## Year2008          -0.1011      0.0839   -1.21   0.2285
## Year2009          -0.0830      0.0703   -1.18   0.2385
## Year2010          -0.0987      0.0791   -1.25   0.2125
## Year2011          -0.1046      0.0748   -1.40   0.1622
## Year2012          -0.1262      0.0732   -1.72   0.0851 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.0336, Adjusted R-squared:  0.0122
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 767 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.176  0.860  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      1.20e-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 1.148 1      1.071
## Year              1.148 16      1.004

```

Residuals from first author



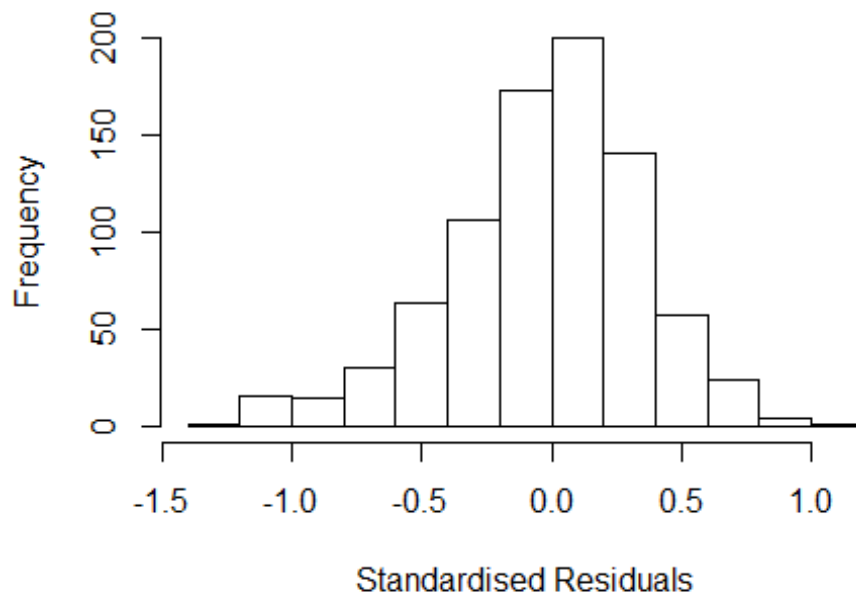
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1970 -0.2342 0.0129 0.2165 1.1342
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1970 0.0632 18.95 <2e-16 ***
## FirstAuthorFemale1 0.0729 0.0263 2.77 0.0057 **
## Year1997 0.1698 0.1005 1.69 0.0916 .
## Year1998 -0.0868 0.0917 -0.95 0.3440
## Year1999 -0.0649 0.0735 -0.88 0.3775
## Year2000 -0.1474 0.0962 -1.53 0.1259
## Year2001 -0.0105 0.0842 -0.12 0.9009
## Year2002 -0.0654 0.0871 -0.75 0.4529
## Year2003 -0.0667 0.1072 -0.62 0.5342
## Year2004 -0.1270 0.0834 -1.52 0.1284
## Year2005 -0.1167 0.0789 -1.48 0.1395
## Year2006 -0.0233 0.0778 -0.30 0.7648
```

```

## Year2007          -0.1153      0.0745   -1.55    0.1217
## Year2008          -0.0989      0.0836   -1.18    0.2370
## Year2009          -0.0830      0.0703   -1.18    0.2379
## Year2010          -0.0998      0.0789   -1.27    0.2062
## Year2011          -0.1030      0.0746   -1.38    0.1674
## Year2012          -0.1260      0.0730   -1.73    0.0848 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.336
## Multiple R-squared:  0.0332, Adjusted R-squared:  0.013
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 764 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.178  0.858  0.953  0.890  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.20e-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 1.072 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
## -1.2328 -0.2345  0.0123  0.2149  1.1841
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2328     0.0620   19.87  <2e-16 ***
## LastAuthorFemale1 -0.0141     0.0302   -0.47    0.641
## Year1997          0.1566     0.1026    1.53    0.128
## Year1998         -0.0808     0.0910   -0.89    0.375
## Year1999         -0.0692     0.0734   -0.94    0.346
## Year2000         -0.1426     0.0980   -1.45    0.146
## Year2001         -0.0176     0.0837   -0.21    0.833
## Year2002         -0.0705     0.0869   -0.81    0.418
## Year2003         -0.0729     0.1095   -0.67    0.506
## Year2004         -0.1439     0.0845   -1.70    0.089 .
## Year2005         -0.1157     0.0794   -1.46    0.146
## Year2006         -0.0153     0.0779   -0.20    0.844
```

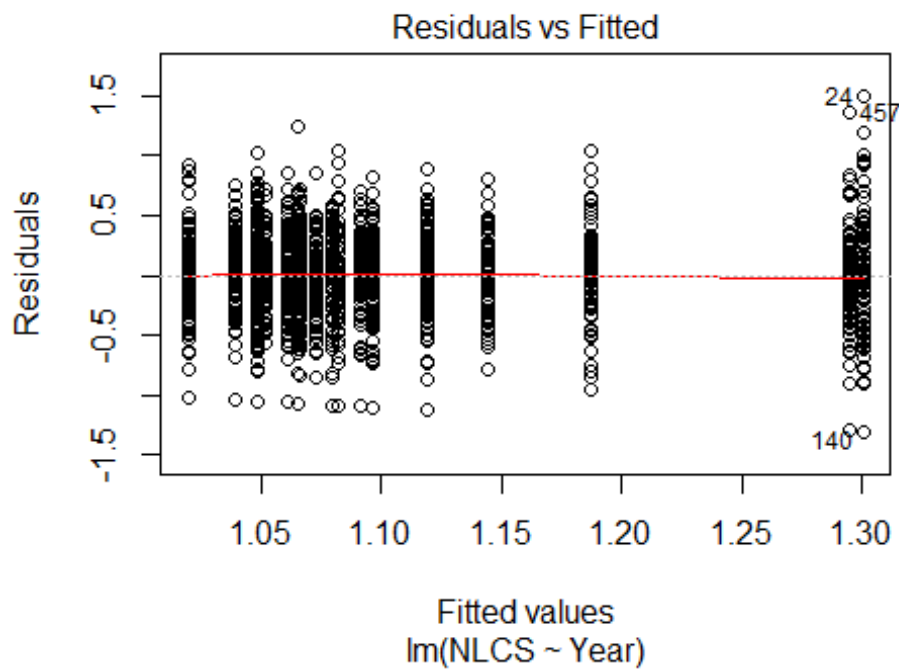
```

## Year2007          -0.1062      0.0751   -1.41    0.158
## Year2008          -0.0992      0.0826   -1.20    0.230
## Year2009          -0.0818      0.0709   -1.15    0.249
## Year2010          -0.0916      0.0792   -1.16    0.248
## Year2011          -0.0923      0.0750   -1.23    0.219
## Year2012          -0.1267      0.0737   -1.72    0.086 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.339
## Multiple R-squared:  0.0242, Adjusted R-squared:  0.00378
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 755 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.159  0.859  0.950  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      1.20e-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: 830"
## [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
## 210 204 147 145 156 163 157 116 127 143 140 188 173 199 205
## 2011 2012
## 186 248
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 91 91 91 86 79 98 94 70 81 91 95 133 121 133 139
## 2011 2012

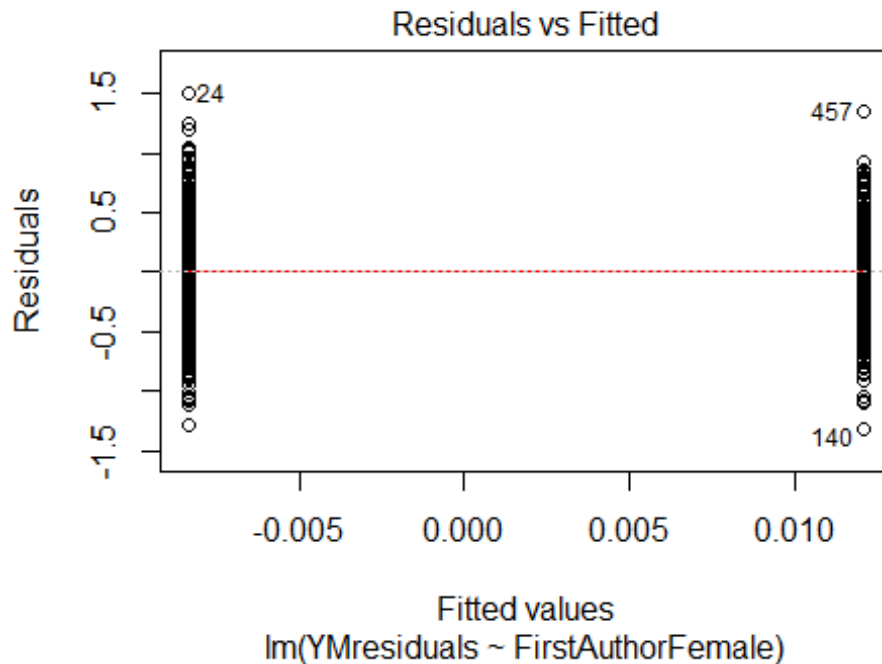
```



```
## 143 186
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 78 86 81 76 72 80 80 62 68 72 82 105 100 108 118
## 2011 2012
## 113 159
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 60, df = 16, p-value = 6e-07
```

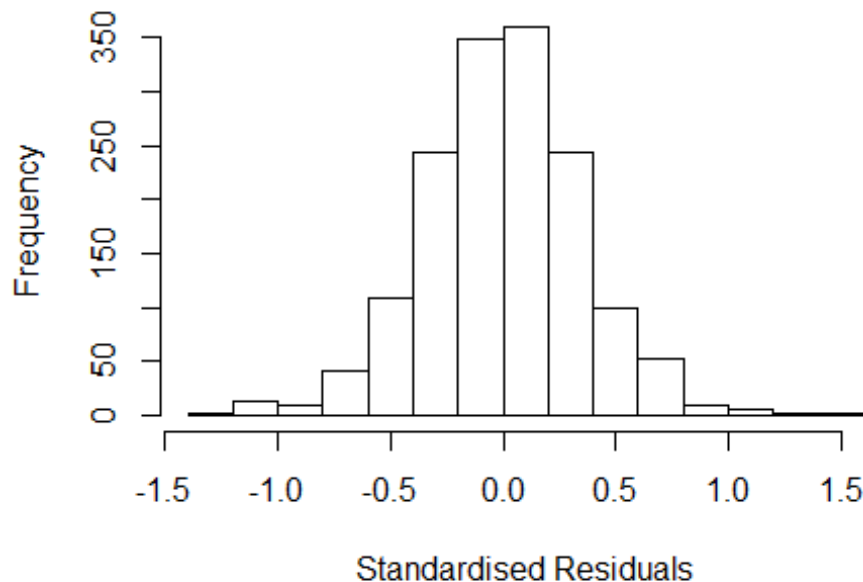


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.6, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 4.62849131887265"
## [1] "Male first author team size 2018 geometric mean: 3.72104308511227"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3200, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.34938053046712"
## [1] "Male last author team size 2018 geometric mean: 4.06746872270026"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.053 1 1.026
## LastAuthorFemale 1.080 1 1.039
## UniqueAuthors 1.313 4 1.035
## Year 1.402 16 1.011
```

Residuals from first and last author and team size



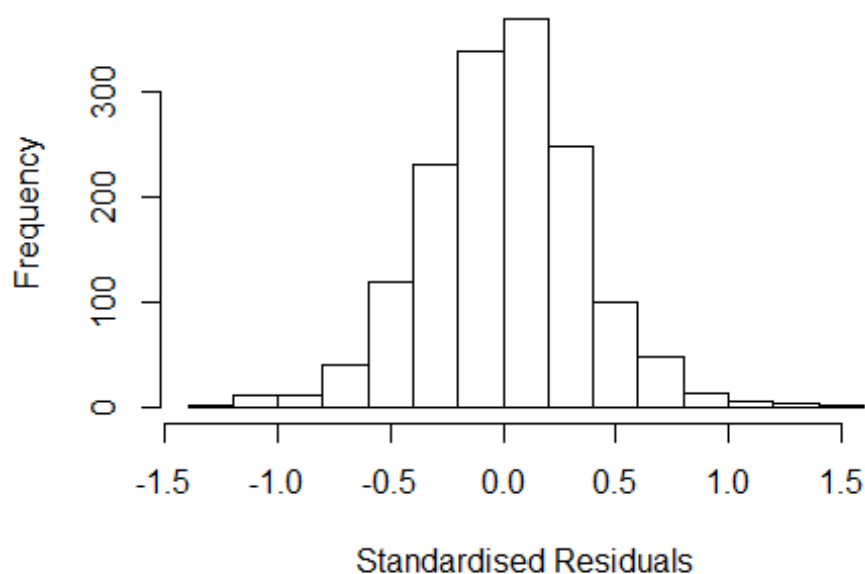
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29106 -0.22405  0.00219  0.21923  1.52039
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19659    0.07714   15.51 < 2e-16 ***
## FirstAuthorFemale1  0.01844    0.01796    1.03  0.30466
## LastAuthorFemale1 -0.00447    0.02089   -0.21  0.83060
## UniqueAuthors2     0.00186    0.05361    0.03  0.97232
## UniqueAuthors3     0.00891    0.05292    0.17  0.86631
## UniqueAuthors4     0.03399    0.05345    0.64  0.52498
## UniqueAuthors5     0.07602    0.05270    1.44  0.14938
## Year1997          0.06837    0.07166    0.95  0.34021
## Year1998         -0.08087    0.07083   -1.14  0.25372
## Year1999         -0.11942    0.06867   -1.74  0.08220 .
```

```

## Year2000      -0.06981    0.07604   -0.92  0.35872
## Year2001      -0.18092    0.07398   -2.45  0.01458 *
## Year2002      -0.13685    0.06846   -2.00  0.04577 *
## Year2003      -0.20275    0.07095   -2.86  0.00432 **
## Year2004      -0.17081    0.07130   -2.40  0.01671 *
## Year2005      -0.18615    0.07187   -2.59  0.00969 **
## Year2006      -0.16818    0.06953   -2.42  0.01569 *
## Year2007      -0.18130    0.06781   -2.67  0.00758 **
## Year2008      -0.25274    0.06765   -3.74  0.00019 ***
## Year2009      -0.13157    0.06821   -1.93  0.05394 .
## Year2010      -0.10810    0.06748   -1.60  0.10939
## Year2011      -0.16945    0.06962   -2.43  0.01505 *
## Year2012      -0.16776    0.06879   -2.44  0.01486 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.329
## Multiple R-squared:  0.0488, Adjusted R-squared:  0.035
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 130 weights are ~= 1. The remaining 1410 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0008 0.8720 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          6.49e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1 1.022
## LastAuthorFemale 1.074 1 1.036
## Year 1.087 16 1.003

```

Residuals from first and last author



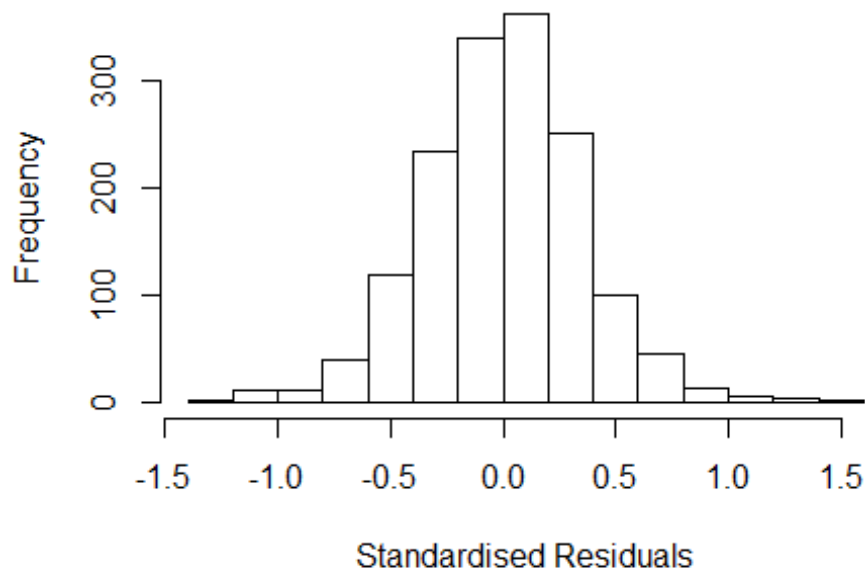
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28568 -0.22098  0.00831  0.22141  1.56979
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22321    0.06075   20.13 < 2e-16 ***
## FirstAuthorFemale1  0.01698    0.01798    0.94  0.34497
## LastAuthorFemale1 -0.00721    0.02104   -0.34  0.73197
## Year1997         0.06247    0.07187    0.87  0.38489
## Year1998        -0.08875    0.07066   -1.26  0.20934
## Year1999        -0.11935    0.06891   -1.73  0.08347 .
## Year2000        -0.06700    0.07564   -0.89  0.37584
## Year2001        -0.17742    0.07374   -2.41  0.01625 *
## Year2002        -0.13161    0.06847   -1.92  0.05476 .
## Year2003        -0.19671    0.07088   -2.78  0.00558 **
## Year2004        -0.15524    0.07135   -2.18  0.02973 *
## Year2005        -0.16875    0.07157   -2.36  0.01852 *
```

```

## Year2006          -0.15432    0.06975   -2.21  0.02707 *
## Year2007          -0.17410    0.06819   -2.55  0.01077 *
## Year2008          -0.24236    0.06802   -3.56  0.00038 ***
## Year2009          -0.11344    0.06838   -1.66  0.09733 .
## Year2010          -0.09538    0.06792   -1.40  0.16046
## Year2011          -0.15703    0.06964   -2.25  0.02429 *
## Year2012          -0.15812    0.06898   -2.29  0.02203 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.33
## Multiple R-squared:  0.0408, Adjusted R-squared:  0.0295
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 11 is an outlier with |weight| = 0 ( < 6.5e-05);
## 141 weights are ~ = 1. The remaining 1398 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0479 0.8700  0.9510  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          6.49e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1          1.014
## Year              1.029 16          1.001

```

Residuals from first author



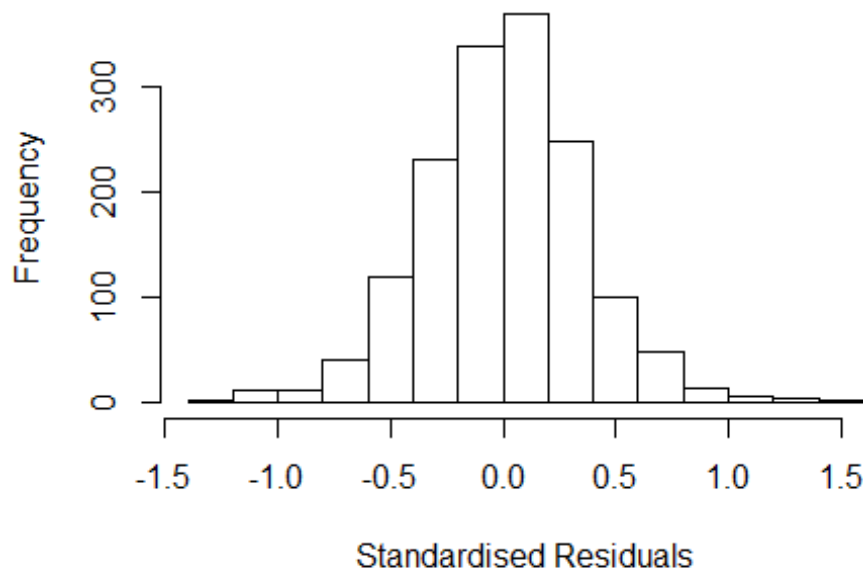
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28498 -0.22144 0.00807 0.22144 1.57120
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2218 0.0607 20.13 < 2e-16 ***
## FirstAuthorFemale1 0.0162 0.0179 0.91 0.36297
## Year1997 0.0632 0.0719 0.88 0.37945
## Year1998 -0.0884 0.0708 -1.25 0.21178
## Year1999 -0.1190 0.0690 -1.73 0.08469 .
## Year2000 -0.0667 0.0757 -0.88 0.37826
## Year2001 -0.1772 0.0739 -2.40 0.01656 *
## Year2002 -0.1321 0.0686 -1.93 0.05414 .
## Year2003 -0.1959 0.0709 -2.76 0.00580 **
## Year2004 -0.1551 0.0714 -2.17 0.03004 *
## Year2005 -0.1685 0.0717 -2.35 0.01891 *
## Year2006 -0.1549 0.0699 -2.22 0.02675 *
```

```

## Year2007          -0.1744      0.0683   -2.55  0.01075 *
## Year2008          -0.2424      0.0681   -3.56  0.00039 ***
## Year2009          -0.1138      0.0685   -1.66  0.09669 .
## Year2010          -0.0956      0.0680   -1.41  0.16007
## Year2011          -0.1572      0.0697   -2.26  0.02419 *
## Year2012          -0.1580      0.0691   -2.29  0.02227 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.33
## Multiple R-squared:  0.0407, Adjusted R-squared:  0.03
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 11 is an outlier with |weight| = 0 ( < 6.5e-05);
## 141 weights are ~= 1. The remaining 1398 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0505 0.8700 0.9510 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.059 1          1.029
## Year          1.059 16          1.002

```


Residuals from last author



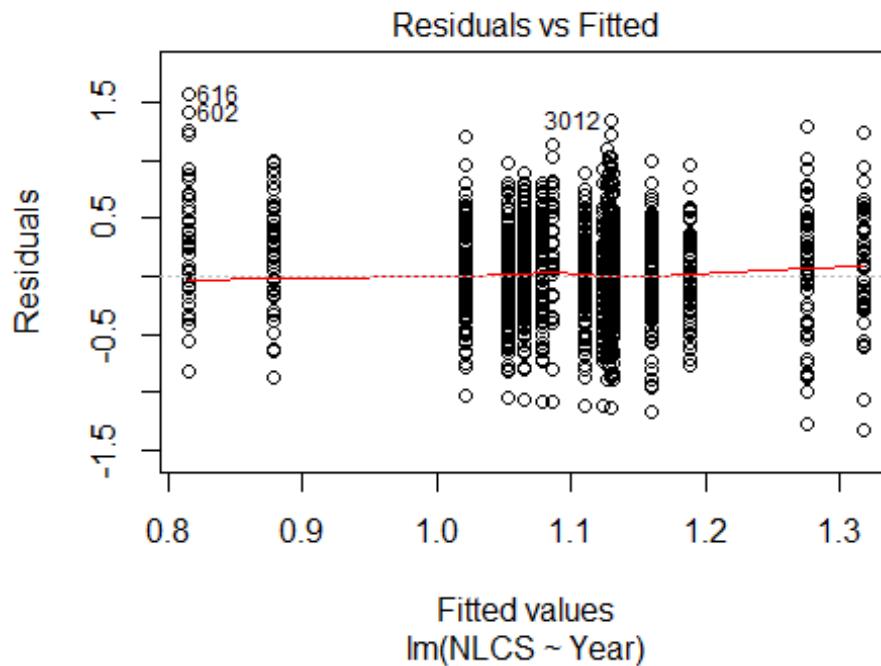
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29192 -0.22082  0.00536  0.21940  1.56567
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22733    0.06065   20.24 < 2e-16 ***
## LastAuthorFemale1 -0.00487    0.02090   -0.23  0.81560
## Year1997         0.06460    0.07172    0.90  0.36793
## Year1998        -0.08646    0.07043   -1.23  0.21976
## Year1999        -0.11763    0.06877   -1.71  0.08737 .
## Year2000        -0.06476    0.07548   -0.86  0.39099
## Year2001        -0.17565    0.07387   -2.38  0.01754 *
## Year2002        -0.13027    0.06841   -1.90  0.05706 .
## Year2003        -0.19485    0.07086   -2.75  0.00603 **
## Year2004        -0.15331    0.07114   -2.16  0.03131 *
## Year2005        -0.16369    0.07127   -2.30  0.02177 *
## Year2006        -0.15234    0.06964   -2.19  0.02886 *
```

```

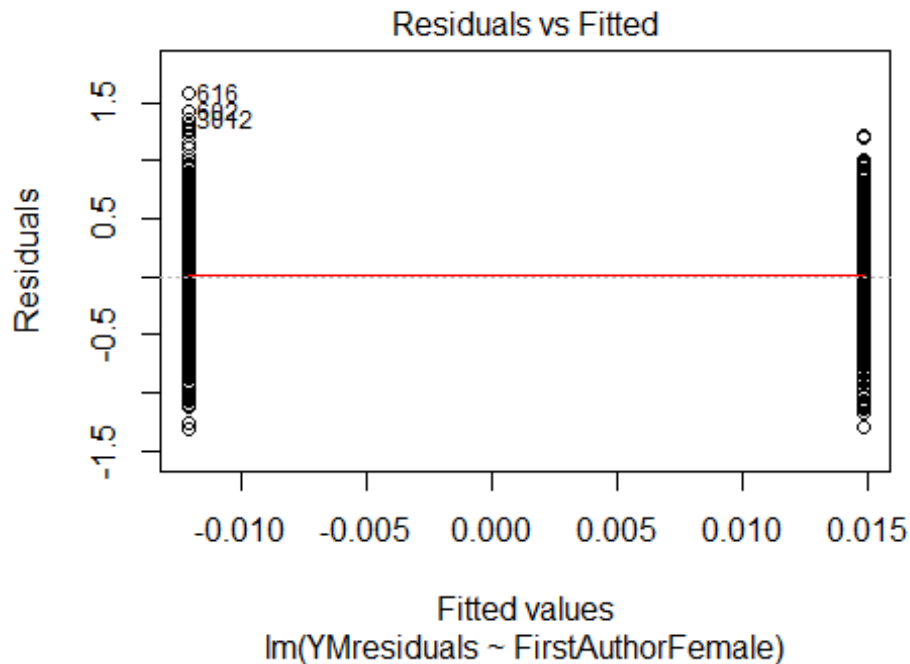
## Year2007          -0.17291      0.06811      -2.54   0.01122  *
## Year2008          -0.23869      0.06773      -3.52   0.00044  ***
## Year2009          -0.11059      0.06819      -1.62   0.10506
## Year2010          -0.09232      0.06780      -1.36   0.17346
## Year2011          -0.15341      0.06941      -2.21   0.02723  *
## Year2012          -0.15460      0.06879      -2.25   0.02476  *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.33
## Multiple R-squared:  0.0402, Adjusted R-squared:  0.0295
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 11 is an outlier with |weight| = 0 ( < 6.5e-05);
## 130 weights are ~= 1. The remaining 1409 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0433 0.8680 0.9510 0.8980 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.49e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1540"
## [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
##   89  106  142   91  112  139   91  105  145  142  141  160  146  189  216
## 2011 2012
##   220  189
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   66   53   62   44   72   66   73   88   98  102  108  131  113  149  170

```

```
## 2011 2012
## 174 152
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 63 44 53 36 70 56 70 85 91 88 99 118 101 134 148
## 2011 2012
## 152 126
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 130, df = 16, p-value <2e-16
```

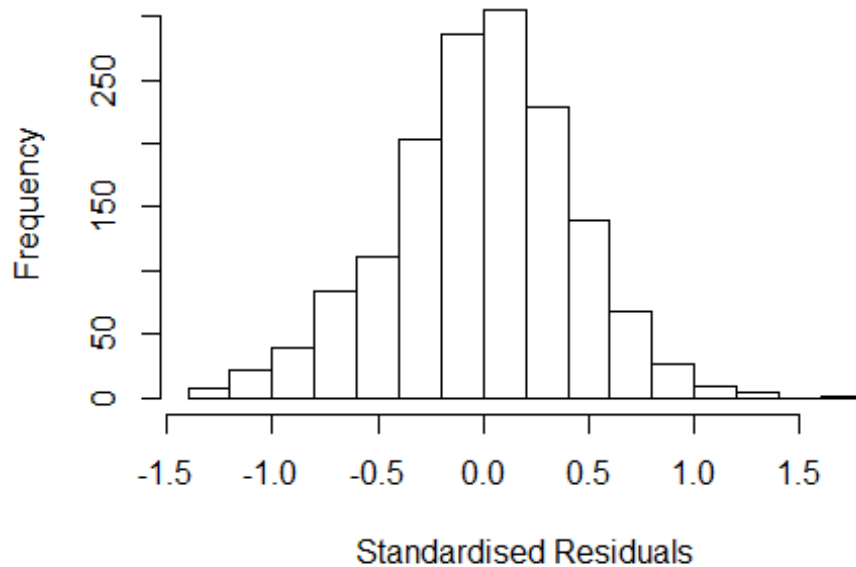


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 10, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 3.95591080496635"
## [1] "Male first author team size 2018 geometric mean: 3.40679152788803"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4700, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.60718515735286"
## [1] "Male last author team size 2018 geometric mean: 3.75944495427504"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3800, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.072 1      1.035
## LastAuthorFemale  1.074 1      1.037
## UniqueAuthors    1.518 4      1.054
## Year              1.639 16     1.016
```

Residuals from first and last author and team size



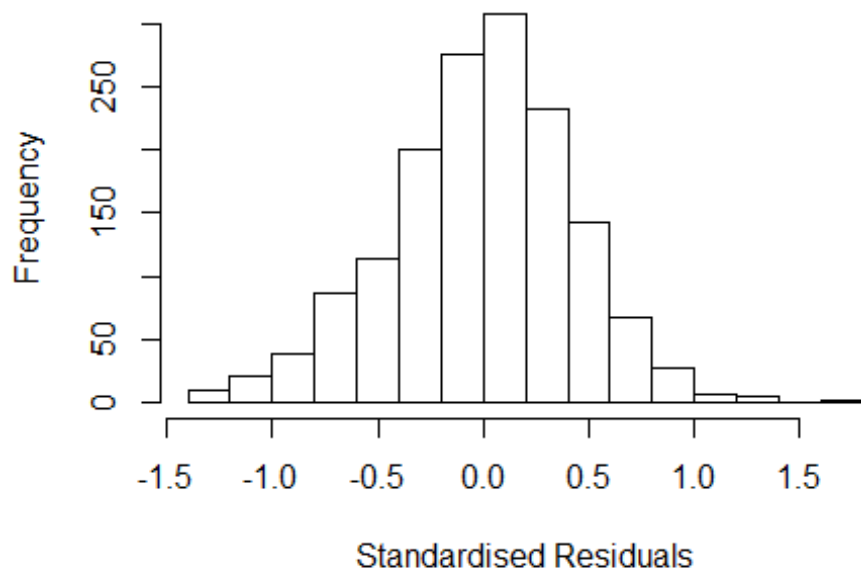
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3458 -0.2719  0.0105  0.2790  1.6812
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8492    0.0956   8.88  <2e-16 ***
## FirstAuthorFemale1 0.0211    0.0228   0.92  0.3556
## LastAuthorFemale1 0.0278    0.0235   1.18  0.2386
## UniqueAuthors2    0.0715    0.0481   1.49  0.1375
## UniqueAuthors3    0.0523    0.0487   1.08  0.2822
## UniqueAuthors4    0.0372    0.0521   0.71  0.4757
## UniqueAuthors5    0.1256    0.0540   2.33  0.0202 *
## Year1997          0.3933    0.1225   3.21  0.0013 **
## Year1998          0.3852    0.1275   3.02  0.0026 **
## Year1999          0.3499    0.1318   2.65  0.0080 **
```

```

## Year2000          -0.1886      0.1422    -1.33    0.1850
## Year2001           0.2441      0.1043     2.34    0.0194 *
## Year2002           0.2191      0.1042     2.10    0.0357 *
## Year2003           0.1908      0.1004     1.90    0.0574 .
## Year2004           0.0909      0.1044     0.87    0.3839
## Year2005           0.1697      0.1030     1.65    0.0996 .
## Year2006           0.2419      0.1025     2.36    0.0185 *
## Year2007           0.1821      0.0977     1.86    0.0626 .
## Year2008           0.2043      0.1002     2.04    0.0416 *
## Year2009           0.1871      0.0962     1.95    0.0519 .
## Year2010           0.1522      0.0990     1.54    0.1246
## Year2011           0.1569      0.0973     1.61    0.1070
## Year2012           0.1905      0.0997     1.91    0.0563 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.0658, Adjusted R-squared:  0.0522
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1405 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0528 0.8590 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          6.52e-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 1.070 1 1.034
## LastAuthorFemale 1.070 1 1.034
## Year 1.105 16 1.003

```

Residuals from first and last author



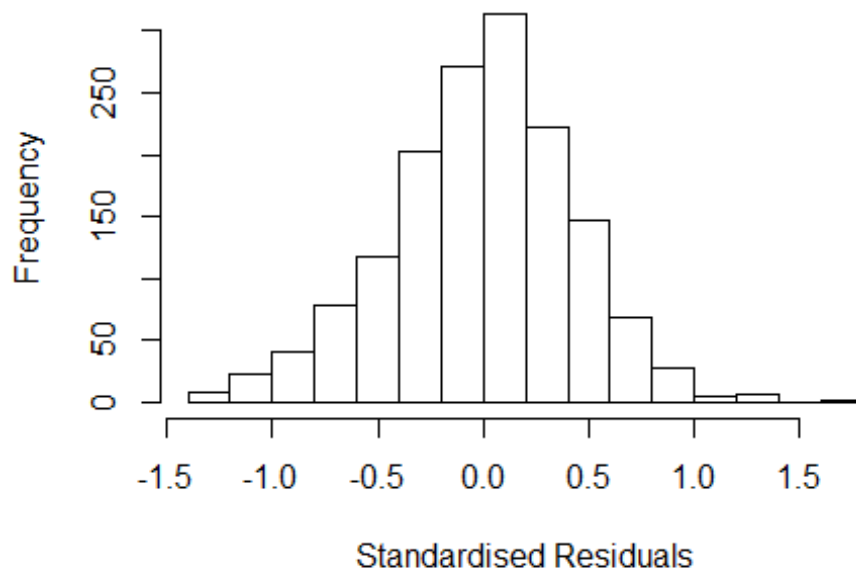
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2960 -0.2743 0.0186 0.2809 1.6642
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8887 0.0901 9.86 <2e-16 ***
## FirstAuthorFemale1 0.0204 0.0229 0.89 0.3739
## LastAuthorFemale1 0.0240 0.0237 1.01 0.3110
## Year1997 0.4027 0.1229 3.28 0.0011 **
## Year1998 0.3869 0.1274 3.04 0.0024 **
## Year1999 0.3578 0.1311 2.73 0.0064 **
## Year2000 -0.1739 0.1430 -1.22 0.2242
## Year2001 0.2638 0.1020 2.59 0.0098 **
## Year2002 0.2452 0.1023 2.40 0.0167 *
## Year2003 0.2180 0.0988 2.21 0.0276 *
## Year2004 0.1153 0.1019 1.13 0.2584
## Year2005 0.2013 0.1008 2.00 0.0459 *
```

```

## Year2006          0.2715      0.1012      2.68    0.0074 **
## Year2007          0.2065      0.0958      2.16    0.0313 *
## Year2008          0.2299      0.0986      2.33    0.0198 *
## Year2009          0.2118      0.0940      2.25    0.0244 *
## Year2010          0.1770      0.0970      1.82    0.0684 .
## Year2011          0.1837      0.0951      1.93    0.0535 .
## Year2012          0.2226      0.0977      2.28    0.0228 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.408
## Multiple R-squared:  0.0599, Adjusted R-squared:  0.0487
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1415 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.059  0.861  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      6.52e-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 1.047 1      1.023
## Year              1.047 16      1.001

```


Residuals from first author



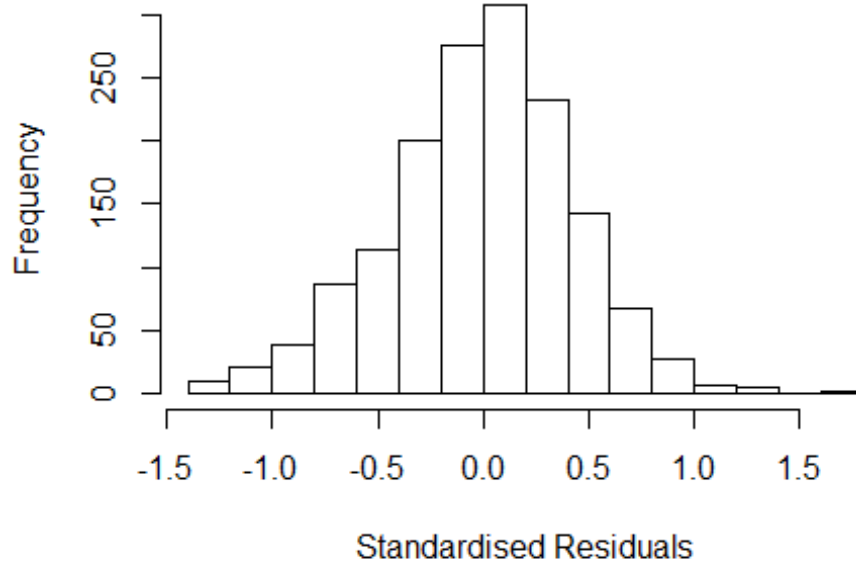
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3083 -0.2765 0.0166 0.2801 1.6585
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8959 0.0897 9.99 < 2e-16 ***
## FirstAuthorFemale1 0.0253 0.0228 1.11 0.26694
## Year1997 0.4082 0.1225 3.33 0.00088 ***
## Year1998 0.3872 0.1274 3.04 0.00242 **
## Year1999 0.3572 0.1314 2.72 0.00666 **
## Year2000 -0.1754 0.1437 -1.22 0.22252
## Year2001 0.2650 0.1019 2.60 0.00940 **
## Year2002 0.2434 0.1022 2.38 0.01740 *
## Year2003 0.2173 0.0988 2.20 0.02796 *
## Year2004 0.1133 0.1020 1.11 0.26675
## Year2005 0.1985 0.1007 1.97 0.04886 *
## Year2006 0.2699 0.1010 2.67 0.00759 **
```

```

## Year2007          0.2050      0.0958      2.14  0.03253 *
## Year2008          0.2279      0.0986      2.31  0.02099 *
## Year2009          0.2114      0.0940      2.25  0.02462 *
## Year2010          0.1758      0.0970      1.81  0.07011 .
## Year2011          0.1808      0.0950      1.90  0.05718 .
## Year2012          0.2198      0.0977      2.25  0.02463 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.408
## Multiple R-squared:  0.0592, Adjusted R-squared:  0.0487
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1418 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0611 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      6.52e-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 1.049 1      1.024
## Year              1.049 16      1.001

```

Residuals from last author



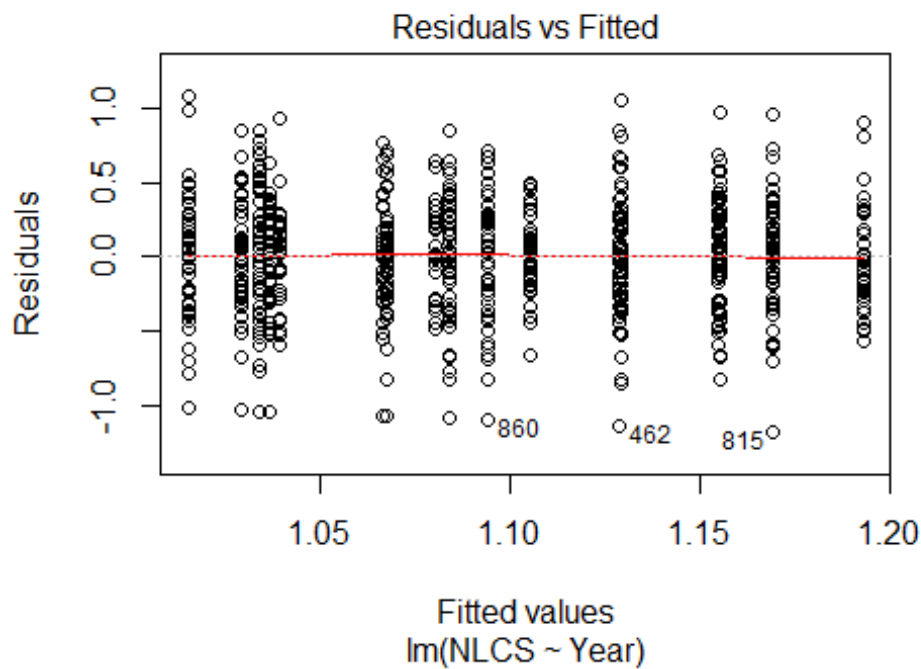
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2977 -0.2775 0.0144 0.2784 1.6574
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8928 0.0899 9.93 < 2e-16 ***
## LastAuthorFemale1 0.0286 0.0235 1.22 0.22394
## Year1997 0.4048 0.1226 3.30 0.00098 ***
## Year1998 0.3882 0.1271 3.06 0.00229 **
## Year1999 0.3587 0.1312 2.73 0.00632 **
## Year2000 -0.1712 0.1433 -1.19 0.23237
## Year2001 0.2673 0.1015 2.63 0.00855 **
## Year2002 0.2486 0.1020 2.44 0.01494 *
## Year2003 0.2242 0.0982 2.28 0.02253 *
## Year2004 0.1211 0.1012 1.20 0.23150
## Year2005 0.2041 0.1004 2.03 0.04232 *
## Year2006 0.2761 0.1007 2.74 0.00619 **
```

```

## Year2007          0.2106      0.0955      2.21  0.02755 *
## Year2008          0.2331      0.0983      2.37  0.01783 *
## Year2009          0.2148      0.0937      2.29  0.02196 *
## Year2010          0.1823      0.0964      1.89  0.05879 .
## Year2011          0.1892      0.0945      2.00  0.04534 *
## Year2012          0.2272      0.0972      2.34  0.01948 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.408
## Multiple R-squared:  0.0595, Adjusted R-squared:  0.049
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1412 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0614 0.8600 0.9480 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.52e-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: 1534"
## [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
##   51   49   51   54   58   58   61   46   31   51   59   59   68   69   59
## 2011 2012
##   58   65
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   35   34   39   42   41   40   48   28   28   37   48   42   52   51   48
## 2011 2012

```

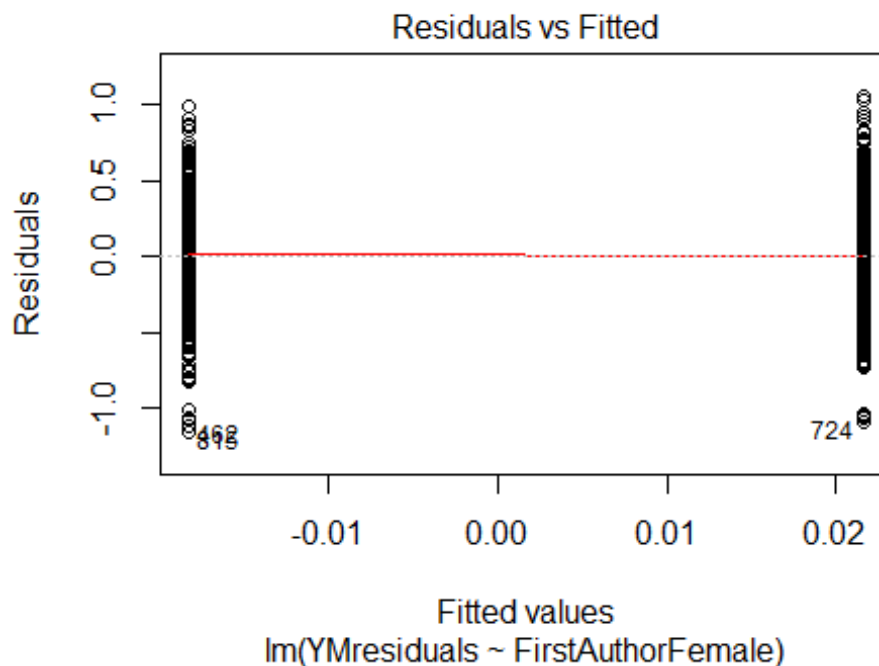
```
## 46 58
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 33 29 33 38 35 34 41 25 21 31 44 34 47 45 45
## 2011 2012
## 37 49
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 24, df = 16, p-value = 0.1
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0054, df = 1, p-value = 0.9
## [1] "Female first author team size 2018 geometric mean: 4.4835237898047"
## [1] "Male first author team size 2018 geometric mean: 2.86852023582314"
## Warning in wilcox.test.default(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 = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.21120935792528"
## [1] "Male last author team size 2018 geometric mean: 3.59944179069414"

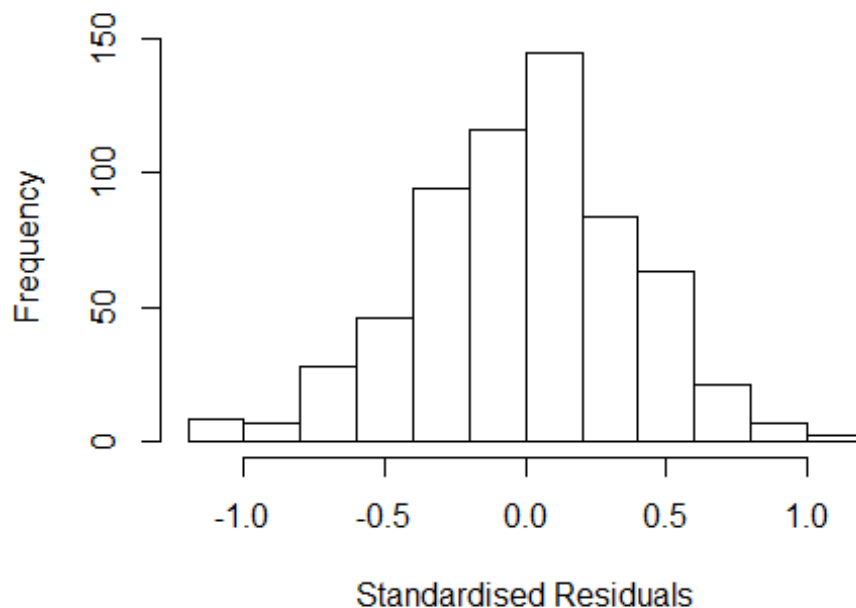
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 520, p-value = 0.5
## 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	1.140	1	1.068
LastAuthorFemale	1.248	1	1.117
UniqueAuthors	1.771	4	1.074
Year	2.094	16	1.023

Residuals from first and last author and team size



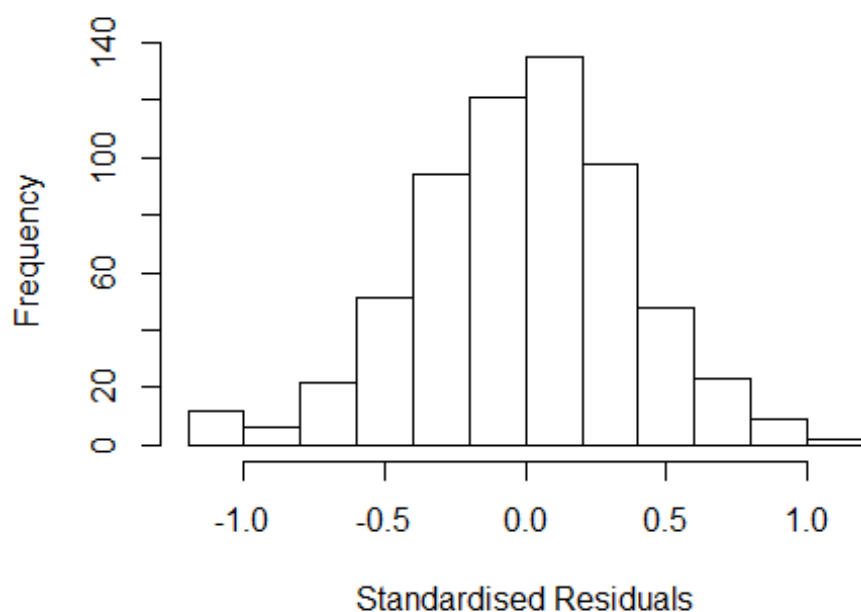
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18487 -0.25416 0.00883 0.25025 1.06900
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8657 0.0848 10.21 <2e-16 ***
## FirstAuthorFemale1 0.0204 0.0323 0.63 0.5281
## LastAuthorFemale1 0.0360 0.0346 1.04 0.2987
## UniqueAuthors2 0.2048 0.0772 2.65 0.0082 **
## UniqueAuthors3 0.1539 0.0751 2.05 0.0409 *
## UniqueAuthors4 0.1743 0.0769 2.27 0.0238 *
## UniqueAuthors5 0.2427 0.0783 3.10 0.0020 **
## Year1997 0.0940 0.0855 1.10 0.2717
## Year1998 0.1143 0.0823 1.39 0.1654
## Year1999 0.0476 0.0725 0.66 0.5117
```

```

## Year2000          0.0478      0.0894      0.53      0.5933
## Year2001         -0.0026      0.0934     -0.03      0.9778
## Year2002          0.0160      0.0913      0.18      0.8610
## Year2003          0.0793      0.1014      0.78      0.4349
## Year2004          0.0305      0.1074      0.28      0.7769
## Year2005          0.0703      0.0903      0.78      0.4363
## Year2006         -0.0593      0.0875     -0.68      0.4984
## Year2007          0.0116      0.0792      0.15      0.8840
## Year2008          0.0914      0.0858      1.07      0.2872
## Year2009          0.0464      0.0930      0.50      0.6177
## Year2010          0.0196      0.0938      0.21      0.8342
## Year2011          0.0750      0.0977      0.77      0.4431
## Year2012         -0.0879      0.0897     -0.98      0.3276
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.374
## Multiple R-squared:  0.0458, Adjusted R-squared:  0.0107
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 560 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.294  0.872  0.947   0.902   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.61e-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 1.114 1      1.056
## LastAuthorFemale  1.232 1      1.110
## Year              1.294 16      1.008

```


Residuals from first and last author



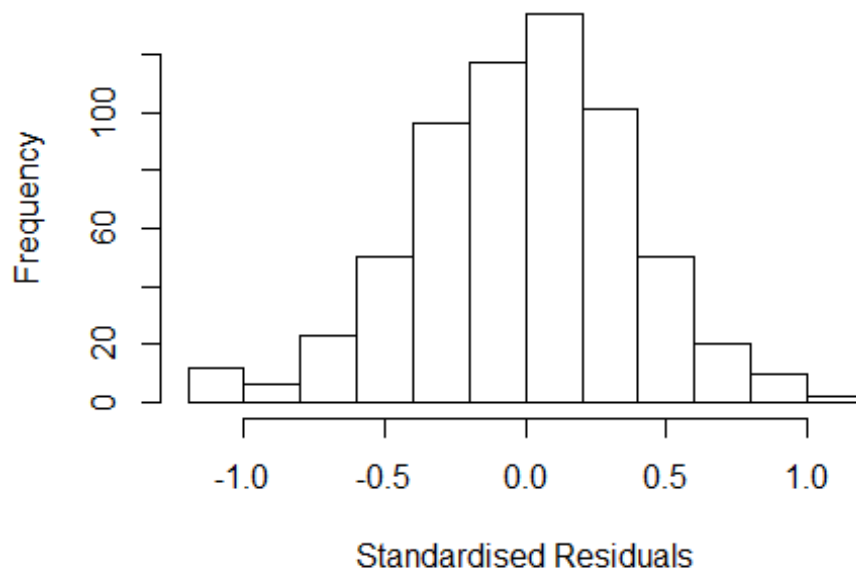
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.15491 -0.26391  0.00166  0.24518  1.07482
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0171     0.0617   16.47  <2e-16 ***
## FirstAuthorFemale1  0.0202     0.0325    0.62   0.53
## LastAuthorFemale1  0.0319     0.0346    0.92   0.36
## Year1997          0.0912     0.0864    1.06   0.29
## Year1998          0.1211     0.0823    1.47   0.14
## Year1999          0.0696     0.0717    0.97   0.33
## Year2000          0.0927     0.0876    1.06   0.29
## Year2001          0.0351     0.0905    0.39   0.70
## Year2002          0.0514     0.0855    0.60   0.55
## Year2003          0.1075     0.0948    1.13   0.26
## Year2004          0.0666     0.1059    0.63   0.53
## Year2005          0.1175     0.0863    1.36   0.17
```

```

## Year2006          -0.0221      0.0864   -0.26      0.80
## Year2007           0.0379      0.0815    0.46      0.64
## Year2008           0.1378      0.0855    1.61      0.11
## Year2009           0.0990      0.0894    1.11      0.27
## Year2010           0.0426      0.0969    0.44      0.66
## Year2011           0.1083      0.0987    1.10      0.27
## Year2012          -0.0483      0.0890   -0.54      0.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.0218, Adjusted R-squared:  -0.0075
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 557 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.314  0.867   0.945   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      1.61e-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 1.097 1      1.047
## Year              1.097 16      1.003

```

Residuals from first author



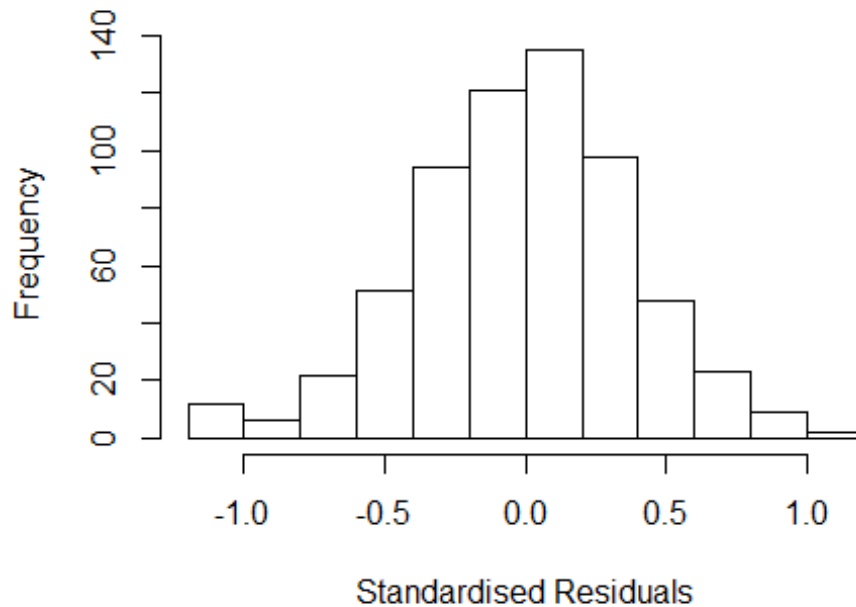
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.16205 -0.27105 0.00708 0.25308 1.05508
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0220 0.0615 16.62 <2e-16 ***
## FirstAuthorFemale1 0.0259 0.0323 0.80 0.42
## Year1997 0.0971 0.0871 1.12 0.27
## Year1998 0.1211 0.0823 1.47 0.14
## Year1999 0.0695 0.0716 0.97 0.33
## Year2000 0.0948 0.0872 1.09 0.28
## Year2001 0.0349 0.0914 0.38 0.70
## Year2002 0.0529 0.0858 0.62 0.54
## Year2003 0.1138 0.0947 1.20 0.23
## Year2004 0.0748 0.1066 0.70 0.48
## Year2005 0.1217 0.0858 1.42 0.16
## Year2006 -0.0129 0.0852 -0.15 0.88
```

```

## Year2007          0.0405      0.0823      0.49      0.62
## Year2008          0.1401      0.0855      1.64      0.10
## Year2009          0.1081      0.0886      1.22      0.22
## Year2010          0.0515      0.0958      0.54      0.59
## Year2011          0.1153      0.0980      1.18      0.24
## Year2012          -0.0413      0.0894     -0.46      0.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.0203, Adjusted R-squared:  -0.00733
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 558 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.307  0.868  0.945  0.898  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      1.61e-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 1.211 1      1.101
## Year              1.211 16      1.006

```

Residuals from last author



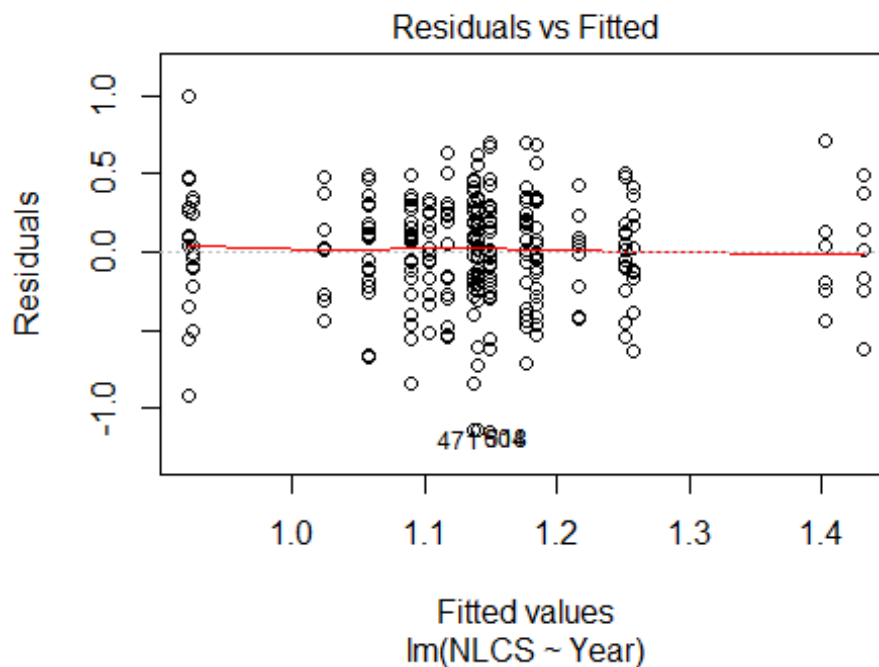
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.16386 -0.26456  0.00861  0.24861  1.08638
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0251     0.0603   16.99  <2e-16 ***
## LastAuthorFemale1  0.0359     0.0343    1.05    0.29
## Year1997          0.0888     0.0854    1.04    0.30
## Year1998          0.1209     0.0821    1.47    0.14
## Year1999          0.0691     0.0715    0.97    0.33
## Year2000          0.0931     0.0877    1.06    0.29
## Year2001          0.0353     0.0902    0.39    0.70
## Year2002          0.0522     0.0854    0.61    0.54
## Year2003          0.1068     0.0948    1.13    0.26
## Year2004          0.0667     0.1065    0.63    0.53
## Year2005          0.1156     0.0868    1.33    0.18
## Year2006         -0.0215     0.0865   -0.25    0.80
```

```

## Year2007          0.0356      0.0814      0.44      0.66
## Year2008          0.1387      0.0860      1.61      0.11
## Year2009          0.0986      0.0897      1.10      0.27
## Year2010          0.0453      0.0975      0.46      0.64
## Year2011          0.1079      0.0985      1.10      0.27
## Year2012         -0.0477      0.0890     -0.54      0.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.0211, Adjusted R-squared:  -0.00652
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 563 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.308  0.871  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      1.61e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 621"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
##   14   22   12   20   21   23   16   21   22   32   20   34   32   35   46
## 2011 2012
##   52   42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   11    8    6   12   15   13   10   12   22   15   27   23   24   33
## 2011 2012

```

```
## 34 24
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5 10 7 5 9 14 10 10 11 19 13 24 20 21 28
## 2011 2012
## 27 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 = 17, df = 16, p-value = 0.4
```



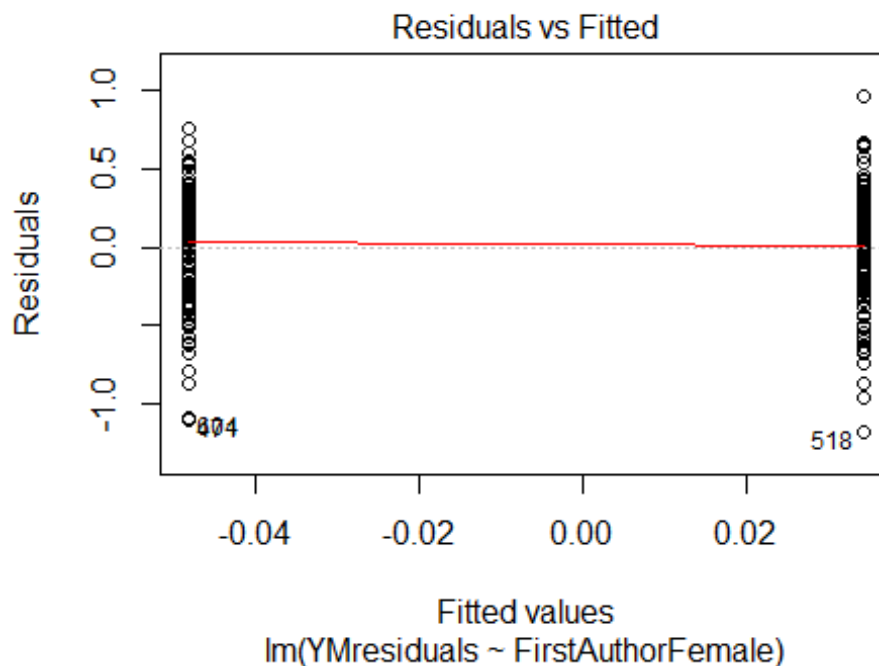
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 6.01586625108859"
## [1] "Male first author team size 2018 geometric mean: 2.99468078411561"

## Warning in wilcox.test.default(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 = 0.005
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.15079753721166"
## [1] "Male last author team size 2018 geometric mean: 4.63655972803564"

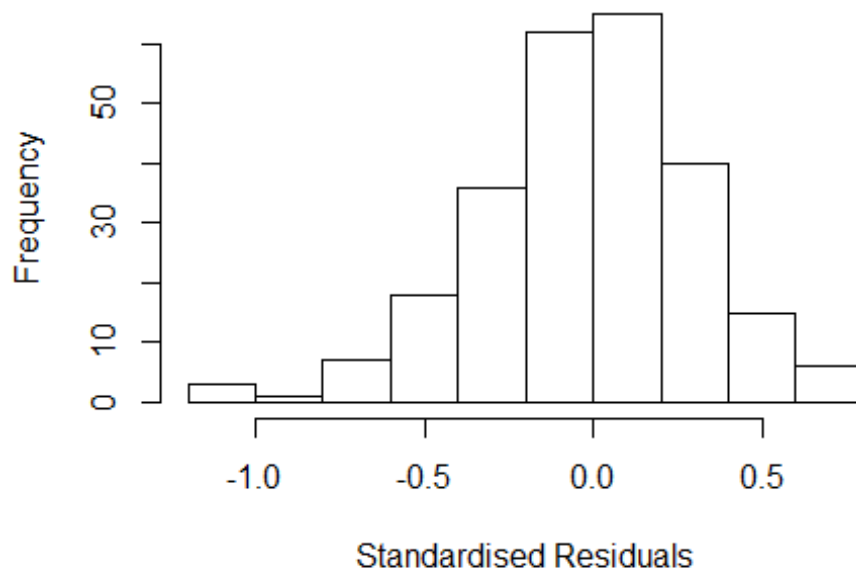
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 74, p-value = 0.9
## 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	1.397	1	1.182
LastAuthorFemale	1.573	1	1.254
UniqueAuthors	3.324	4	1.162
Year	5.399	16	1.054

Residuals from first and last author and team size



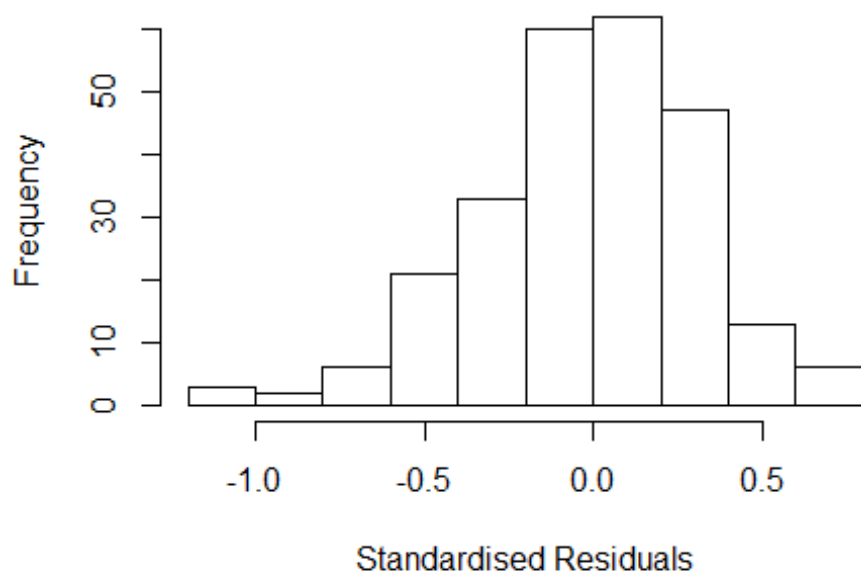
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.148360 -0.208140 -0.000137 0.192191 0.781293
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.40674 0.23775 5.92 1.2e-08 ***
## FirstAuthorFemale1 0.05775 0.04554 1.27 0.206
## LastAuthorFemale1 0.08648 0.04538 1.91 0.058 .
## UniqueAuthors2 0.02424 0.10149 0.24 0.811
## UniqueAuthors3 0.00437 0.10094 0.04 0.965
## UniqueAuthors4 0.06791 0.10870 0.62 0.533
## UniqueAuthors5 0.11280 0.09835 1.15 0.253
## Year1997 -0.19390 0.23468 -0.83 0.410
## Year1998 -0.43001 0.24529 -1.75 0.081 .
## Year1999 -0.04360 0.25949 -0.17 0.867
```

```

## Year2000      -0.33769    0.23053   -1.46    0.144
## Year2001      -0.33648    0.25580   -1.32    0.190
## Year2002      -0.22585    0.24594   -0.92    0.359
## Year2003      -0.55933    0.23723   -2.36    0.019 *
## Year2004      -0.48119    0.26401   -1.82    0.070 .
## Year2005      -0.36994    0.22561   -1.64    0.102
## Year2006      -0.46381    0.22866   -2.03    0.044 *
## Year2007      -0.29223    0.22660   -1.29    0.198
## Year2008      -0.45742    0.22840   -2.00    0.046 *
## Year2009      -0.48884    0.22362   -2.19    0.030 *
## Year2010      -0.36426    0.22252   -1.64    0.103
## Year2011      -0.34898    0.22304   -1.56    0.119
## Year2012      -0.37119    0.23509   -1.58    0.116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.308
## Multiple R-squared:  0.134, Adjusted R-squared:  0.051
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 229 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.135  0.868  0.947  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          3.95e-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 1.339 1          1.157
## LastAuthorFemale  1.482 1          1.217
## Year              1.968 16          1.021

```

Residuals from first and last author



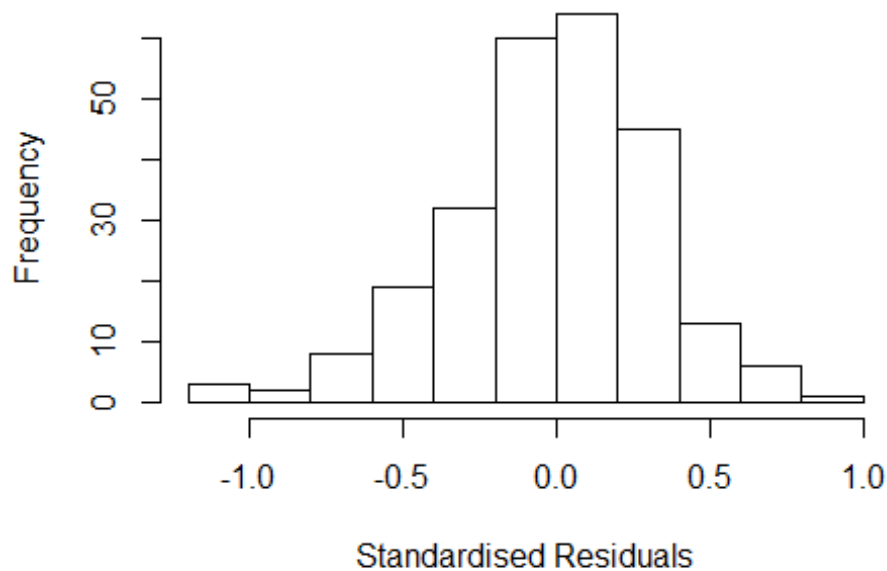
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.13804 -0.21730 0.00843 0.21270 0.78096
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4549 0.2329 6.25 2e-09 ***
## FirstAuthorFemale1 0.0486 0.0466 1.04 0.299
## LastAuthorFemale1 0.0757 0.0445 1.70 0.090 .
## Year1997 -0.1953 0.2488 -0.78 0.433
## Year1998 -0.4418 0.2565 -1.72 0.086 .
## Year1999 -0.0604 0.2767 -0.22 0.827
## Year2000 -0.3377 0.2435 -1.39 0.167
## Year2001 -0.3462 0.2618 -1.32 0.187
## Year2002 -0.2194 0.2543 -0.86 0.389
## Year2003 -0.5519 0.2488 -2.22 0.028 *
## Year2004 -0.4412 0.2809 -1.57 0.118
## Year2005 -0.3508 0.2427 -1.45 0.150
```

```

## Year2006          -0.4629      0.2416    -1.92     0.057 .
## Year2007          -0.2982      0.2423    -1.23     0.220
## Year2008          -0.4435      0.2421    -1.83     0.068 .
## Year2009          -0.4749      0.2401    -1.98     0.049 *
## Year2010          -0.3363      0.2387    -1.41     0.160
## Year2011          -0.3359      0.2384    -1.41     0.160
## Year2012          -0.3470      0.2492    -1.39     0.165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.311
## Multiple R-squared:  0.117, Adjusted R-squared:  0.0494
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 230 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.152  0.871  0.946  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      3.95e-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 1.355 1      1.164
## Year              1.355 16      1.010

```

Residuals from first author



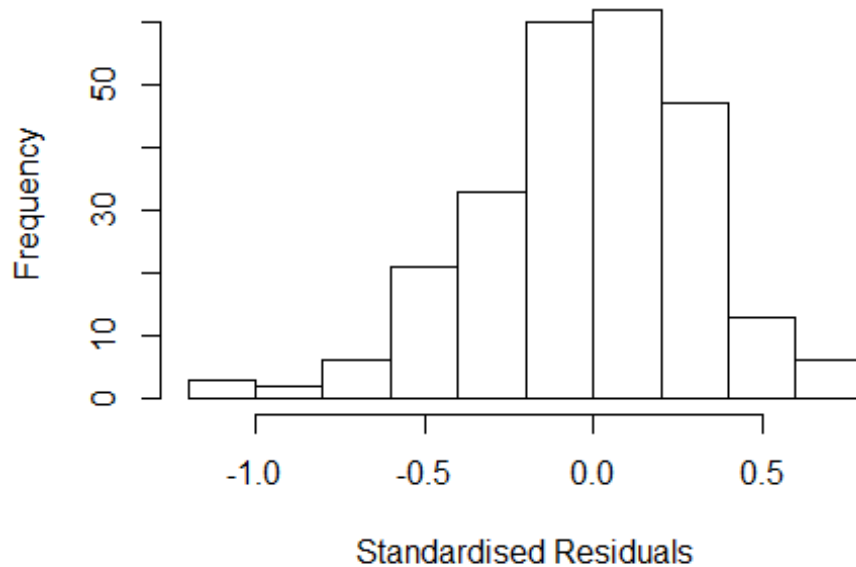
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.14101 -0.20077  0.00819  0.20699  0.80440
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5078     0.2413   6.25 1.9e-09 ***
## FirstAuthorFemale1  0.0713     0.0484   1.47  0.142
## Year1997         -0.2457     0.2568  -0.96  0.340
## Year1998         -0.4685     0.2680  -1.75  0.082 .
## Year1999         -0.0762     0.2951  -0.26  0.796
## Year2000         -0.3482     0.2561  -1.36  0.175
## Year2001         -0.3921     0.2687  -1.46  0.146
## Year2002         -0.2444     0.2633  -0.93  0.354
## Year2003         -0.6071     0.2558  -2.37  0.018 *
## Year2004         -0.4645     0.2911  -1.60  0.112
## Year2005         -0.3618     0.2540  -1.42  0.156
## Year2006         -0.4940     0.2515  -1.96  0.051 .
```

```

## Year2007          -0.3370      0.2523   -1.34    0.183
## Year2008          -0.4647      0.2531   -1.84    0.068 .
## Year2009          -0.5074      0.2511   -2.02    0.044 *
## Year2010          -0.3668      0.2490   -1.47    0.142
## Year2011          -0.3680      0.2484   -1.48    0.140
## Year2012          -0.3774      0.2595   -1.45    0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.309
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0438
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 232 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.143  0.879   0.942   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      3.95e-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 1.427 1          1.195
## Year            1.427 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
## -1.14328 -0.20328 -0.00358 0.20034 0.80459
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4456 0.2312 6.25 1.9e-09 ***
## LastAuthorFemale1 0.0893 0.0453 1.97 0.050 *
## Year1997 -0.1473 0.2449 -0.60 0.548
## Year1998 -0.4188 0.2587 -1.62 0.107
## Year1999 -0.0468 0.2719 -0.17 0.864
## Year2000 -0.3163 0.2433 -1.30 0.195
## Year2001 -0.3249 0.2608 -1.25 0.214
## Year2002 -0.1936 0.2504 -0.77 0.440
## Year2003 -0.5241 0.2457 -2.13 0.034 *
## Year2004 -0.4205 0.2783 -1.51 0.132
## Year2005 -0.3210 0.2414 -1.33 0.185
## Year2006 -0.4282 0.2377 -1.80 0.073 .
```

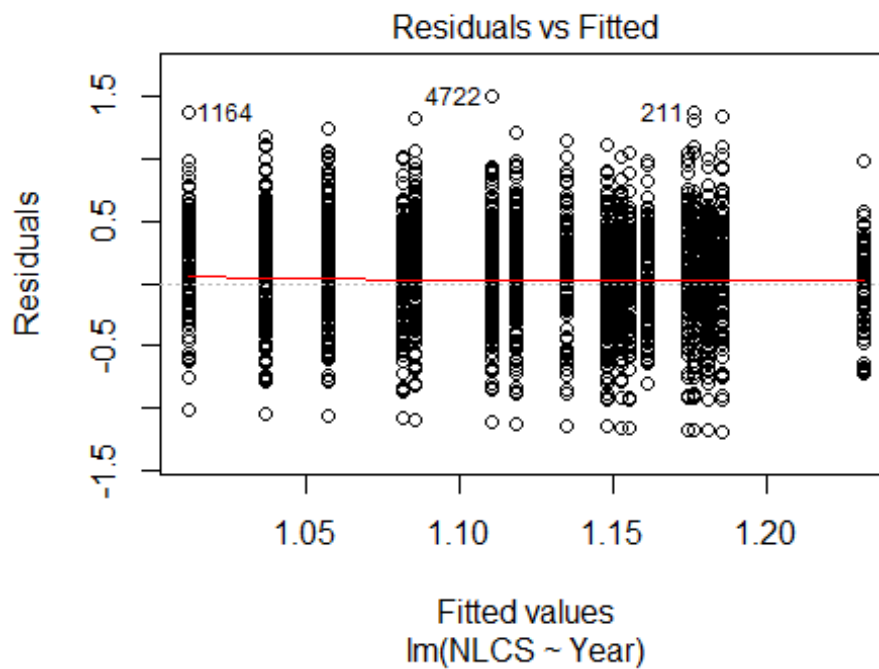
```

## Year2007          -0.2600      0.2376   -1.09    0.275
## Year2008          -0.4070      0.2374   -1.71    0.088 .
## Year2009          -0.4424      0.2364   -1.87    0.063 .
## Year2010          -0.3024      0.2349   -1.29    0.199
## Year2011          -0.3050      0.2345   -1.30    0.195
## Year2012          -0.3027      0.2413   -1.25    0.211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.311
## Multiple R-squared:  0.113, Adjusted R-squared:  0.0492
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 229 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.148  0.868  0.946  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      3.95e-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: 253"
## [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
## 167 224 224 221 217 178 184 187 214 242 276 296 351 393 386
## 2011 2012
## 408 402
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 103 127 105 128 120 87 126 124 153 181 199 209 261 261 269
## 2011 2012

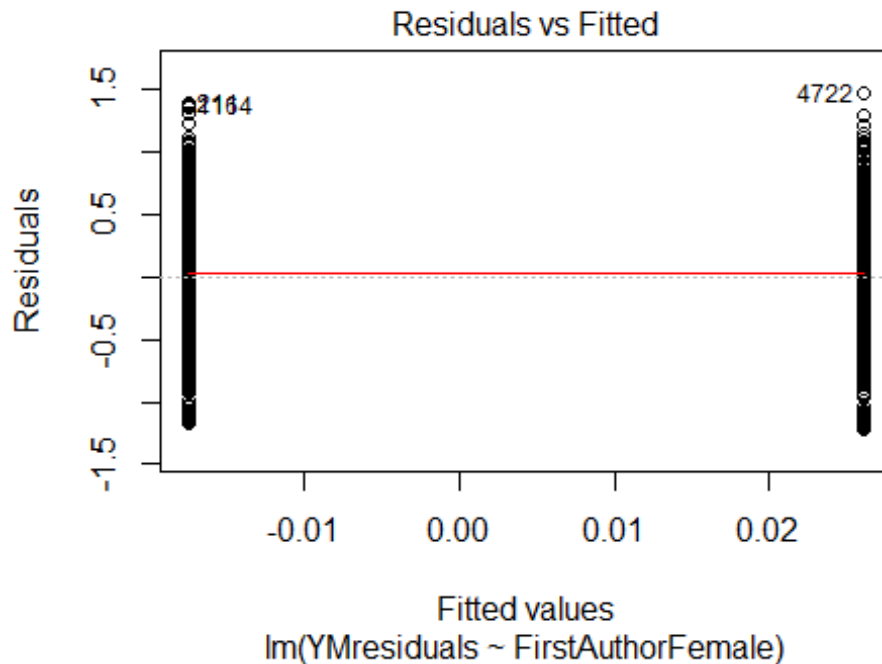
```



```
## 284 298
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 97 118 90 111 109 75 113 112 123 154 163 172 218 230 240
## 2011 2012
## 246 248
## [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 = 6e-15
```

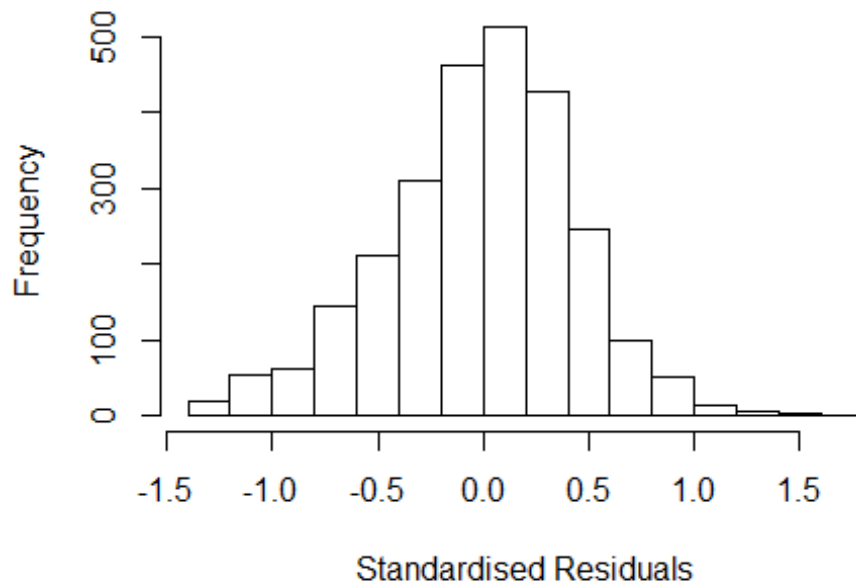


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.8, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 4.95654864952969"
## [1] "Male first author team size 2018 geometric mean: 4.62261412375842"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.9963387981578"
## [1] "Male last author team size 2018 geometric mean: 4.72191952942263"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "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  1.059 1      1.029
## UniqueAuthors    1.204 4      1.023
## Year             1.266 16      1.007
```

Residuals from first and last author and team size



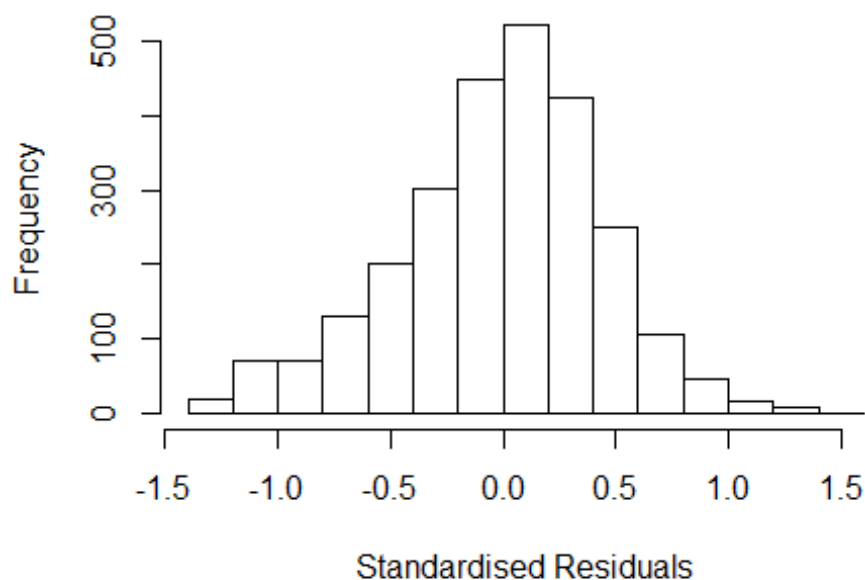
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3682 -0.2761  0.0206  0.2796  1.6125
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8478    0.0639   13.27 < 2e-16 ***
## FirstAuthorFemale1 0.0273    0.0179    1.53  0.1265
## LastAuthorFemale1 0.0794    0.0194    4.10 4.2e-05 ***
## UniqueAuthors2    0.2956    0.0490    6.03 1.9e-09 ***
## UniqueAuthors3    0.3419    0.0467    7.32 3.4e-13 ***
## UniqueAuthors4    0.3336    0.0477    6.99 3.5e-12 ***
## UniqueAuthors5    0.4235    0.0456    9.30 < 2e-16 ***
## Year1997          0.0263    0.0682    0.39  0.7001
## Year1998          0.0728    0.0744    0.98  0.3285
## Year1999          0.0697    0.0672    1.04  0.2995
```

```

## Year2000          -0.1713      0.0815    -2.10    0.0357 *
## Year2001           0.0212      0.0622     0.34    0.7336
## Year2002          -0.0335      0.0606    -0.55    0.5801
## Year2003          -0.0757      0.0599    -1.26    0.2062
## Year2004          -0.1330      0.0612    -2.17    0.0298 *
## Year2005          -0.0865      0.0581    -1.49    0.1365
## Year2006          -0.0516      0.0602    -0.86    0.3908
## Year2007          -0.0351      0.0567    -0.62    0.5359
## Year2008          -0.0934      0.0574    -1.63    0.1039
## Year2009          -0.1126      0.0589    -1.91    0.0563 .
## Year2010          -0.1731      0.0592    -2.92    0.0035 **
## Year2011          -0.1067      0.0567    -1.88    0.0602 .
## Year2012          -0.1766      0.0590    -2.99    0.0028 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.0801, Adjusted R-squared:  0.0723
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 224 weights are ~= 1. The remaining 2395 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0858 0.8560 0.9490 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.82e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1 1.021
## LastAuthorFemale 1.042 1 1.021
## Year 1.085 16 1.003

```

Residuals from first and last author



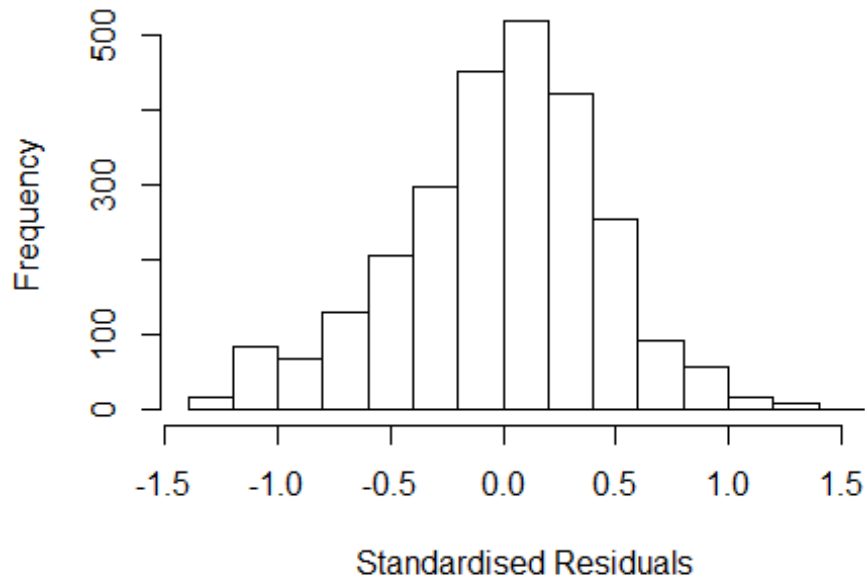
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3234 -0.2777  0.0259  0.2843  1.4066
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13707    0.05856   19.42  < 2e-16 ***
## FirstAuthorFemale1 0.04340    0.01814    2.39  0.01677 *
## LastAuthorFemale1 0.07177    0.01947    3.69  0.00023 ***
## Year1997         0.03022    0.07463    0.40  0.68555
## Year1998         0.07112    0.08167    0.87  0.38394
## Year1999         0.07890    0.07433    1.06  0.28857
## Year2000        -0.14429    0.08640   -1.67  0.09503 .
## Year2001         0.07165    0.06828    1.05  0.29408
## Year2002         0.00974    0.06834    0.14  0.88672
## Year2003        -0.01563    0.06784   -0.23  0.81782
## Year2004        -0.10017    0.06980   -1.44  0.15138
## Year2005        -0.02550    0.06561   -0.39  0.69754
```

```

## Year2006          0.00362    0.06734    0.05  0.95718
## Year2007          0.01479    0.06488    0.23  0.81968
## Year2008         -0.02990    0.06525   -0.46  0.64680
## Year2009         -0.05596    0.06642   -0.84  0.39964
## Year2010         -0.11408    0.06715   -1.70  0.08942 .
## Year2011         -0.04981    0.06494   -0.77  0.44315
## Year2012         -0.11202    0.06699   -1.67  0.09461 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.414
## Multiple R-squared:  0.0258, Adjusted R-squared:  0.019
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 223 weights are ~= 1. The remaining 2396 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.225  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      3.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1         1.022
## Year              1.044 16         1.001

```

Residuals from first author



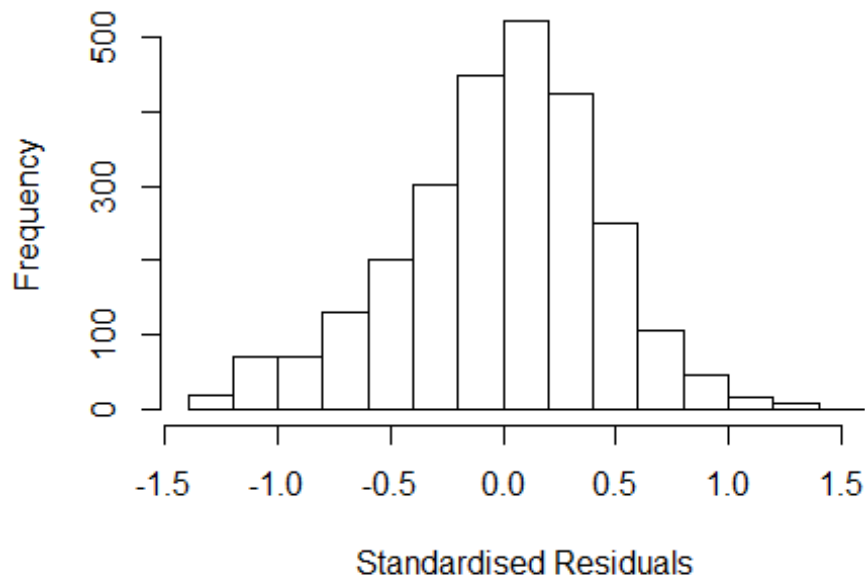
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2799 -0.2837 0.0201 0.2823 1.4543
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14615 0.05867 19.53 <2e-16 ***
## FirstAuthorFemale1 0.04960 0.01818 2.73 0.0064 **
## Year1997 0.03998 0.07450 0.54 0.5915
## Year1998 0.07952 0.08201 0.97 0.3323
## Year1999 0.08413 0.07464 1.13 0.2598
## Year2000 -0.12823 0.08655 -1.48 0.1386
## Year2001 0.08063 0.06839 1.18 0.2386
## Year2002 0.00984 0.06851 0.14 0.8858
## Year2003 -0.00560 0.06779 -0.08 0.9342
## Year2004 -0.09151 0.07025 -1.30 0.1928
## Year2005 -0.02400 0.06568 -0.37 0.7148
## Year2006 0.01559 0.06710 0.23 0.8163
```

```

## Year2007          0.02604    0.06491    0.40    0.6883
## Year2008          -0.02509    0.06539   -0.38    0.7012
## Year2009          -0.04848    0.06644   -0.73    0.4657
## Year2010          -0.10676    0.06729   -1.59    0.1127
## Year2011          -0.04108    0.06492   -0.63    0.5270
## Year2012          -0.10333    0.06708   -1.54    0.1236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.414
## Multiple R-squared:  0.0206, Adjusted R-squared:  0.0142
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 229 weights are ~= 1. The remaining 2390 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.192  0.859  0.949  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      3.82e-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 1.044 1          1.022
## Year            1.044 16          1.001

```


Residuals from last author



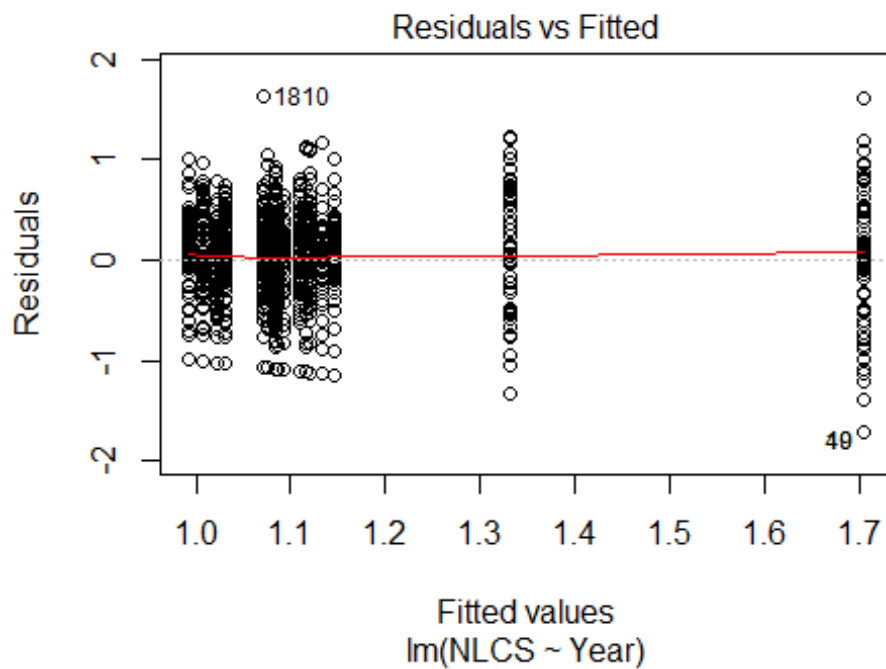
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2977 -0.2821 0.0238 0.2822 1.4259
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14946 0.05887 19.53 < 2e-16 ***
## LastAuthorFemale1 0.07638 0.01953 3.91 9.4e-05 ***
## Year1997 0.03164 0.07514 0.42 0.674
## Year1998 0.07188 0.08245 0.87 0.383
## Year1999 0.07882 0.07452 1.06 0.290
## Year2000 -0.14277 0.08662 -1.65 0.099 .
## Year2001 0.07406 0.06859 1.08 0.280
## Year2002 0.01294 0.06863 0.19 0.850
## Year2003 -0.01104 0.06832 -0.16 0.872
## Year2004 -0.09812 0.07031 -1.40 0.163
## Year2005 -0.02329 0.06624 -0.35 0.725
## Year2006 0.00952 0.06771 0.14 0.888
```

```

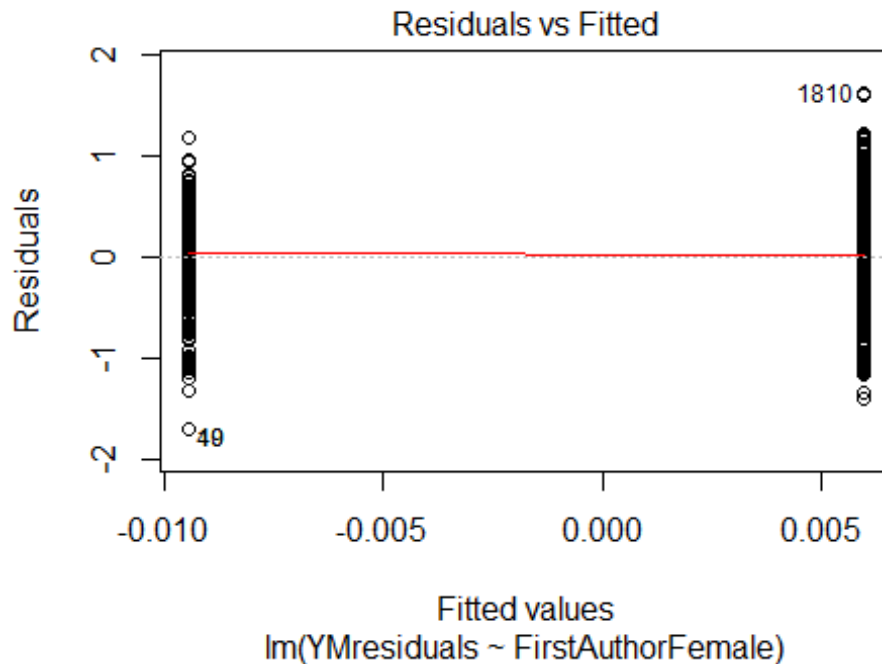
## Year2007      0.01881    0.06534    0.29    0.773
## Year2008     -0.02622    0.06576   -0.40    0.690
## Year2009     -0.05187    0.06688   -0.78    0.438
## Year2010     -0.10721    0.06744   -1.59    0.112
## Year2011     -0.04275    0.06528   -0.65    0.513
## Year2012     -0.10635    0.06755   -1.57    0.116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.415
## Multiple R-squared:  0.0233, Adjusted R-squared:  0.0169
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 214 weights are ~= 1. The remaining 2405 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.213  0.857  0.948  0.891  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2619"
## [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
##  145  134   70   96   81   88  115   98   73   75   98  113  127  133  125
## 2011 2012
##  121  116
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   67   55   43   70   46   50   79   74   51   52   71   84   98  102   91
## 2011 2012

```

```
## 99 96
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 55 47 36 64 39 40 70 65 43 41 66 68 87 90 78
## 2011 2012
## 90 83
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 95, df = 16, p-value = 4e-13
```



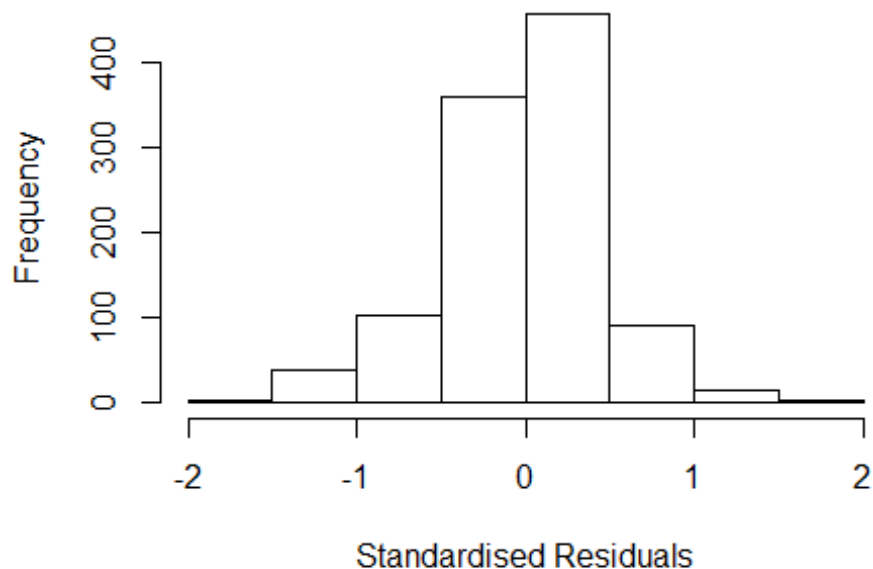
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3, df = 1, p-value = 0.08
## [1] "Female first author team size 2018 geometric mean: 4.07238521723688"
## [1] "Male first author team size 2018 geometric mean: 3.43179731632551"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 900, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.46798462963291"
## [1] "Male last author team size 2018 geometric mean: 3.49504964920412"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 760, 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	1.110	1	1.054
LastAuthorFemale	1.101	1	1.049
UniqueAuthors	1.489	4	1.051
Year	1.700	16	1.017

Residuals from first and last author and team size



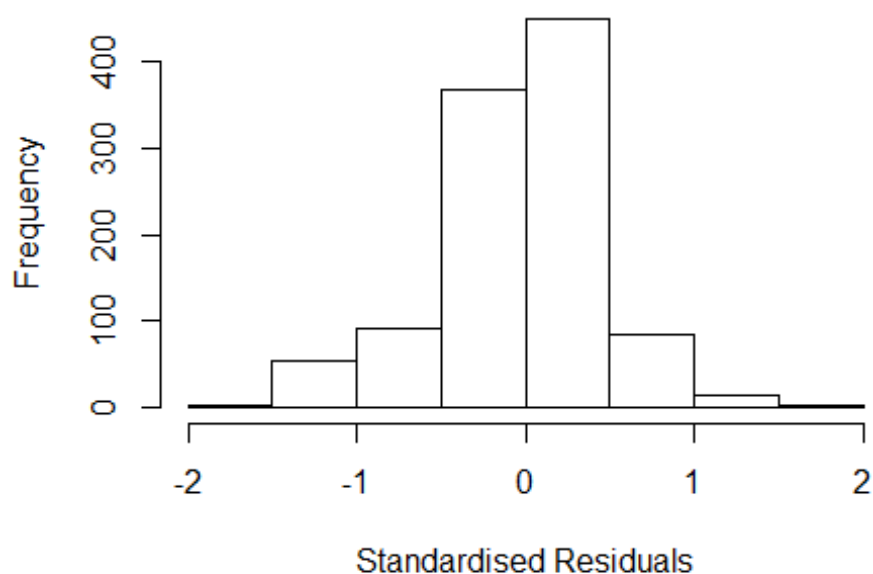
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8772 -0.2715 0.0234 0.2753 1.6769
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5049 0.1265 11.90 < 2e-16 ***
## FirstAuthorFemale1 0.0143 0.0276 0.52 0.60464
## LastAuthorFemale1 -0.0136 0.0297 -0.46 0.64857
## UniqueAuthors2 0.2439 0.0661 3.69 0.00024 ***
## UniqueAuthors3 0.2635 0.0669 3.94 8.6e-05 ***
## UniqueAuthors4 0.2999 0.0697 4.30 1.9e-05 ***
## UniqueAuthors5 0.3580 0.0689 5.20 2.4e-07 ***
## Year1997 -0.4138 0.1687 -2.45 0.01436 *
## Year1998 -0.6027 0.1227 -4.91 1.0e-06 ***
## Year1999 -0.6135 0.1180 -5.20 2.4e-07 ***
```

```

## Year2000          -0.5448      0.1293    -4.21  2.7e-05 ***
## Year2001          -0.5463      0.1225    -4.46  9.0e-06 ***
## Year2002          -0.6093      0.1194    -5.10  4.0e-07 ***
## Year2003          -0.6995      0.1204    -5.81  8.3e-09 ***
## Year2004          -0.7078      0.1314    -5.39  8.8e-08 ***
## Year2005          -0.6712      0.1151    -5.83  7.3e-09 ***
## Year2006          -0.6395      0.1132    -5.65  2.1e-08 ***
## Year2007          -0.6665      0.1137    -5.86  6.2e-09 ***
## Year2008          -0.6779      0.1150    -5.89  5.1e-09 ***
## Year2009          -0.7267      0.1140    -6.37  2.8e-10 ***
## Year2010          -0.7026      0.1164    -6.04  2.2e-09 ***
## Year2011          -0.7639      0.1117    -6.84  1.4e-11 ***
## Year2012          -0.7247      0.1157    -6.27  5.4e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared:  0.136, Adjusted R-squared:  0.118
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 8 is an outlier with |weight| <= 3.4e-07 ( < 9.4e-05);
## 81 weights are ~= 1. The remaining 980 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.041  0.863  0.950  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          9.42e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500          50          2          1          1000          200
## trace.lev mts compute.rd
##           0          1000          0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.135 1          1.065
## LastAuthorFemale  1.094 1          1.046
## Year              1.218 16          1.006

```

Residuals from first and last author



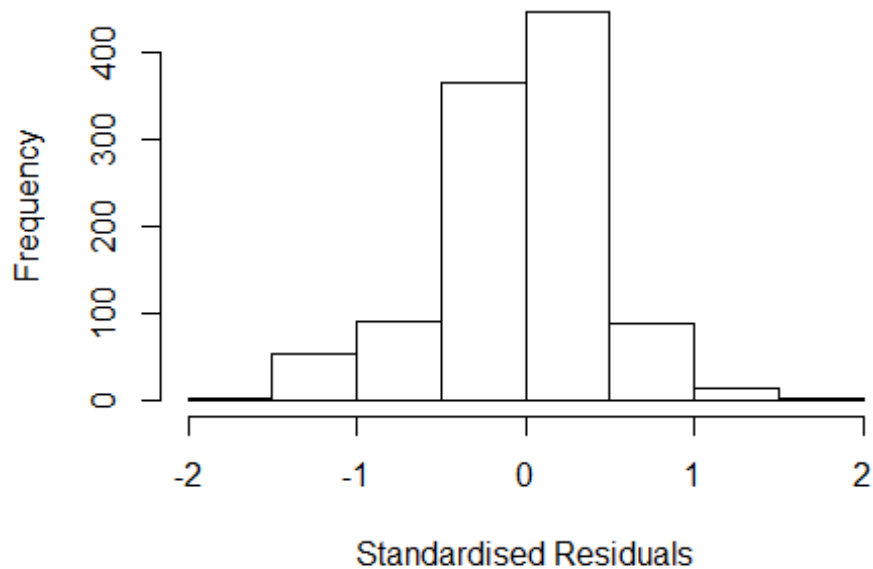
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7496 -0.2779 0.0253 0.2649 1.6493
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.7307 0.1080 16.03 < 2e-16 ***
## FirstAuthorFemale1 0.0189 0.0284 0.67 0.505
## LastAuthorFemale1 -0.0165 0.0303 -0.55 0.585
## Year1997 -0.4177 0.1740 -2.40 0.017 *
## Year1998 -0.5981 0.1225 -4.88 1.2e-06 ***
## Year1999 -0.6301 0.1189 -5.30 1.4e-07 ***
## Year2000 -0.5349 0.1298 -4.12 4.1e-05 ***
## Year2001 -0.5396 0.1246 -4.33 1.6e-05 ***
## Year2002 -0.5944 0.1209 -4.92 1.0e-06 ***
## Year2003 -0.6829 0.1236 -5.52 4.2e-08 ***
## Year2004 -0.6552 0.1347 -4.86 1.3e-06 ***
## Year2005 -0.6154 0.1162 -5.30 1.4e-07 ***
```

```

## Year2006          -0.6096      0.1132   -5.38  9.0e-08 ***
## Year2007          -0.6182      0.1141   -5.42  7.5e-08 ***
## Year2008          -0.6354      0.1153   -5.51  4.5e-08 ***
## Year2009          -0.6810      0.1151   -5.92  4.5e-09 ***
## Year2010          -0.6614      0.1174   -5.64  2.2e-08 ***
## Year2011          -0.7133      0.1125   -6.34  3.5e-10 ***
## Year2012          -0.6775      0.1171   -5.78  9.6e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.104, Adjusted R-squared:  0.0885
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 982 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.020  0.865   0.950   0.886   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.128 1          1.062
## Year              1.128 16          1.004

```


Residuals from first author



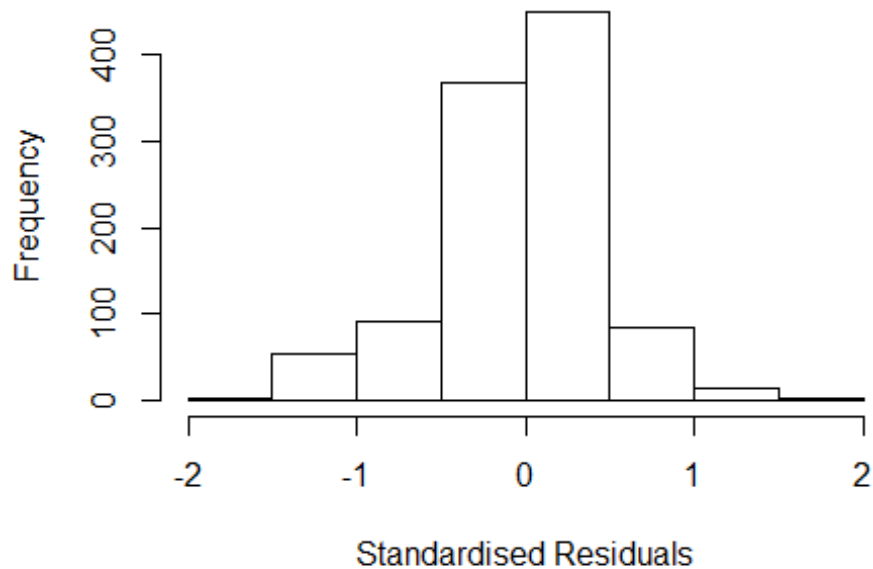
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.744 -0.277 0.023 0.259 1.653
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.7273 0.1076 16.06 < 2e-16 ***
## FirstAuthorFemale1 0.0171 0.0283 0.60 0.546
## Year1997 -0.4171 0.1742 -2.39 0.017 *
## Year1998 -0.5974 0.1221 -4.89 1.2e-06 ***
## Year1999 -0.6312 0.1186 -5.32 1.2e-07 ***
## Year2000 -0.5361 0.1294 -4.14 3.7e-05 ***
## Year2001 -0.5410 0.1242 -4.36 1.5e-05 ***
## Year2002 -0.5955 0.1206 -4.94 9.1e-07 ***
## Year2003 -0.6853 0.1233 -5.56 3.5e-08 ***
## Year2004 -0.6559 0.1345 -4.88 1.2e-06 ***
## Year2005 -0.6147 0.1157 -5.31 1.3e-07 ***
## Year2006 -0.6114 0.1129 -5.42 7.5e-08 ***
```

```

## Year2007          -0.6188      0.1138   -5.44  6.8e-08 ***
## Year2008          -0.6371      0.1148   -5.55  3.7e-08 ***
## Year2009          -0.6811      0.1148   -5.93  4.1e-09 ***
## Year2010          -0.6620      0.1171   -5.65  2.0e-08 ***
## Year2011          -0.7149      0.1121   -6.38  2.7e-10 ***
## Year2012          -0.6783      0.1168   -5.81  8.5e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.104, Adjusted R-squared:  0.0892
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 993 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0218 0.8650 0.9510 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      9.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.088 1      1.043
## Year              1.088 16      1.003

```

Residuals from last author



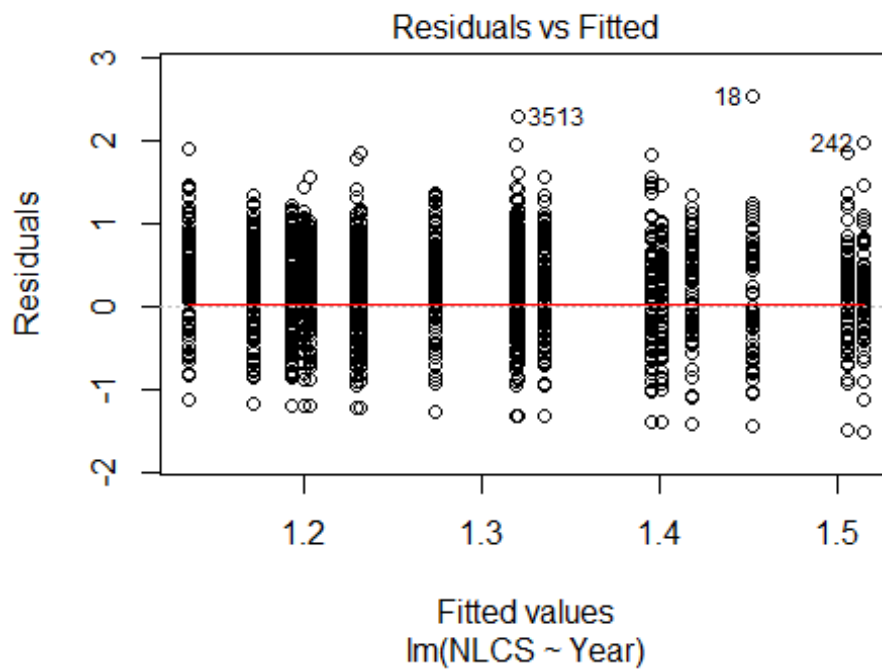
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7393 -0.2731 0.0243 0.2682 1.6419
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.7393 0.1063 16.37 < 2e-16 ***
## LastAuthorFemale1 -0.0142 0.0302 -0.47 0.639
## Year1997 -0.4235 0.1728 -2.45 0.014 *
## Year1998 -0.6012 0.1225 -4.91 1.1e-06 ***
## Year1999 -0.6329 0.1190 -5.32 1.3e-07 ***
## Year2000 -0.5346 0.1297 -4.12 4.1e-05 ***
## Year2001 -0.5436 0.1240 -4.38 1.3e-05 ***
## Year2002 -0.5960 0.1209 -4.93 9.6e-07 ***
## Year2003 -0.6858 0.1233 -5.56 3.4e-08 ***
## Year2004 -0.6569 0.1350 -4.86 1.3e-06 ***
## Year2005 -0.6168 0.1162 -5.31 1.3e-07 ***
## Year2006 -0.6129 0.1130 -5.42 7.3e-08 ***
```

```

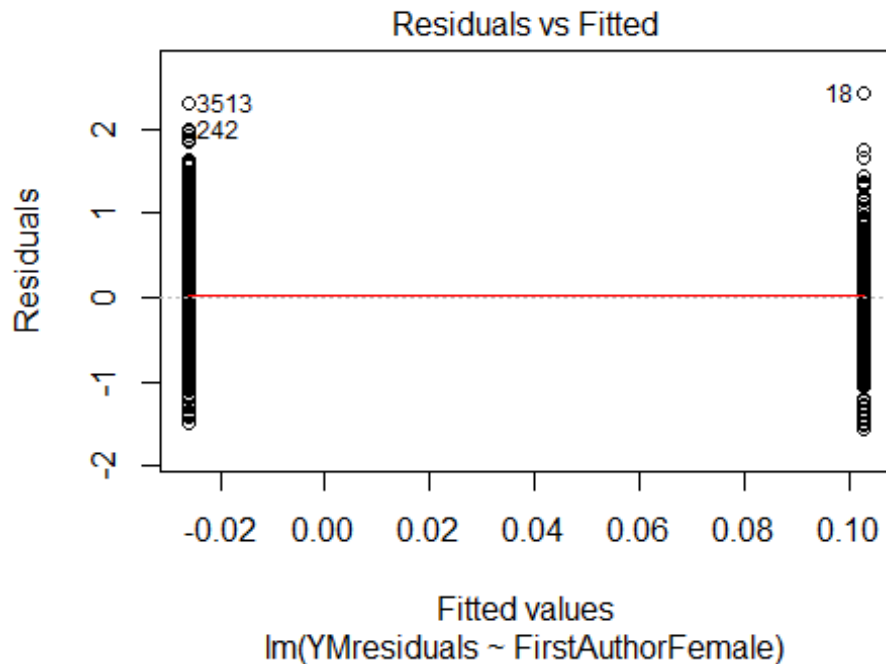
## Year2007          -0.6198      0.1141    -5.43  6.9e-08 ***
## Year2008          -0.6367      0.1154    -5.52  4.3e-08 ***
## Year2009          -0.6822      0.1153    -5.92  4.4e-09 ***
## Year2010          -0.6626      0.1175    -5.64  2.2e-08 ***
## Year2011          -0.7156      0.1124    -6.37  2.9e-10 ***
## Year2012          -0.6766      0.1175    -5.76  1.1e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.104, Adjusted R-squared:  0.0892
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 990 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0228 0.8660 0.9510 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      9.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1062"
## [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
## 117 139 134 155 175 151 174 143 166 180 214 229 230 222 213
## 2011 2012
## 176 194
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 79 89 71 102 105 75 152 127 142 151 184 202 194 196 184
## 2011 2012

```

```
## 157 173
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 75 83 68 93 97 72 144 122 132 143 171 184 175 179 175
## 2011 2012
## 140 152
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 27, df = 16, p-value = 0.04
```

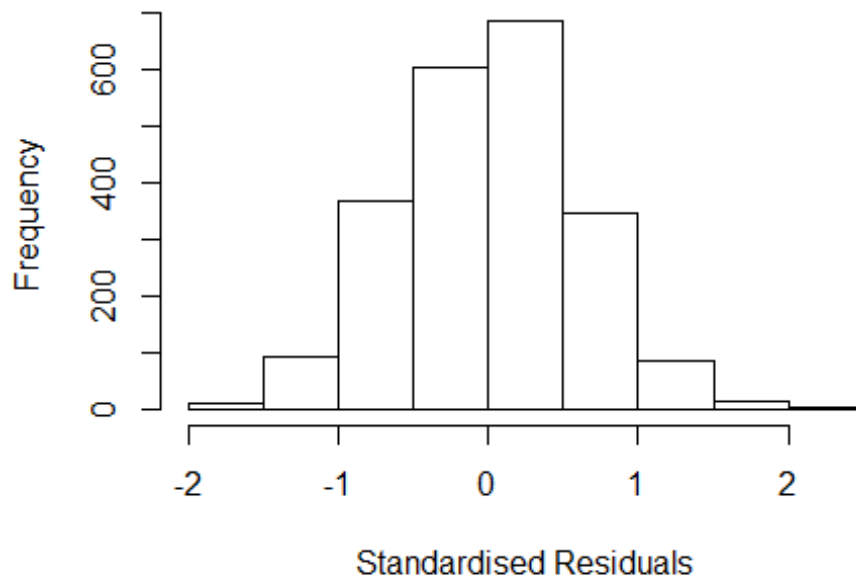


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.2, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 3.23291666876346"
## [1] "Male first author team size 2018 geometric mean: 2.39164080857053"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.05537981211697"
## [1] "Male last author team size 2018 geometric mean: 3.0358859718109"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.251 1          1.118
## LastAuthorFemale  1.298 1          1.139
## UniqueAuthors    1.173 4          1.020
## Year             1.213 16          1.006
```

Residuals from first and last author and team size



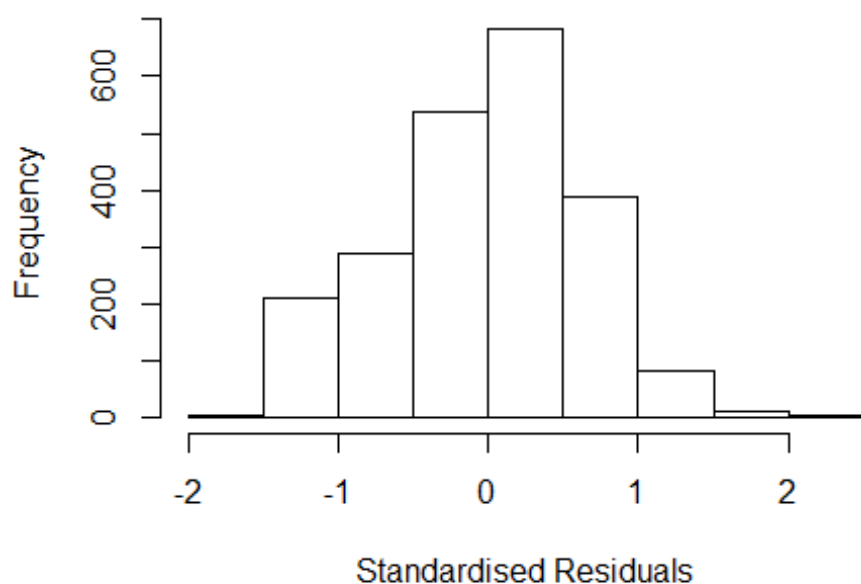
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.790 -0.413 0.024 0.421 2.326
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2480 0.0977 12.78 < 2e-16 ***
## FirstAuthorFemale1 -0.0988 0.0352 -2.80 0.00508 **
## LastAuthorFemale1 0.0193 0.0335 0.58 0.56497
## UniqueAuthors2 0.4036 0.0400 10.09 < 2e-16 ***
## UniqueAuthors3 0.4329 0.0418 10.37 < 2e-16 ***
## UniqueAuthors4 0.5226 0.0439 11.92 < 2e-16 ***
## UniqueAuthors5 0.6047 0.0403 15.02 < 2e-16 ***
## Year1997 0.1320 0.1142 1.16 0.24805
## Year1998 0.0251 0.1222 0.21 0.83707
## Year1999 -0.0172 0.1092 -0.16 0.87463
```

```

## Year2000          -0.0265      0.1139    -0.23    0.81603
## Year2001          -0.0198      0.1283    -0.15    0.87713
## Year2002          -0.2464      0.1060    -2.32    0.02026 *
## Year2003          -0.2739      0.1037    -2.64    0.00832 **
## Year2004          -0.3848      0.1086    -3.54    0.00041 ***
## Year2005          -0.3328      0.1023    -3.25    0.00117 **
## Year2006          -0.2728      0.0992    -2.75    0.00603 **
## Year2007          -0.3556      0.0969    -3.67    0.00025 ***
## Year2008          -0.3317      0.0994    -3.34    0.00086 ***
## Year2009          -0.2197      0.0995    -2.21    0.02741 *
## Year2010          -0.2082      0.0998    -2.09    0.03702 *
## Year2011          -0.2686      0.1054    -2.55    0.01089 *
## Year2012          -0.2580      0.1036    -2.49    0.01285 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.617
## Multiple R-squared:  0.151, Adjusted R-squared:  0.143
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 180 weights are ~= 1. The remaining 2025 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.124  0.859  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          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 1.208 1          1.099
## LastAuthorFemale  1.228 1          1.108
## Year              1.081 16          1.002

```


Residuals from first and last author



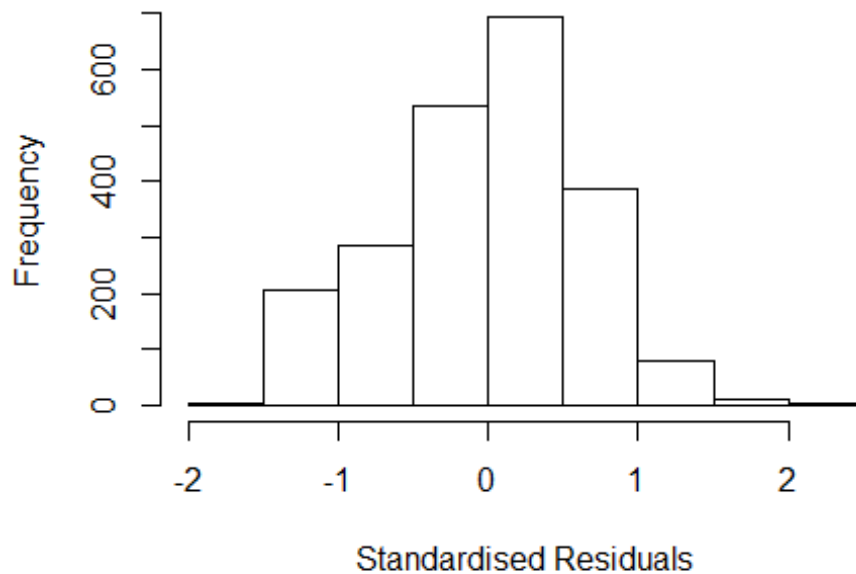
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5615 -0.4426 0.0504 0.4463 2.4226
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.555428 0.096652 16.09 <2e-16 ***
## FirstAuthorFemale1 -0.084064 0.038662 -2.17 0.0298 *
## LastAuthorFemale1 -0.061545 0.035297 -1.74 0.0814 .
## Year1997 0.073569 0.112736 0.65 0.5141
## Year1998 0.073186 0.122211 0.60 0.5493
## Year1999 -0.018930 0.110025 -0.17 0.8634
## Year2000 0.006105 0.115853 0.05 0.9580
## Year2001 -0.000179 0.129238 0.00 0.9989
## Year2002 -0.220881 0.106137 -2.08 0.0375 *
## Year2003 -0.220598 0.106459 -2.07 0.0384 *
## Year2004 -0.324644 0.113219 -2.87 0.0042 **
## Year2005 -0.208813 0.104058 -2.01 0.0449 *
```

```

## Year2006          -0.211804    0.099848    -2.12    0.0340 *
## Year2007          -0.287242    0.099528    -2.89    0.0039 **
## Year2008          -0.254300    0.100082    -2.54    0.0111 *
## Year2009          -0.124839    0.101342    -1.23    0.2181
## Year2010          -0.081220    0.100209    -0.81    0.4177
## Year2011          -0.163369    0.106969    -1.53    0.1268
## Year2012          -0.114744    0.107110    -1.07    0.2842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.642
## Multiple R-squared:  0.032, Adjusted R-squared:  0.024
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 188 weights are ~= 1. The remaining 2017 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.124  0.861  0.948  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      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 1.044 1      1.022
## Year              1.044 16      1.001

```

Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5359 -0.4484 0.0538 0.4461 2.4511
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.52695 0.09424 16.20 <2e-16 ***
## FirstAuthorFemale1 -0.11442 0.03567 -3.21 0.0014 **
## Year1997 0.07937 0.11227 0.71 0.4797
## Year1998 0.08564 0.12105 0.71 0.4793
## Year1999 -0.00911 0.10922 -0.08 0.9336
## Year2000 0.00894 0.11533 0.08 0.9382
## Year2001 0.00408 0.12903 0.03 0.9748
## Year2002 -0.21403 0.10589 -2.02 0.0434 *
## Year2003 -0.21316 0.10589 -2.01 0.0442 *
## Year2004 -0.31709 0.11281 -2.81 0.0050 **
## Year2005 -0.19805 0.10373 -1.91 0.0564 .
## Year2006 -0.20218 0.09921 -2.04 0.0417 *
```

```

## Year2007          -0.28066    0.09917   -2.83    0.0047 **
## Year2008          -0.24756    0.09956   -2.49    0.0130 *
## Year2009          -0.11851    0.10095   -1.17    0.2405
## Year2010          -0.07837    0.09996   -0.78    0.4331
## Year2011          -0.15602    0.10636   -1.47    0.1425
## Year2012          -0.10672    0.10667   -1.00    0.3172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.642
## Multiple R-squared:  0.0307, Adjusted R-squared:  0.0232
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 188 weights are ~= 1. The remaining 2017 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.113  0.860  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      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 1.059 1      1.029
## Year      1.059 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
## -1.5218 -0.4485  0.0534  0.4509  2.4677

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5103    0.0928   16.27 <2e-16 ***
## LastAuthorFemale1 -0.0966    0.0327   -2.95  0.0032 **
## Year1997          0.0734    0.1126    0.65  0.5148
## Year1998          0.0744    0.1229    0.61  0.5449
## Year1999         -0.0142    0.1098   -0.13  0.8968
## Year2000          0.0115    0.1160    0.10  0.9212
## Year2001          0.0053    0.1287    0.04  0.9672
## Year2002         -0.2162    0.1060   -2.04  0.0415 *
## Year2003         -0.2137    0.1062   -2.01  0.0444 *
## Year2004         -0.3173    0.1129   -2.81  0.0050 **
## Year2005         -0.2060    0.1039   -1.98  0.0476 *
## Year2006         -0.2056    0.0997   -2.06  0.0393 *
## Year2007         -0.2809    0.0993   -2.83  0.0047 **
## Year2008         -0.2466    0.0998   -2.47  0.0136 *
## Year2009         -0.1245    0.1013   -1.23  0.2189
## Year2010         -0.0799    0.1001   -0.80  0.4247
## Year2011         -0.1636    0.1070   -1.53  0.1264
## Year2012         -0.1126    0.1070   -1.05  0.2926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.642
## Multiple R-squared:  0.0299, Adjusted R-squared:  0.0223
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 192 weights are ~= 1. The remaining 2013 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.107  0.860  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      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: 2205"
## [1] ""

```

```

## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2901"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    2    4    1    2    9    4    8   12   14   12    2   12   25   14
## 2011 2012
##   18   25
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    1    1    1    1    7    3    5   10   11    6    1   11   19   11
## 2011 2012
##   12   21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    1    1    1    1    7    3    5   10   11    6    1   11   19   10
## 2011 2012
##   11   18
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3.30300227762026"
## [1] "Male first author team size 2018 geometric mean: 2.90833658738187"

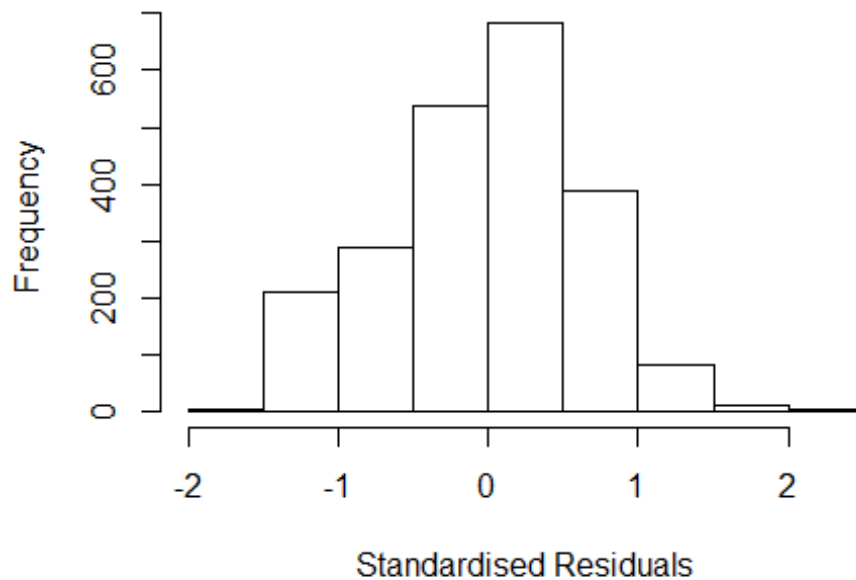
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## 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: 2.88476401472054"
## [1] "Male last author team size 2018 geometric mean: 3.80348332736773"

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

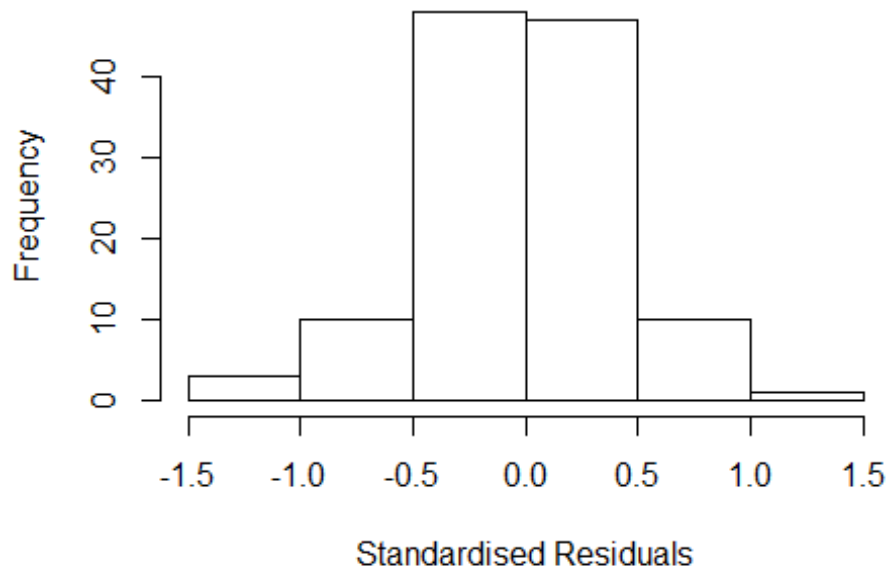
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 64, 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 1.526e+15  1      3.907e+07
## LastAuthorFemale  3.379e+14  1      1.838e+07
## UniqueAuthors    7.729e+42  4      2.296e+05
## Year              1.764e+44 16      2.414e+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
## -1.40e+00 -1.96e-01 1.55e-15 1.90e-01 1.01e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1758 0.1238 9.49 1.8e-15 ***
## FirstAuthorFemale1 -0.1972 0.1159 -1.70 0.0921 .
## LastAuthorFemale1 -0.0101 0.1220 -0.08 0.9342
## UniqueAuthors2 0.0290 0.1169 0.25 0.8049
## UniqueAuthors3 0.3005 0.0908 3.31 0.0013 **
## UniqueAuthors4 0.0670 0.1222 0.55 0.5847
## UniqueAuthors5 0.0125 0.1444 0.09 0.9313
## Year1997 0.4625 0.1072 4.32 3.9e-05 ***
## Year1998 0.6195 0.1072 5.78 9.2e-08 ***
## Year1999 0.4885 0.1469 3.33 0.0013 **
```

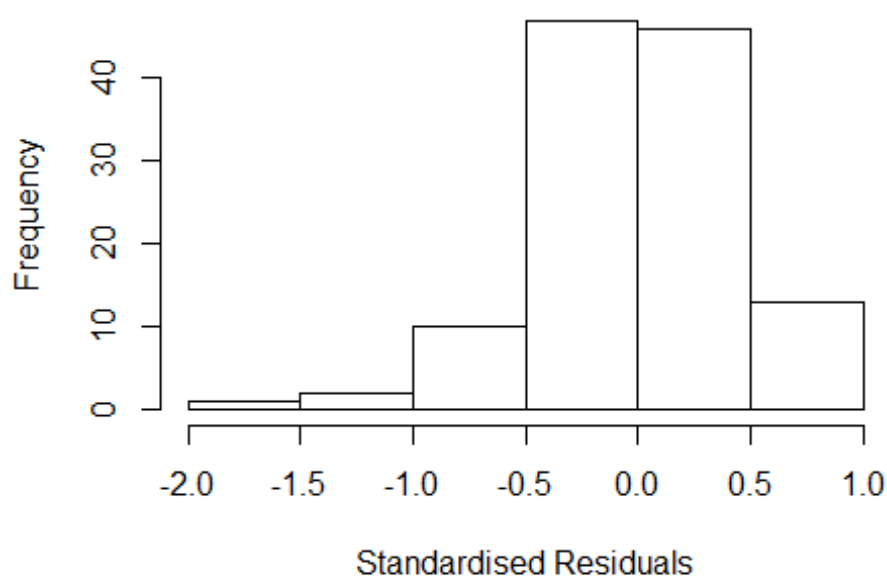


```

## Year2000          -0.2971      0.1806    -1.65    0.1032
## Year2001           0.4331      0.3065     1.41    0.1609
## Year2002          -0.1162      0.3470    -0.33    0.7385
## Year2003           0.1522      0.1358     1.12    0.2652
## Year2004          -0.1767      0.1354    -1.30    0.1951
## Year2005          -0.2029      0.1510    -1.34    0.1823
## Year2006           0.3697      0.0890     4.15  7.1e-05 ***
## Year2007           0.7003      0.1565     4.48  2.1e-05 ***
## Year2008          -0.0519      0.1900    -0.27    0.7853
## Year2009          -0.0122      0.1355    -0.09    0.9282
## Year2010           0.1760      0.1746     1.01    0.3161
## Year2011          -0.0150      0.1552    -0.10    0.9230
## Year2012          -0.2302      0.1184    -1.94    0.0548 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.326, Adjusted R-squared:  0.172
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 103 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.147  0.875   0.959   0.898   0.988   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 -5.907e+15  1      NaN
## LastAuthorFemale  -7.316e+15  1      NaN
## Year              1.895e+30 16      8.834

```

Residuals from first and last author



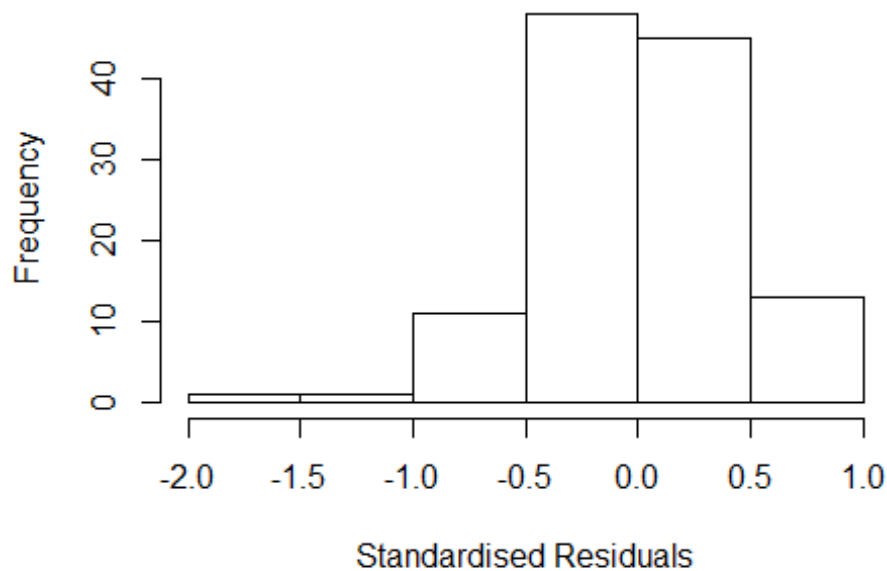
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.52e+00 -2.24e-01 1.72e-15 2.45e-01 9.26e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1865 0.0940 12.63 < 2e-16 ***
## FirstAuthorFemale1 -0.1583 0.1244 -1.27 0.20595
## LastAuthorFemale1 -0.0245 0.1370 -0.18 0.85853
## Year1997 0.4274 0.1001 4.27 4.5e-05 ***
## Year1998 0.5844 0.1001 5.84 6.6e-08 ***
## Year1999 0.5204 0.1001 5.20 1.1e-06 ***
## Year2000 -0.3341 0.0848 -3.94 0.00015 ***
## Year2001 0.5126 0.2575 1.99 0.04921 *
## Year2002 -0.1318 0.3389 -0.39 0.69825
## Year2003 0.1224 0.1375 0.89 0.37556
## Year2004 -0.1228 0.1407 -0.87 0.38495
## Year2005 -0.1327 0.1643 -0.81 0.42118
```

```

## Year2006          0.4514      0.1005      4.49  1.9e-05 ***
## Year2007          0.7330      0.1520      4.82  5.1e-06 ***
## Year2008         -0.0136      0.2281     -0.06  0.95259
## Year2009          0.0505      0.1332      0.38  0.70522
## Year2010          0.2170      0.1531      1.42  0.15962
## Year2011          0.0106      0.1535      0.07  0.94526
## Year2012         -0.1811      0.1220     -1.48  0.14084
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.399
## Multiple R-squared:  0.269, Adjusted R-squared:  0.138
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 105 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.117 0.871 0.955 0.901 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      8.40e-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 -1.879e+14 1      NaN
## Year              -1.879e+14 16      NaN

```

Residuals from first author



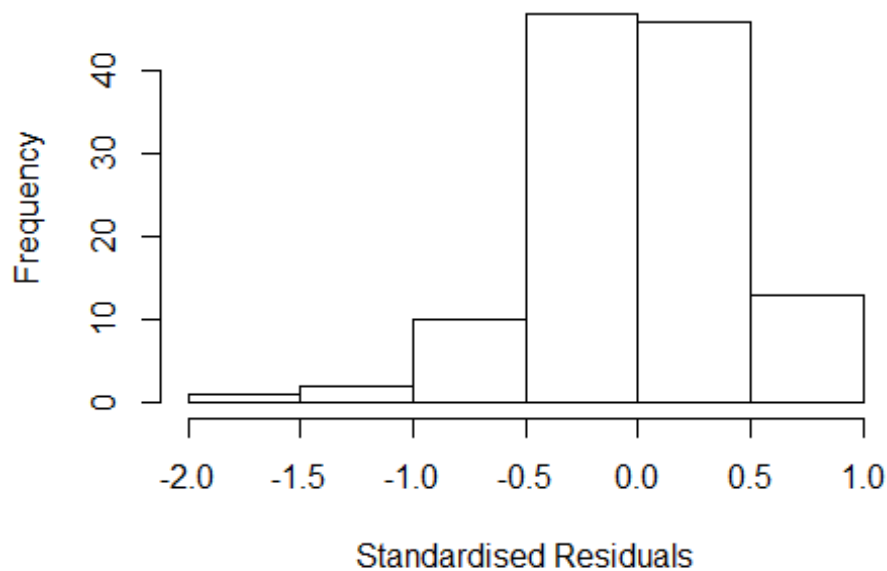
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.53e+00 -2.19e-01 -2.22e-16 2.42e-01 9.08e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18429 0.09511 12.45 < 2e-16 ***
## FirstAuthorFemale1 -0.16719 0.11802 -1.42 0.160
## Year1997 0.41390 0.06559 6.31 7.5e-09 ***
## Year1998 0.57090 0.06559 8.70 6.3e-14 ***
## Year1999 0.50690 0.06559 7.73 8.3e-12 ***
## Year2000 -0.32310 0.06559 -4.93 3.3e-06 ***
## Year2001 0.51647 0.25190 2.05 0.043 *
## Year2002 -0.14634 0.33409 -0.44 0.662
## Year2003 0.11298 0.11728 0.96 0.338
## Year2004 -0.13373 0.12714 -1.05 0.295
## Year2005 -0.14266 0.15154 -0.94 0.349
## Year2006 0.44857 0.10506 4.27 4.4e-05 ***
```

```

## Year2007          0.71071      0.09511      7.47 2.9e-11 ***
## Year2008         -0.03007      0.17353     -0.17  0.863
## Year2009          0.04199      0.11967      0.35  0.726
## Year2010          0.20800      0.13756      1.51  0.134
## Year2011          0.00858      0.15614      0.05  0.956
## Year2012         -0.18903      0.10539     -1.79  0.076 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.396
## Multiple R-squared:  0.271, Adjusted R-squared:  0.148
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 105 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.100  0.872  0.958  0.900  0.990  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 -3.25e+15  1      NaN
## Year             -3.25e+15 16      NaN

```

Residuals from last author



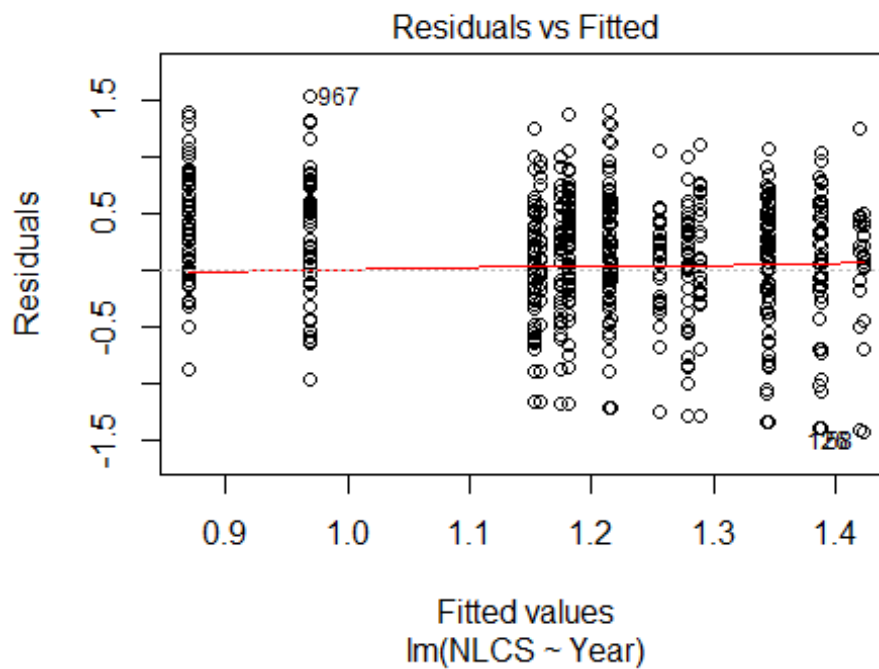
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.57e+00 -2.26e-01 7.22e-16 2.50e-01 8.76e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0966 0.0903 12.14 < 2e-16 ***
## LastAuthorFemale1 -0.0741 0.1328 -0.56 0.57826
## Year1997 0.4085 0.1178 3.47 0.00077 ***
## Year1998 0.5655 0.1178 4.80 5.5e-06 ***
## Year1999 0.5015 0.1178 4.26 4.6e-05 ***
## Year2000 -0.4026 0.0903 -4.46 2.2e-05 ***
## Year2001 0.5433 0.2403 2.26 0.02591 *
## Year2002 -0.1511 0.3440 -0.44 0.66152
## Year2003 0.0960 0.1586 0.61 0.54638
## Year2004 -0.1103 0.1382 -0.80 0.42645
## Year2005 -0.1433 0.1670 -0.86 0.39278
## Year2006 0.4342 0.1188 3.65 0.00041 ***
```

```

## Year2007          0.8725      0.1178      7.41  4.0e-11 ***
## Year2008          0.0346      0.2226      0.16  0.87692
## Year2009          0.0211      0.1469      0.14  0.88601
## Year2010          0.2208      0.1796      1.23  0.22177
## Year2011          0.0108      0.1729      0.06  0.95047
## Year2012         -0.1715      0.1387     -1.24  0.21928
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.398
## Multiple R-squared:  0.26,   Adjusted R-squared:  0.135
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 106 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0863 0.8820 0.9600 0.9000 0.9910 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 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
##   22   19   33   21   50   40   52   34   46   33   64   49   64   95   89
## 2011 2012
##   93   104
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   16   16   21   13   30   25   38   30   40   30   55   44   57   78   77
## 2011 2012

```

```
##      82      88
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   16   14   17   11   29   18   34   25   37   27   49   40   49   71   72
## 2011 2012
##    75    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 = 20, df = 16, p-value = 0.2
```

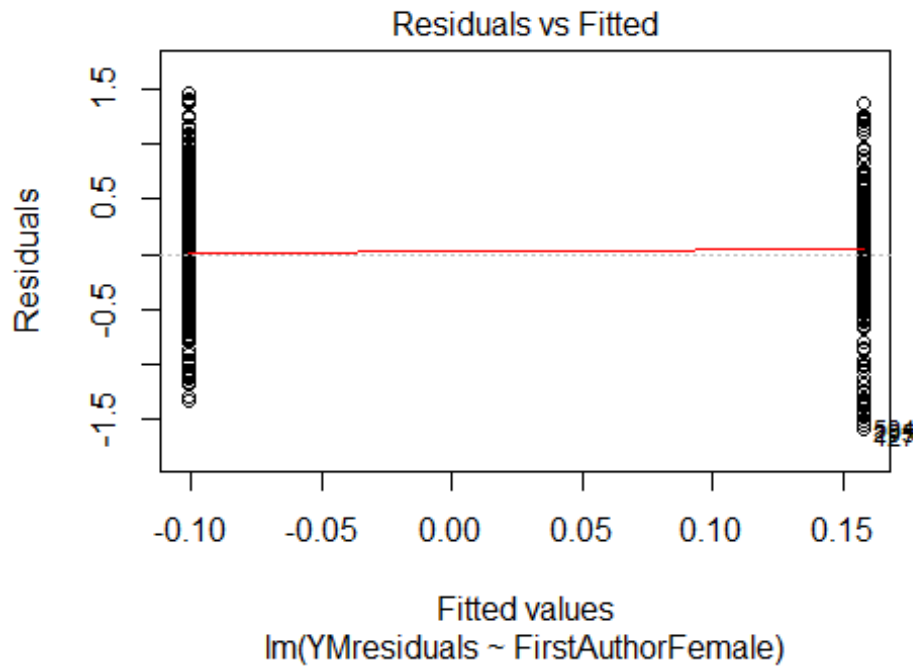


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5, df = 1, p-value = 0.02
##
## [1] "Female first author team size 2018 geometric mean: 2.83307219135855"
## [1] "Male first author team size 2018 geometric mean: 3.84096711843155"
##
## Wilcoxon rank sum test with continuity correction
##
## data:  FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 380, p-value = 0.1
```



```
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.68688138474089"
## [1] "Male last author team size 2018 geometric mean: 3.9443357658014"

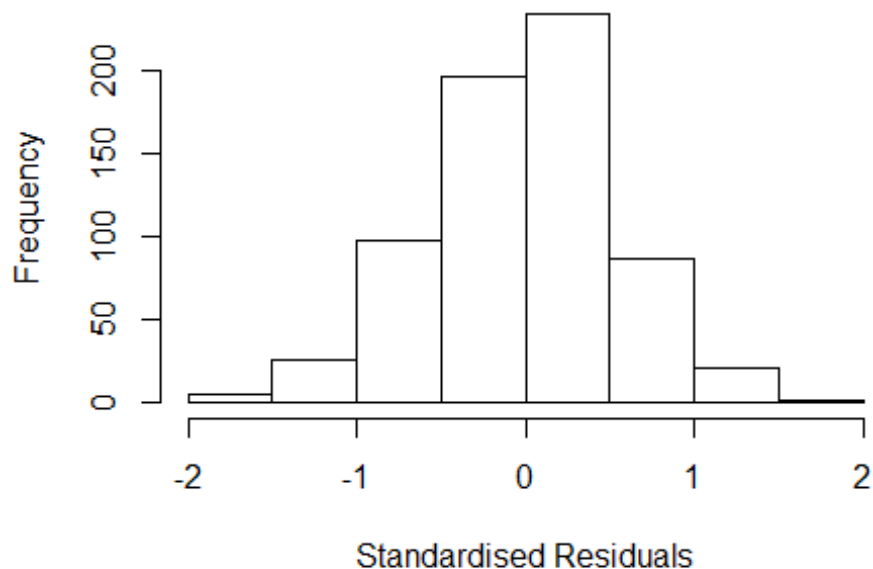
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 420, p-value = 0.05
## 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	1.216	1	1.103
LastAuthorFemale	1.440	1	1.200
UniqueAuthors	2.228	4	1.105
Year	2.304	16	1.026

Residuals from first and last author and team size



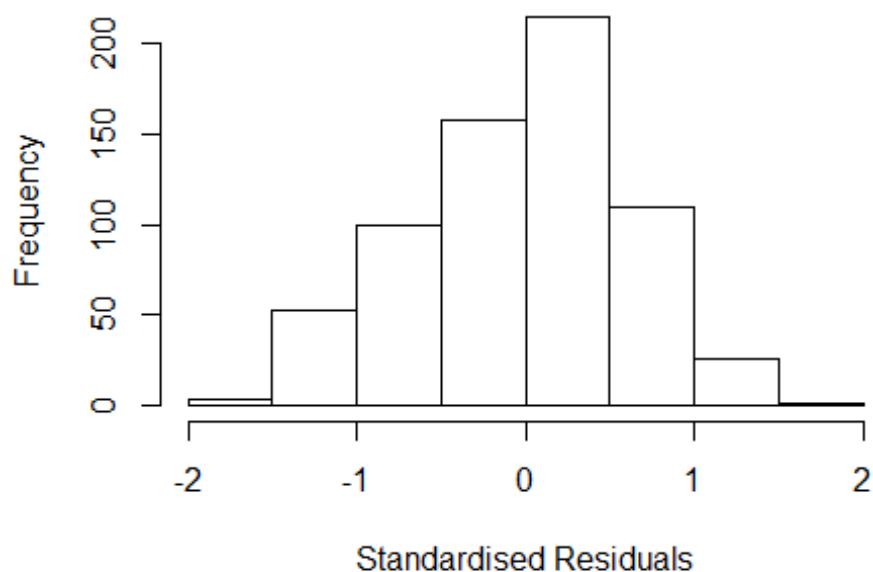
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8458 -0.3790  0.0194  0.3475  1.7626
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.02212    0.15974   6.40 3.0e-10 ***
## FirstAuthorFemale1 -0.13139    0.04611  -2.85 0.00452 **
## LastAuthorFemale1 -0.02020    0.05040  -0.40 0.68872
## UniqueAuthors2     0.30936    0.08831   3.50 0.00049 ***
## UniqueAuthors3     0.60237    0.09799   6.15 1.4e-09 ***
## UniqueAuthors4     0.72003    0.09098   7.91 1.1e-14 ***
## UniqueAuthors5     0.86847    0.07306  11.89 < 2e-16 ***
## Year1997         -0.18113    0.19665  -0.92 0.35735
## Year1998         -0.30998    0.22258  -1.39 0.16421
## Year1999         -0.02455    0.22312  -0.11 0.91241
```

```

## Year2000      -0.12599    0.19117   -0.66  0.51012
## Year2001      -0.30585    0.19469   -1.57  0.11669
## Year2002       0.00603    0.17704    0.03  0.97285
## Year2003      -0.18695    0.16826   -1.11  0.26696
## Year2004      -0.32933    0.17065   -1.93  0.05407 .
## Year2005      -0.30941    0.17734   -1.74  0.08152 .
## Year2006      -0.38033    0.16410   -2.32  0.02078 *
## Year2007      -0.30024    0.16805   -1.79  0.07448 .
## Year2008      -0.22338    0.16161   -1.38  0.16738
## Year2009      -0.24646    0.16224   -1.52  0.12922
## Year2010      -0.28807    0.16505   -1.75  0.08141 .
## Year2011      -0.56784    0.15667   -3.62  0.00031 ***
## Year2012      -0.49149    0.15996   -3.07  0.00221 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.528
## Multiple R-squared:  0.343, Adjusted R-squared:  0.321
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 612 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.871  0.953  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      1.50e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.190 1      1.091
## LastAuthorFemale  1.183 1      1.088
## Year              1.213 16      1.006

```

Residuals from first and last author



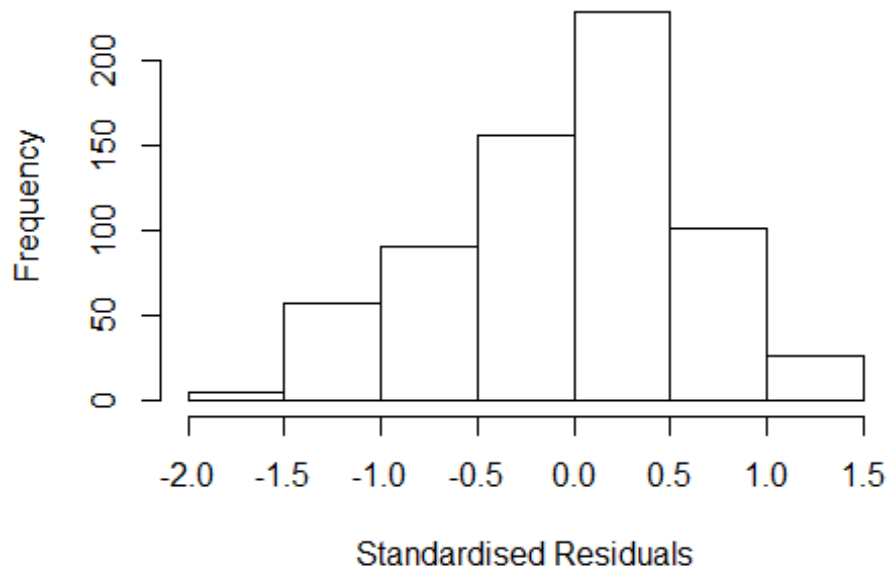
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6733 -0.4225 0.0645 0.4235 1.5862
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.62940 0.15422 10.57 < 2e-16 ***
## FirstAuthorFemale1 -0.20818 0.05298 -3.93 9.4e-05 ***
## LastAuthorFemale1 -0.23024 0.05277 -4.36 1.5e-05 ***
## Year1997 -0.01387 0.19896 -0.07 0.9444
## Year1998 -0.32285 0.24652 -1.31 0.1908
## Year1999 -0.09897 0.25346 -0.39 0.6963
## Year2000 -0.05139 0.19509 -0.26 0.7923
## Year2001 -0.15170 0.22460 -0.68 0.4997
## Year2002 0.04389 0.19594 0.22 0.8228
## Year2003 -0.04171 0.17736 -0.24 0.8141
## Year2004 -0.14634 0.17589 -0.83 0.4057
## Year2005 -0.20311 0.19689 -1.03 0.3027
```

```

## Year2006          -0.21411      0.17297      -1.24      0.2162
## Year2007          -0.18715      0.17866      -1.05      0.2952
## Year2008           0.00333      0.17135       0.02      0.9845
## Year2009          -0.15475      0.17182      -0.90      0.3681
## Year2010          -0.16952      0.17198      -0.99      0.3246
## Year2011          -0.47936      0.17500      -2.74      0.0063 **
## Year2012          -0.47589      0.17244      -2.76      0.0059 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.628
## Multiple R-squared:  0.141, Adjusted R-squared:  0.117
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 622 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.458  0.885   0.953   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      1.50e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0           1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.137 1      1.066
## Year              1.137 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
## -1.6027 -0.4449 0.0613 0.4256 1.4824
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5744 0.1486 10.60 < 2e-16 ***
## FirstAuthorFemale1 -0.2924 0.0520 -5.62 2.8e-08 ***
## Year1997 -0.0261 0.1889 -0.14 0.8902
## Year1998 -0.3107 0.2448 -1.27 0.2049
## Year1999 -0.0591 0.2731 -0.22 0.8287
## Year2000 -0.0625 0.1955 -0.32 0.7494
## Year2001 -0.1408 0.2213 -0.64 0.5248
## Year2002 0.0283 0.1912 0.15 0.8825
## Year2003 -0.0582 0.1776 -0.33 0.7435
## Year2004 -0.1314 0.1727 -0.76 0.4470
## Year2005 -0.2204 0.1954 -1.13 0.2599
## Year2006 -0.2032 0.1682 -1.21 0.2274
```

```

## Year2007          -0.1988      0.1750   -1.14    0.2564
## Year2008          -0.0271      0.1656   -0.16    0.8700
## Year2009          -0.1662      0.1698   -0.98    0.3282
## Year2010          -0.1938      0.1677   -1.16    0.2484
## Year2011          -0.4905      0.1707   -2.87    0.0042 **
## Year2012          -0.4901      0.1718   -2.85    0.0045 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.628
## Multiple R-squared:  0.117, Adjusted R-squared:  0.0934
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 615 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.494  0.860  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      1.50e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.12  1      1.058
## Year              1.12 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
## -1.5904 -0.4328  0.0625  0.4135  1.7499

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.55516    0.15818    9.83 < 2e-16 ***
## LastAuthorFemale1 -0.30193    0.05243   -5.76 1.3e-08 ***
## Year1997         0.02441    0.20554    0.12  0.9055
## Year1998        -0.30482    0.24703   -1.23  0.2177
## Year1999        -0.04391    0.25391   -0.17  0.8627
## Year2000        -0.07043    0.19828   -0.36  0.7225
## Year2001        -0.12444    0.23633   -0.53  0.5987
## Year2002         0.03527    0.20077    0.18  0.8606
## Year2003        -0.05996    0.18192   -0.33  0.7418
## Year2004        -0.17153    0.17975   -0.95  0.3403
## Year2005        -0.22538    0.19935   -1.13  0.2586
## Year2006        -0.24176    0.17818   -1.36  0.1753
## Year2007        -0.22567    0.18528   -1.22  0.2237
## Year2008        -0.00563    0.17822   -0.03  0.9748
## Year2009        -0.17991    0.17768   -1.01  0.3116
## Year2010        -0.16993    0.17854   -0.95  0.3416
## Year2011        -0.49712    0.18012   -2.76  0.0059 **
## Year2012        -0.52372    0.17789   -2.94  0.0034 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.627
## Multiple R-squared:  0.123, Adjusted R-squared:  0.1
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 622 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.416  0.872  0.951  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.50e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 665"
## [1] ""

```



```

## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2903"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    4    8    5    8    3    7    5    7    3    4    3    6   11    9
## 2011 2012
##    5    9
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    1    3    3    2    2    3    2    7    3    3    2    5    9    8
## 2011 2012
##    4    8
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    1    3    3    2    2    3    2    5    3    3    2    4    8    8
## 2011 2012
##    4    6
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.55084900125158"
## [1] "Male first author team size 2018 geometric mean: 1.67392930873899"

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

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.30779843345719"
## [1] "Male last author team size 2018 geometric mean: 1.34800615459728"

## Warning in wilcox.test.default(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 = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

```

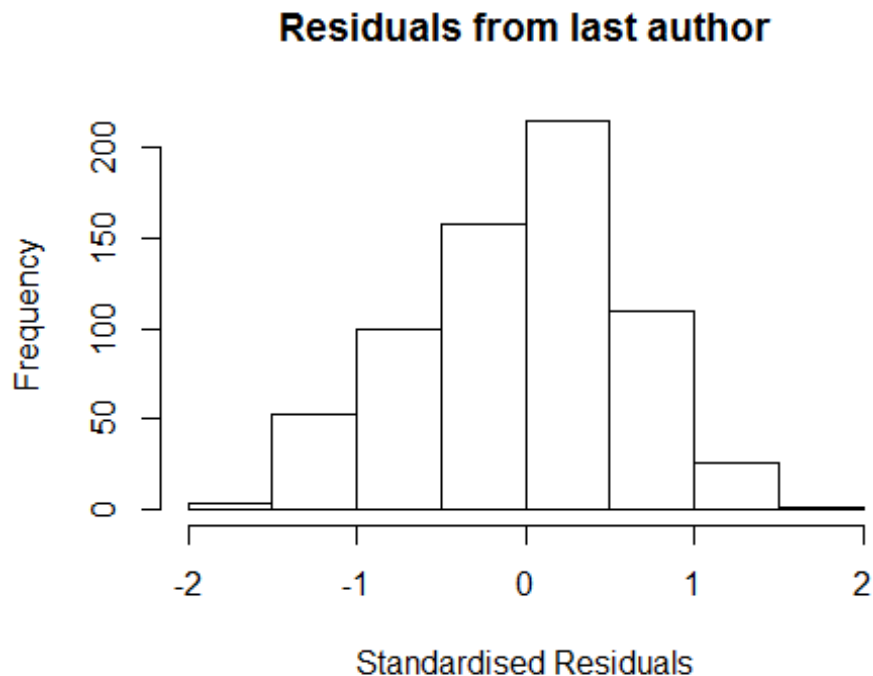
```
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## 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"

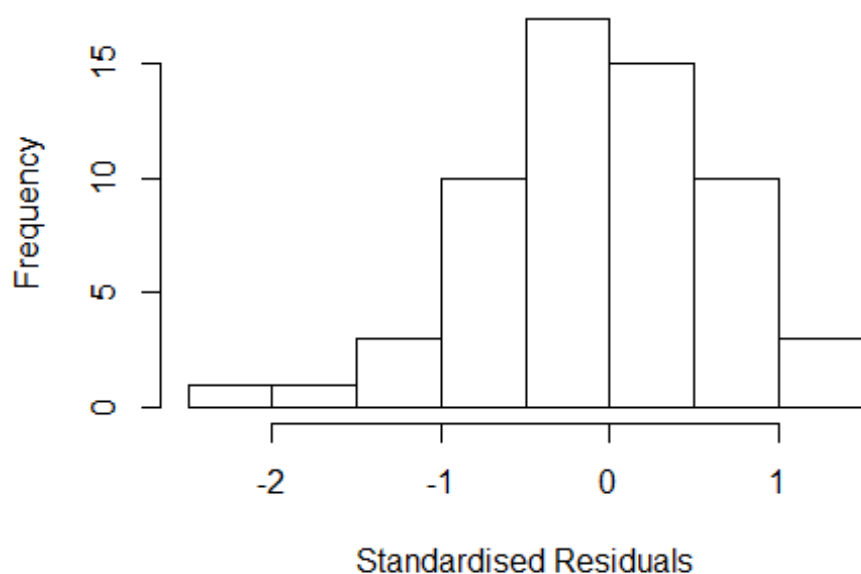
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## 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
## Year	NaN	16	NaN

Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -2.02e+00 -4.67e-01 -3.63e-16 4.54e-01 1.42e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.58e-01 3.56e-01 1.85 0.07187 .
## FirstAuthorFemale1 -8.51e-01 4.83e-01 -1.76 0.08522 .
## LastAuthorFemale1 1.94e-01 4.27e-01 0.45 0.65271
## Year1997 -5.15e-15 0.00e+00 -Inf < 2e-16 ***
## Year1998 2.35e-01 6.79e-01 0.35 0.73107
## Year1999 7.35e-01 4.25e-01 1.73 0.09165 .
## Year2000 8.83e-01 8.32e-01 1.06 0.29466
## Year2001 7.73e-01 4.79e-01 1.61 0.11415
## Year2002 1.36e+00 1.23e+00 1.11 0.27295
## Year2003 9.51e-01 2.80e-01 3.40 0.00151 **
## Year2004 5.50e-01 3.28e-01 1.68 0.10057
## Year2005 -3.22e-15 0.00e+00 -Inf < 2e-16 ***
```

```

## Year2006          1.04e+00  7.01e-01  1.48  0.14586
## Year2007          1.67e+00  4.39e-01  3.80  0.00047 ***
## Year2008          1.84e+00  3.05e-01  6.03  4e-07 ***
## Year2009          8.67e-01  3.53e-01  2.46  0.01825 *
## Year2010          1.18e+00  3.34e-01  3.54  0.00101 **
## Year2011          7.45e-01  2.98e-01  2.50  0.01663 *
## Year2012          1.02e+00  3.03e-01  3.38  0.00162 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.778
## Multiple R-squared:  0.406, Adjusted R-squared:  0.145
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 53 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.479  0.890  0.961  0.924  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      1.67e-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 1800 1      42.430
## Year              1800 16      1.264

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

```

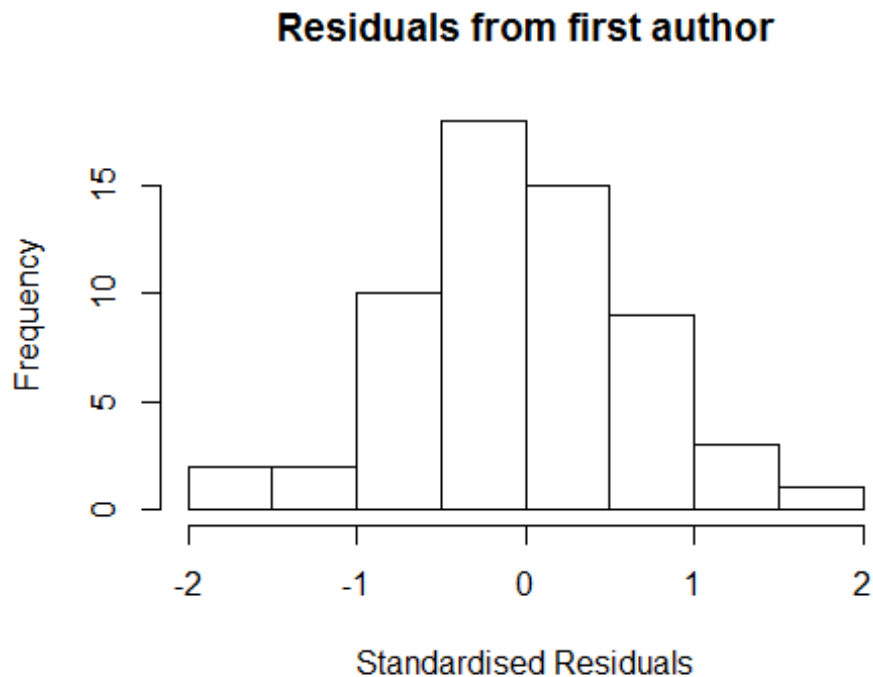
```

## -2.00e+00 -4.65e-01 -1.71e-16 4.50e-01 1.56e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.89e-01  3.00e-01   2.30  0.02676 *
## FirstAuthorFemale1 -6.89e-01  3.00e-01  -2.30  0.02676 *
## Year1997          2.45e-15  2.83e-08   0.00  1.00000
## Year1998          2.59e-01  7.51e-01   0.35  0.73145
## Year1999          7.11e-01  3.89e-01   1.83  0.07441 .
## Year2000          8.83e-01  7.99e-01   1.10  0.27561
## Year2001          8.38e-01  3.87e-01   2.17  0.03608 *
## Year2002          1.31e+00  1.13e+00   1.16  0.25266
## Year2003          9.51e-01  2.78e-01   3.42  0.00141 **
## Year2004          5.31e-01  2.95e-01   1.80  0.07869 .
## Year2005          1.71e-16  3.00e-08   0.00  1.00000
## Year2006          1.03e+00  6.61e-01   1.55  0.12804
## Year2007          1.65e+00  4.23e-01   3.91  0.00033 ***
## Year2008          1.78e+00  2.92e-01   6.08   3e-07 ***
## Year2009          8.43e-01  3.29e-01   2.56  0.01420 *
## Year2010          1.18e+00  3.17e-01   3.74  0.00055 ***
## Year2011          7.37e-01  2.89e-01   2.55  0.01447 *
## Year2012          9.75e-01  2.84e-01   3.43  0.00137 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.827
## Multiple R-squared:  0.395, Adjusted R-squared:  0.15
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 54 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.540  0.904  0.966  0.933  0.985  0.999
## Algorithmic parameters:
##           tuning.chi              bb          tuning.psi          refine.tol
##           1.55e+00             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-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
```



```
##              GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale NaN 1              NaN
## Year            NaN 16             NaN

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.78e+00 -3.98e-01  5.91e-16  3.97e-01  2.28e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.57e-01   3.31e-01    1.38  0.1743
## LastAuthorFemale1 -4.57e-01   3.31e-01   -1.38  0.1743
```

```

## Year1997          -7.74e-16   0.00e+00   -Inf   < 2e-16 ***
## Year1998          2.24e-01   9.04e-01   0.25   0.8056
## Year1999          8.79e-01   4.77e-01   1.84   0.0726 .
## Year2000          8.83e-01   8.28e-01   1.07   0.2923
## Year2001          1.30e+00   1.82e-01   7.14   9.3e-09 ***
## Year2002          1.32e+00   1.02e+00   1.30   0.2010
## Year2003          9.51e-01   2.80e-01   3.40   0.0015 **
## Year2004          6.74e-01   3.63e-01   1.86   0.0703 .
## Year2005          -1.92e-15   0.00e+00   -Inf   < 2e-16 ***
## Year2006          1.14e+00   6.81e-01   1.67   0.1024
## Year2007          1.77e+00   3.54e-01   5.00   1.1e-05 ***
## Year2008          1.73e+00   3.81e-01   4.54   4.7e-05 ***
## Year2009          9.50e-01   4.11e-01   2.31   0.0258 *
## Year2010          1.36e+00   2.95e-01   4.61   3.8e-05 ***
## Year2011          7.89e-01   3.31e-01   2.39   0.0216 *
## Year2012          1.03e+00   3.42e-01   3.01   0.0045 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.783
## Multiple R-squared:  0.362, Adjusted R-squared:  0.103
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 52 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.375  0.889  0.956  0.915  0.986  0.998
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.67e-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: 60"
## [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]"
##
## 1996 1997 1998 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    1    1    1    3    1    4    6    7    5    2    1    9   10    5    6
##
## 1996 1997 1998 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    0    0    0    0    1    4    6    7    3    0    1    8    9    4    6
##
## 1996 1997 1998 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    0    0    0    0    1    4    4    6    3    0    1    7    8    4    4
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4.37572387551628"
## [1] "Male first author team size 2018 geometric mean: 2.28942848510666"

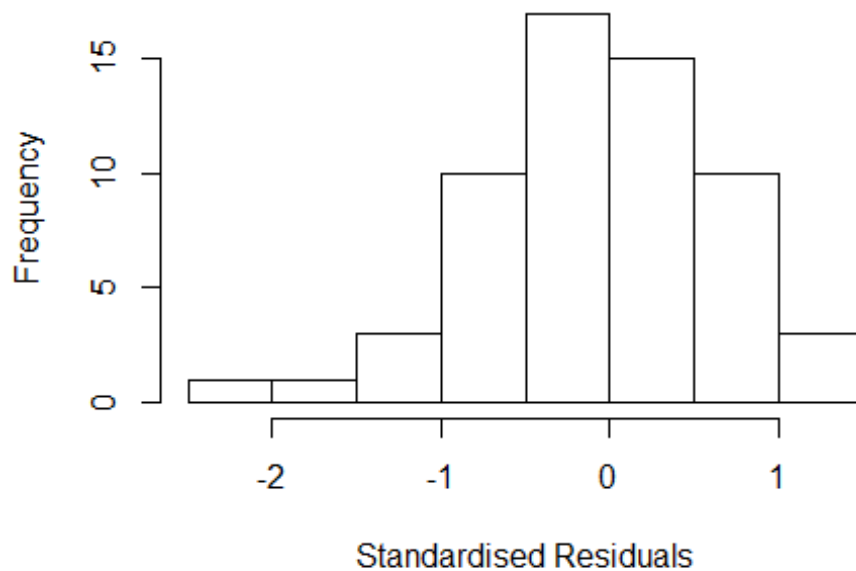
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.29314551673089"
## [1] "Male last author team size 2018 geometric mean: 2.78315768371374"

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

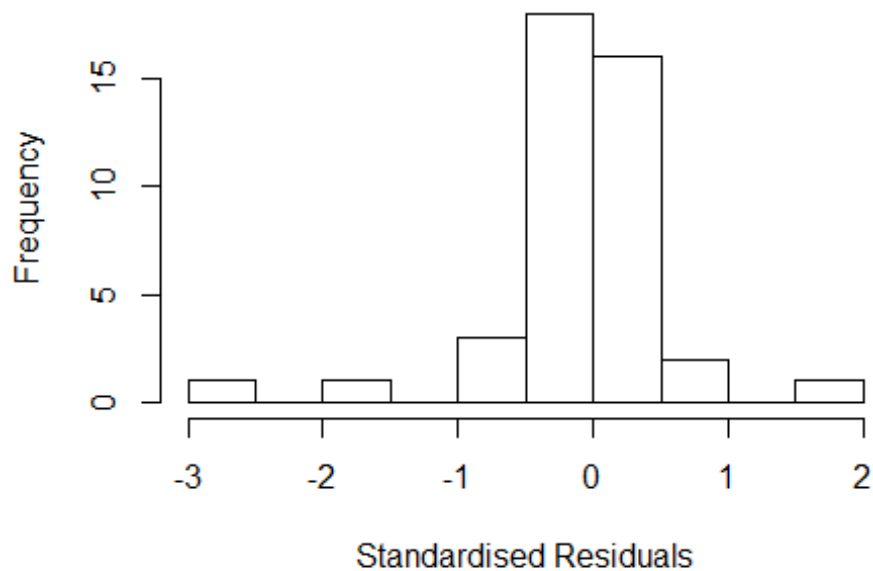
```


Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 18, 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 -3.397e+15  1      NaN
## LastAuthorFemale  -5.435e+00  1      NaN
## UniqueAuthors    -2.869e+17  4      NaN
## Year              -2.260e+18  9      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
## 52 77955456564 0.393 2010      2719      4    -2.672
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6717 -0.1917 -0.0277  0.2062  1.7622
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8700    0.2866   3.04  0.0054 **
## FirstAuthorFemale1 -0.6306    0.1641  -3.84  0.0007 ***
## LastAuthorFemale1  0.0827    0.1936   0.43  0.6729
## UniqueAuthors2    0.1069    0.3068   0.35  0.7302
## UniqueAuthors3   -0.7572    0.3026  -2.50  0.0190 *
## UniqueAuthors4    1.2143    0.2057   5.90 3.2e-06 ***
## UniqueAuthors5    0.5755    0.2755   2.09  0.0467 *
## Year2003         -0.6723    0.4104  -1.64  0.1135
## Year2004          0.2405    0.3267   0.74  0.4683
## Year2005         -0.1306    0.1876  -0.70  0.4925
```

```

## Year2006          -0.0889      0.4445   -0.20   0.8431
## Year2008          0.6569      0.2672    2.46   0.0209 *
## Year2009          0.6653      0.3173    2.10   0.0459 *
## Year2010          0.8977      0.2903    3.09   0.0047 **
## Year2011         -0.1801      0.2317   -0.78   0.4440
## Year2012          0.4646      0.3185    1.46   0.1566
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.458
## Multiple R-squared:  0.743, Adjusted R-squared:  0.594
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 29 is an outlier with |weight| = 0 ( < 0.0024);
## 2 weights are ~ 1. The remaining 39 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.106  0.928  0.982  0.916  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] "Regression 2: First author gender, Last author gender, Year as
factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 9.007e+15  1      9.491e+07
## LastAuthorFemale  4.102e+01  1      6.405e+00
## Year              2.025e+17  9      9.151e+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"

```

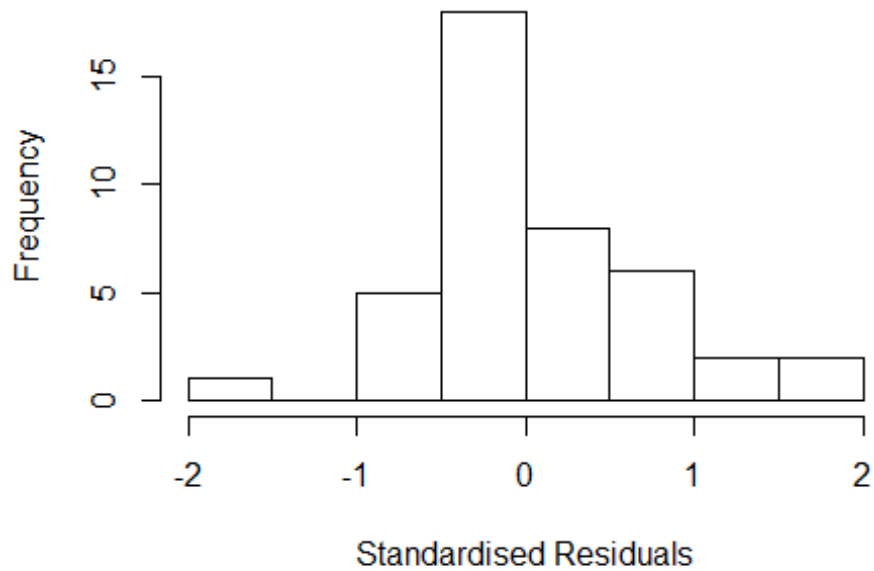
```

## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6404 -0.3271 -0.0494  0.3738  1.6131
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.372      0.594   3.99 0.00039 ***
## FirstAuthorFemale1 -0.271      0.458  -0.59 0.55896
## LastAuthorFemale1 -0.205      0.594  -0.34 0.73256
## Year2003          -2.025      0.638  -3.17 0.00347 **
## Year2004          -1.478      0.567  -2.60 0.01417 *
## Year2005          -1.338      0.538  -2.49 0.01869 *
## Year2006          -1.485      0.680  -2.19 0.03683 *
## Year2008          -0.917      0.458  -2.00 0.05444 .
## Year2009          -0.731      0.933  -0.78 0.43916
## Year2010          -0.872      0.525  -1.66 0.10748
## Year2011          -1.391      0.635  -2.19 0.03637 *
## Year2012          -0.676      0.485  -1.39 0.17328
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.354, Adjusted R-squared:  0.117
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~ = 1. The remaining 35 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.364  0.866  0.952  0.878  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          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"

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

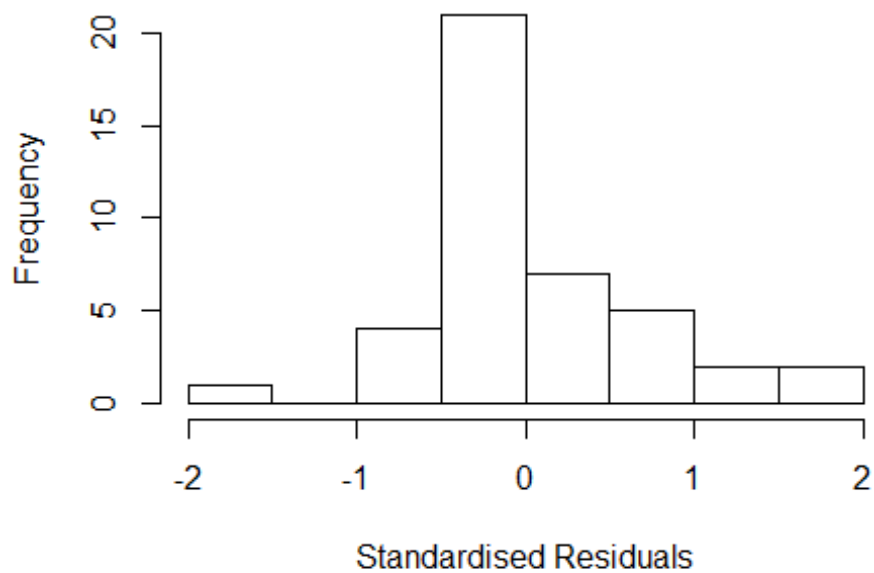
```

Residuals from first and last author



##		GVIF	Df	$GVIF^{1/(2 \cdot Df)}$
##	FirstAuthorFemale	2.729e+14	1	1.652e+07
##	Year	2.729e+14	9	6.339e+00

Residuals from first author



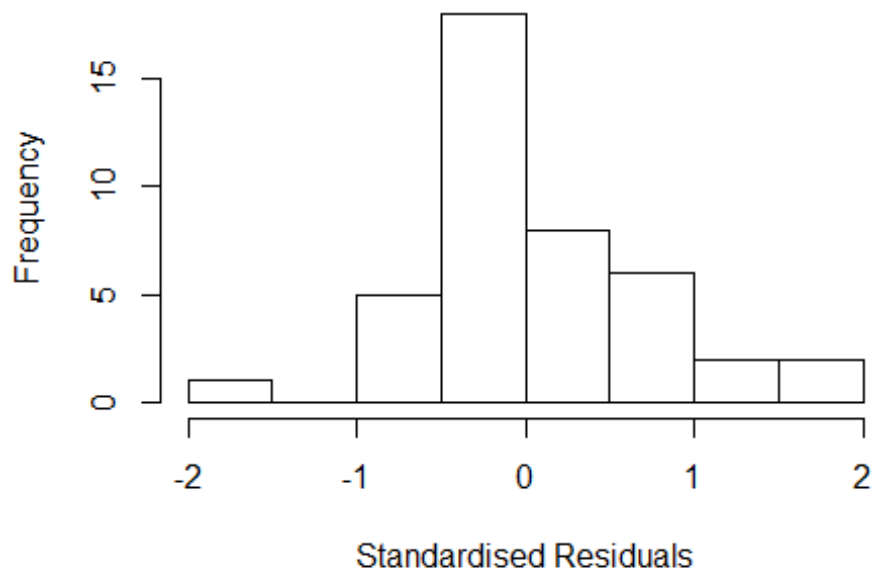
```

## [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
## -1.658 -0.341 -0.060 0.412 1.694
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.167 0.000 Inf < 2e-16 ***
## FirstAuthorFemale1 -0.372 0.370 -1.00 0.32310
## Year2003 -1.824 0.420 -4.35 0.00014 ***
## Year2004 -1.296 0.183 -7.07 6.2e-08 ***
## Year2005 -1.164 0.148 -7.88 6.9e-09 ***
## Year2006 -1.284 0.299 -4.29 0.00016 ***
## Year2008 -0.816 0.370 -2.20 0.03520 *
## Year2009 -0.509 0.434 -1.17 0.24933
## Year2010 -0.839 0.719 -1.17 0.25226
## Year2011 -1.238 0.608 -2.04 0.05037 .
## Year2012 -0.549 0.316 -1.74 0.09256 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared: 0.355, Adjusted R-squared: 0.146
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 5 weights are ~ 1. The remaining 37 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.328 0.887 0.961 0.881 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] "Regression 4: Last author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 34.89 1      5.907
## Year             34.89 9      1.218
```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4543 -0.2532 -0.0279  0.3669  1.6362
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   2.56e+00   5.44e-01  4.70e+00   5e-05 ***
## LastAuthorFemale1 -3.91e-01   5.44e-01 -7.20e-01  0.47762
## Year2003        -2.23e+00   5.81e-01 -3.84e+00  0.00056 ***
## Year2004        -1.68e+00   4.83e-01 -3.49e+00  0.00149 **
```

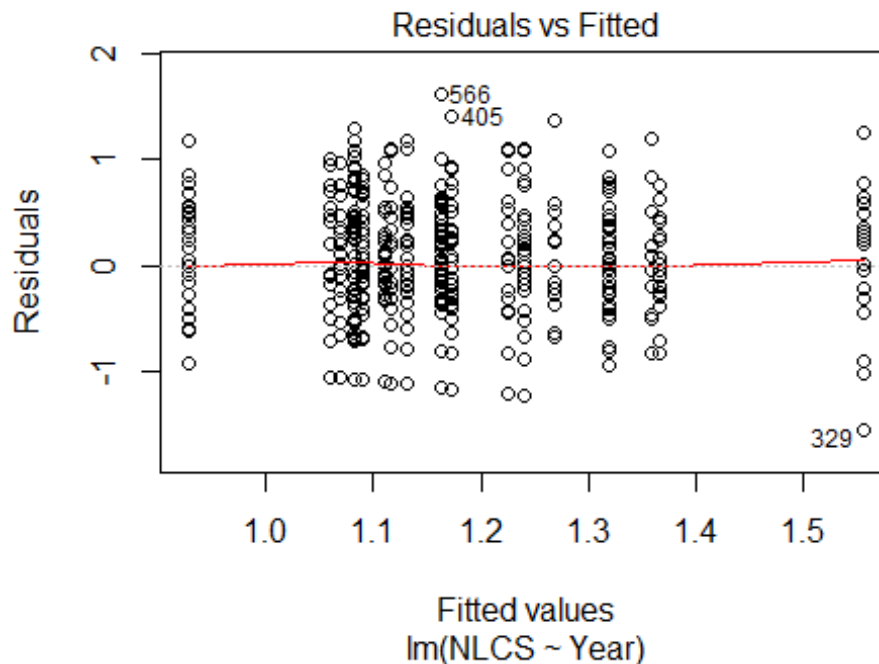
```

## Year2005          -1.55e+00   4.16e-01 -3.73e+00   0.00077 ***
## Year2006          -1.76e+00   5.52e-01 -3.20e+00   0.00320 **
## Year2008          -1.19e+00   2.80e-09 -4.24e+08   < 2e-16 ***
## Year2009          -1.10e+00   6.00e-01 -1.84e+00   0.07552 .
## Year2010          -1.03e+00   4.05e-01 -2.54e+00   0.01619 *
## Year2011          -1.70e+00   4.16e-01 -4.10e+00   0.00028 ***
## Year2012          -9.26e-01   2.11e-01 -4.39e+00   0.00012 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.343, Adjusted R-squared:  0.131
## Convergence in 30 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~1. The remaining 34 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.335  0.844  0.932  0.864  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      2.38e-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 2905"
## [1] "#####"
## [1] "Sample sizes for all years [All, 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  38  24  18  22  38  41  28  23  45  23  41  34  53  62
## 2011 2012
##   58  59
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19  35  21  16  12  19  36  26  18  38  20  38  30  43  59

```



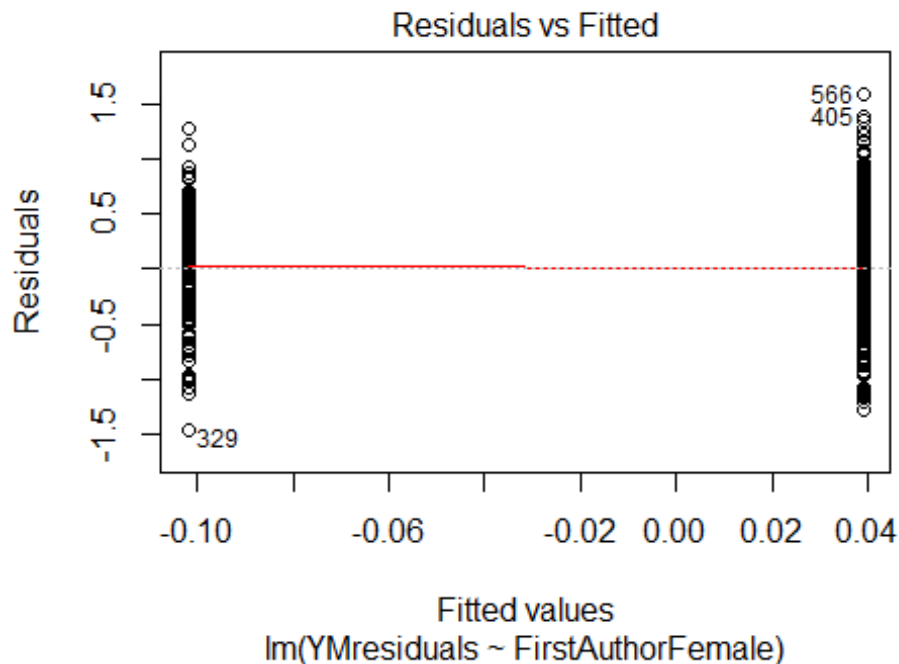
```
## 2011 2012
## 54 51
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 17 35 20 16 12 19 33 24 17 35 20 35 28 37 56
## 2011 2012
## 53 49
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6
```



```
##
## 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: 3.64753417007087"
## [1] "Male first author team size 2018 geometric mean: 3.39676265912991"
## Warning in wilcox.test.default(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: 3.48056804355461"
## [1] "Male last author team size 2018 geometric mean: 3.97537247221972"

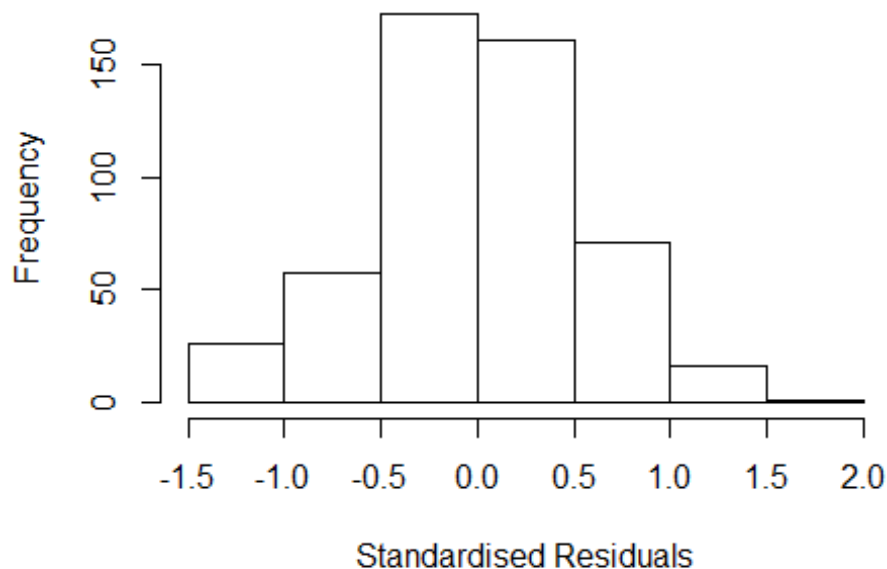
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.5
## 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	1.248	1	1.117
UniqueAuthors	1.488	4	1.051
Year	1.812	16	1.019

Residuals from first and last author and team size



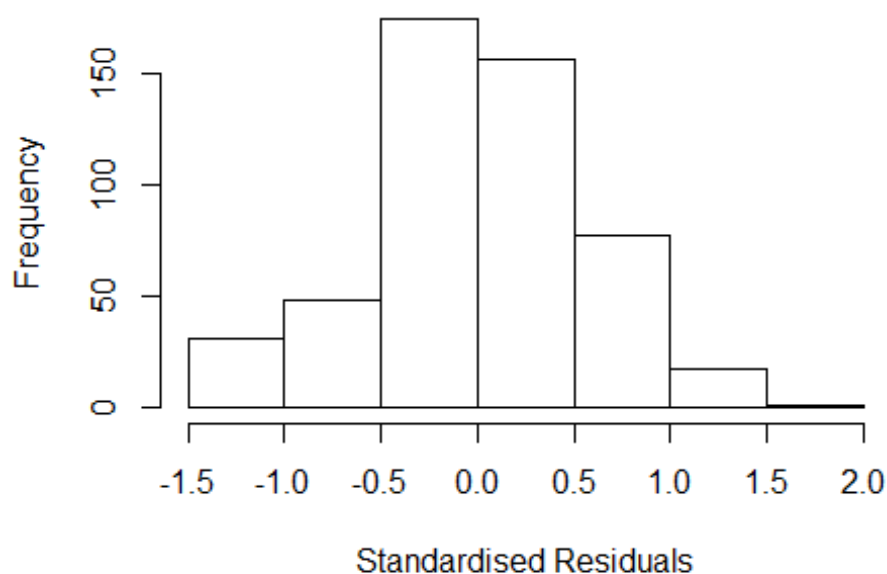
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4186 -0.3407 -0.0117  0.3394  1.6598
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.1232    0.2096   5.36 1.3e-07 ***
## FirstAuthorFemale1  0.0503    0.0574   0.88  0.3812
## LastAuthorFemale1  0.1208    0.0567   2.13  0.0336 *
## UniqueAuthors2     0.0711    0.0725   0.98  0.3273
## UniqueAuthors3     0.2397    0.0741   3.23  0.0013 **
## UniqueAuthors4     0.0593    0.0871   0.68  0.4962
## UniqueAuthors5     0.0657    0.0805   0.82  0.4148
## Year1997         -0.0572    0.2203  -0.26  0.7952
## Year1998         -0.1729    0.2511  -0.69  0.4913
## Year1999         -0.1072    0.2314  -0.46  0.6433
```

```

## Year2000          -0.0240      0.2627   -0.09   0.9272
## Year2001          -0.2429      0.2525   -0.96   0.3364
## Year2002          -0.3698      0.2198   -1.68   0.0931 .
## Year2003          -0.1878      0.2274   -0.83   0.4093
## Year2004           0.0977      0.2263    0.43   0.6662
## Year2005          -0.2113      0.2187   -0.97   0.3345
## Year2006           0.2297      0.2504    0.92   0.3595
## Year2007          -0.1562      0.2151   -0.73   0.4681
## Year2008          -0.2057      0.2189   -0.94   0.3479
## Year2009           0.0202      0.2152    0.09   0.9254
## Year2010          -0.1721      0.2107   -0.82   0.4144
## Year2011          -0.2218      0.2217   -1.00   0.3175
## Year2012          -0.2379      0.2131   -1.12   0.2648
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.537
## Multiple R-squared:  0.0855, Adjusted R-squared:  0.0439
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 461 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.319  0.873  0.957  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      1.98e-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 1.208 1      1.099
## LastAuthorFemale  1.227 1      1.108
## Year              1.242 16      1.007

```

Residuals from first and last author



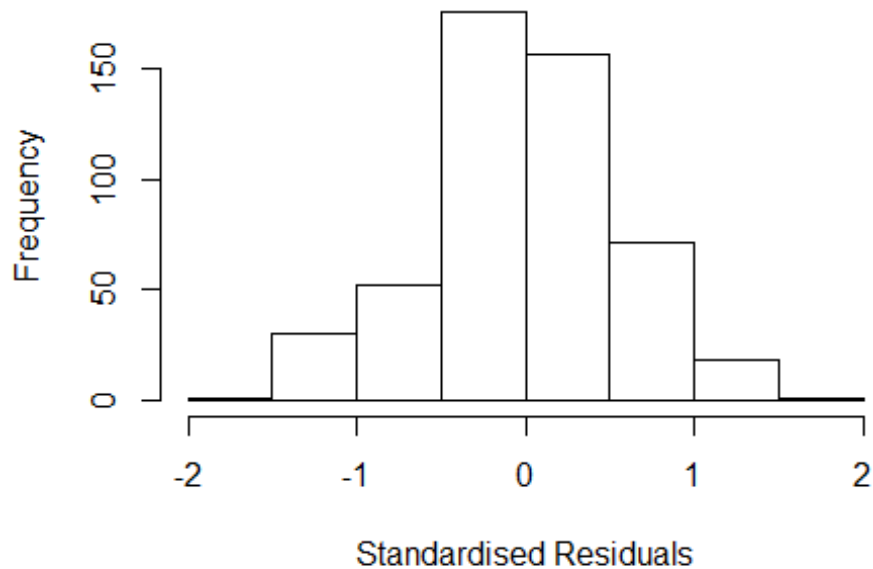
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.463602 -0.337712 -0.000327 0.338036 1.574395
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1868 0.1970 6.02 3.4e-09 ***
## FirstAuthorFemale1 0.0665 0.0570 1.17 0.245
## LastAuthorFemale1 0.1044 0.0572 1.82 0.069 .
## Year1997 -0.0738 0.2130 -0.35 0.729
## Year1998 -0.1846 0.2455 -0.75 0.452
## Year1999 -0.0966 0.2243 -0.43 0.667
## Year2000 -0.0027 0.2598 -0.01 0.992
## Year2001 -0.2029 0.2488 -0.82 0.415
## Year2002 -0.3525 0.2130 -1.65 0.099 .
## Year2003 -0.1818 0.2166 -0.84 0.402
## Year2004 0.0867 0.2192 0.40 0.693
## Year2005 -0.1756 0.2066 -0.85 0.396
```

```

## Year2006          0.2768      0.2375      1.17      0.244
## Year2007          -0.1264      0.2066     -0.61      0.541
## Year2008          -0.1971      0.2091     -0.94      0.346
## Year2009           0.0323      0.2081      0.16      0.877
## Year2010          -0.1501      0.2020     -0.74      0.458
## Year2011          -0.1926      0.2142     -0.90      0.369
## Year2012          -0.1934      0.2037     -0.95      0.343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.0677, Adjusted R-squared:  0.0333
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 464 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.376  0.868  0.956  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      1.98e-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 1.089 1          1.043
## Year              1.089 16          1.003

```

Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5183 -0.3446 -0.0256 0.3567 1.5879
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2091 0.1865 6.48 2.2e-10 ***
## FirstAuthorFemale1 0.1024 0.0551 1.86 0.064 .
## Year1997 -0.0541 0.2040 -0.27 0.791
## Year1998 -0.1707 0.2375 -0.72 0.473
## Year1999 -0.0761 0.2174 -0.35 0.727
## Year2000 0.0330 0.2573 0.13 0.898
## Year2001 -0.1859 0.2442 -0.76 0.447
## Year2002 -0.3372 0.2045 -1.65 0.100 .
## Year2003 -0.1699 0.2079 -0.82 0.414
## Year2004 0.1060 0.2088 0.51 0.612
## Year2005 -0.1531 0.1984 -0.77 0.441
## Year2006 0.3092 0.2309 1.34 0.181
```

```

## Year2007          -0.1064      0.1962   -0.54    0.588
## Year2008          -0.1660      0.1989   -0.83    0.404
## Year2009           0.0660      0.1989    0.33    0.740
## Year2010          -0.1173      0.1911   -0.61    0.540
## Year2011          -0.1766      0.2059   -0.86    0.391
## Year2012          -0.1826      0.1942   -0.94    0.348
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.0617, Adjusted R-squared:  0.029
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 465 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.370  0.870  0.956  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      1.98e-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 1.121 1      1.059
## Year              1.121 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
## -1.5028 -0.3535 -0.0107  0.3434  1.5820

```



```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22649    0.19904    6.16 1.5e-09 ***
## LastAuthorFemale1 0.12529    0.05503    2.28  0.023 *
## Year1997       -0.07355    0.22131   -0.33  0.740
## Year1998       -0.19169    0.25159   -0.76  0.446
## Year1999       -0.09339    0.23130   -0.40  0.687
## Year2000       -0.00554    0.26649   -0.02  0.983
## Year2001       -0.21439    0.25847   -0.83  0.407
## Year2002       -0.36291    0.21969   -1.65  0.099 .
## Year2003       -0.18986    0.22355   -0.85  0.396
## Year2004        0.08296    0.22702    0.37  0.715
## Year2005       -0.18343    0.21334   -0.86  0.390
## Year2006        0.27629    0.24467    1.13  0.259
## Year2007       -0.13180    0.21466   -0.61  0.540
## Year2008       -0.21179    0.21558   -0.98  0.326
## Year2009        0.02176    0.21508    0.10  0.919
## Year2010       -0.15182    0.21046   -0.72  0.471
## Year2011       -0.19827    0.22119   -0.90  0.370
## Year2012       -0.19748    0.21175   -0.93  0.351
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.539
## Multiple R-squared:  0.0657, Adjusted R-squared:  0.0332
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 459 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.370  0.867  0.954  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      1.98e-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: 506"
## [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
##    6   10    9    8   11   12   12    8   10    9   14   11   10   16    8
## 2011 2012
##   16   26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    3    4    3    3    6    6    5   10    9   14   10    8   16    7
## 2011 2012
##   15   24
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    3    4    3    3    6    6    5   10    7   11   10    8   15    7
## 2011 2012
##   14   23
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.28942848510666"
## [1] "Male first author team size 2018 geometric mean: 3.00410118433536"

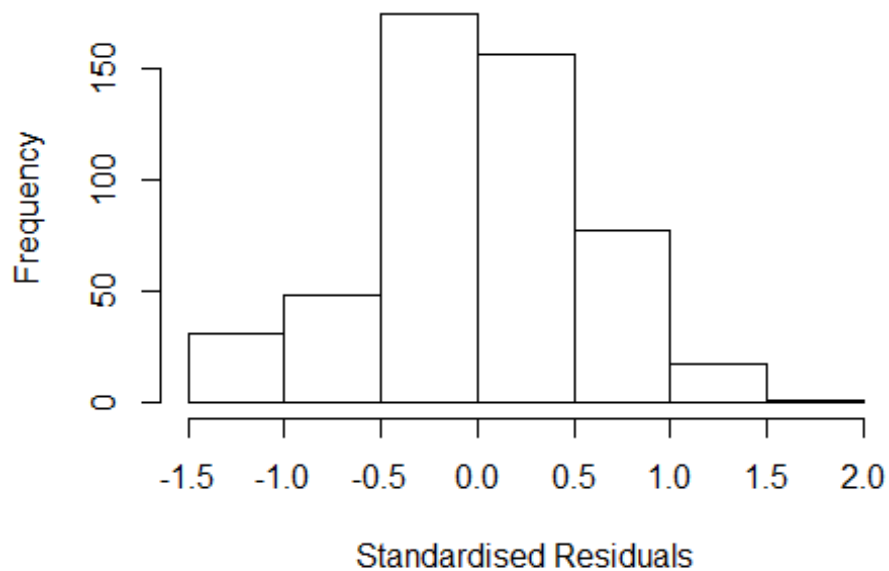
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.28646055545368"
## [1] "Male last author team size 2018 geometric mean: 1.91293118277239"

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

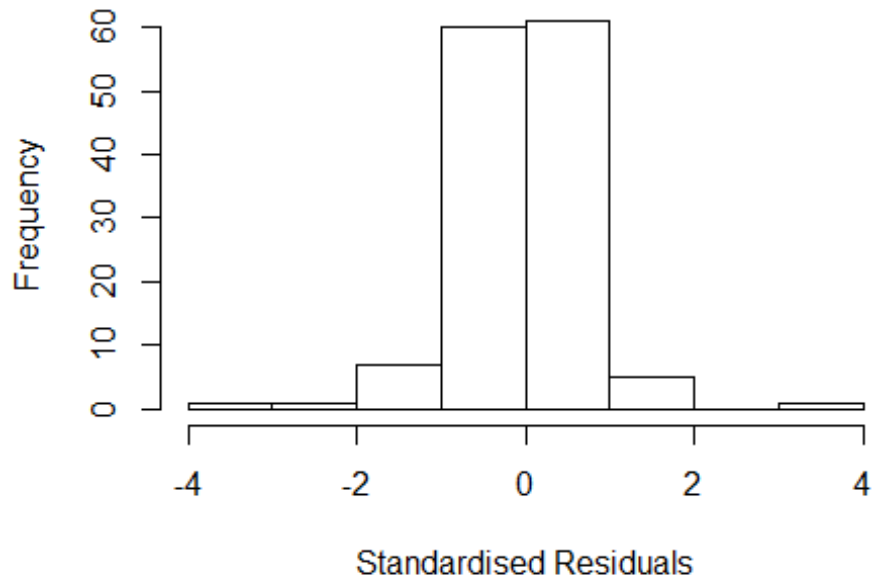
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 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 480.64 1      21.923
## LastAuthorFemale  26.29 1       5.127
## UniqueAuthors    1349.82 4       2.462
## Year              59208.72 16      1.410
```

Residuals from first and last author and team size



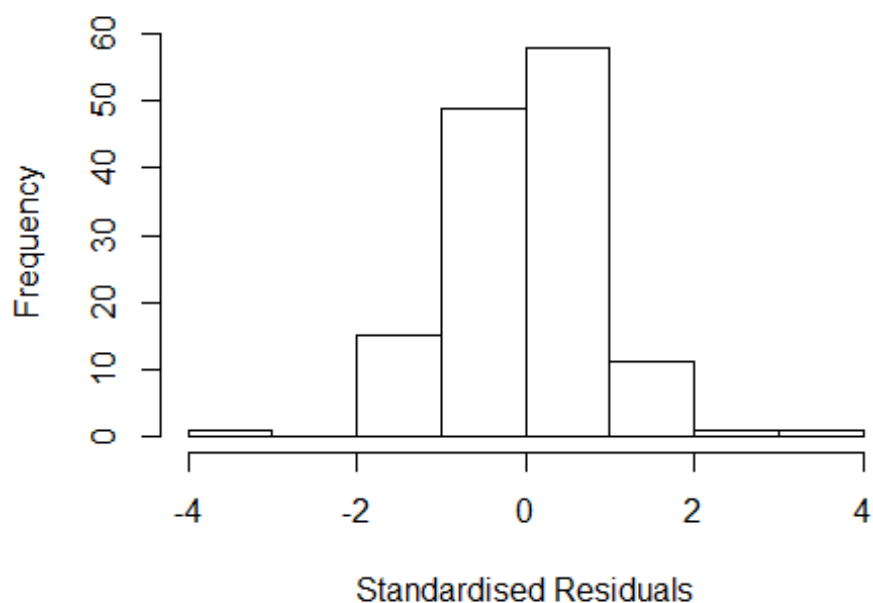
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 17 85026136563 0.000 1997    2902     5   -3.557
## 19 0032013169 3.837 1998    2906     1    3.374
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -3.55702 -0.41962 -0.00251  0.43143  3.37350
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.139      0.416   -0.33  0.73926
## FirstAuthorFemale1  0.438      0.348    1.26  0.21102
## LastAuthorFemale1 -0.299      0.153   -1.96  0.05266 .
## UniqueAuthors2    0.534      0.230    2.32  0.02202 *
## UniqueAuthors3    0.971      0.285    3.40  0.00093 ***
## UniqueAuthors4    0.973      0.245    3.98  0.00012 ***
## UniqueAuthors5    0.945      0.207    4.57  1.3e-05 ***
## Year1997         3.023      0.413    7.31  4.0e-11 ***
## Year1998         0.463      0.869    0.53  0.59469
```

```

## Year1999          0.628      0.973      0.65  0.51972
## Year2000          1.415      0.730      1.94  0.05503 .
## Year2001          0.297      0.455      0.65  0.51411
## Year2002          1.569      0.555      2.83  0.00551 **
## Year2003          1.022      0.698      1.46  0.14611
## Year2004          1.284      0.315      4.08  8.4e-05 ***
## Year2005          0.784      0.426      1.84  0.06807 .
## Year2006          0.702      0.262      2.68  0.00847 **
## Year2007          0.970      0.265      3.66  0.00039 ***
## Year2008          0.858      0.351      2.45  0.01600 *
## Year2009          0.684      0.225      3.05  0.00289 **
## Year2010          1.398      0.241      5.80  6.1e-08 ***
## Year2011          1.051      0.223      4.71  7.1e-06 ***
## Year2012          0.847      0.216      3.92  0.00015 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.625
## Multiple R-squared:  0.447, Adjusted R-squared:  0.34
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 2 observations c(4,5) are outliers with |weight| = 0 ( < 0.00074);
## 15 weights are ~= 1. The remaining 119 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.166  0.872  0.950  0.896  0.981  0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          7.35e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##           0          1000          0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 6.255 1 2.501
## LastAuthorFemale 2.591 1 1.610
## Year 14.115 16 1.086

```

Residuals from first and last author



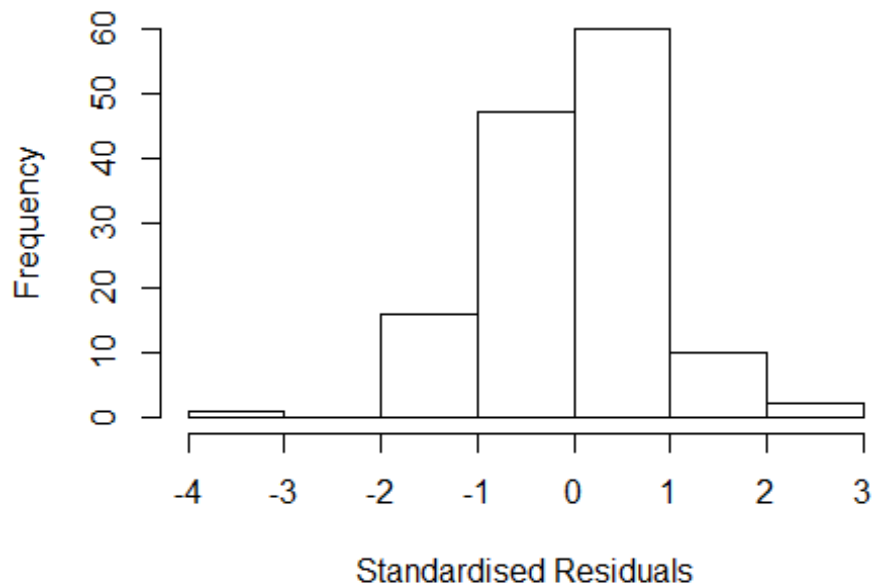
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 17 85026136563 0.000 1997    2902     5   -3.438
## 19 0032013169 3.837 1998    2906     1    3.245
## 71 0038050385 3.718 2003    2906     1    2.587
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4383 -0.4108  0.0423  0.4650  3.2455
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.222     0.509   -0.44  0.6639
## FirstAuthorFemale1  0.651     0.443    1.47  0.1450
## LastAuthorFemale1 -0.429     0.150   -2.86  0.0050 **
## Year1997        3.438     0.540    6.36 4.0e-09 ***
## Year1998         0.592     0.844    0.70  0.4846
## Year1999         1.158     0.947    1.22  0.2236
## Year2000         1.798     1.362    1.32  0.1892
## Year2001         0.880     0.554    1.59  0.1150
## Year2002         1.752     0.787    2.22  0.0280 *
## Year2003         1.131     0.386    2.93  0.0041 **
```

```

## Year2004          1.768      0.312      5.67  1.0e-07 ***
## Year2005          1.162      0.604      1.92   0.0569 .
## Year2006          1.377      0.168      8.19  3.7e-13 ***
## Year2007          1.666      0.194      8.60  4.2e-14 ***
## Year2008          1.714      0.263      6.51  2.0e-09 ***
## Year2009          1.418      0.122     11.63 < 2e-16 ***
## Year2010          2.171      0.174     12.46 < 2e-16 ***
## Year2011          1.801      0.147     12.26 < 2e-16 ***
## Year2012          1.542      0.182      8.49  7.5e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.686
## Multiple R-squared:  0.284, Adjusted R-squared:  0.174
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## 2 observations c(4,5) are outliers with |weight| = 0 ( < 0.00074);
## 12 weights are ~ = 1. The remaining 122 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.123  0.851  0.953   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          7.35e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 17.28 1           4.157
## Year              17.28 16           1.093

```

Residuals from first author



```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 17 85026136563 0.000 1997    2902     5   -3.438
## 19 0032013169 3.837 1998    2906     1    3.245
## 71 0038050385 3.718 2003    2906     1    2.587
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5198 -0.4937  0.0356  0.4661  2.9225
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.518     0.378   -1.37   0.1733
## FirstAuthorFemale1  0.518     0.378    1.37   0.1733
## Year1997       3.520     0.391    9.00 4.5e-15 ***
## Year1998       0.915     1.313    0.70  0.4876
## Year1999       1.174     0.872    1.35  0.1808
## Year2000       1.794     1.667    1.08  0.2839
## Year2001       1.139     0.485    2.35  0.0205 *
## Year2002       1.792     0.576    3.11  0.0023 **
## Year2003       1.290     0.499    2.58  0.0110 *
## Year2004       1.844     0.338    5.45 2.8e-07 ***
```

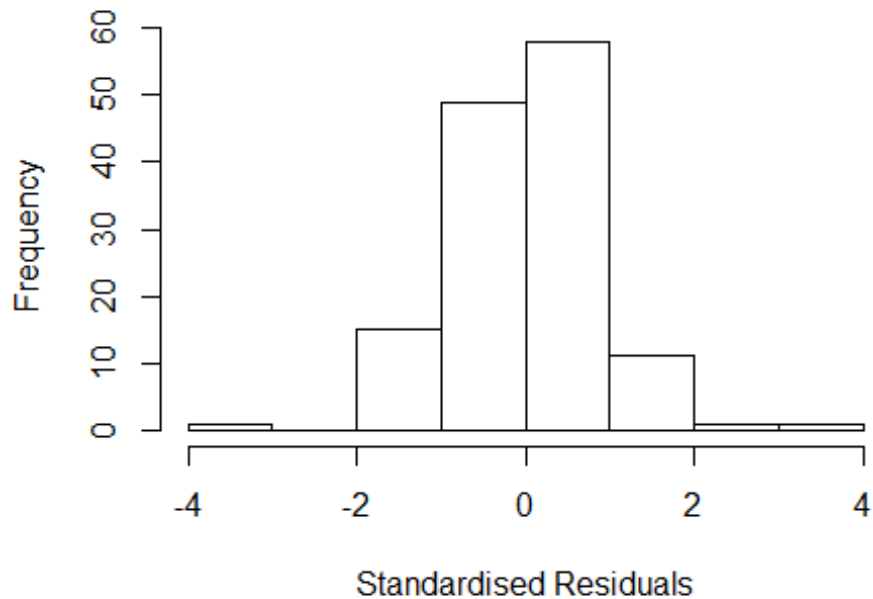


```

## Year2005          1.149      0.552      2.08      0.0395 *
## Year2006          1.456      0.179      8.15      4.4e-13 ***
## Year2007          1.753      0.193      9.08      3.0e-15 ***
## Year2008          1.783      0.236      7.55      1.0e-11 ***
## Year2009          1.416      0.124     11.44      < 2e-16 ***
## Year2010          2.215      0.187     11.86      < 2e-16 ***
## Year2011          1.863      0.154     12.08      < 2e-16 ***
## Year2012          1.637      0.175      9.37      6.4e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.738
## Multiple R-squared:  0.232, Adjusted R-squared:  0.121
## Convergence in 29 IRWLS iterations
##
## Robustness weights:
## observation 4 is an outlier with |weight| = 0 ( < 0.00074);
## 15 weights are ~= 1. The remaining 120 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.082  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          7.35e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 2.953  1           1.718
## Year             2.953 16           1.034

```

Residuals from last author



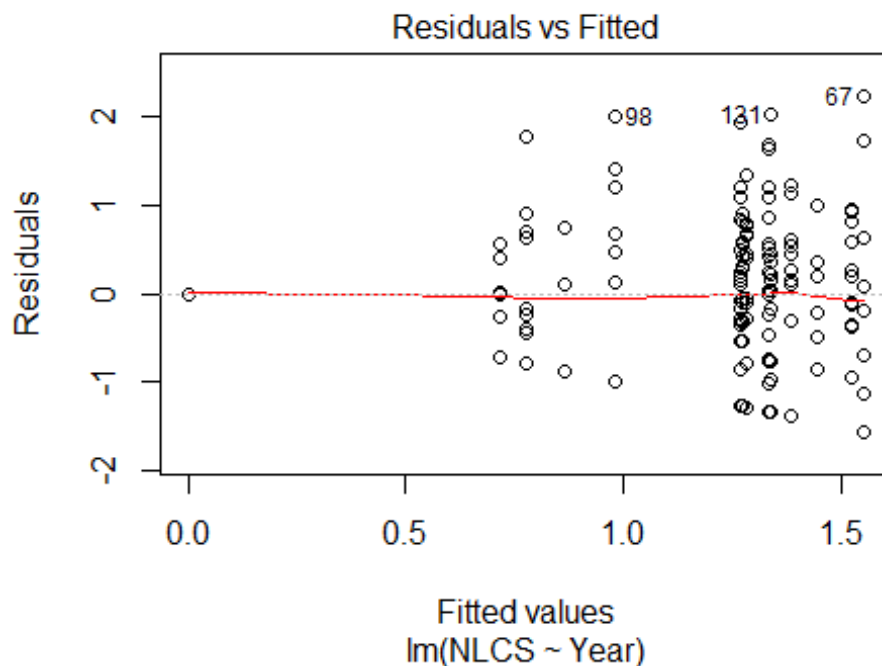
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 17 85026136563 0.000 1997    2902     5   -3.438
## 19 0032013169 3.837 1998    2906     1    3.245
## 71 0038050385 3.718 2003    2906     1    2.587
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6590 -0.4275  0.0322  0.4746  3.6086
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.394      0.219   1.80  0.0742 .
## LastAuthorFemale1 -0.394      0.219  -1.80  0.0742 .
## Year1997         2.659      0.588   4.52 1.5e-05 ***
## Year1998         0.228      0.881   0.26  0.7959
## Year1999         1.159      0.942   1.23  0.2210
## Year2000         1.552      0.799   1.94  0.0545 .
## Year2001         0.591      0.607   0.97  0.3321
## Year2002         1.884      0.675   2.79  0.0061 **
## Year2003         0.935      0.244   3.83  0.0002 ***
## Year2004         1.776      0.313   5.67 1.0e-07 ***
```

```

## Year2005          1.161      0.601      1.93      0.0559 .
## Year2006          1.384      0.171      8.08      6.3e-13 ***
## Year2007          1.618      0.164      9.84      < 2e-16 ***
## Year2008          1.764      0.264      6.68      8.1e-10 ***
## Year2009          1.418      0.122     11.62      < 2e-16 ***
## Year2010          2.088      0.203     10.27      < 2e-16 ***
## Year2011          1.806      0.150     12.02      < 2e-16 ***
## Year2012          1.433      0.159      9.02      4.0e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.688
## Multiple R-squared:  0.268, Adjusted R-squared:  0.162
## Convergence in 27 IRWLS iterations
##
## Robustness weights:
## observation 5 is an outlier with |weight| = 0 ( < 0.00074);
## 11 weights are ~= 1. The remaining 124 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0651 0.8520 0.9480 0.8830 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      7.35e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 136"
## [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
##    3    8    7   16   13   13   17   11   13   13   11    6   17   19   13
## 2011 2012
##   17   17
##

```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    4    3    6    6   12    9    8   12   12   10    6   14   16   12
## 2011 2012
##   14   13
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    4    3    5    6   12    8    8   12   12    8    6    9   13   12
## 2011 2012
##   14   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 = Inf, df = 16, p-value <2e-16
```

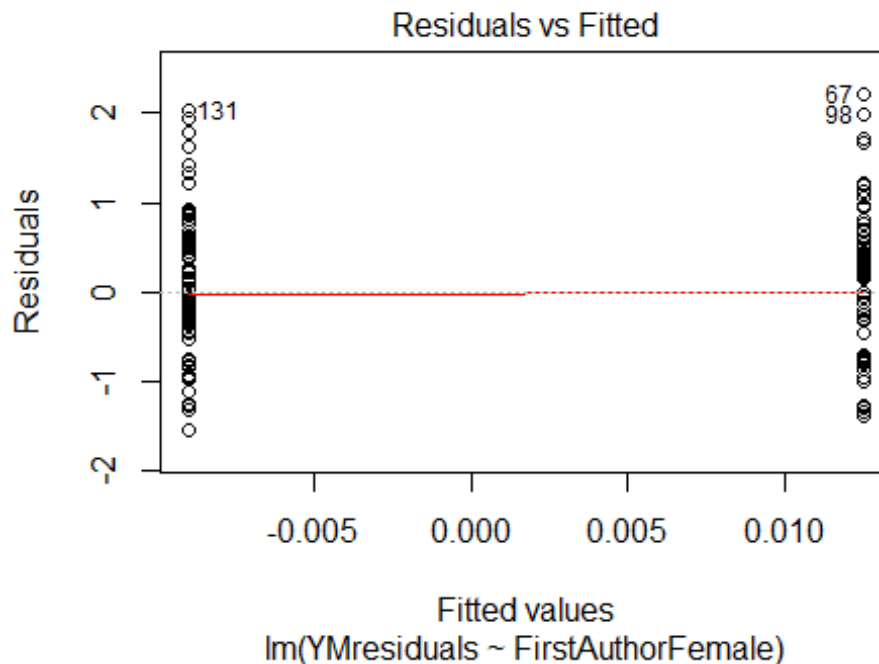


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.3, df = 1, p-value = 0.1
## [1] "Female first author team size 2018 geometric mean: 4.0356543087298"
## [1] "Male first author team size 2018 geometric mean: 4.82134352483793"
```

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

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 28, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.65905165331721"
## [1] "Male last author team size 2018 geometric mean: 5.02807592269694"

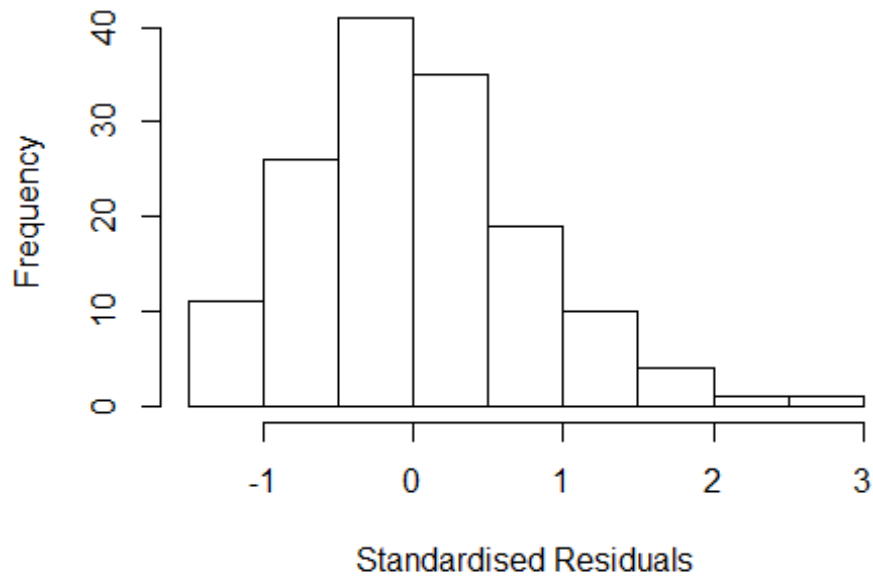
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 18, p-value = 0.06
## 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 2.763 1 1.662
```

## LastAuthorFemale	3.080	1	1.755
## UniqueAuthors	10.998	4	1.349
## Year	15.176	16	1.089

Residuals from first and last author and team size



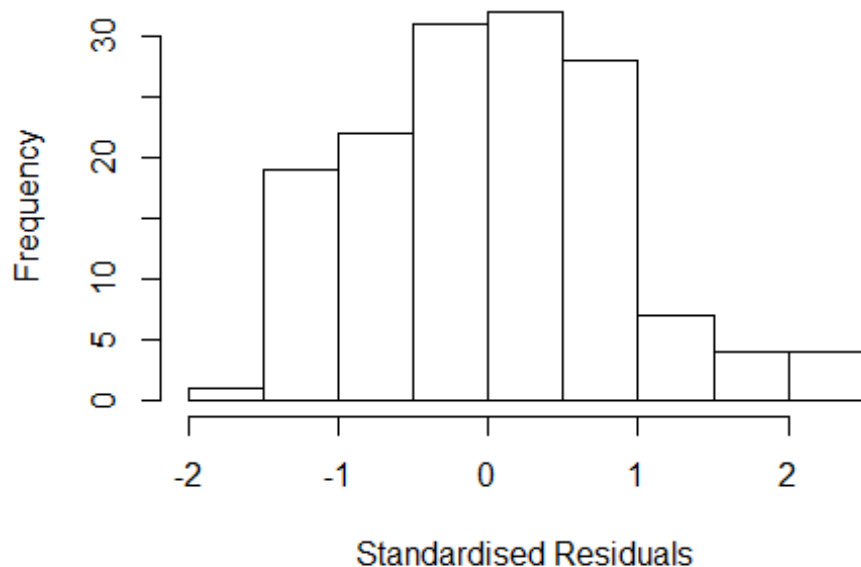
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4 0030622475 3.192 1997    2711      2      2.592
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.306 -0.502 -0.023  0.466  2.592
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.2332    0.2827   0.82   0.411
## FirstAuthorFemale1 0.1312    0.1940   0.68   0.500
## LastAuthorFemale1 0.0863    0.2180   0.40   0.693
## UniqueAuthors2    0.6289    0.3482   1.81   0.073 .
## UniqueAuthors3    0.8184    0.1922   4.26 4.0e-05 ***
## UniqueAuthors4    0.8482    0.2006   4.23 4.5e-05 ***
```

```

## UniqueAuthors5      1.2831      0.2038      6.30 4.7e-09 ***
## Year1997             -0.2621      0.7539     -0.35  0.729
## Year1998             -0.5842      0.3379     -1.73  0.086 .
## Year1999             -0.2811      0.3068     -0.92  0.361
## Year2000              0.6606      0.4787      1.38  0.170
## Year2001             -0.2177      0.3103     -0.70  0.484
## Year2002              0.7981      0.5926      1.35  0.180
## Year2003              0.3921      0.4644      0.84  0.400
## Year2004             -0.3481      0.5123     -0.68  0.498
## Year2005              0.2219      0.3214      0.69  0.491
## Year2006              0.2243      0.3386      0.66  0.509
## Year2007              0.3388      0.2990      1.13  0.259
## Year2008              0.0820      0.4000      0.21  0.838
## Year2009              0.3924      0.3218      1.22  0.225
## Year2010              0.4536      0.3576      1.27  0.207
## Year2011              0.2263      0.3286      0.69  0.492
## Year2012              0.5610      0.2955      1.90  0.060 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.743
## Multiple R-squared:  0.362, Adjusted R-squared:  0.25
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 134 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.199  0.877  0.953  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      6.76e-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 2.341 1          1.530
## LastAuthorFemale  2.418 1          1.555
## Year              2.743 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
## -1.5641 -0.6566 0.0348 0.5743 2.3487
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8889 0.4026 2.21 0.029 *
## FirstAuthorFemale1 0.0945 0.2181 0.43 0.666
## LastAuthorFemale1 -0.1435 0.2280 -0.63 0.530
## Year1997 0.2048 0.8848 0.23 0.817
## Year1998 -0.8726 0.4021 -2.17 0.032 *
## Year1999 -0.0652 0.4571 -0.14 0.887
## Year2000 0.4214 0.5003 0.84 0.401
## Year2001 -0.1833 0.4756 -0.39 0.701
## Year2002 0.6752 0.7085 0.95 0.342
## Year2003 0.3580 0.6022 0.59 0.553
## Year2004 0.0233 0.5687 0.04 0.967
## Year2005 0.5572 0.4835 1.15 0.251
```

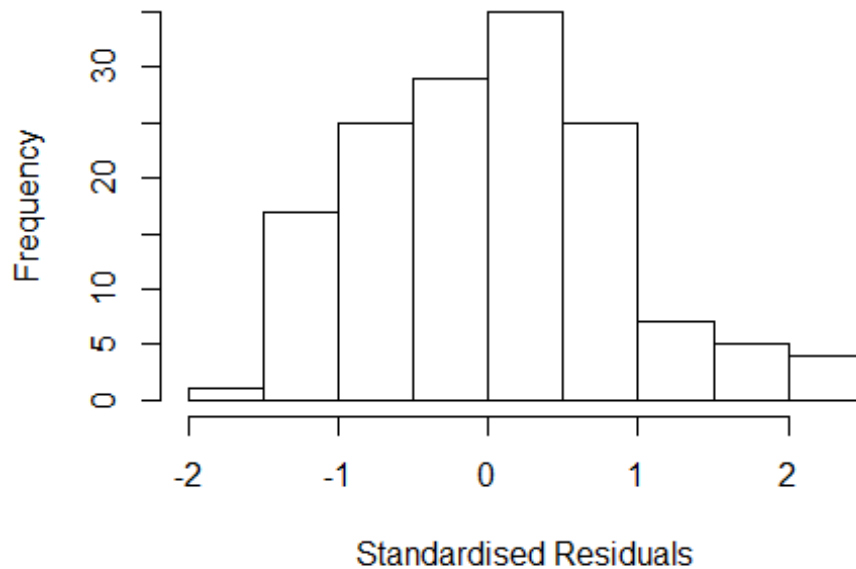


```

## Year2006          0.2749      0.5166      0.53      0.596
## Year2007          0.5118      0.4841      1.06      0.292
## Year2008          0.4156      0.4947      0.84      0.402
## Year2009          0.5791      0.4718      1.23      0.222
## Year2010          0.3913      0.4621      0.85      0.399
## Year2011          0.4721      0.4815      0.98      0.329
## Year2012          0.6643      0.4513      1.47      0.143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.874
## Multiple R-squared:  0.119, Adjusted R-squared:  -0.00434
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 136 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.450  0.881  0.946  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.76e-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 1.616 1          1.271
## Year              1.616 16          1.015

```

Residuals from first author



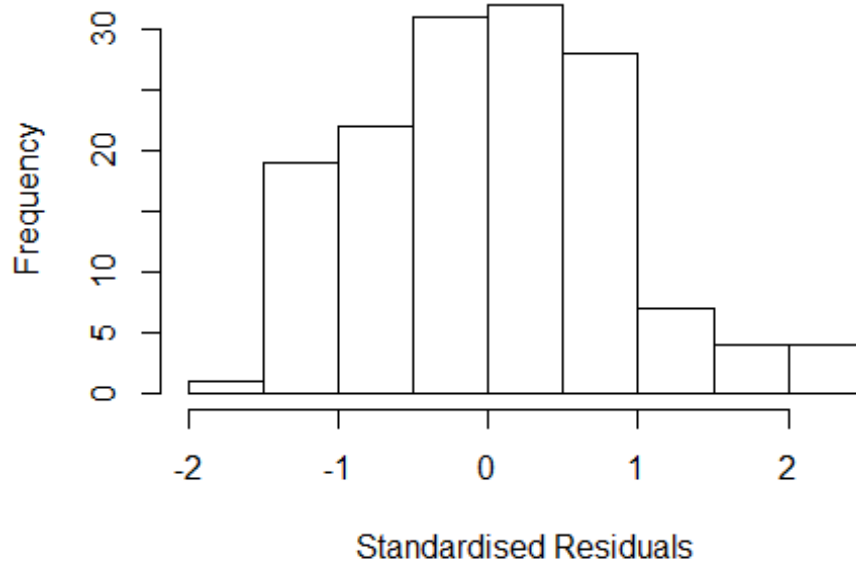
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.560 -0.710 0.015 0.581 2.224
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.87183 0.41896 2.08 0.039 *
## FirstAuthorFemale1 0.00881 0.18405 0.05 0.962
## Year1997 0.18595 0.88542 0.21 0.834
## Year1998 -0.87476 0.41968 -2.08 0.039 *
## Year1999 -0.09767 0.46514 -0.21 0.834
## Year2000 0.42199 0.51584 0.82 0.415
## Year2001 -0.15656 0.48767 -0.32 0.749
## Year2002 0.68803 0.70935 0.97 0.334
## Year2003 0.35552 0.60349 0.59 0.557
## Year2004 0.01205 0.57869 0.02 0.983
## Year2005 0.56238 0.50577 1.11 0.268
## Year2006 0.27305 0.53236 0.51 0.609
```

```

## Year2007          0.56082    0.50051    1.12    0.265
## Year2008          0.41470    0.51291    0.81    0.420
## Year2009          0.57636    0.48642    1.18    0.238
## Year2010          0.39901    0.48015    0.83    0.407
## Year2011          0.43731    0.49223    0.89    0.376
## Year2012          0.65223    0.46499    1.40    0.163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.882
## Multiple R-squared:  0.114, Adjusted R-squared:  -0.00236
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 138 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.504  0.890   0.942   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      6.76e-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 1.628 1      1.276
## Year              1.628 16      1.015

```

Residuals from last author



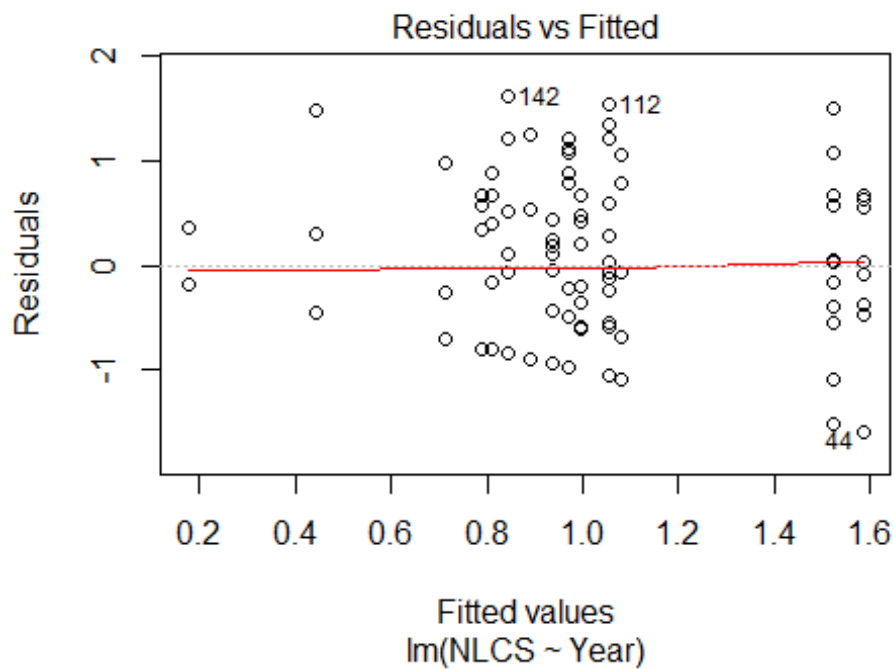
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5846 -0.6677 0.0389 0.5435 2.2763
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8967 0.3913 2.29 0.024 *
## LastAuthorFemale1 -0.0776 0.1881 -0.41 0.681
## Year1997 0.2386 0.8624 0.28 0.782
## Year1998 -0.8709 0.3908 -2.23 0.028 *
## Year1999 -0.0752 0.4429 -0.17 0.865
## Year2000 0.4195 0.4900 0.86 0.394
## Year2001 -0.1515 0.4538 -0.33 0.739
## Year2002 0.6879 0.6870 1.00 0.319
## Year2003 0.3721 0.5854 0.64 0.526
## Year2004 0.0255 0.5485 0.05 0.963
## Year2005 0.5601 0.4770 1.17 0.242
## Year2006 0.2735 0.5077 0.54 0.591
```

```

## Year2007          0.5566      0.4663      1.19      0.235
## Year2008          0.4284      0.4840      0.89      0.378
## Year2009          0.5780      0.4617      1.25      0.213
## Year2010          0.3964      0.4556      0.87      0.386
## Year2011          0.4660      0.4714      0.99      0.325
## Year2012          0.6634      0.4412      1.50      0.135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.909
## Multiple R-squared:  0.113, Adjusted R-squared:  -0.00283
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 134 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.510  0.883  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      6.76e-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: 148"
## [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   13    9    3    7   10   10   10    6    6    5   12   19   13   10
## 2012
##    9
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    6    5    0    3    9    7    8    4    5    3   12   16   12    9
## 2012

```

```
##      8
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    6    5    0    3    9    7    8    3    4    3   12   14   10    8
## 2012
##    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 = 9.1, df = 13, p-value = 0.8
```



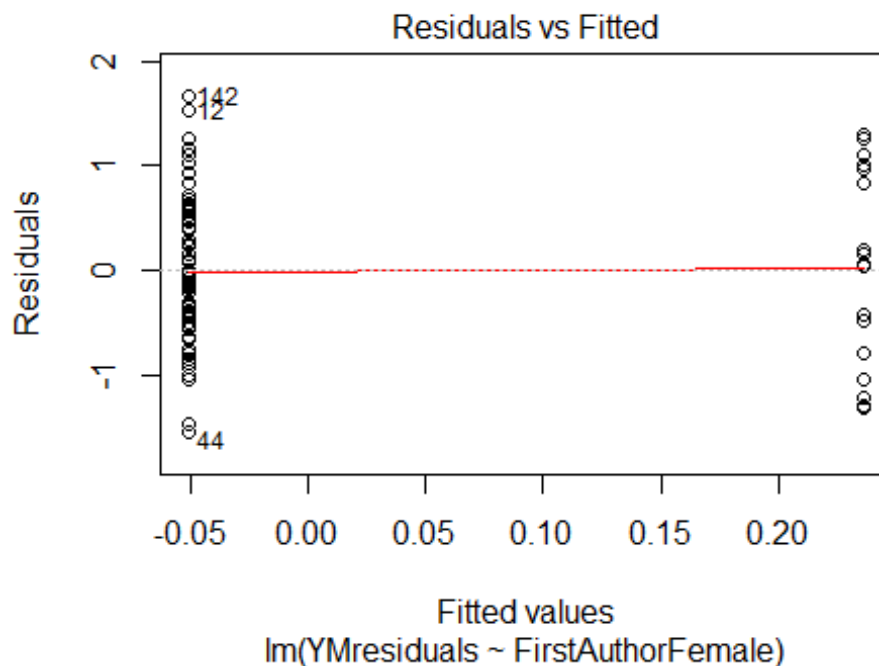
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.7, df = 1, p-value = 0.06

## [1] "Female first author team size 2018 geometric mean: 2.42080815083869"
## [1] "Male first author team size 2018 geometric mean: 1.43096908110526"

## Warning in wilcox.test.default(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 = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.3284355309218"
## [1] "Male last author team size 2018 geometric mean: 1.73851050644476"

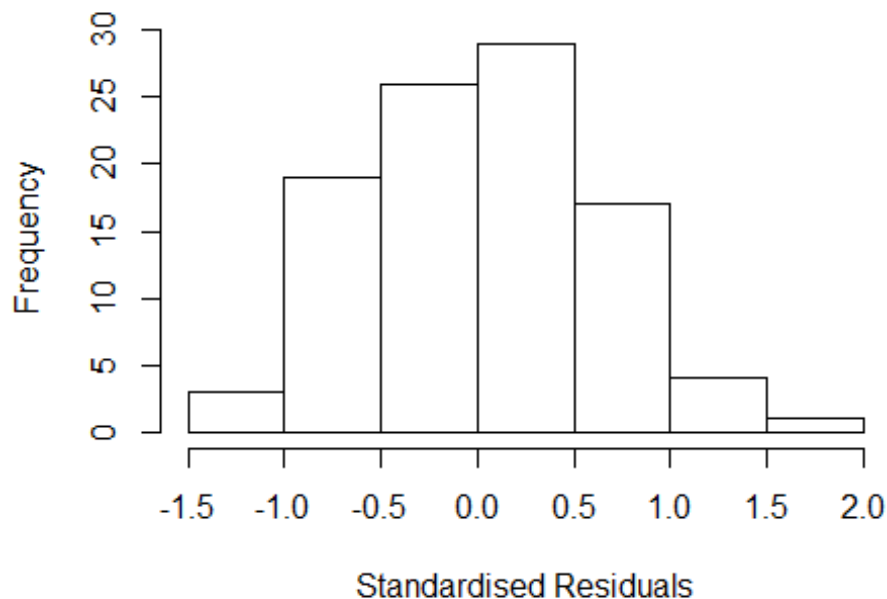
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 36, p-value = 0.4
## 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	6.030	1	2.456
LastAuthorFemale	3.433	1	1.853
UniqueAuthors	10.519	4	1.342
Year	22.146	13	1.127

Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17839 -0.42396 0.00697 0.40468 1.64514
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.1329 0.2999 -0.44 0.65879
## FirstAuthorFemale1 -0.1266 0.3632 -0.35 0.72829
## LastAuthorFemale1 0.1418 0.2474 0.57 0.56816
## UniqueAuthors2 0.5136 0.1701 3.02 0.00341 **
## UniqueAuthors3 0.8938 0.2844 3.14 0.00235 **
## UniqueAuthors4 0.8122 0.2820 2.88 0.00511 **
## UniqueAuthors5 1.4626 0.2392 6.11 3.5e-08 ***
## Year1999 0.6961 0.3243 2.15 0.03489 *
## Year2001 0.7635 0.5416 1.41 0.16249
## Year2002 0.8577 0.3248 2.64 0.00996 **
```

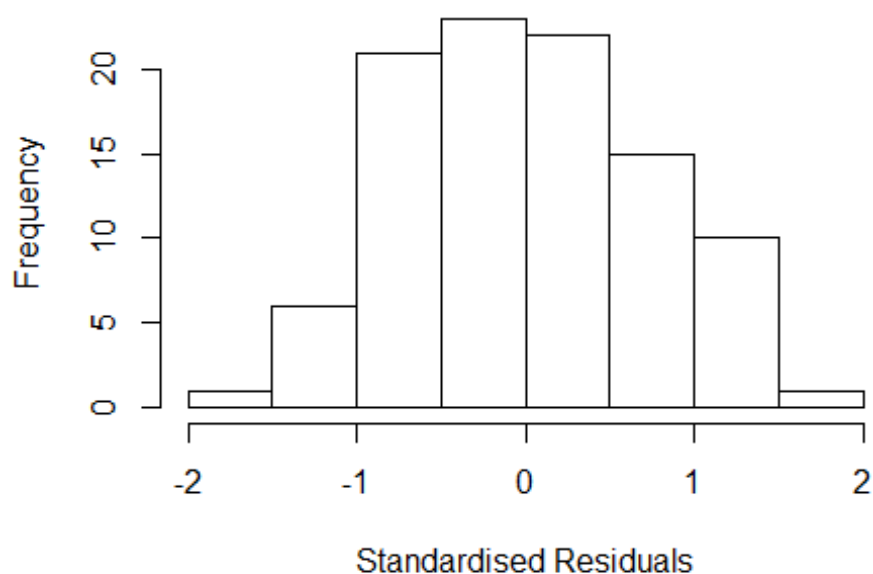


```

## Year2003          0.6226      0.3984      1.56  0.12210
## Year2004          0.5639      0.2526      2.23  0.02839 *
## Year2005         -0.0246      0.7973     -0.03  0.97549
## Year2006          0.7753      0.3406      2.28  0.02552 *
## Year2007          0.2961      0.2315      1.28  0.20457
## Year2008          0.4839      0.3666      1.32  0.19058
## Year2009          0.4847      0.2326      2.08  0.04044 *
## Year2010          1.0870      0.2841      3.83  0.00026 ***
## Year2011          0.2886      0.2568      1.12  0.26457
## Year2012          0.6138      0.2417      2.54  0.01308 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.602
## Multiple R-squared:  0.464, Adjusted R-squared:  0.335
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 92 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.436  0.871  0.940  0.910  0.978  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          1.01e-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 1.580 1          1.257
## LastAuthorFemale  1.521 1          1.233
## Year              1.744 13          1.022

```

Residuals from first and last author



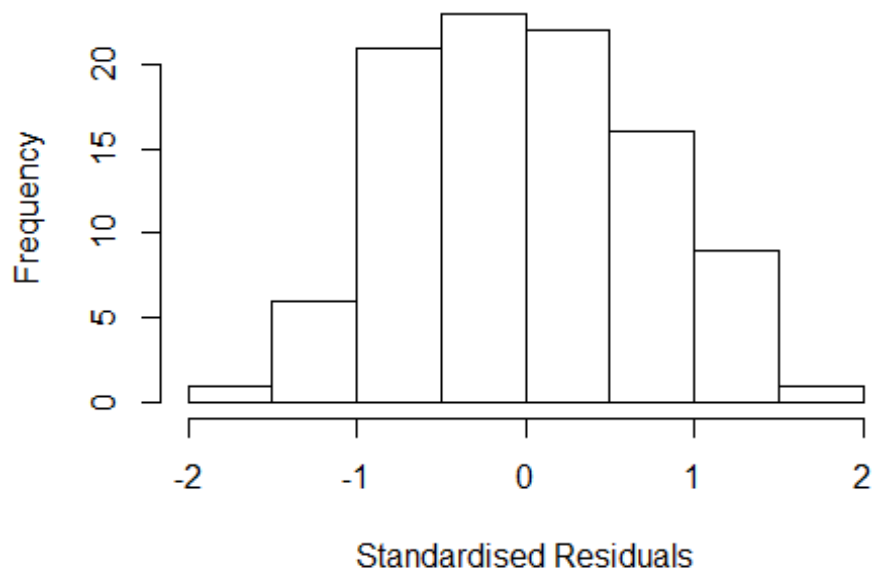
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6426 -0.5455 -0.0346 0.5613 1.6238
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6130 0.3599 1.70 0.09231 .
## FirstAuthorFemale1 -0.3544 0.3216 -1.10 0.27373
## LastAuthorFemale1 0.0407 0.2761 0.15 0.88327
## Year1999 0.5075 0.4351 1.17 0.24672
## Year2001 0.3796 0.5341 0.71 0.47928
## Year2002 1.3434 0.3624 3.71 0.00038 ***
## Year2003 0.4235 0.3696 1.15 0.25515
## Year2004 0.7011 0.3191 2.20 0.03078 *
## Year2005 0.2135 0.6203 0.34 0.73161
## Year2006 0.7549 0.7569 1.00 0.32152
## Year2007 -0.2298 0.3272 -0.70 0.48444
## Year2008 0.5980 0.4123 1.45 0.15067
```

```

## Year2009          0.5203      0.3257      1.60  0.11402
## Year2010          1.0554      0.3839      2.75  0.00733 **
## Year2011          0.2937      0.3667      0.80  0.42549
## Year2012          0.5456      0.3039      1.80  0.07626 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.783
## Multiple R-squared:  0.2,   Adjusted R-squared:  0.0559
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 92 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.639  0.884  0.948  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      1.01e-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 1.37 1          1.170
## Year              1.37 13          1.012

```

Residuals from first author



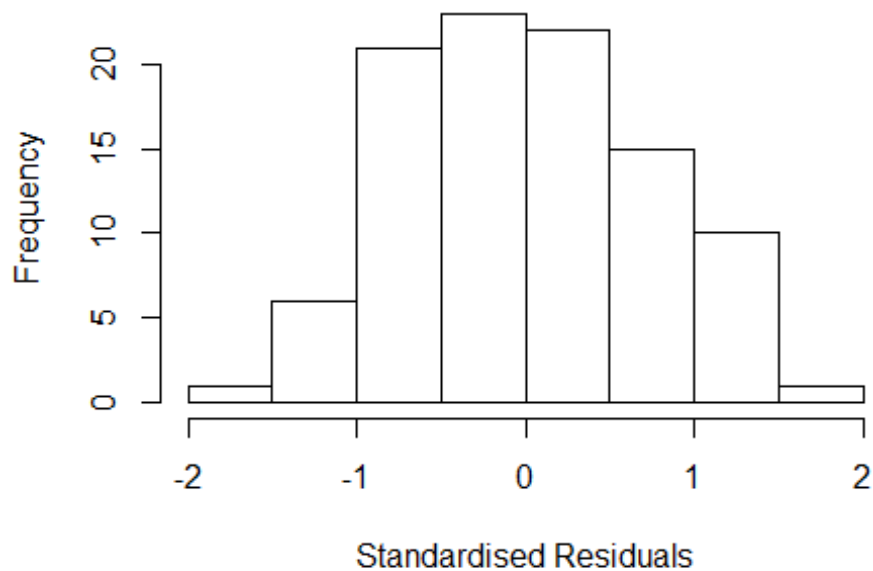
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6444 -0.5601 -0.0364 0.5601 1.6304
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.626 0.344 1.82 0.07260 .
## FirstAuthorFemale1 -0.333 0.307 -1.09 0.28048
## Year1999 0.515 0.436 1.18 0.24159
## Year2001 0.385 0.535 0.72 0.47375
## Year2002 1.352 0.363 3.72 0.00036 ***
## Year2003 0.425 0.372 1.14 0.25644
## Year2004 0.702 0.320 2.19 0.03117 *
## Year2005 0.220 0.637 0.35 0.73034
## Year2006 0.761 0.773 0.99 0.32742
## Year2007 -0.230 0.331 -0.69 0.49057
## Year2008 0.600 0.414 1.45 0.15087
## Year2009 0.523 0.328 1.60 0.11405
```

```

## Year2010          1.062      0.380      2.79  0.00650 **
## Year2011          0.287      0.363      0.79  0.43144
## Year2012          0.544      0.302      1.80  0.07563 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.772
## Multiple R-squared:  0.202, Adjusted R-squared:  0.0688
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 92 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.629  0.882  0.949  0.918  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      1.01e-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 1.346 1          1.160
## Year            1.346 13          1.012

```

Residuals from last author



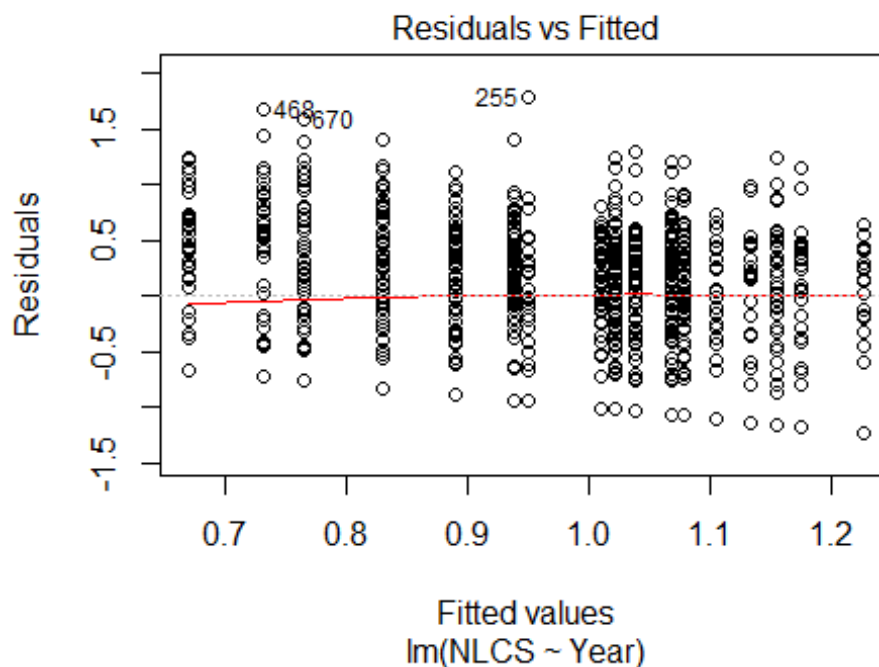
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6451 -0.5399 -0.0632 0.5597 1.6044
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.459 0.351 1.31 0.19538
## LastAuthorFemale1 -0.124 0.251 -0.49 0.62265
## Year1999 0.473 0.453 1.04 0.29974
## Year2001 0.343 0.550 0.62 0.53488
## Year2002 1.311 0.384 3.42 0.00098 ***
## Year2003 0.441 0.388 1.13 0.25975
## Year2004 0.645 0.344 1.87 0.06433 .
## Year2005 0.204 0.787 0.26 0.79594
## Year2006 0.747 0.714 1.05 0.29867
## Year2007 -0.202 0.335 -0.60 0.54820
## Year2008 0.577 0.429 1.35 0.18223
## Year2009 0.547 0.353 1.55 0.12501
```

```

## Year2010          1.082      0.415      2.60  0.01090 *
## Year2011          0.222      0.389      0.57  0.57038
## Year2012          0.514      0.333      1.54  0.12663
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.768
## Multiple R-squared:  0.187, Adjusted R-squared:  0.0517
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 93 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.625  0.879  0.953  0.916  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      1.01e-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: 99"
## [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
##   31   51   36   35   47   71   69   76   71   97   75   80   63   92   91
## 2011 2012
##   86   82
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   27   40   27   23   28   33   62   64   57   83   70   76   55   76   85
## 2011 2012
##   80   76
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 24 39 26 22 27 30 55 58 53 75 61 68 51 70 77
## 2011 2012
## 74 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
```

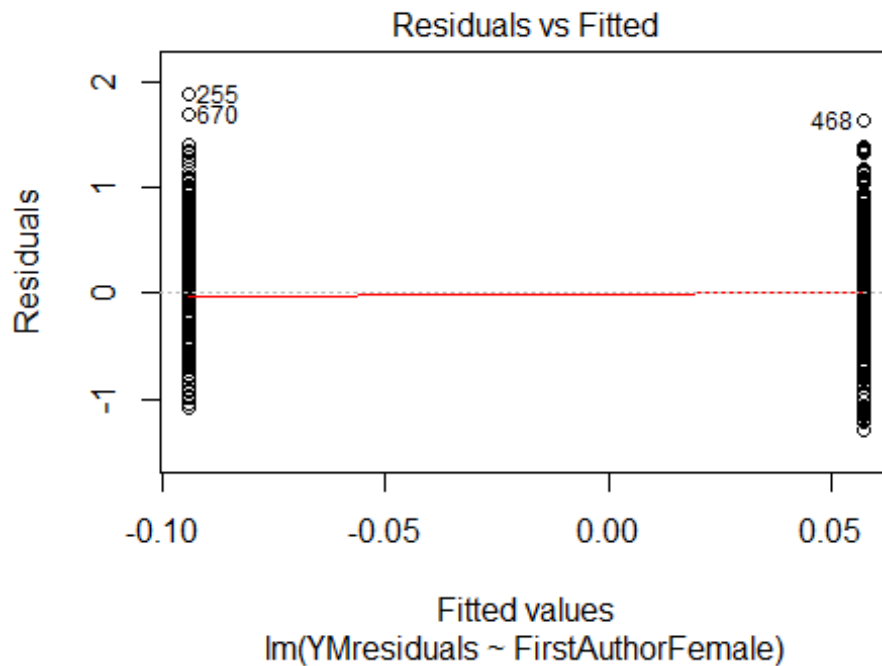


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.4, df = 1, p-value = 0.01
##
## [1] "Female first author team size 2018 geometric mean: 3.63066847106221"
## [1] "Male first author team size 2018 geometric mean: 3.08733789977856"
##
## 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] "Female last author team size 2018 geometric mean: 3.50147500206034"
## [1] "Male last author team size 2018 geometric mean: 3.4558109434229"

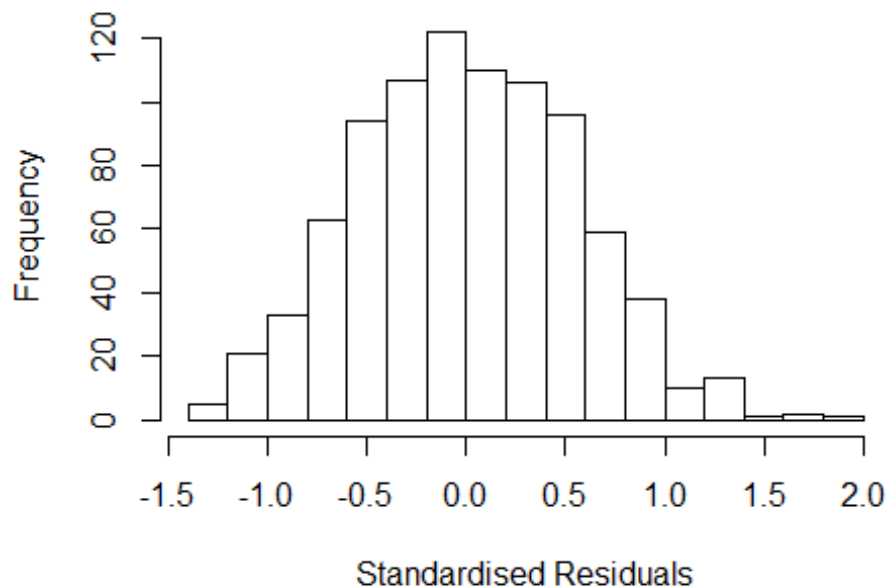
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 830, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	1.265	1	1.125
LastAuthorFemale	1.268	1	1.126
UniqueAuthors	1.524	4	1.054
Year	1.613	16	1.015

Residuals from first and last author and team size



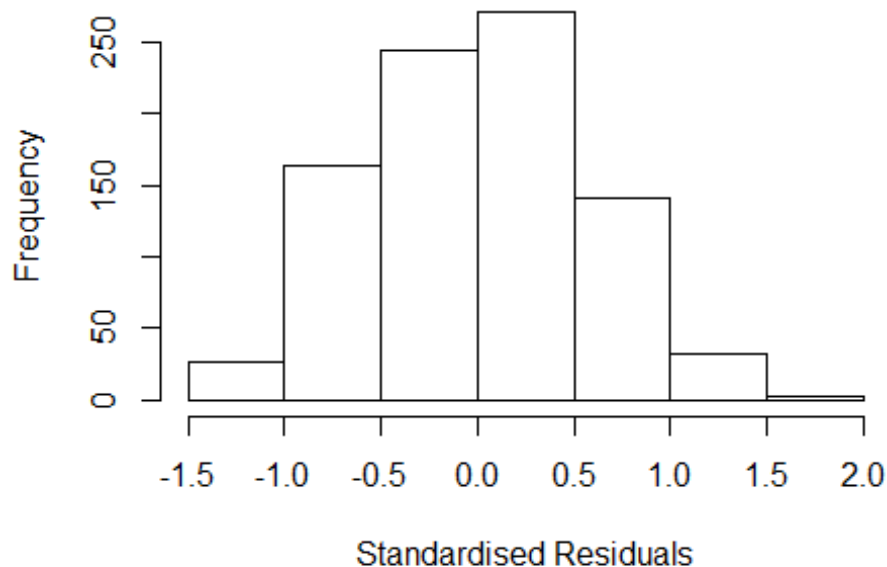
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2686 -0.3873 -0.0172 0.3976 1.8027
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.86738 0.14314 6.06 2.0e-09 ***
## FirstAuthorFemale1 0.11866 0.04607 2.58 0.01018 *
## LastAuthorFemale1 0.09906 0.04400 2.25 0.02460 *
## UniqueAuthors2 0.15754 0.05959 2.64 0.00835 **
## UniqueAuthors3 0.40771 0.06268 6.50 1.3e-10 ***
## UniqueAuthors4 0.39034 0.06557 5.95 3.8e-09 ***
## UniqueAuthors5 0.40445 0.06399 6.32 4.2e-10 ***
## Year1997 0.00933 0.15833 0.06 0.95304
## Year1998 -0.06677 0.17149 -0.39 0.69712
## Year1999 -0.07825 0.16812 -0.47 0.64173
```

```

## Year2000      -0.14500    0.16408   -0.88   0.37710
## Year2001      -0.34150    0.16380   -2.08   0.03738 *
## Year2002      -0.18348    0.14832   -1.24   0.21639
## Year2003      -0.45481    0.16593   -2.74   0.00625 **
## Year2004      -0.54596    0.15661   -3.49   0.00052 ***
## Year2005      -0.45525    0.15027   -3.03   0.00252 **
## Year2006      -0.55264    0.16148   -3.42   0.00065 ***
## Year2007      -0.39042    0.15314   -2.55   0.01096 *
## Year2008      -0.30581    0.15505   -1.97   0.04889 *
## Year2009      -0.33360    0.15420   -2.16   0.03078 *
## Year2010      -0.22422    0.14862   -1.51   0.13175
## Year2011      -0.26717    0.15016   -1.78   0.07556 .
## Year2012      -0.24840    0.14725   -1.69   0.09198 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.545
## Multiple R-squared:  0.173, Adjusted R-squared:  0.152
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 816 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.252  0.872  0.948  0.907  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.14e-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 1.184 1          1.088
## LastAuthorFemale  1.191 1          1.091
## Year              1.119 16          1.004

```

Residuals from first and last author



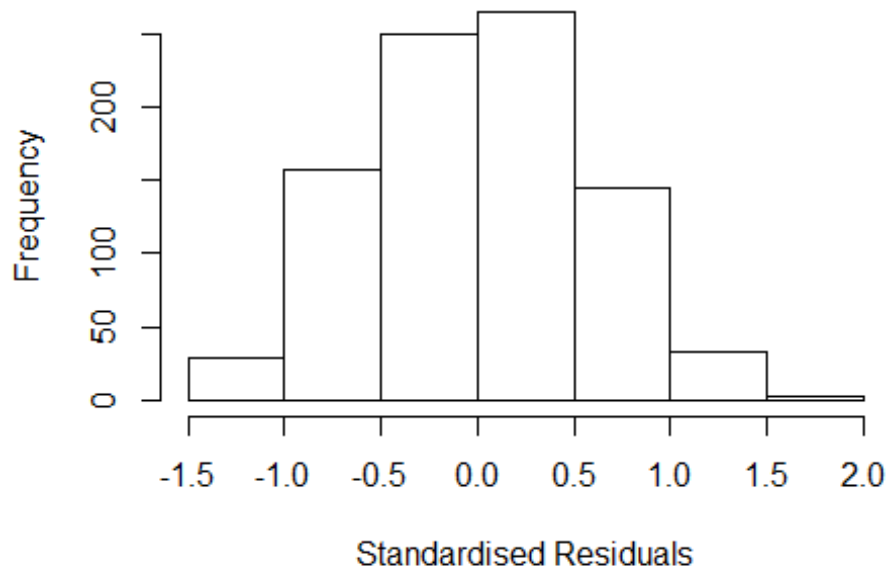
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2330 -0.4436 0.0109 0.4113 1.9193
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0688 0.1240 8.62 < 2e-16 ***
## FirstAuthorFemale1 0.1668 0.0476 3.50 0.00049 ***
## LastAuthorFemale1 0.0632 0.0452 1.40 0.16192
## Year1997 -0.0420 0.1502 -0.28 0.77986
## Year1998 -0.0658 0.1625 -0.41 0.68544
## Year1999 -0.0145 0.1577 -0.09 0.92654
## Year2000 -0.0925 0.1503 -0.62 0.53820
## Year2001 -0.2551 0.1531 -1.67 0.09613 .
## Year2002 -0.1891 0.1367 -1.38 0.16696
## Year2003 -0.4829 0.1625 -2.97 0.00305 **
## Year2004 -0.5941 0.1558 -3.81 0.00015 ***
## Year2005 -0.4112 0.1412 -2.91 0.00368 **
```

```

## Year2006          -0.5001      0.1569   -3.19  0.00149 **
## Year2007          -0.3169      0.1414   -2.24  0.02526 *
## Year2008          -0.1843      0.1392   -1.32  0.18589
## Year2009          -0.2597      0.1419   -1.83  0.06759 .
## Year2010          -0.1565      0.1369   -1.14  0.25305
## Year2011          -0.2014      0.1351   -1.49  0.13628
## Year2012          -0.1469      0.1344   -1.09  0.27464
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.107, Adjusted R-squared:  0.088
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 814 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.224  0.868  0.938  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      1.14e-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 1.062 1          1.031
## Year              1.062 16          1.002

```

Residuals from first author



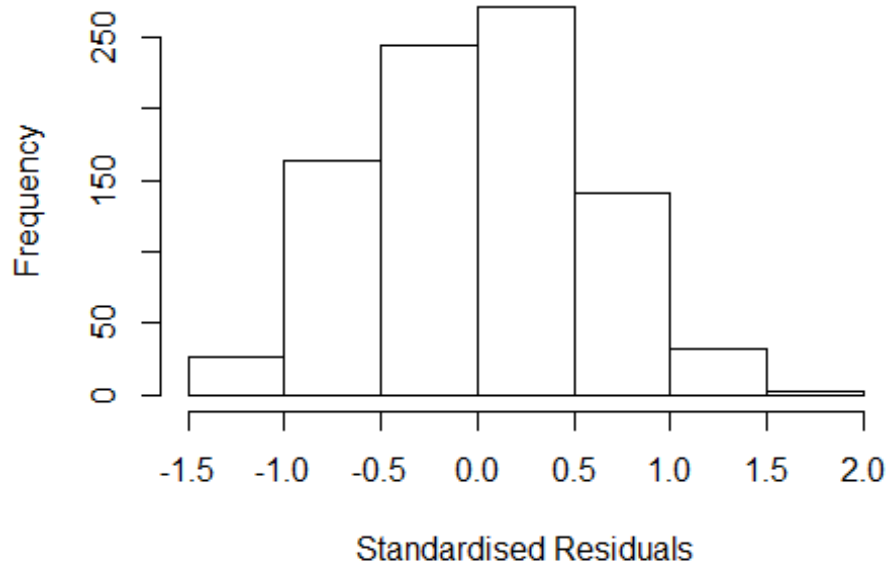
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2650 -0.4373 0.0139 0.4189 1.8984
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0882 0.1209 9.00 < 2e-16 ***
## FirstAuthorFemale1 0.1895 0.0457 4.15 3.7e-05 ***
## Year1997 -0.0415 0.1488 -0.28 0.78028
## Year1998 -0.0703 0.1608 -0.44 0.66181
## Year1999 -0.0127 0.1562 -0.08 0.93506
## Year2000 -0.0932 0.1480 -0.63 0.52909
## Year2001 -0.2536 0.1526 -1.66 0.09684 .
## Year2002 -0.1849 0.1350 -1.37 0.17127
## Year2003 -0.4837 0.1618 -2.99 0.00288 **
## Year2004 -0.6001 0.1543 -3.89 0.00011 ***
## Year2005 -0.4115 0.1403 -2.93 0.00344 **
## Year2006 -0.5047 0.1556 -3.24 0.00122 **
```

```

## Year2007          -0.3144      0.1389   -2.26  0.02387 *
## Year2008          -0.1681      0.1370   -1.23  0.22003
## Year2009          -0.2611      0.1410   -1.85  0.06447 .
## Year2010          -0.1476      0.1345   -1.10  0.27288
## Year2011          -0.1988      0.1333   -1.49  0.13627
## Year2012          -0.1450      0.1323   -1.10  0.27322
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.105, Adjusted R-squared:  0.0874
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 819 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.232  0.872  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      1.14e-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 1.069 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
## -1.2550 -0.4208 0.0107 0.4177 1.8622
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1364 0.1280 8.88 < 2e-16 ***
## LastAuthorFemale1 0.1185 0.0436 2.72 0.00669 **
## Year1997 -0.0353 0.1555 -0.23 0.82026
## Year1998 -0.0473 0.1686 -0.28 0.77895
## Year1999 0.0213 0.1581 0.13 0.89295
## Year2000 -0.1005 0.1552 -0.65 0.51731
## Year2001 -0.2656 0.1599 -1.66 0.09702 .
## Year2002 -0.1933 0.1416 -1.36 0.17264
## Year2003 -0.5074 0.1682 -3.02 0.00264 **
## Year2004 -0.5990 0.1628 -3.68 0.00025 ***
## Year2005 -0.4054 0.1482 -2.74 0.00636 **
## Year2006 -0.4961 0.1657 -2.99 0.00283 **
```

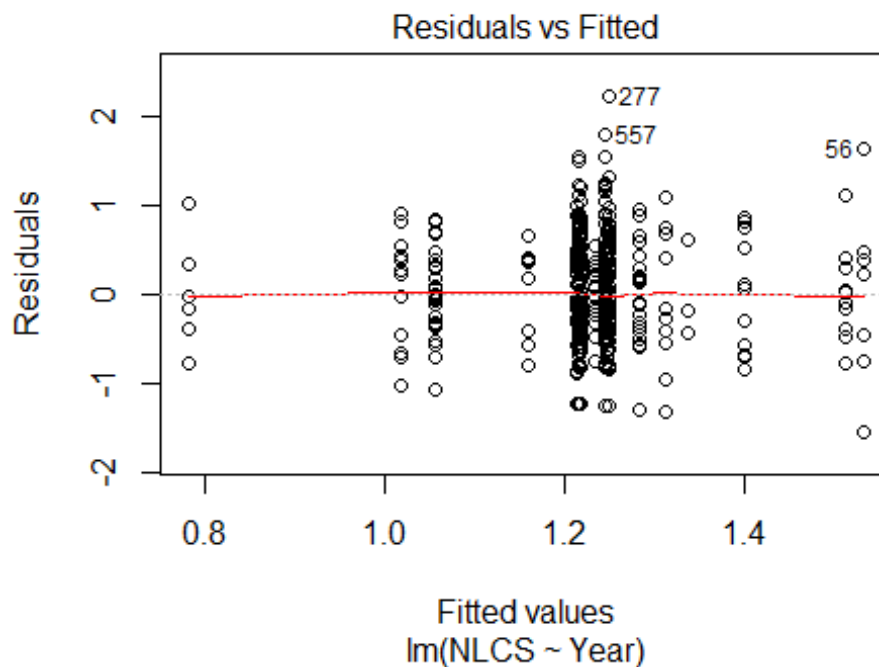


```

## Year2007          -0.3097      0.1491   -2.08   0.03808 *
## Year2008          -0.1757      0.1454   -1.21   0.22732
## Year2009          -0.2538      0.1480   -1.71   0.08671 .
## Year2010          -0.1312      0.1430   -0.92   0.35935
## Year2011          -0.1752      0.1414   -1.24   0.21554
## Year2012          -0.1217      0.1401   -0.87   0.38537
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.564
## Multiple R-squared:  0.0921, Adjusted R-squared:  0.0743
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 812 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.253  0.861  0.937  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      1.14e-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: 881"
## [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
##    7    4   12   13    9   19   16   13   23   28   17   22   33   82   74
## 2011 2012
##   79   88
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    6    3   11   10    7   16   11   10   20   26   15   16   26   70   67
## 2011 2012

```

```
## 70 76
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 3 10 10 7 16 11 10 16 24 13 16 24 67 61
## 2011 2012
## 67 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 = 0.1
```



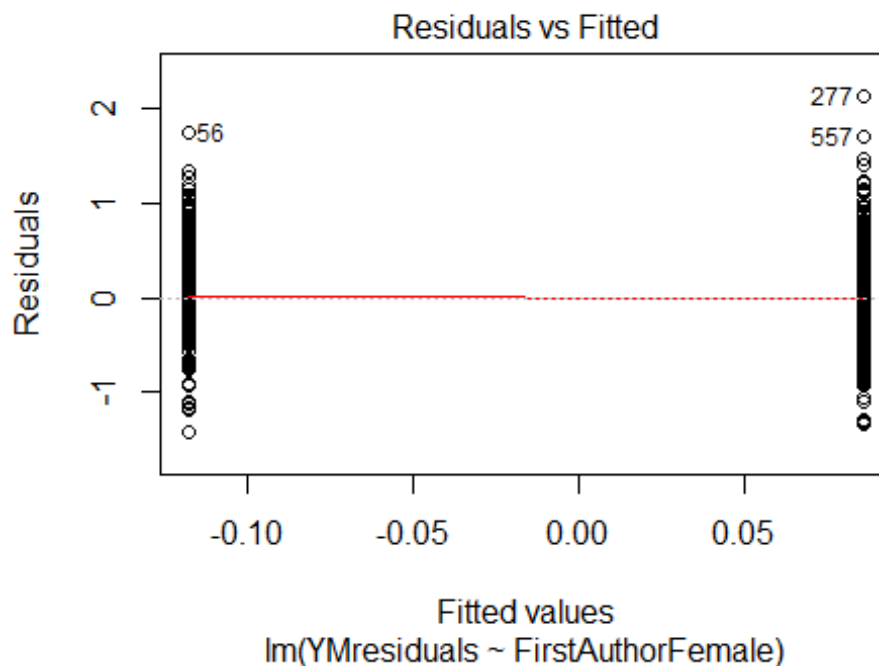
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.041, df = 1, p-value = 0.8

## [1] "Female first author team size 2018 geometric mean: 1.6416639893404"
## [1] "Male first author team size 2018 geometric mean: 2.39709706677529"

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

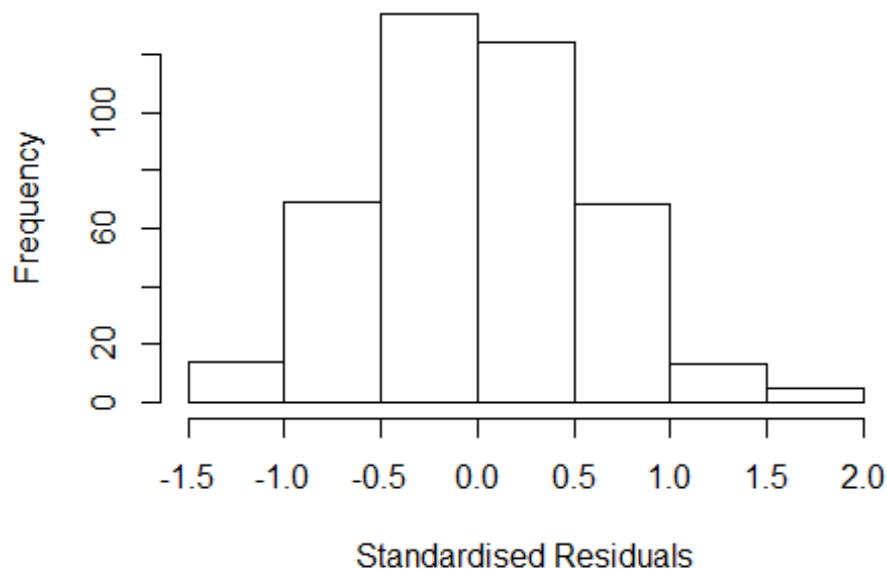
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 210, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.96886784678839"
## [1] "Male last author team size 2018 geometric mean: 1.71766791043423"

## Warning in wilcox.test.default(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 = 0.6
## 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 1.933  1      1.390
## LastAuthorFemale  1.792  1      1.339
## UniqueAuthors    1.876  4      1.082
## Year              2.048 16      1.023
```

Residuals from first and last author and team size



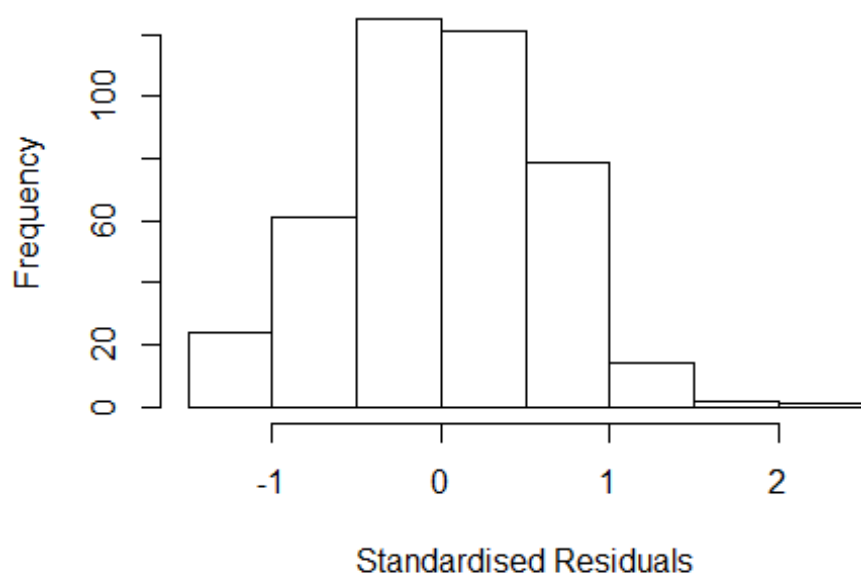
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4301 -0.3962 -0.0116  0.3839  1.9278
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4030     0.2871    1.40  0.16126
## FirstAuthorFemale1 0.1078     0.0790    1.36  0.17322
## LastAuthorFemale1 0.0531     0.0761    0.70  0.48544
## UniqueAuthors2    0.2637     0.0738    3.57  0.00039 ***
## UniqueAuthors3    0.2695     0.0953    2.83  0.00491 **
## UniqueAuthors4    0.4451     0.1192    3.73  0.00022 ***
## UniqueAuthors5    0.4631     0.1040    4.45  1.1e-05 ***
## Year1997          0.7589     0.3917    1.94  0.05339 .
## Year1998          0.8166     0.3583    2.28  0.02317 *
## Year1999          0.8668     0.3227    2.69  0.00753 **
```

```

## Year2000          0.8232      0.4254      1.94  0.05364 .
## Year2001          0.4776      0.3258      1.47  0.14343
## Year2002          0.7461      0.3579      2.08  0.03772 *
## Year2003          0.6234      0.3165      1.97  0.04956 *
## Year2004          0.7472      0.3113      2.40  0.01683 *
## Year2005          0.3872      0.2921      1.33  0.18568
## Year2006          0.5885      0.2896      2.03  0.04283 *
## Year2007          0.6011      0.3093      1.94  0.05267 .
## Year2008          0.6298      0.2891      2.18  0.02992 *
## Year2009          0.5967      0.2818      2.12  0.03484 *
## Year2010          0.5541      0.2807      1.97  0.04907 *
## Year2011          0.4987      0.2869      1.74  0.08292 .
## Year2012          0.5330      0.2848      1.87  0.06202 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.589
## Multiple R-squared:  0.129, Adjusted R-squared:  0.082
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 35 weights are ~= 1. The remaining 392 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.262  0.880  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      2.34e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.908 1      1.381
## LastAuthorFemale  1.866 1      1.366
## Year              1.209 16      1.006

```

Residuals from first and last author



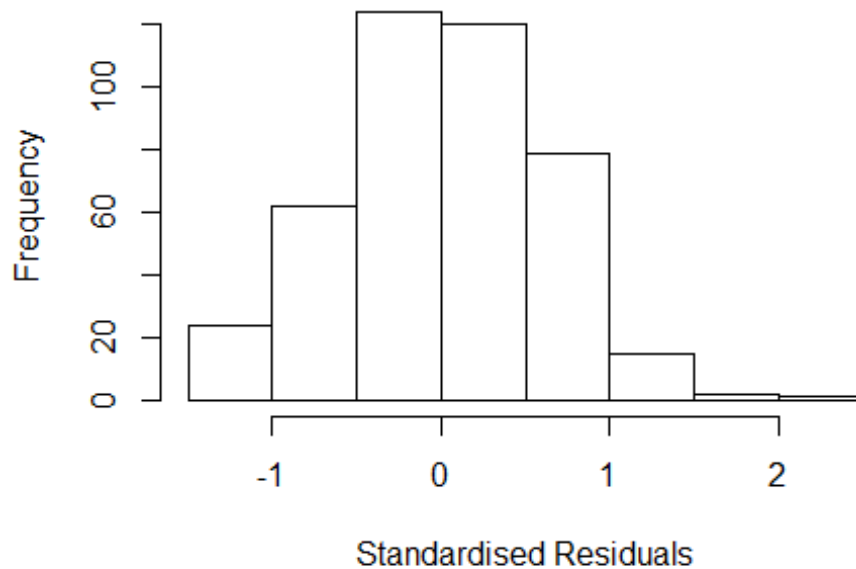
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3845 -0.4098 0.0179 0.4452 2.1230
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5741 0.2987 1.92 0.055 .
## FirstAuthorFemale1 0.1972 0.0831 2.37 0.018 *
## LastAuthorFemale1 0.0446 0.0816 0.55 0.585
## Year1997 0.5087 0.3979 1.28 0.202
## Year1998 0.7460 0.3706 2.01 0.045 *
## Year1999 0.7657 0.3423 2.24 0.026 *
## Year2000 0.7242 0.4778 1.52 0.130
## Year2001 0.3665 0.3533 1.04 0.300
## Year2002 0.6479 0.3969 1.63 0.103
## Year2003 0.4259 0.3272 1.30 0.194
## Year2004 0.7019 0.3225 2.18 0.030 *
## Year2005 0.2816 0.3066 0.92 0.359
```

```

## Year2006          0.5753      0.3057      1.88      0.061 .
## Year2007          0.5088      0.3205      1.59      0.113
## Year2008          0.5685      0.3109      1.83      0.068 .
## Year2009          0.5240      0.3020      1.74      0.083 .
## Year2010          0.5079      0.3018      1.68      0.093 .
## Year2011          0.4748      0.3038      1.56      0.119
## Year2012          0.5202      0.3019      1.72      0.086 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.63
## Multiple R-squared:  0.0633, Adjusted R-squared:  0.022
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 396 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.232  0.891  0.952  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      2.34e-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 1.125 1      1.061
## Year              1.125 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
## -1.3745 -0.4157 0.0168 0.4422 2.1299
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5853 0.3004 1.95 0.05204 .
## FirstAuthorFemale1 0.2271 0.0638 3.56 0.00041 ***
## Year1997 0.5126 0.3986 1.29 0.19918
## Year1998 0.7383 0.3714 1.99 0.04747 *
## Year1999 0.7547 0.3427 2.20 0.02824 *
## Year2000 0.7289 0.4796 1.52 0.12929
## Year2001 0.3598 0.3544 1.02 0.31055
## Year2002 0.6386 0.3952 1.62 0.10687
## Year2003 0.4246 0.3290 1.29 0.19753
## Year2004 0.7006 0.3233 2.17 0.03083 *
## Year2005 0.2756 0.3082 0.89 0.37170
## Year2006 0.5701 0.3092 1.84 0.06588 .
```

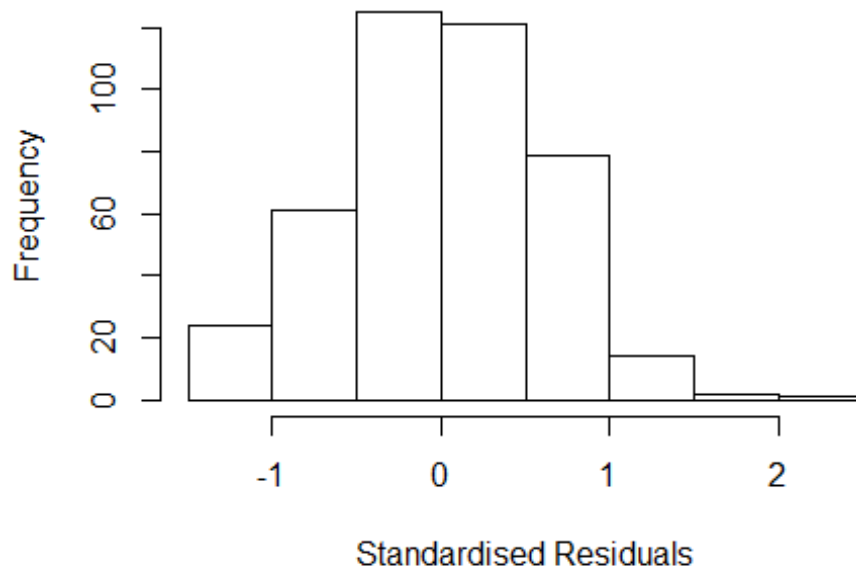


```

## Year2007          0.5049      0.3217      1.57  0.11738
## Year2008          0.5621      0.3131      1.80  0.07333 .
## Year2009          0.5207      0.3040      1.71  0.08753 .
## Year2010          0.5019      0.3036      1.65  0.09906 .
## Year2011          0.4678      0.3054      1.53  0.12634
## Year2012          0.5197      0.3036      1.71  0.08776 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.638
## Multiple R-squared:  0.0624, Adjusted R-squared:  0.0234
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 394 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.242  0.894   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      2.34e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.104 1          1.051
## Year              1.104 16          1.003

```

Residuals from last author



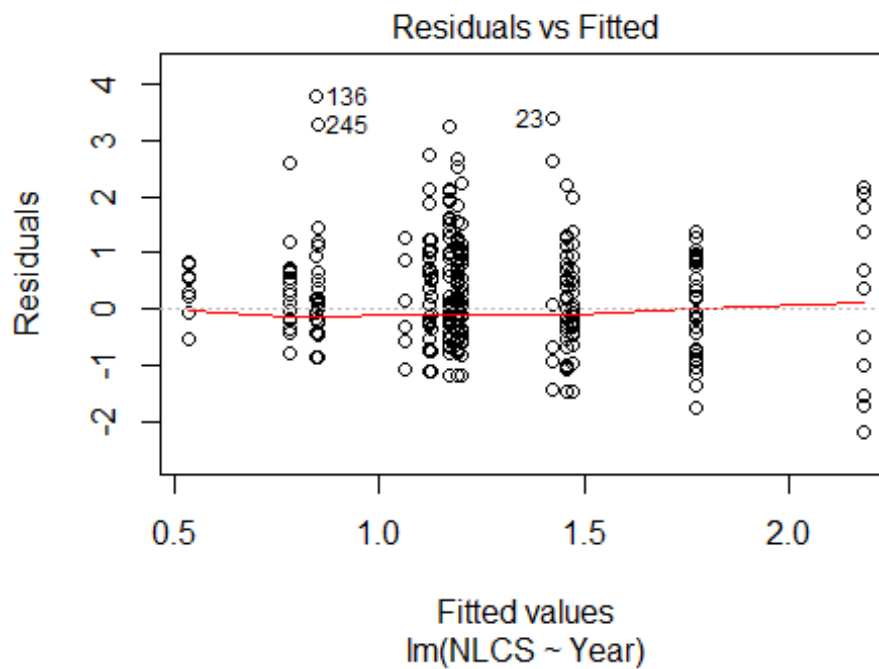
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3595 -0.4078 -0.0128 0.4180 2.1528
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6298 0.2718 2.32 0.0210 *
## LastAuthorFemale1 0.1731 0.0632 2.74 0.0065 **
## Year1997 0.5217 0.3807 1.37 0.1713
## Year1998 0.7602 0.3497 2.17 0.0303 *
## Year1999 0.7864 0.3153 2.49 0.0130 *
## Year2000 0.7297 0.4740 1.54 0.1245
## Year2001 0.3419 0.3332 1.03 0.3054
## Year2002 0.6389 0.3849 1.66 0.0977 .
## Year2003 0.4178 0.3062 1.36 0.1732
## Year2004 0.6658 0.3004 2.22 0.0272 *
## Year2005 0.2897 0.2842 1.02 0.3085
## Year2006 0.5698 0.2766 2.06 0.0400 *
```

```

## Year2007          0.5078      0.2995      1.70      0.0908 .
## Year2008          0.5508      0.2881      1.91      0.0566 .
## Year2009          0.5073      0.2778      1.83      0.0686 .
## Year2010          0.4980      0.2780      1.79      0.0740 .
## Year2011          0.4730      0.2804      1.69      0.0924 .
## Year2012          0.4898      0.2769      1.77      0.0777 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.628
## Multiple R-squared:  0.0505, Adjusted R-squared:  0.011
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 389 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.215  0.880   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      2.34e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 427"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
##   19   29   29   25   20   29   43   29   33   29   23   38   49   38   49
## 2011 2012
##   37   33
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    9   13    7   16   11   13   41   29   30   25   17   34   47   34   47
## 2011 2012

```

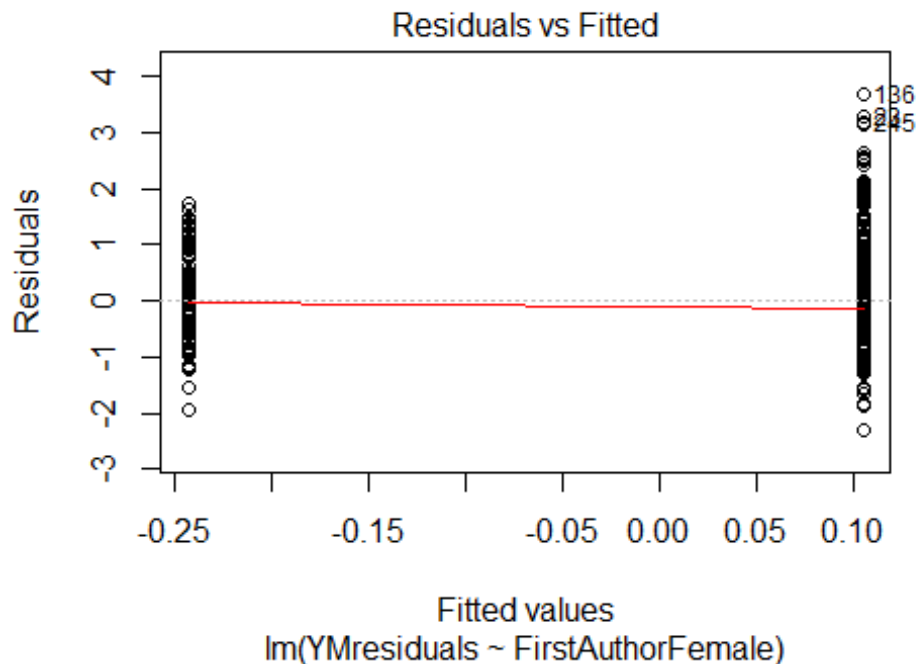
```
## 34 30
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 9 13 7 14 11 13 37 28 28 21 17 31 45 29 40
## 2011 2012
## 33 23
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 44, df = 16, p-value = 2e-04
```



```
##
## 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: 2.5622731277084"
## [1] "Male first author team size 2018 geometric mean: 2.85363852822755"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

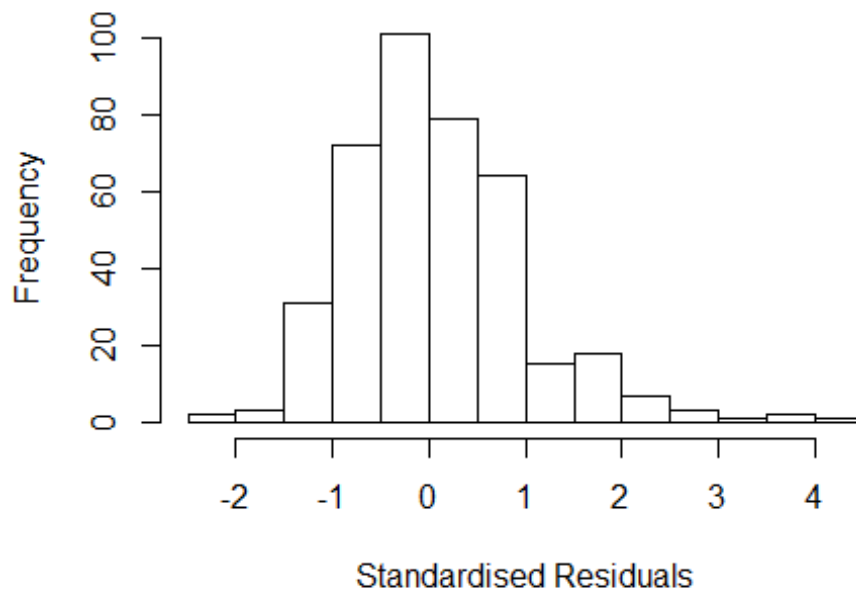
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 50, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.63544997154647"
## [1] "Male last author team size 2018 geometric mean: 2.60517108469735"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 49, 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 2.010  1      1.418
## LastAuthorFemale  1.750  1      1.323
## UniqueAuthors    2.381  4      1.115
## Year              3.553 16      1.040
```

Residuals from first and last author and team size



```
## [1] "List of 7 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 21    0030140457 4.055 1996    2911      1    3.107
## 23    0030527276 4.797 1996    2911      1    4.289
## 136   0033832472 4.625 2000    2911      1    3.602
## 273   4043062313 3.864 2004    2911      1    2.956
## 289  26944446190 3.371 2005    2911      1    2.555
## 335  33744500990 4.409 2006    2911      1    3.672
## 370  34250303065 3.877 2007    2911      1    2.946
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.129 -0.518 -0.032  0.589  4.289
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.2065     0.2551   0.81   0.4187
## FirstAuthorFemale1 0.1027     0.1139   0.90   0.3677
## LastAuthorFemale1 0.1988     0.1076   1.85   0.0655 .
## UniqueAuthors2    0.2024     0.1206   1.68   0.0940 .
```

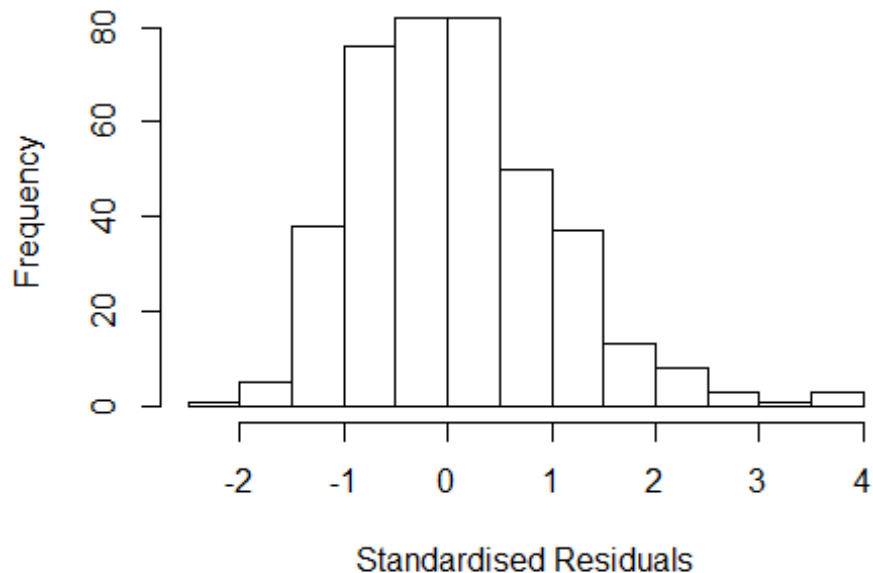
```

## UniqueAuthors3      0.4401      0.1337      3.29      0.0011 **
## UniqueAuthors4      0.7501      0.1529      4.91      1.4e-06 ***
## UniqueAuthors5      0.6002      0.1430      4.20      3.4e-05 ***
## Year1997             0.5501      0.3556      1.55      0.1227
## Year1998             0.2132      0.3579      0.60      0.5517
## Year1999             0.0108      0.2858      0.04      0.9700
## Year2000            -0.2355      0.2934     -0.80      0.4228
## Year2001             1.4184      0.9067      1.56      0.1186
## Year2002             0.4147      0.2836      1.46      0.1444
## Year2003             0.2048      0.2761      0.74      0.4586
## Year2004             0.4002      0.3334      1.20      0.2308
## Year2005             0.1057      0.2768      0.38      0.7027
## Year2006             0.2249      0.3218      0.70      0.4849
## Year2007             0.2206      0.2840      0.78      0.4378
## Year2008             0.3042      0.2992      1.02      0.3099
## Year2009             0.6064      0.2839      2.14      0.0333 *
## Year2010             0.8874      0.2904      3.06      0.0024 **
## Year2011             0.3016      0.2809      1.07      0.2836
## Year2012             0.5738      0.2914      1.97      0.0497 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.814
## Multiple R-squared:  0.25,   Adjusted R-squared:  0.206
## Convergence in 27 IRWLS iterations
##
## Robustness weights:
## observation 9 is an outlier with |weight| = 0 ( < 0.00025);
## 37 weights are ~= 1. The remaining 361 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0052 0.8780 0.9480 0.8930 0.9820 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.51e-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	1.869	1	1.367
## LastAuthorFemale	1.672	1	1.293
## Year	1.659	16	1.016

Residuals from first and last author



```
## [1] "List of 7 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 21    0030140457 4.055 1996    2911      1    3.203
## 23    0030527276 4.797 1996    2911      1    3.945
## 136   0033832472 4.625 2000    2911      1    3.991
## 272   2342667497 3.709 2004    2911      1    2.743
## 273   4043062313 3.864 2004    2911      1    2.686
## 335   33744500990 4.409 2006    2911      1    3.611
## 370   34250303065 3.877 2007    2911      1    2.751
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.286 -0.557 -0.059  0.661  3.991
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4523    0.2539   1.78  0.0756 .
## FirstAuthorFemale1  0.1877    0.1186   1.58  0.1143
```



```

## LastAuthorFemale1    0.2115    0.1137    1.86    0.0635 .
## Year1997             0.4797    0.3525    1.36    0.1744
## Year1998             0.3072    0.3931    0.78    0.4349
## Year1999            -0.0577    0.2896   -0.20    0.8421
## Year2000            -0.2174    0.3088   -0.70    0.4819
## Year2001             1.4346    0.8417    1.70    0.0891 .
## Year2002             0.4520    0.2990    1.51    0.1315
## Year2003             0.1044    0.2816    0.37    0.7111
## Year2004             0.3263    0.3208    1.02    0.3098
## Year2005             0.0862    0.2945    0.29    0.7699
## Year2006             0.1583    0.3499    0.45    0.6512
## Year2007             0.2748    0.2891    0.95    0.3423
## Year2008             0.3604    0.3075    1.17    0.2419
## Year2009             0.6655    0.3014    2.21    0.0279 *
## Year2010             0.9811    0.3052    3.21    0.0014 **
## Year2011             0.3912    0.2924    1.34    0.1817
## Year2012             0.7094    0.2976    2.38    0.0177 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.891
## Multiple R-squared:  0.172, Adjusted R-squared:  0.133
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 376 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0074 0.8780 0.9520 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      2.51e-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 1.385 1          1.177
## Year              1.385 16          1.010

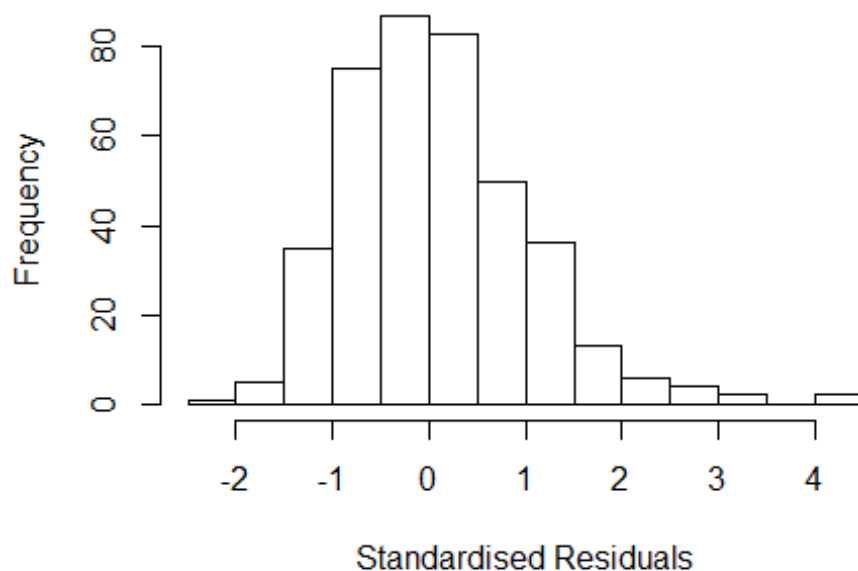
```

```
## [1] "List of 7 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 21    0030140457 4.055 1996    2911      1    3.203
## 23    0030527276 4.797 1996    2911      1    3.945
## 136   0033832472 4.625 2000    2911      1    3.991
## 272   2342667497 3.709 2004    2911      1    2.743
## 273   4043062313 3.864 2004    2911      1    2.686
## 335   33744500990 4.409 2006    2911      1    3.611
## 370   34250303065 3.877 2007    2911      1    2.751
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2914 -0.5870 -0.0402  0.6269  4.0789
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.4240     0.2283   1.86  0.0640 .
## FirstAuthorFemale1 0.2941     0.1028   2.86  0.0045 **
## Year1997          0.5498     0.3194   1.72  0.0861 .
## Year1998          0.4719     0.3765   1.25  0.2108
## Year1999          0.0157     0.2621   0.06  0.9523
## Year2000         -0.1311     0.2948  -0.44  0.6568
## Year2001          1.5733     0.8781   1.79  0.0740 .
## Year2002          0.5395     0.2699   2.00  0.0464 *
## Year2003          0.1654     0.2592   0.64  0.5238
## Year2004          0.4088     0.2921   1.40  0.1625
## Year2005          0.1497     0.2698   0.55  0.5793
## Year2006          0.2671     0.3325   0.80  0.4222
## Year2007          0.3578     0.2689   1.33  0.1842
## Year2008          0.4640     0.2800   1.66  0.0983 .
## Year2009          0.7578     0.2755   2.75  0.0062 **
## Year2010          1.0847     0.2763   3.93  0.0001 ***
## Year2011          0.4908     0.2631   1.87  0.0629 .
## Year2012          0.7755     0.2736   2.83  0.0048 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.898
## Multiple R-squared:  0.164, Adjusted R-squared:  0.127
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 384 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0036  0.8730  0.9610  0.9060  0.9860  0.9990
```

```
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.51e-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"

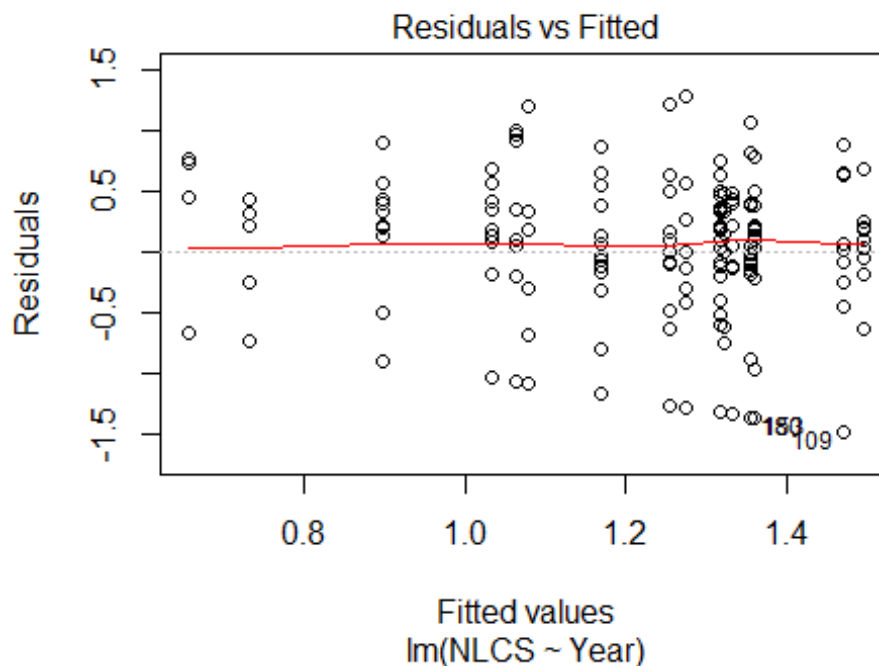
## Warning in lmrob.S(x, y, control = control, mf = mf): S refinements did
not
## converge (to refine.tol=1e-07) in 1000 (= k.max) steps
```

Residuals from first author



```
## [1] "Sample size for the above analysis: 399"
## [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
##    5   14   23   18   19    9   17   13    7    7   14   12   13   13   24
## 2011 2012
##   16   14
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    9   10    8    8    5    9    7    7    6   12   10   12   11   22
## 2011 2012
##   12   12
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    9   10    6    8    5    9    7    6    5   11   10   12   11   19
## 2011 2012
##   12   10
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 11, df = 16, p-value = 0.8
```



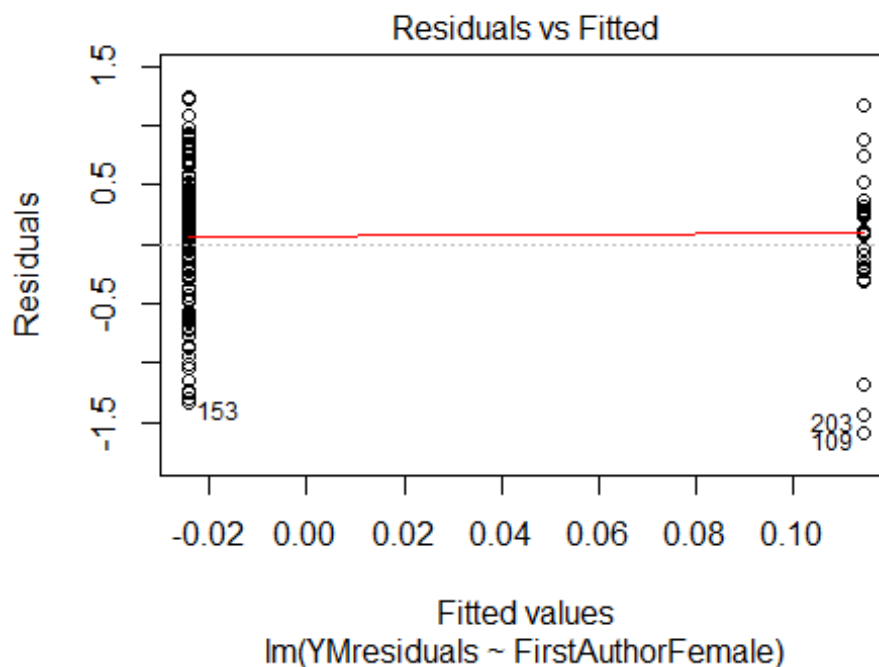
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.046, df = 1, p-value = 0.8

## [1] "Female first author team size 2018 geometric mean: 3.11337964201891"
## [1] "Male first author team size 2018 geometric mean: 1.69838132956495"

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

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 54, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.96119153546871"
## [1] "Male last author team size 2018 geometric mean: 2.10397901101729"

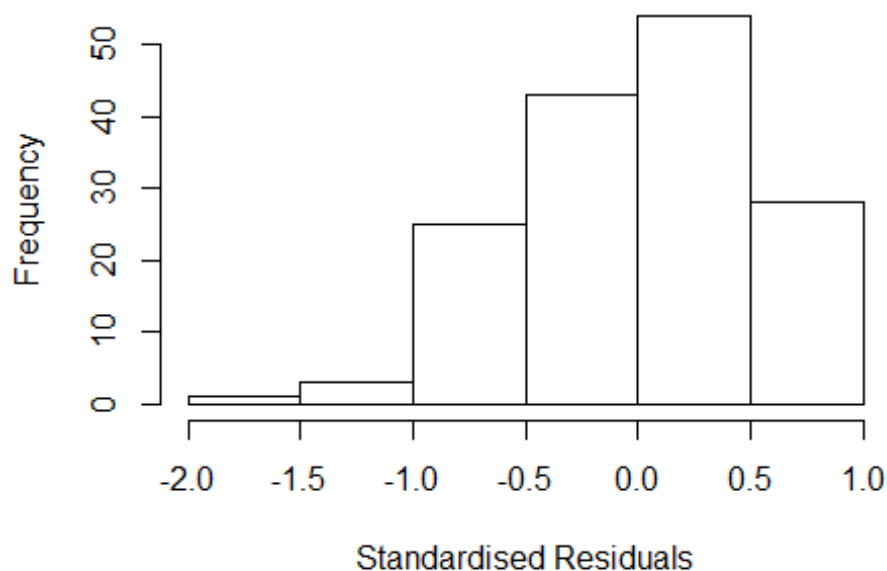
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



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

```
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 48, p-value = 0.5
## 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 2.121 1          1.457
## LastAuthorFemale 1.800 1          1.341
## UniqueAuthors    5.334 4          1.233
## Year             14.393 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
## -1.6264 -0.3405 0.0173 0.3783 0.9568
##
```

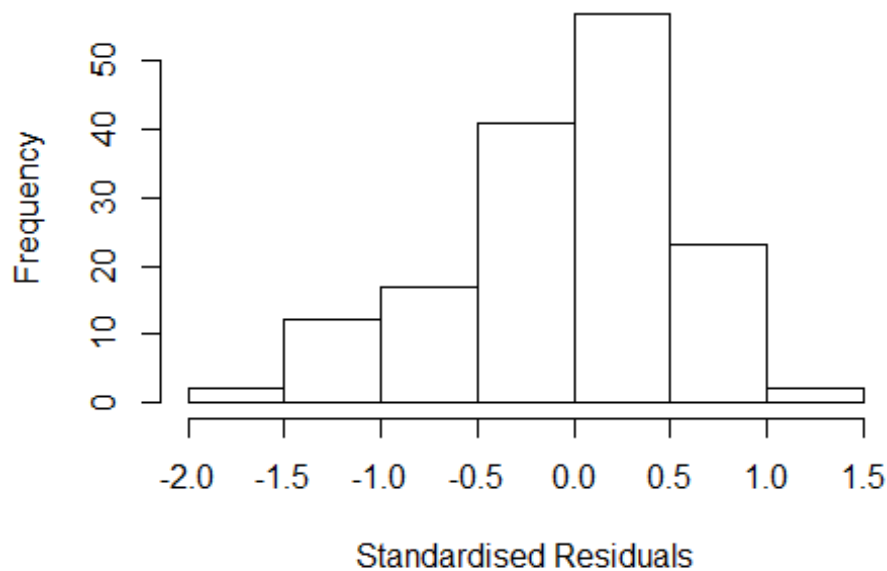
```

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4306    0.1781    2.42  0.01700 *
## FirstAuthorFemale1 -0.1194    0.1229   -0.97  0.33298
## LastAuthorFemale1  0.0555    0.1009    0.55  0.58361
## UniqueAuthors2    0.3903    0.1480    2.64  0.00938 **
## UniqueAuthors3    0.5856    0.1549    3.78  0.00024 ***
## UniqueAuthors4    0.7803    0.1648    4.74  5.6e-06 ***
## UniqueAuthors5    0.5246    0.2771    1.89  0.06052 .
## Year1997          0.5468    0.2884    1.90  0.06018 .
## Year1998          0.5442    0.2656    2.05  0.04242 *
## Year1999          0.5633    0.2793    2.02  0.04578 *
## Year2000          0.5795    0.2680    2.16  0.03241 *
## Year2001          0.2296    0.2816    0.82  0.41637
## Year2002          0.6108    0.2669    2.29  0.02372 *
## Year2003          0.5764    0.1945    2.96  0.00361 **
## Year2004          0.1032    0.1779    0.58  0.56260
## Year2005         -0.0268    0.2568   -0.10  0.91705
## Year2006          0.5834    0.1881    3.10  0.00236 **
## Year2007          0.6303    0.2011    3.13  0.00213 **
## Year2008          0.5035    0.2054    2.45  0.01554 *
## Year2009          0.3367    0.1991    1.69  0.09327 .
## Year2010          0.6102    0.2232    2.73  0.00712 **
## Year2011          0.2138    0.2053    1.04  0.29957
## Year2012          0.4796    0.2020    2.37  0.01903 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.317, Adjusted R-squared:  0.202
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 140 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.341  0.865  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           6.49e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats

```

```
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.860 1          1.364
## LastAuthorFemale  1.983 1          1.408
## Year              3.496 16          1.040
```

Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7809 -0.3477  0.0475  0.3504  1.3822
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7802     0.2213   3.53  0.00058 ***
## FirstAuthorFemale1 -0.1659     0.1280  -1.30  0.19721
## LastAuthorFemale1 -0.0495     0.1142  -0.43  0.66517
```

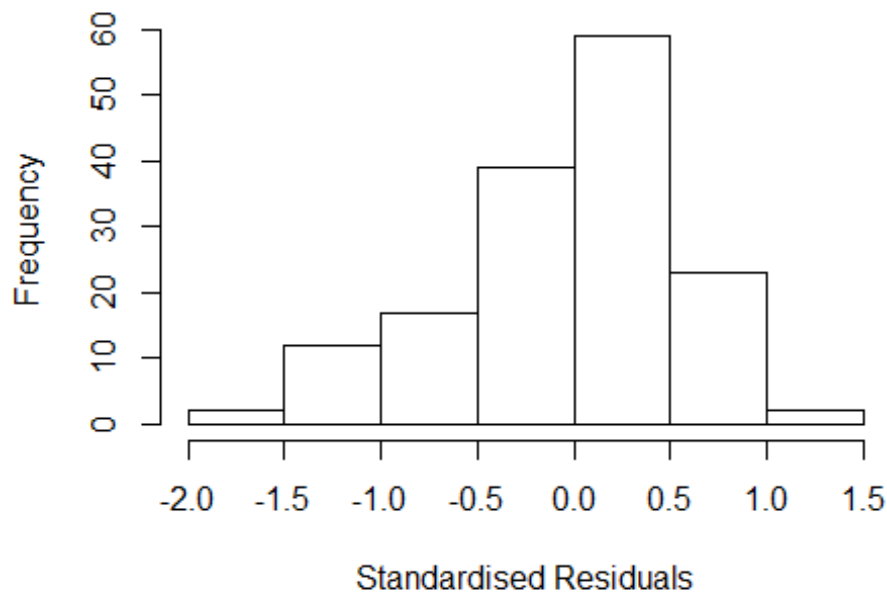


```

## Year1997      0.6026      0.2842      2.12  0.03580 *
## Year1998      0.5000      0.4388      1.14  0.25651
## Year1999      0.5430      0.3306      1.64  0.10285
## Year2000      0.8019      0.2561      3.13  0.00214 **
## Year2001      0.5129      0.2271      2.26  0.02554 *
## Year2002      1.0006      0.3117      3.21  0.00166 **
## Year2003      0.7214      0.2558      2.82  0.00552 **
## Year2004      0.3414      0.3308      1.03  0.30391
## Year2005     -0.1229      0.3989     -0.31  0.75858
## Year2006      0.7998      0.2534      3.16  0.00197 **
## Year2007      0.8818      0.2292      3.85  0.00018 ***
## Year2008      0.7441      0.2768      2.69  0.00809 **
## Year2009      0.5144      0.2651      1.94  0.05445 .
## Year2010      0.7705      0.2491      3.09  0.00240 **
## Year2011      0.3683      0.2918      1.26  0.20915
## Year2012      0.5240      0.2742      1.91  0.05813 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.539
## Multiple R-squared:  0.174, Adjusted R-squared:  0.0635
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 142 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.253  0.860  0.950  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      6.49e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.02 1          1.421
## Year              2.02 16          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
## -1.7590 -0.3369 0.0451 0.3582 1.3627
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.752 0.205 3.67 0.00035 ***
## FirstAuthorFemale1 -0.182 0.139 -1.31 0.19314
## Year1997 0.606 0.288 2.11 0.03684 *
## Year1998 0.499 0.438 1.14 0.25627
## Year1999 0.560 0.325 1.72 0.08706 .
## Year2000 0.800 0.252 3.17 0.00188 **
## Year2001 0.521 0.226 2.31 0.02258 *
## Year2002 1.007 0.309 3.26 0.00143 **
## Year2003 0.719 0.256 2.80 0.00578 **
## Year2004 0.336 0.329 1.02 0.30851
## Year2005 -0.120 0.411 -0.29 0.77066
## Year2006 0.802 0.254 3.16 0.00195 **
```

```

## Year2007          0.889      0.231      3.84  0.00019 ***
## Year2008          0.763      0.266      2.86  0.00485 **
## Year2009          0.519      0.269      1.93  0.05564 .
## Year2010          0.767      0.246      3.12  0.00223 **
## Year2011          0.369      0.293      1.26  0.21092
## Year2012          0.538      0.272      1.97  0.05045 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.173, Adjusted R-squared:  0.0695
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 142 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.258  0.862  0.951  0.892  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      6.49e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.973  1      1.405
## Year              1.973 16      1.021

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.6281 -0.3215  0.0333  0.3492  1.3394

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7269    0.2472   2.94  0.0039 **
## LastAuthorFemale1 -0.0875    0.1202  -0.73  0.4679
## Year1997          0.5524    0.3220   1.72  0.0885 .
## Year1998          0.4583    0.4564   1.00  0.3171
## Year1999          0.4521    0.3519   1.29  0.2010
## Year2000          0.7715    0.2866   2.69  0.0080 **
## Year2001          0.4929    0.2630   1.87  0.0630 .
## Year2002          0.9012    0.3286   2.74  0.0069 **
## Year2003          0.6929    0.2991   2.32  0.0220 *
## Year2004          0.2673    0.3499   0.76  0.4463
## Year2005         -0.2059    0.4131  -0.50  0.6190
## Year2006          0.7781    0.2883   2.70  0.0078 **
## Year2007          0.8320    0.2649   3.14  0.0021 **
## Year2008          0.7047    0.3073   2.29  0.0234 *
## Year2009          0.4333    0.2880   1.50  0.1347
## Year2010          0.7016    0.2724   2.58  0.0111 *
## Year2011          0.2904    0.3122   0.93  0.3540
## Year2012          0.4922    0.3114   1.58  0.1162
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.536
## Multiple R-squared:  0.164, Adjusted R-squared:  0.06
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 137 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.335  0.840  0.949  0.888  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      6.49e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0        1000         0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 154"
## [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
##    9    8    5    7    3    7    5    3    9    5   11   16   13   12    7
## 2011 2012
##   14   21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2    3    2    2    1    3    5    3    9    5   10   13    8   12    7
## 2011 2012
##   12   18
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2    3    2    2    1    3    5    3    9    4    7   13    8   12    7
## 2011 2012
##   11   17
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.91934962773421"
## [1] "Male first author team size 2018 geometric mean: 4"

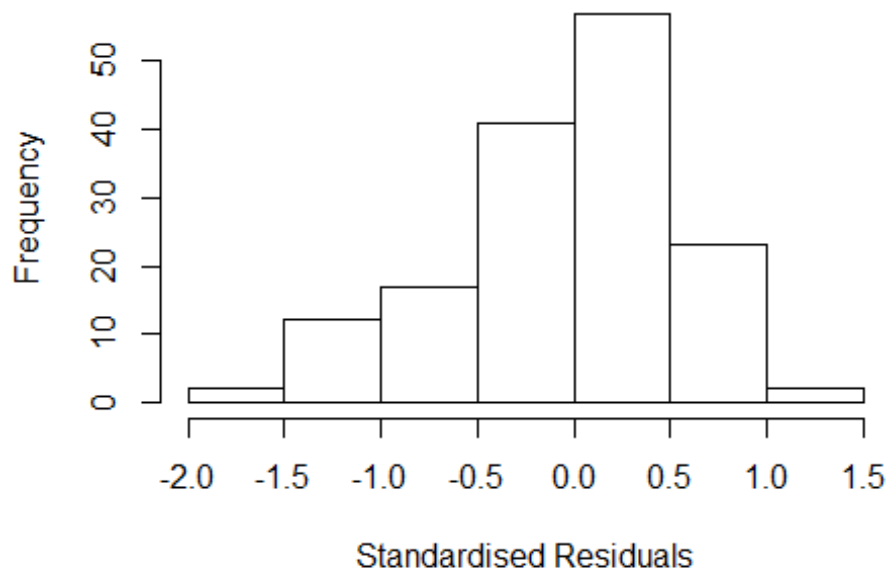
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.07278283365345"
## [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

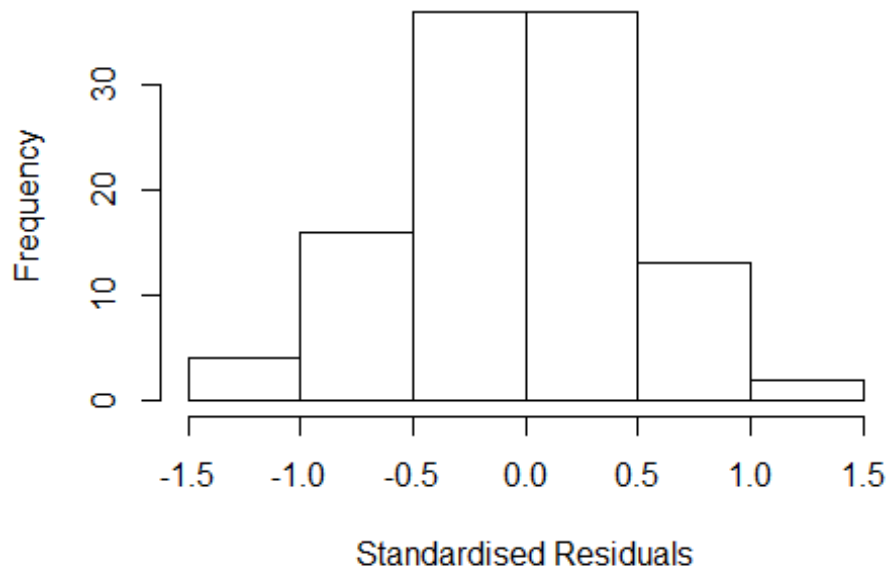
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 18, 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    2.388  1      1.545
## LastAuthorFemale     5.931  1      2.435
## UniqueAuthors    438.948  4      2.139
## Year           1478.968 16      1.256
```

Residuals from first and last author and team size



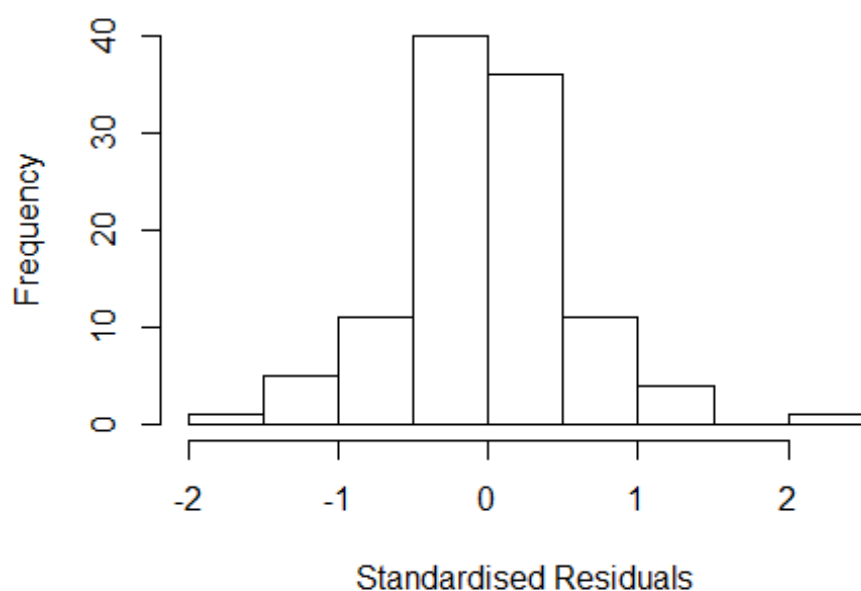
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.39138 -0.32394 -0.00376 0.34907 1.04731
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9340 0.5116 1.83 0.0714 .
## FirstAuthorFemale1 0.8199 0.4316 1.90 0.0609 .
## LastAuthorFemale1 -0.2039 0.2144 -0.95 0.3442
## UniqueAuthors2 0.0134 0.2193 0.06 0.9513
## UniqueAuthors3 0.1202 0.2456 0.49 0.6259
## UniqueAuthors4 0.3332 0.2792 1.19 0.2360
## UniqueAuthors5 0.5971 0.2518 2.37 0.0200 *
## Year1997 -1.2394 0.4295 -2.89 0.0049 **
## Year1998 0.0865 0.4734 0.18 0.8555
## Year1999 0.8174 0.1985 4.12 8.8e-05 ***
```

```

## Year2000          -1.5634      0.2080    -7.52  5.0e-11 ***
## Year2001          -0.1575      0.2219    -0.71  0.4797
## Year2002           0.6350      0.7901     0.80  0.4238
## Year2003          -0.2933      0.5881    -0.50  0.6193
## Year2004           0.3856      0.2158     1.79  0.0775 .
## Year2005           0.2720      0.1753     1.55  0.1245
## Year2006          -0.3022      0.2724    -1.11  0.2704
## Year2007          -0.3920      0.2978    -1.32  0.1916
## Year2008          -0.0563      0.2444    -0.23  0.8184
## Year2009          -0.4843      0.2032    -2.38  0.0194 *
## Year2010          -0.0773      0.4224    -0.18  0.8552
## Year2011          -0.2835      0.2070    -1.37  0.1745
## Year2012          -0.2937      0.2910    -1.01  0.3157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.514, Adjusted R-squared:  0.389
## Convergence in 34 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 97 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.335  0.838  0.943  0.883  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.17e-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 2.458 1          1.568
## LastAuthorFemale  2.548 1          1.596
## Year              3.991 16          1.044

```


Residuals from first and last author



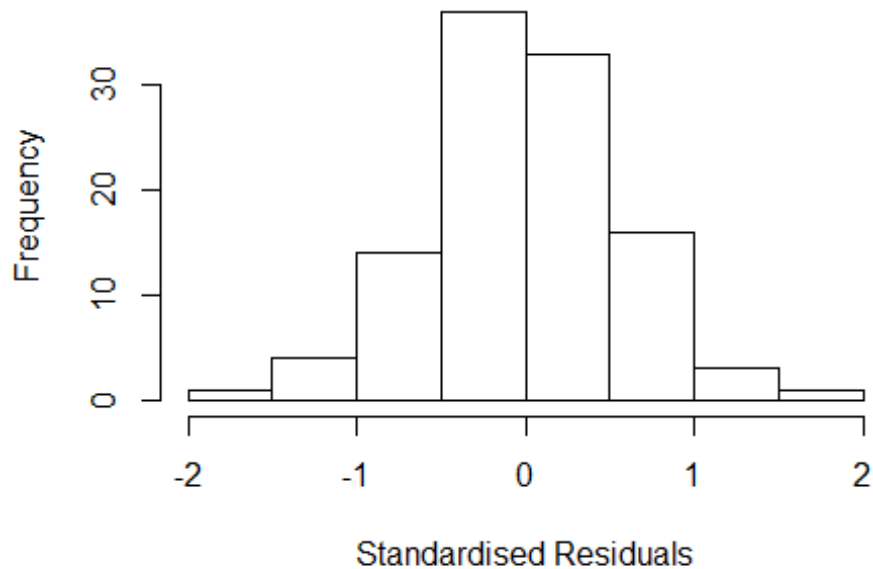
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7272 -0.3378 -0.0282 0.3733 2.2088
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8171 0.3489 2.34 0.0214 *
## FirstAuthorFemale1 1.0622 0.3641 2.92 0.0045 **
## LastAuthorFemale1 -0.2693 0.1963 -1.37 0.1736
## Year1997 -1.2714 0.3990 -3.19 0.0020 **
## Year1998 0.0865 0.3844 0.23 0.8225
## Year1999 0.7640 0.1776 4.30 4.3e-05 ***
## Year2000 -1.6100 0.1166 -13.80 < 2e-16 ***
## Year2001 -0.0365 0.2351 -0.16 0.8770
## Year2002 0.2263 0.6733 0.34 0.7375
## Year2003 0.0306 0.5059 0.06 0.9519
## Year2004 0.4847 0.1575 3.08 0.0028 **
## Year2005 0.3329 0.1834 1.81 0.0729 .
```

```

## Year2006          -0.2634      0.3177    -0.83    0.4092
## Year2007          -0.1827      0.2428    -0.75    0.4538
## Year2008           0.1037      0.2577     0.40    0.6884
## Year2009          -0.2122      0.1987    -1.07    0.2883
## Year2010           0.0428      0.3193     0.13    0.8936
## Year2011          -0.1492      0.1944    -0.77    0.4448
## Year2012           0.1172      0.1803     0.65    0.5173
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.447, Adjusted R-squared:  0.337
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 100 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0286 0.8720 0.9530 0.8920 0.9920 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.17e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.795 1      1.672
## Year              2.795 16      1.033

```

Residuals from first author



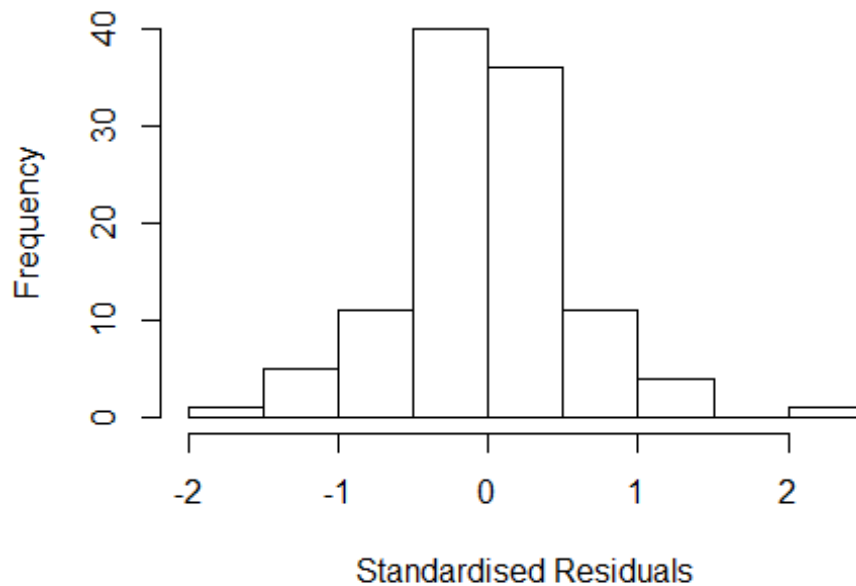
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7723 -0.3428 -0.0036 0.2894 1.8476
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7716 0.4256 1.81 0.0731 .
## FirstAuthorFemale1 0.8384 0.4093 2.05 0.0434 *
## Year1997 -1.2661 0.3943 -3.21 0.0018 **
## Year1998 0.0865 0.3800 0.23 0.8204
## Year1999 0.7640 0.1773 4.31 4.2e-05 ***
## Year2000 -1.6100 0.1165 -13.82 < 2e-16 ***
## Year2001 0.0552 0.1837 0.30 0.7645
## Year2002 0.3638 0.9097 0.40 0.6901
## Year2003 0.0414 0.5184 0.08 0.9366
## Year2004 0.5442 0.1670 3.26 0.0016 **
## Year2005 0.3322 0.1832 1.81 0.0730 .
## Year2006 -0.2646 0.2926 -0.90 0.3682
```

```

## Year2007          -0.1497      0.2781   -0.54   0.5916
## Year2008           0.1876      0.2269    0.83   0.4105
## Year2009          -0.2072      0.2000   -1.04   0.3028
## Year2010           0.2129      0.2723    0.78   0.4364
## Year2011          -0.1450      0.2031   -0.71   0.4772
## Year2012           0.1623      0.1793    0.90   0.3680
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.417, Adjusted R-squared:  0.309
## Convergence in 34 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 99 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.214  0.857  0.963  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      9.17e-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 2.231 1          1.494
## Year            2.231 16          1.025

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.80757 -0.34428 -0.00817 0.27317 1.41390
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5566 0.2554 6.09 2.6e-08 ***
## LastAuthorFemale1 0.0534 0.2273 0.24 0.8146
## Year1997 -1.2657 0.3941 -3.21 0.0018 **
## Year1998 0.0865 0.3797 0.23 0.8203
## Year1999 0.7640 0.1773 4.31 4.1e-05 ***
## Year2000 -1.6100 0.1165 -13.82 < 2e-16 ***
## Year2001 0.0737 0.1984 0.37 0.7112
## Year2002 0.4915 0.5170 0.95 0.3442
## Year2003 -0.1887 0.2775 -0.68 0.4982
## Year2004 0.5262 0.1907 2.76 0.0070 **
## Year2005 0.3322 0.1832 1.81 0.0730 .
## Year2006 -0.3230 0.3693 -0.87 0.3841
```

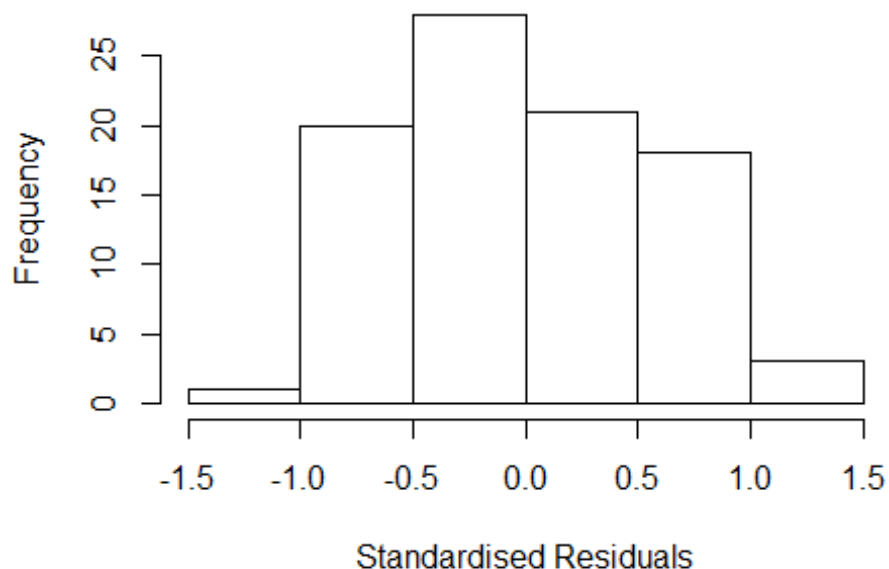
```

## Year2007          -0.3762      0.3162    -1.19    0.2373
## Year2008           0.2510      0.2374     1.06    0.2931
## Year2009          -0.2855      0.1895    -1.51    0.1354
## Year2010           0.2476      0.3121     0.79    0.4296
## Year2011          -0.2185      0.1750    -1.25    0.2152
## Year2012           0.1113      0.1905     0.58    0.5605
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.539
## Multiple R-squared:  0.342, Adjusted R-squared:  0.219
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 99 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.238  0.837  0.968  0.884  0.990  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.17e-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: 109"
## [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    3    7    6    6    8    3    6    5    8    5   11   16   15
## 2011 2012
##   14   10
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2    0    0    1    4    2    7    3    5    4    7    5   10   12   15
## 2011 2012

```

```
## 12 10
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2 0 0 1 4 2 6 3 4 3 7 5 9 10 15
## 2011 2012
## 10 10
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.29564966066123"
## [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 3.140 1 1.772
## LastAuthorFemale 4.716 1 2.172
## UniqueAuthors 44.636 4 1.608
## Year 37.524 14 1.138
```

Residuals from first and last author and team size



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

```

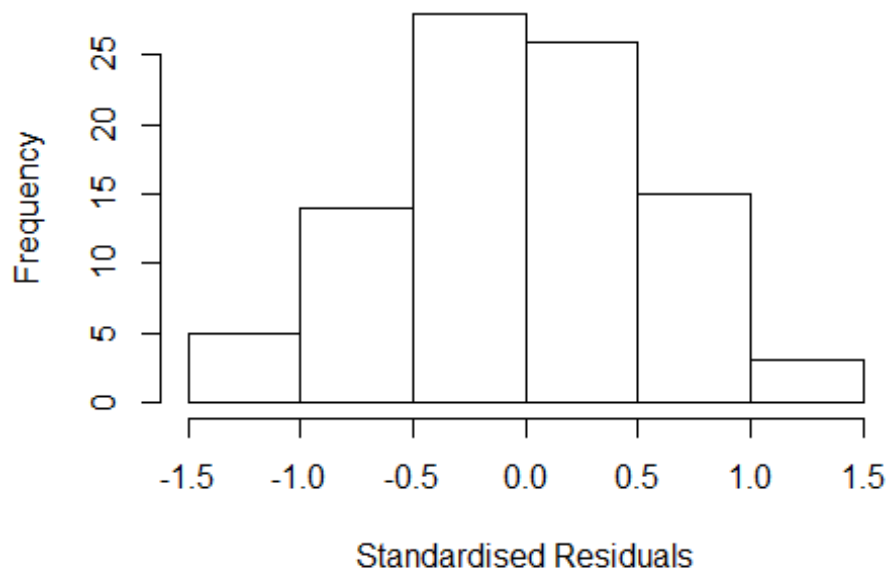
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.1855 -0.4509 -0.0548  0.4605  1.1003
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.2344      0.4694   0.50   0.6191
## FirstAuthorFemale1 0.3988      0.3229   1.24   0.2210
## LastAuthorFemale1  0.1876      0.2250   0.83   0.4073
## UniqueAuthors2     0.2644      0.2287   1.16   0.2515
## UniqueAuthors3     0.5834      0.2196   2.66   0.0098 **
## UniqueAuthors4     0.6047      0.2463   2.45   0.0166 *
## UniqueAuthors5     0.4209      0.4046   1.04   0.3018
## Year1999           0.0702      0.3207   0.22   0.8274
## Year2000           0.2846      0.3626   0.78   0.4351
## Year2001          -0.3418      0.4925  -0.69   0.4900
## Year2002           0.1187      0.4178   0.28   0.7772
## Year2003          -0.2456      0.4311  -0.57   0.5707
## Year2004          -0.1070      0.5254  -0.20   0.8392
## Year2005           0.3321      0.5761   0.58   0.5661
## Year2006          -0.5758      0.3185  -1.81   0.0750 .
## Year2007          -0.0555      0.4833  -0.11   0.9089
## Year2008           0.4228      0.3732   1.13   0.2611
## Year2009           0.2967      0.3397   0.87   0.3854
## Year2010           0.0346      0.3825   0.09   0.9283
## Year2011           0.0356      0.3706   0.10   0.9238
## Year2012          -0.0479      0.4096  -0.12   0.9072
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.677
## Multiple R-squared:  0.273, Adjusted R-squared:  0.065
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 83 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      0.740  0.886  0.949  0.934  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      1.10e-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 2.295 1           1.515
## LastAuthorFemale 2.481 1           1.575
## Year              3.190 14          1.042
```

Residuals from first and last author



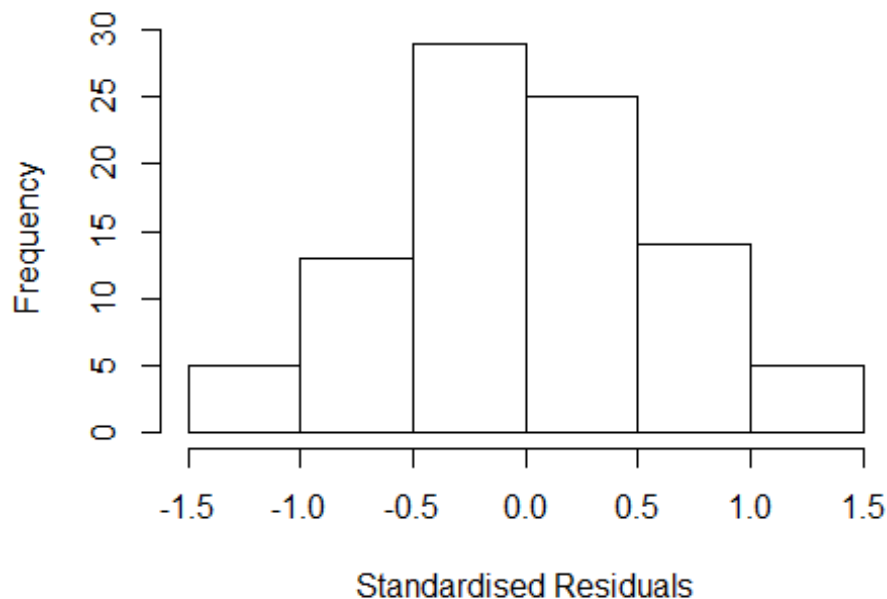
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2115 -0.3688 -0.0125 0.4111 1.3332
##
## Coefficients:
```

```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.5182    0.3296   1.57   0.120
## FirstAuthorFemale1 0.5130    0.3045   1.69   0.096 .
## LastAuthorFemale1 -0.0782    0.2021  -0.39   0.700
## Year1999          -0.0620    0.1924  -0.32   0.748
## Year2000           0.1678    0.3222   0.52   0.604
## Year2001          -0.4740    0.4174  -1.14   0.260
## Year2002           0.0800    0.3902   0.21   0.838
## Year2003          -0.0366    0.4659  -0.08   0.938
## Year2004           0.1598    0.4598   0.35   0.729
## Year2005           0.5871    0.2768   2.12   0.037 *
## Year2006          -0.4506    0.2427  -1.86   0.067 .
## Year2007          -0.0826    0.4337  -0.19   0.849
## Year2008           0.4416    0.2785   1.59   0.117
## Year2009           0.4526    0.2220   2.04   0.045 *
## Year2010           0.1803    0.2940   0.61   0.542
## Year2011           0.0587    0.2889   0.20   0.840
## Year2012          -0.0372    0.3352  -0.11   0.912
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.706
## Multiple R-squared:  0.177, Adjusted R-squared:  -0.0011
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 81 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.701  0.890  0.961  0.928  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      1.10e-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 1.675 1      1.294
## Year              1.675 14      1.019

```

Residuals from first author



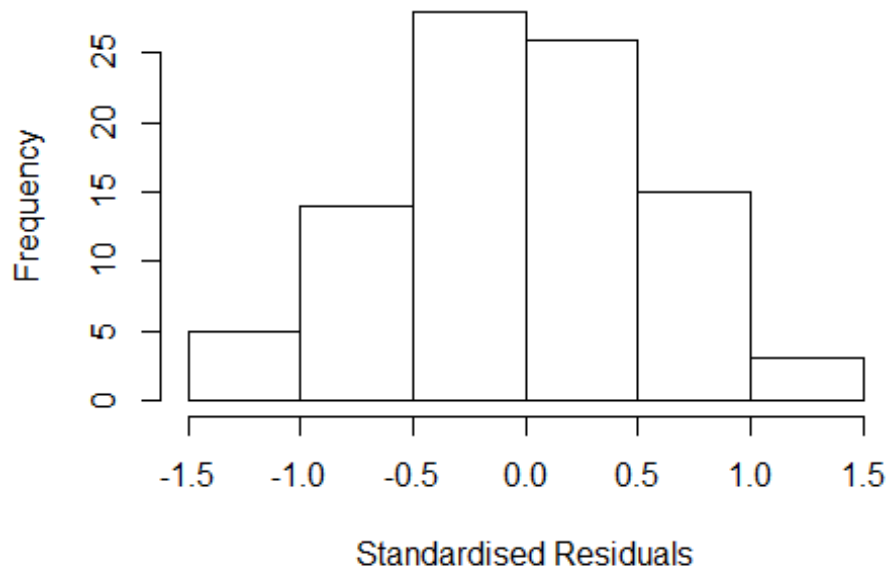
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.14464 -0.38374 -0.00692 0.39811 1.31020
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.48028 0.32661 1.47 0.146
## FirstAuthorFemale1 0.47272 0.26371 1.79 0.077 .
## Year1999 -0.06200 0.19270 -0.32 0.749
## Year2000 0.19754 0.30154 0.66 0.514
## Year2001 -0.47400 0.41970 -1.13 0.262
## Year2002 0.09322 0.40304 0.23 0.818
## Year2003 -0.00239 0.46531 -0.01 0.996
## Year2004 0.16148 0.46323 0.35 0.728
## Year2005 0.57399 0.27914 2.06 0.043 *
## Year2006 -0.42726 0.23639 -1.81 0.075 .
## Year2007 -0.05251 0.44787 -0.12 0.907
## Year2008 0.44980 0.27606 1.63 0.107
```

```

## Year2009          0.46356    0.21965    2.11    0.038 *
## Year2010          0.19164    0.29583    0.65    0.519
## Year2011          0.07392    0.28548    0.26    0.796
## Year2012         -0.02444    0.33933   -0.07    0.943
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.683
## Multiple R-squared:  0.176, Adjusted R-squared:  0.0117
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 81 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.693  0.883  0.960  0.924  0.986  0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.852  1          1.361
## Year            1.852 14          1.022

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1118 -0.4367 0.0282 0.4221 1.3514
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8822 0.2679 3.29 0.0015 **
## LastAuthorFemale1 0.0708 0.1867 0.38 0.7055
## Year1999 -0.0620 0.1921 -0.32 0.7478
## Year2000 0.1186 0.3512 0.34 0.7366
## Year2001 -0.4740 0.4161 -1.14 0.2582
## Year2002 0.1024 0.4028 0.25 0.8000
## Year2003 0.0198 0.4827 0.04 0.9673
## Year2004 0.1588 0.4577 0.35 0.7297
## Year2005 0.4279 0.3766 1.14 0.2594
## Year2006 -0.4071 0.2457 -1.66 0.1017
## Year2007 -0.1197 0.4754 -0.25 0.8019
## Year2008 0.4628 0.2794 1.66 0.1018
```

```

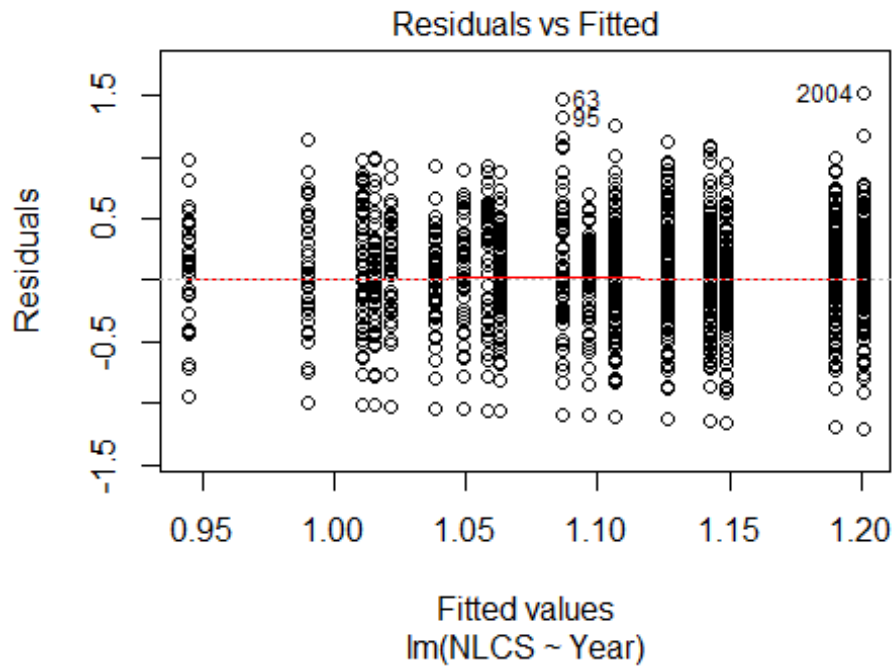
## Year2009          0.4303      0.2253      1.91      0.0600 .
## Year2010          0.1140      0.2944      0.39      0.6996
## Year2011          0.0026      0.2795      0.01      0.9926
## Year2012         -0.0397      0.3183     -0.12      0.9011
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.72
## Multiple R-squared:  0.137, Adjusted R-squared:  -0.0358
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 83 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.705  0.882  0.960  0.930  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] "Sample size for the above analysis: 91"
## [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 2012
##    1    1
##
## 2011 2012
##    1    1
##
## 2011 2012
##    1    1
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"

```

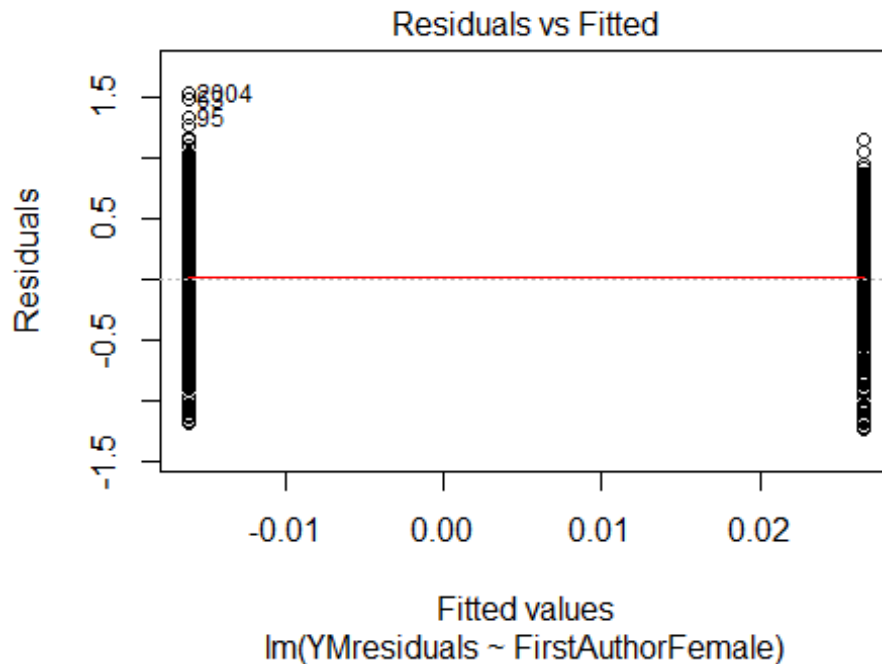
```

## [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: 2"
## [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
##    92    80    77    85    92    94    89    94    88    96   129   157   201   232   235
## 2011 2012
##   242   249
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    64    51    42    54    48    32    55    62    66    64    91   123   158   166   175
## 2011 2012
##   178   181
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    51    39    37    46    42    25    43    52    56    57    77   106   135   141   153
## 2011 2012
##   158   157
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 66, df = 16, p-value = 5e-08

```

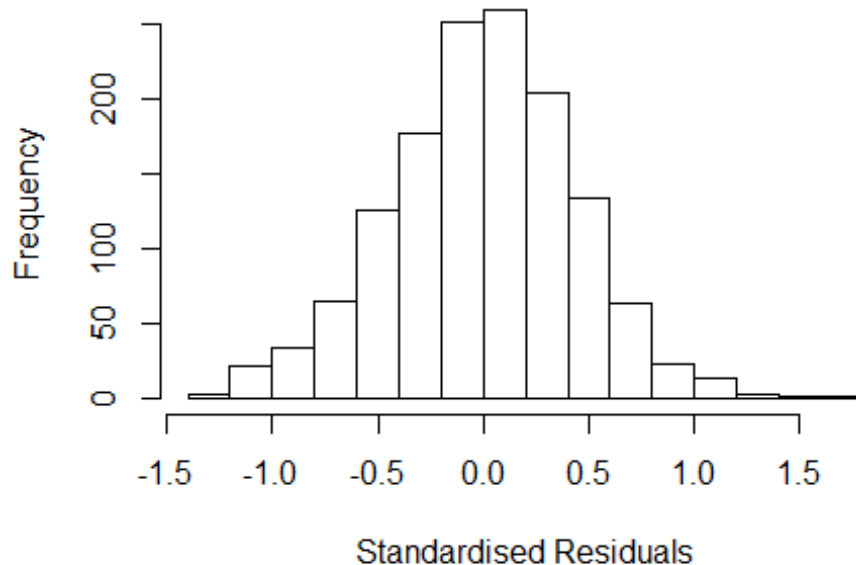


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.17, df = 1, p-value = 0.7
```

```
## [1] "Female first author team size 2018 geometric mean: 4.74829750445481"
## [1] "Male first author team size 2018 geometric mean: 4.39598671462739"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3800, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.24308540973806"
## [1] "Male last author team size 2018 geometric mean: 5.02855325063673"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3500, 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 1.164 1 1.079
## LastAuthorFemale 1.195 1 1.093
## UniqueAuthors 1.423 4 1.045
## Year 1.472 16 1.012
```

Residuals from first and last author and team size



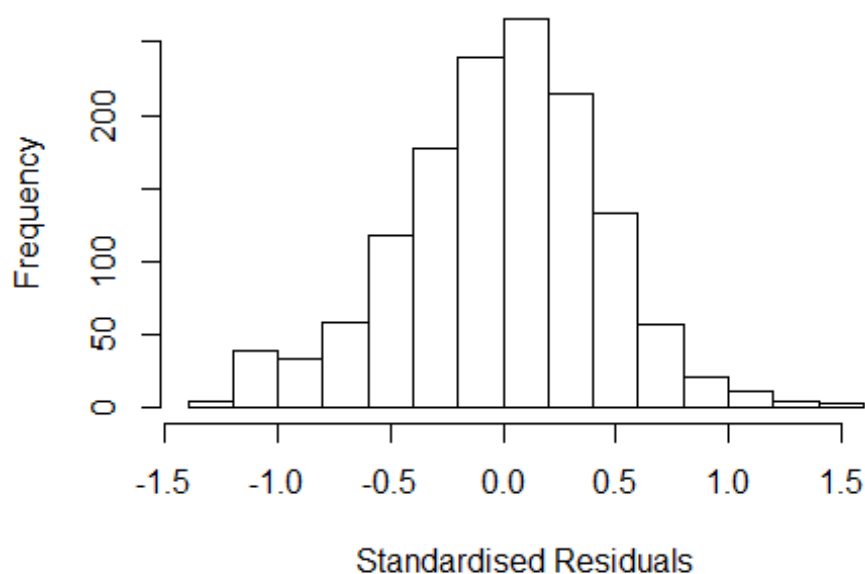
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3450 -0.2773 0.0116 0.2797 1.6221
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.82789 0.13764 6.01 2.3e-09 ***
## FirstAuthorFemale1 -0.03778 0.02575 -1.47 0.14259
## LastAuthorFemale1 -0.09353 0.02533 -3.69 0.00023 ***
## UniqueAuthors2 0.33607 0.07299 4.60 4.5e-06 ***
## UniqueAuthors3 0.33899 0.06944 4.88 1.2e-06 ***
## UniqueAuthors4 0.35138 0.07109 4.94 8.7e-07 ***
## UniqueAuthors5 0.40702 0.06920 5.88 5.1e-09 ***
## Year1997 0.03678 0.17226 0.21 0.83098
## Year1998 -0.08656 0.15453 -0.56 0.57545
## Year1999 -0.03228 0.14198 -0.23 0.82018
```

```

## Year2000      -0.09157      0.14650      -0.63      0.53202
## Year2001      -0.14788      0.16708      -0.89      0.37626
## Year2002      -0.06772      0.13743      -0.49      0.62227
## Year2003      -0.07375      0.12886      -0.57      0.56723
## Year2004      -0.15111      0.14054      -1.08      0.28247
## Year2005       0.01403      0.12886      0.11      0.91331
## Year2006      -0.05460      0.12790      -0.43      0.66955
## Year2007       0.02375      0.12758      0.19      0.85233
## Year2008       0.01770      0.12718      0.14      0.88932
## Year2009       0.06122      0.12578      0.49      0.62652
## Year2010       0.04533      0.12581      0.36      0.71867
## Year2011      -0.00934      0.12659      -0.07      0.94122
## Year2012       0.11008      0.12578      0.88      0.38164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.415
## Multiple R-squared:  0.0952, Adjusted R-squared:  0.0805
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1257 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0928 0.8710 0.9500 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          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 1.118 1      1.057
## LastAuthorFemale  1.141 1      1.068
## Year              1.087 16      1.003

```

Residuals from first and last author



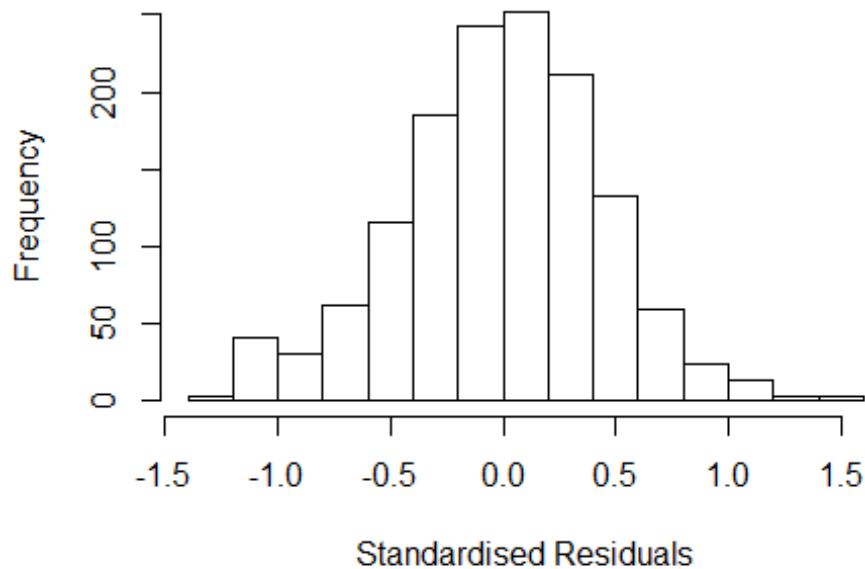
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.289 -0.293  0.017  0.283  1.600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1464    0.1437   7.98 3.2e-15 ***
## FirstAuthorFemale1 -0.0287    0.0258  -1.11   0.27
## LastAuthorFemale1 -0.1059    0.0254  -4.16 3.3e-05 ***
## Year1997          0.0129    0.1902   0.07   0.95
## Year1998         -0.1125    0.1731  -0.65   0.52
## Year1999         -0.0192    0.1618  -0.12   0.91
## Year2000         -0.0595    0.1633  -0.36   0.72
## Year2001         -0.1693    0.2071  -0.82   0.41
## Year2002         -0.0554    0.1587  -0.35   0.73
## Year2003         -0.0519    0.1496  -0.35   0.73
## Year2004         -0.1363    0.1591  -0.86   0.39
## Year2005          0.0511    0.1487   0.34   0.73
```

```

## Year2006          -0.0401      0.1482    -0.27      0.79
## Year2007           0.0411      0.1473      0.28      0.78
## Year2008           0.0461      0.1471      0.31      0.75
## Year2009           0.1043      0.1467      0.71      0.48
## Year2010           0.0749      0.1462      0.51      0.61
## Year2011           0.0281      0.1462      0.19      0.85
## Year2012           0.1423      0.1456      0.98      0.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.416
## Multiple R-squared:  0.0445, Adjusted R-squared:  0.0318
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 111 weights are ~= 1. The remaining 1264 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.107  0.869  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      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 1.033 1      1.017
## Year              1.033 16      1.001

```

Residuals from first author



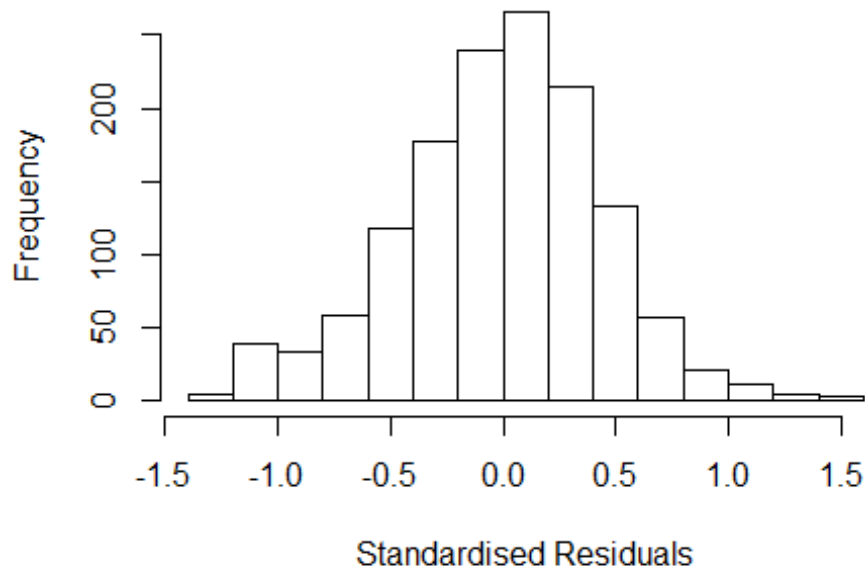
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.25390 -0.29298  0.00402  0.29287  1.55302
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10465    0.13837   7.98   3e-15 ***
## FirstAuthorFemale1 -0.06371    0.02498  -2.55   0.011 *
## Year1997         0.01606    0.19133   0.08   0.933
## Year1998        -0.10430    0.17054  -0.61   0.541
## Year1999         0.00206    0.15775   0.01   0.990
## Year2000        -0.04581    0.15967  -0.29   0.774
## Year2001        -0.16292    0.20203  -0.81   0.420
## Year2002        -0.03071    0.15476  -0.20   0.843
## Year2003        -0.02657    0.14540  -0.18   0.855
## Year2004        -0.11691    0.15553  -0.75   0.452
## Year2005         0.04398    0.14474   0.30   0.761
## Year2006        -0.01992    0.14396  -0.14   0.890
```

```

## Year2007          0.05294    0.14313    0.37    0.712
## Year2008          0.05737    0.14305    0.40    0.688
## Year2009          0.12204    0.14227    0.86    0.391
## Year2010          0.08835    0.14191    0.62    0.534
## Year2011          0.04005    0.14181    0.28    0.778
## Year2012          0.14925    0.14146    1.06    0.292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.421
## Multiple R-squared:  0.0316, Adjusted R-squared:  0.0194
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 123 weights are ~= 1. The remaining 1252 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.144  0.866  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      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 1.052 1      1.026
## Year              1.052 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
## -1.2761 -0.2930  0.0122  0.2765  1.5944
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1341     0.1453   7.81 1.2e-14 ***
## LastAuthorFemale1 -0.1150     0.0244  -4.70 2.8e-06 ***
## Year1997          0.0108     0.1912   0.06  0.95
## Year1998         -0.1096     0.1743  -0.63  0.53
## Year1999         -0.0219     0.1634  -0.13  0.89
## Year2000         -0.0627     0.1647  -0.38  0.70
## Year2001         -0.1684     0.2071  -0.81  0.42
## Year2002         -0.0567     0.1602  -0.35  0.72
## Year2003         -0.0571     0.1512  -0.38  0.71
## Year2004         -0.1352     0.1607  -0.84  0.40
## Year2005          0.0519     0.1504   0.35  0.73
## Year2006         -0.0399     0.1499  -0.27  0.79
```

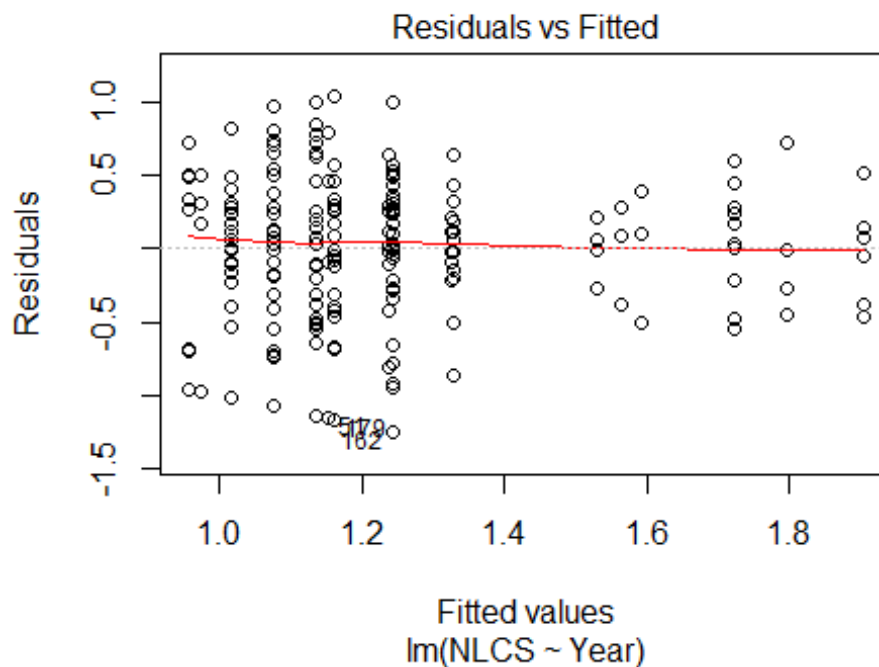


```

## Year2007          0.0417      0.1491      0.28      0.78
## Year2008          0.0448      0.1489      0.30      0.76
## Year2009          0.1025      0.1484      0.69      0.49
## Year2010          0.0720      0.1478      0.49      0.63
## Year2011          0.0259      0.1479      0.17      0.86
## Year2012          0.1420      0.1474      0.96      0.34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.416
## Multiple R-squared:  0.0436, Adjusted R-squared:  0.0317
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 1267 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.109  0.871  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      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 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
##    8   11   11    8    7    7   11    7    5   10   15   17   32   23   31
## 2011 2012
##   28   48
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    4    4    7    3    4   10    5    4    8   11   14   29   22   25
## 2011 2012

```

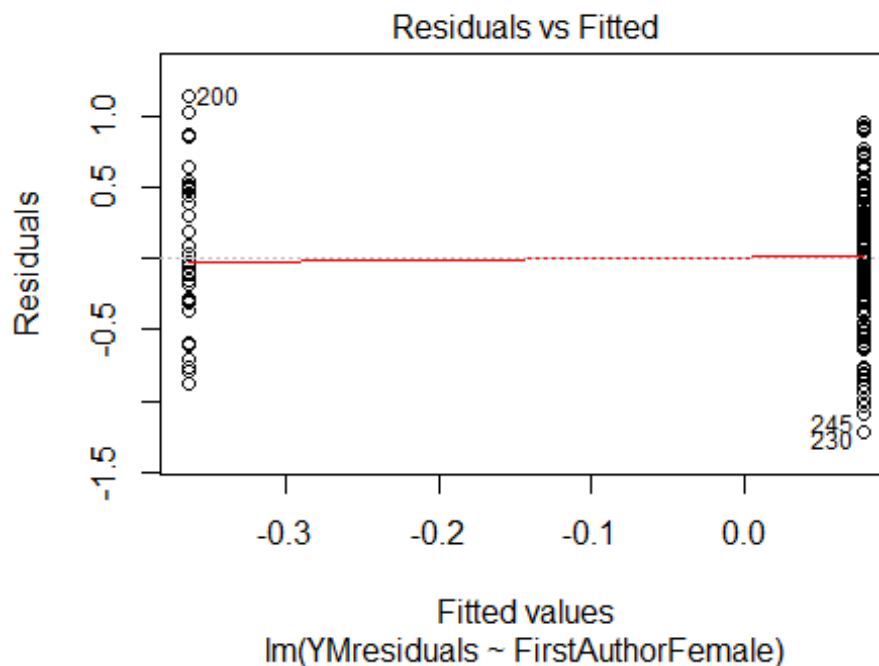
```
## 19 37
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3 4 3 7 2 4 9 5 3 7 9 13 26 21 22
## 2011 2012
## 18 37
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 20, df = 16, p-value = 0.2
```



```
##
## 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: 4.11250590766219"
## [1] "Male first author team size 2018 geometric mean: 3.16227766016838"
## Warning in wilcox.test.default(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 = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.14660639350465"
## [1] "Male last author team size 2018 geometric mean: 4.88073329330349"

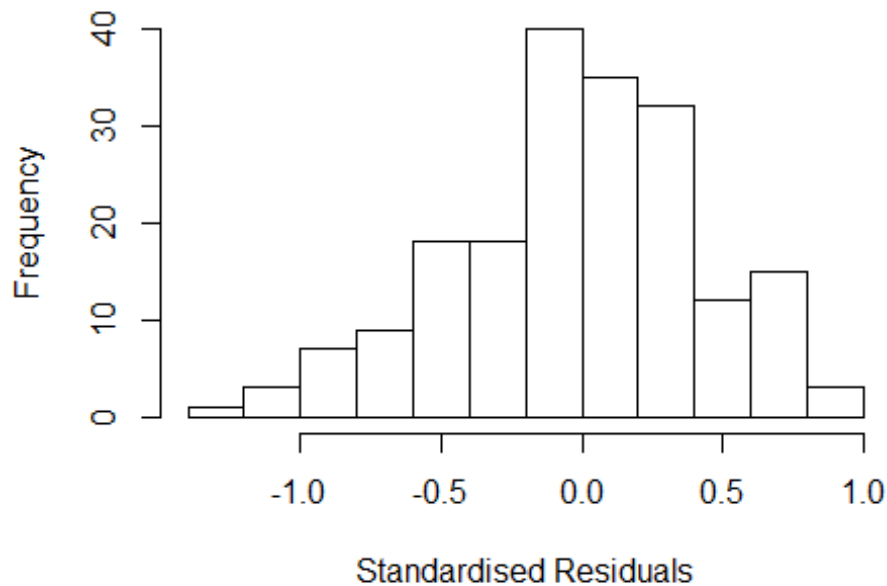
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 40, p-value = 0.5
## 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	2.276	1	1.509
LastAuthorFemale	1.879	1	1.371
UniqueAuthors	8.425	4	1.305
Year	12.738	16	1.083

Residuals from first and last author and team size



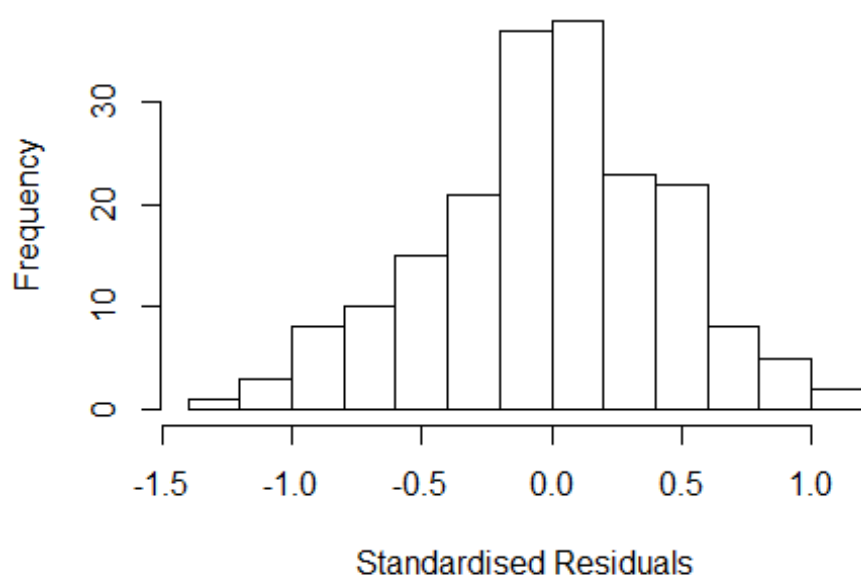
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26132 -0.28458 0.00194 0.28490 0.96535
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8752 0.1977 4.43 1.7e-05 ***
## FirstAuthorFemale1 0.4156 0.1266 3.28 0.0012 **
## LastAuthorFemale1 0.0613 0.0861 0.71 0.4774
## UniqueAuthors2 0.2605 0.1127 2.31 0.0220 *
## UniqueAuthors3 0.4484 0.1420 3.16 0.0019 **
## UniqueAuthors4 0.3015 0.1126 2.68 0.0081 **
## UniqueAuthors5 0.3373 0.1112 3.03 0.0028 **
## Year1997 0.4229 0.3461 1.22 0.2234
## Year1998 0.1813 0.2002 0.91 0.3664
## Year1999 0.5090 0.2516 2.02 0.0446 *
```

```

## Year2000          0.3979      0.1935      2.06      0.0413 *
## Year2001          -0.2069      0.3792     -0.55      0.5862
## Year2002          -0.3652      0.2521     -1.45      0.1494
## Year2003          -0.3130      0.1865     -1.68      0.0952 .
## Year2004          -0.4700      0.2580     -1.82      0.0702 .
## Year2005          -0.2811      0.2279     -1.23      0.2191
## Year2006           0.2226      0.2284      0.97      0.3313
## Year2007          -0.1634      0.2387     -0.68      0.4945
## Year2008          -0.1925      0.1883     -1.02      0.3080
## Year2009          -0.3420      0.1886     -1.81      0.0716 .
## Year2010          -0.3354      0.2164     -1.55      0.1229
## Year2011          -0.5391      0.2056     -2.62      0.0095 **
## Year2012          -0.4095      0.1960     -2.09      0.0382 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.452
## Multiple R-squared:  0.363, Adjusted R-squared:  0.28
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 178 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.416  0.875   0.954   0.914   0.989   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.18e-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 1.717 1      1.310
## LastAuthorFemale  1.831 1      1.353
## Year              1.712 16      1.017

```

Residuals from first and last author



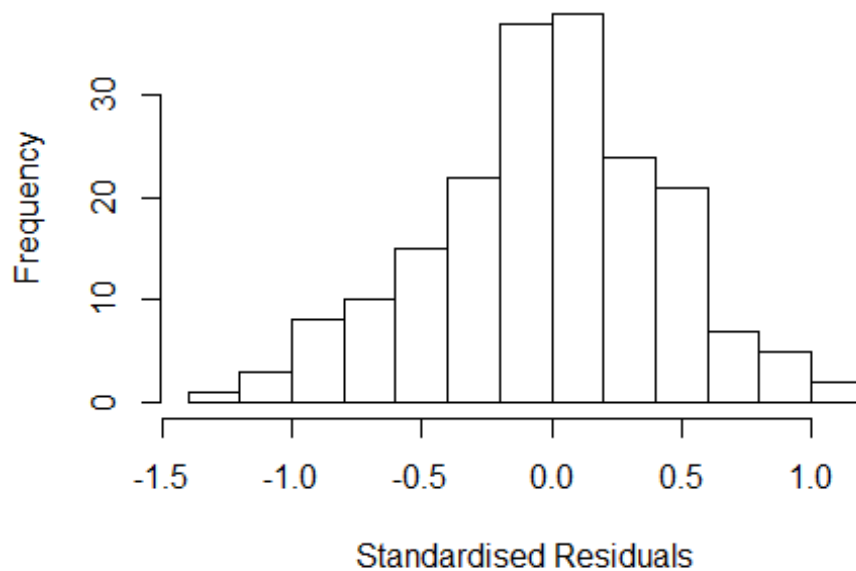
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28474 -0.26486 0.00235 0.32371 1.15621
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0689 0.2625 4.07 7e-05 ***
## FirstAuthorFemale1 0.5081 0.1313 3.87 0.00015 ***
## LastAuthorFemale1 0.0259 0.0940 0.28 0.78344
## Year1997 0.3252 0.3369 0.97 0.33587
## Year1998 0.0173 0.2414 0.07 0.94280
## Year1999 0.3014 0.2653 1.14 0.25742
## Year2000 0.1471 0.2451 0.60 0.54906
## Year2001 -0.2927 0.4052 -0.72 0.47105
## Year2002 -0.2890 0.2795 -1.03 0.30248
## Year2003 -0.2725 0.2466 -1.10 0.27071
## Year2004 -0.4889 0.3483 -1.40 0.16216
## Year2005 -0.3087 0.2976 -1.04 0.30099
```

```

## Year2006          0.1809      0.2667      0.68  0.49856
## Year2007          -0.2110      0.2658     -0.79  0.42836
## Year2008          -0.2573      0.2546     -1.01  0.31362
## Year2009          -0.2973      0.2540     -1.17  0.24347
## Year2010          -0.3181      0.2726     -1.17  0.24479
## Year2011          -0.5354      0.2612     -2.05  0.04192 *
## Year2012          -0.3632      0.2566     -1.42  0.15878
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.483
## Multiple R-squared:  0.302, Adjusted R-squared:  0.23
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 175 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.460  0.877   0.952   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      5.18e-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 1.284 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
## -1.27960 -0.26497  0.00732  0.31880  1.15204
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0784     0.2619   4.12 5.9e-05 ***
## FirstAuthorFemale1 0.5246     0.1145   4.58 8.8e-06 ***
## Year1997          0.3221     0.3367    0.96  0.340
## Year1998          0.0172     0.2417    0.07  0.943
## Year1999          0.3012     0.2656    1.13  0.258
## Year2000          0.1470     0.2454    0.60  0.550
## Year2001         -0.3021     0.4023   -0.75  0.454
## Year2002         -0.2898     0.2796   -1.04  0.301
## Year2003         -0.2779     0.2456   -1.13  0.259
## Year2004         -0.4911     0.3514   -1.40  0.164
## Year2005         -0.3138     0.2988   -1.05  0.295
## Year2006          0.1773     0.2679    0.66  0.509
```



```

## Year2007          -0.2184      0.2641   -0.83    0.409
## Year2008          -0.2601      0.2547   -1.02    0.309
## Year2009          -0.3038      0.2535   -1.20    0.232
## Year2010          -0.3234      0.2721   -1.19    0.236
## Year2011          -0.5451      0.2582   -2.11    0.036 *
## Year2012          -0.3715      0.2550   -1.46    0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.48
## Multiple R-squared:  0.303, Adjusted R-squared:  0.235
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 176 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.457  0.876   0.952   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      5.18e-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 1.418 1          1.191
## Year              1.418 16          1.011
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.2316 -0.3050 -0.0102  0.2797  0.9520

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.373      0.253   5.43 1.9e-07 ***
## LastAuthorFemale1 0.229      0.094   2.44  0.016 *
## Year1997          0.233      0.330   0.71  0.481
## Year1998          0.018      0.240   0.07  0.941
## Year1999          0.302      0.264   1.14  0.254
## Year2000          0.148      0.244   0.61  0.546
## Year2001         -0.243      0.506  -0.48  0.632
## Year2002         -0.325      0.272  -1.20  0.232
## Year2003         -0.232      0.253  -0.92  0.360
## Year2004         -0.556      0.440  -1.26  0.209
## Year2005         -0.395      0.312  -1.27  0.207
## Year2006          0.166      0.266   0.62  0.533
## Year2007         -0.181      0.261  -0.69  0.489
## Year2008         -0.294      0.259  -1.13  0.259
## Year2009         -0.353      0.257  -1.37  0.172
## Year2010         -0.371      0.280  -1.32  0.188
## Year2011         -0.480      0.259  -1.85  0.066 .
## Year2012         -0.410      0.253  -1.62  0.107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.212, Adjusted R-squared:  0.136
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 178 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.519  0.873   0.959   0.912   0.989   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.18e-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: 193"
## [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
##      8   21   10   16    8   13   12    4   10   13   21   23   23   16   23
## 2011 2012
##     24   39
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      1    6    4   10    5   10   11    3   10   11   16   21   21   15   22
## 2011 2012
##     23   34
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      1    6    4   10    5    9   10    3   10   10   13   20   20   15   21
## 2011 2012
##     22   31
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.34381894849665"
## [1] "Male first author team size 2018 geometric mean: 3.13016916014657"

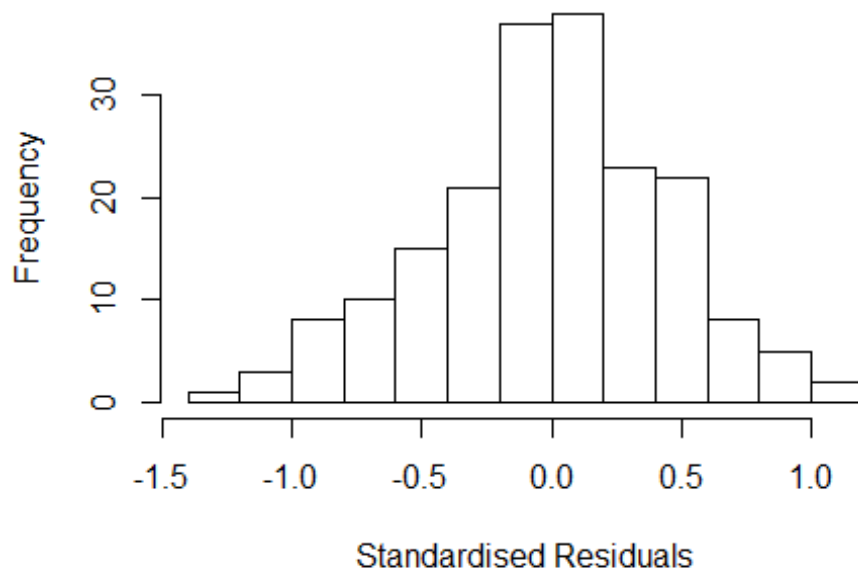
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.32909993477734"
## [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

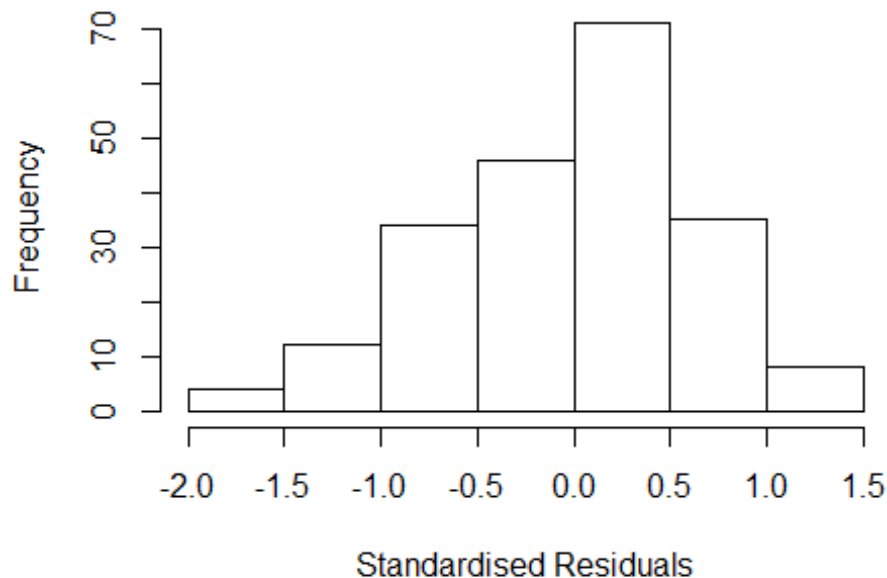
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 33, 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 2.399 1      1.549
## LastAuthorFemale  2.122 1      1.457
## UniqueAuthors    23.484 4      1.484
## Year              34.292 16     1.117
```

Residuals from first and last author and team size



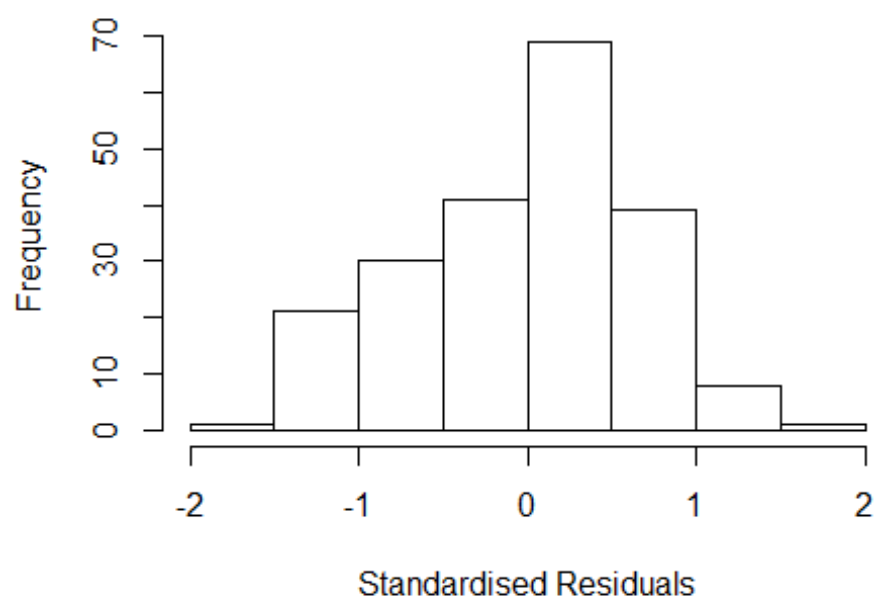
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6938 -0.4606 0.0487 0.4251 1.2078
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5478 0.2580 6.00 1.0e-08 ***
## FirstAuthorFemale1 0.4449 0.1841 2.42 0.01660 *
## LastAuthorFemale1 -0.0810 0.1247 -0.65 0.51685
## UniqueAuthors2 -0.0408 0.1659 -0.25 0.80613
## UniqueAuthors3 0.3655 0.1607 2.27 0.02409 *
## UniqueAuthors4 0.4462 0.1898 2.35 0.01978 *
## UniqueAuthors5 0.7061 0.1446 4.88 2.2e-06 ***
## Year1997 -0.5374 0.4243 -1.27 0.20695
## Year1998 -0.6266 0.5684 -1.10 0.27175
## Year1999 -0.0517 0.2954 -0.18 0.86122
```

```

## Year2000          -0.4004      0.2615    -1.53    0.12747
## Year2001          -0.4681      0.2262    -2.07    0.03988 *
## Year2002          -0.2497      0.4250    -0.59    0.55765
## Year2003          -1.0406      0.3724    -2.79    0.00575 **
## Year2004          -0.2331      0.2548    -0.91    0.36138
## Year2005          -0.7160      0.2131    -3.36    0.00095 ***
## Year2006          -0.4841      0.1669    -2.90    0.00418 **
## Year2007          -0.8774      0.2132    -4.12    5.8e-05 ***
## Year2008          -0.7056      0.1546    -4.56    9.1e-06 ***
## Year2009          -1.2629      0.2300    -5.49    1.3e-07 ***
## Year2010          -0.5194      0.1862    -2.79    0.00584 **
## Year2011          -0.8948      0.1245    -7.19    1.5e-11 ***
## Year2012          -0.9241      0.1721    -5.37    2.3e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.643
## Multiple R-squared:  0.279, Adjusted R-squared:  0.195
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 189 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.467  0.863  0.947   0.911  0.989   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 2.276 1      1.509
## LastAuthorFemale 2.233 1      1.494
## Year      2.492 16      1.029

```

Residuals from first and last author



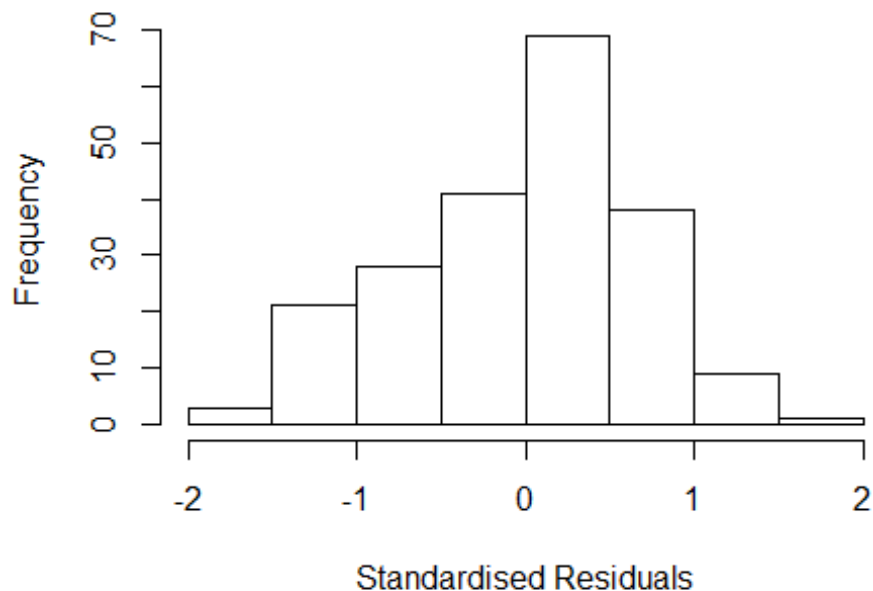
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7096 -0.4728  0.0624  0.4343  1.8148
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4037     0.1612   8.71 1.5e-15 ***
## FirstAuthorFemale1  0.6126     0.1871   3.27 0.00126 **
## LastAuthorFemale1 -0.1453     0.1402  -1.04 0.30160
## Year1997          -0.3561     0.4813  -0.74 0.46031
## Year1998          -0.3790     0.4499  -0.84 0.40067
## Year1999           0.1186     0.2179   0.54 0.58684
## Year2000          -0.3292     0.1787  -1.84 0.06704 .
## Year2001          -0.2849     0.2150  -1.32 0.18679
## Year2002          -0.0902     0.3877  -0.23 0.81622
## Year2003          -0.7362     0.3922  -1.88 0.06206 .
## Year2004           0.0453     0.1867   0.24 0.80837
## Year2005          -0.4375     0.1603  -2.73 0.00696 **
```

```

## Year2006          -0.3166      0.1761    -1.80    0.07375 .
## Year2007          -0.5467      0.2645    -2.07    0.04009 *
## Year2008          -0.4183      0.1288    -3.25    0.00137 **
## Year2009          -0.8446      0.2363    -3.57    0.00045 ***
## Year2010          -0.2034      0.1600    -1.27    0.20516
## Year2011          -0.6419      0.1187    -5.41    1.9e-07 ***
## Year2012          -0.4127      0.1477    -2.79    0.00574 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.645
## Multiple R-squared:  0.156, Adjusted R-squared:  0.0764
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 189 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.409  0.849  0.935  0.896  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.76e-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 1.635 1      1.279
## Year              1.635 16      1.015

```


Residuals from first author



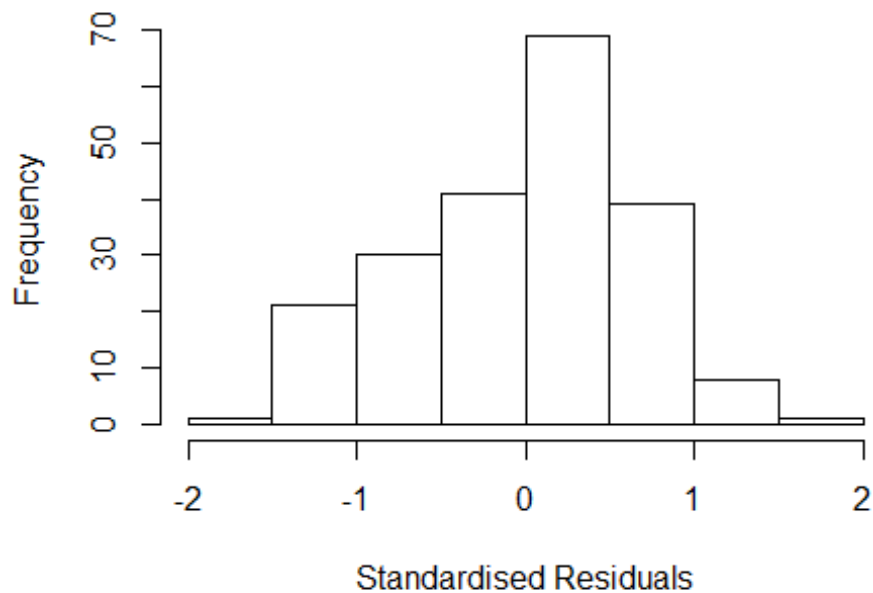
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7303 -0.4799 0.0813 0.4460 1.7079
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3472 0.1563 8.62 2.5e-15 ***
## FirstAuthorFemale1 0.5238 0.1563 3.35 0.00097 ***
## Year1997 -0.3597 0.4495 -0.80 0.42455
## Year1998 -0.4132 0.4547 -0.91 0.36465
## Year1999 0.1393 0.2180 0.64 0.52375
## Year2000 -0.3094 0.1741 -1.78 0.07709 .
## Year2001 -0.2601 0.2057 -1.26 0.20760
## Year2002 -0.0720 0.3826 -0.19 0.85085
## Year2003 -0.6822 0.4324 -1.58 0.11624
## Year2004 0.0634 0.2036 0.31 0.75580
## Year2005 -0.4310 0.1606 -2.68 0.00794 **
## Year2006 -0.3024 0.1713 -1.77 0.07914 .
```

```

## Year2007          -0.5221      0.2751   -1.90  0.05923 .
## Year2008          -0.3898      0.1260   -3.09  0.00228 **
## Year2009          -0.8302      0.2302   -3.61  0.00040 ***
## Year2010          -0.1838      0.1613   -1.14  0.25574
## Year2011          -0.6071      0.1173   -5.17  5.7e-07 ***
## Year2012          -0.3678      0.1386   -2.65  0.00864 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.643
## Multiple R-squared:  0.151, Adjusted R-squared:  0.0755
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 193 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.450  0.846  0.942  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      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 1.423 1          1.193
## Year            1.423 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
## -1.706 -0.473 0.061 0.489 1.523
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.8155 0.1276 14.23 < 2e-16 ***
## LastAuthorFemale1 0.0555 0.1276 0.43 0.66405
## Year1997 -0.5758 0.3172 -1.82 0.07105 .
## Year1998 -0.6338 0.5024 -1.26 0.20869
## Year1999 0.1149 0.2285 0.50 0.61568
## Year2000 -0.4773 0.2361 -2.02 0.04459 *
## Year2001 -0.2904 0.2333 -1.24 0.21466
## Year2002 -0.1220 0.2981 -0.41 0.68276
## Year2003 -0.7997 0.6059 -1.32 0.18849
## Year2004 0.0299 0.2342 0.13 0.89842
## Year2005 -0.4809 0.1613 -2.98 0.00324 **
## Year2006 -0.3796 0.1769 -2.15 0.03311 *
```

```

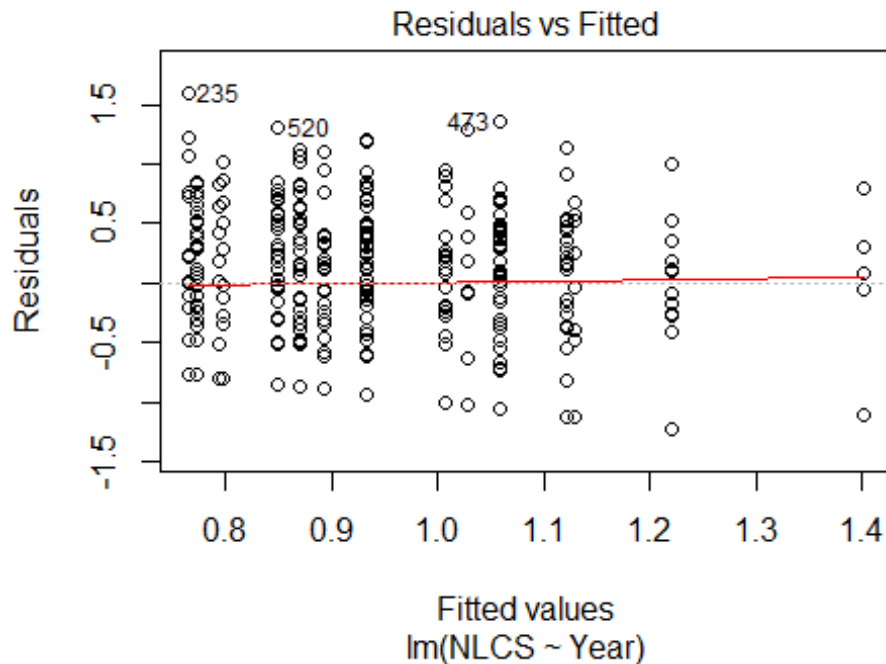
## Year2007          -0.6018      0.2445    -2.46   0.01473 *
## Year2008          -0.4087      0.1344    -3.04   0.00270 **
## Year2009          -0.8978      0.2348    -3.82   0.00018 ***
## Year2010          -0.2868      0.1683    -1.70   0.08995 .
## Year2011          -0.5862      0.1203    -4.87   2.3e-06 ***
## Year2012          -0.4024      0.1501    -2.68   0.00797 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.72
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0219
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 192 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.554  0.857  0.952  0.913  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 2920"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1997 1998 1999 2000 2002 2003 2005 2006 2007 2008 2009 2010 2011 2012
##    1    3    1    2    2    2    3    2    1    5    2    2    3    1
##
## 1997 1998 1999 2000 2002 2003 2005 2006 2007 2008 2009 2010 2011 2012
##    0    3    1    1    1    2    0    2    1    3    2    2    3    1
##
## 1997 1998 1999 2000 2002 2003 2005 2006 2007 2008 2009 2010 2011 2012
##    0    2    1    1    1    2    0    1    1    3    2    2    3    0

```

```

## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1.41421356237309"
## [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: 19"
## [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
## 24 26 24 19 19 11 20 15 14 29 26 29 43 62 62
## 2011 2012
## 53 53
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 9 13 12 11 5 13 9 8 24 19 24 38 53 55
## 2011 2012
## 43 44
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 9 12 12 10 5 12 6 6 21 16 21 35 49 50
## 2011 2012
## 39 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 = 7.2, df = 16, p-value = 1

```



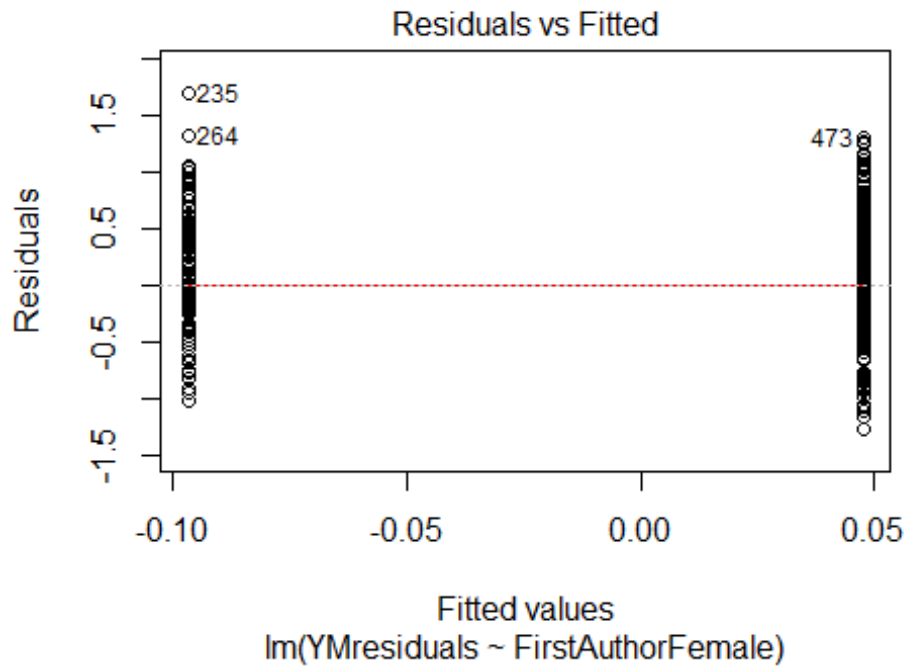
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.26, df = 1, p-value = 0.6

## [1] "Female first author team size 2018 geometric mean: 3.27748533887161"
## [1] "Male first author team size 2018 geometric mean: 4.41703772222759"

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

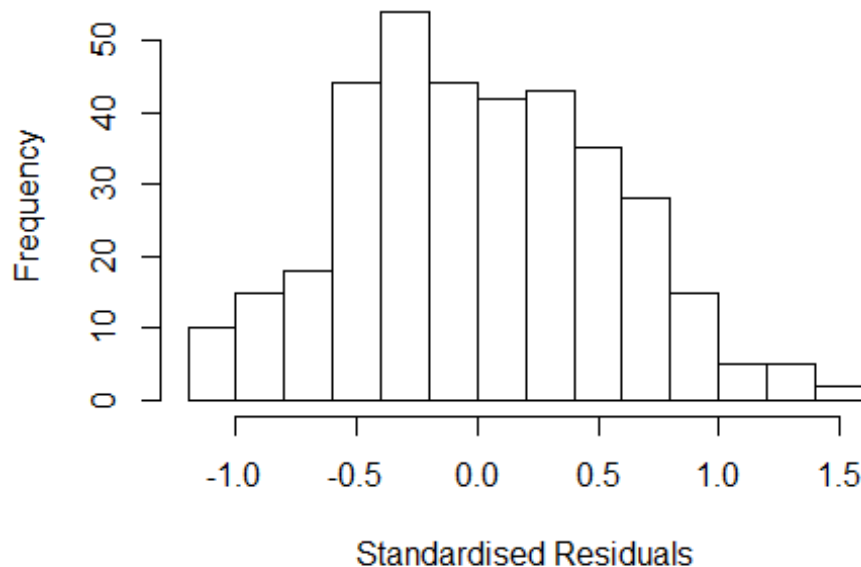
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.41380124107933"
## [1] "Male last author team size 2018 geometric mean: 3.70254939790312"

## Warning in wilcox.test.default(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 = 0.8
## 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 1.703 1          1.305
## LastAuthorFemale  1.791 1          1.338
## UniqueAuthors    1.954 4          1.087
## Year             2.764 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
## -1.1119 -0.3951 -0.0147 0.3935 1.5104
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.63e-01 1.67e-01 2.17 0.0304 *
## FirstAuthorFemale1 1.23e-01 7.89e-02 1.56 0.1198
## LastAuthorFemale1 1.76e-01 7.91e-02 2.22 0.0272 *
## UniqueAuthors2 2.20e-01 9.35e-02 2.35 0.0192 *
## UniqueAuthors3 3.30e-01 9.03e-02 3.65 0.0003 ***
## UniqueAuthors4 5.60e-01 9.94e-02 5.63 3.8e-08 ***
## UniqueAuthors5 4.34e-01 9.78e-02 4.44 1.2e-05 ***
## Year1997 3.15e-01 2.82e-01 1.12 0.2641
## Year1998 4.20e-01 2.15e-01 1.95 0.0518 .
## Year1999 3.94e-01 1.92e-01 2.05 0.0411 *
```

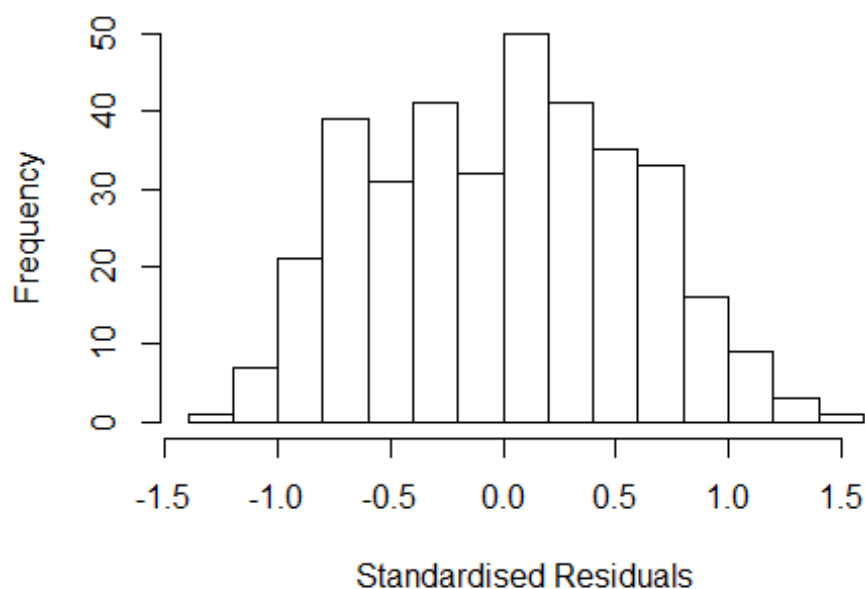


```

## Year2000          1.90e-01    2.46e-01    0.77    0.4420
## Year2001          6.96e-01    3.14e-01    2.22    0.0271 *
## Year2002          3.69e-01    2.28e-01    1.62    0.1068
## Year2003         -6.84e-04    2.30e-01    0.00    0.9976
## Year2004          2.27e-01    3.31e-01    0.69    0.4936
## Year2005          1.86e-01    1.99e-01    0.93    0.3505
## Year2006         -2.92e-01    2.18e-01   -1.34    0.1814
## Year2007         -3.09e-05    1.81e-01    0.00    0.9999
## Year2008          3.28e-02    1.71e-01    0.19    0.8479
## Year2009          1.33e-01    1.66e-01    0.80    0.4229
## Year2010          2.30e-01    1.74e-01    1.32    0.1869
## Year2011          6.65e-03    1.77e-01    0.04    0.9701
## Year2012          1.23e-02    1.83e-01    0.07    0.9463
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.59
## Multiple R-squared:  0.182, Adjusted R-squared:  0.128
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 332 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.492  0.893  0.955  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      2.78e-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 1.500 1      1.225
## LastAuthorFemale  1.537 1      1.240
## Year              1.280 16      1.008

```

Residuals from first and last author



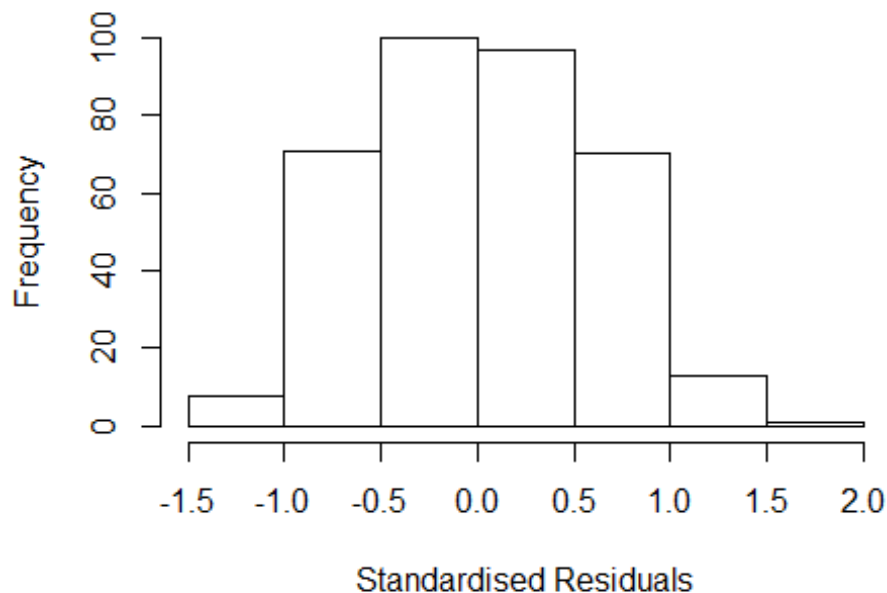
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2362 -0.4401 0.0365 0.4660 1.4586
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5053 0.1759 2.87 0.0043 **
## FirstAuthorFemale1 0.1342 0.0827 1.62 0.1058
## LastAuthorFemale1 0.1678 0.0799 2.10 0.0364 *
## Year1997 0.2918 0.2987 0.98 0.3293
## Year1998 0.4289 0.2207 1.94 0.0527 .
## Year1999 0.3501 0.2095 1.67 0.0956 .
## Year2000 0.1557 0.2814 0.55 0.5804
## Year2001 0.6372 0.3327 1.92 0.0563 .
## Year2002 0.3552 0.2588 1.37 0.1709
## Year2003 0.0646 0.3036 0.21 0.8316
## Year2004 0.3644 0.3395 1.07 0.2838
## Year2005 0.2526 0.2018 1.25 0.2116
```

```

## Year2006          -0.1548      0.2304   -0.67   0.5023
## Year2007           0.1584      0.1972    0.80   0.4222
## Year2008           0.1204      0.1874    0.64   0.5208
## Year2009           0.2569      0.1822    1.41   0.1594
## Year2010           0.3621      0.1868    1.94   0.0533 .
## Year2011           0.1658      0.1927    0.86   0.3900
## Year2012           0.1984      0.1946    1.02   0.3087
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.646
## Multiple R-squared:  0.0897, Adjusted R-squared:  0.0417
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 336 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.589  0.897  0.952  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      2.78e-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 1.133 1      1.065
## Year              1.133 16      1.004

```

Residuals from first author



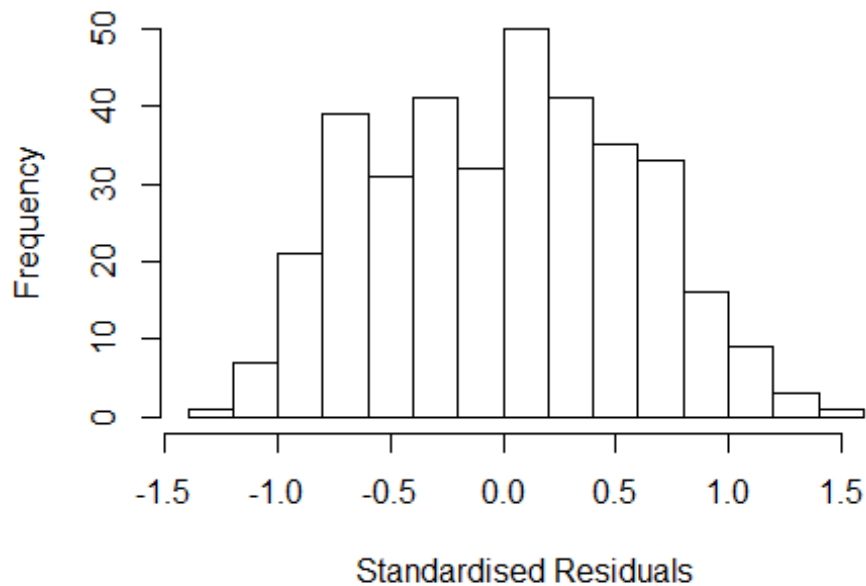
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24281 -0.41157 0.00168 0.44090 1.56823
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5836 0.1722 3.39 0.00078 ***
## FirstAuthorFemale1 0.2166 0.0732 2.96 0.00330 **
## Year1997 0.2656 0.3015 0.88 0.37894
## Year1998 0.4426 0.2216 2.00 0.04657 *
## Year1999 0.3638 0.2114 1.72 0.08618 .
## Year2000 0.1320 0.2754 0.48 0.63192
## Year2001 0.6431 0.3319 1.94 0.05348 .
## Year2002 0.3453 0.2561 1.35 0.17848
## Year2003 0.0683 0.3013 0.23 0.82094
## Year2004 0.3222 0.3317 0.97 0.33202
## Year2005 0.2137 0.2045 1.05 0.29675
## Year2006 -0.1748 0.2351 -0.74 0.45752
```

```

## Year2007          0.1467      0.1956      0.75  0.45372
## Year2008          0.0778      0.1873      0.42  0.67788
## Year2009          0.2260      0.1804      1.25  0.21113
## Year2010          0.3154      0.1852      1.70  0.08943 .
## Year2011          0.1130      0.1892      0.60  0.55071
## Year2012          0.1509      0.1929      0.78  0.43481
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.653
## Multiple R-squared:  0.0785, Adjusted R-squared:  0.0326
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 335 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.544  0.895   0.951   0.927   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      2.78e-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 1.159 1          1.077
## Year              1.159 16          1.005

```

Residuals from last author



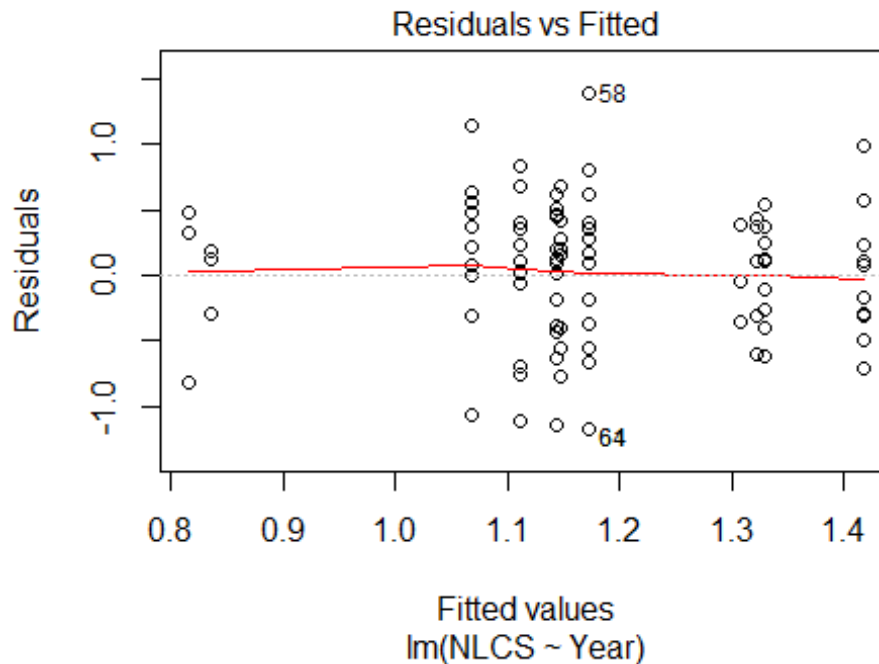
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2244 -0.4107 0.0377 0.4271 1.4286
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5690 0.1715 3.32 0.001 **
## LastAuthorFemale1 0.2321 0.0700 3.31 0.001 **
## Year1997 0.2842 0.2895 0.98 0.327
## Year1998 0.4233 0.2174 1.95 0.052 .
## Year1999 0.3251 0.2041 1.59 0.112
## Year2000 0.1636 0.2784 0.59 0.557
## Year2001 0.6437 0.3323 1.94 0.054 .
## Year2002 0.3485 0.2571 1.36 0.176
## Year2003 0.0343 0.2916 0.12 0.907
## Year2004 0.3299 0.3341 0.99 0.324
## Year2005 0.2488 0.2018 1.23 0.219
## Year2006 -0.1664 0.2273 -0.73 0.465
```

```

## Year2007          0.1475      0.1976      0.75      0.456
## Year2008          0.0988      0.1865      0.53      0.597
## Year2009          0.2458      0.1831      1.34      0.180
## Year2010          0.3514      0.1866      1.88      0.061 .
## Year2011          0.1574      0.1928      0.82      0.415
## Year2012          0.1809      0.1948      0.93      0.354
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.644
## Multiple R-squared:  0.0825, Adjusted R-squared:  0.0369
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 342 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.601  0.889  0.958  0.927  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.78e-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: 360"
## [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]"
##
## 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##   4   8   10   4   3   7   13   15   11   14   19
##
## 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##   3   8   9   3   3   5   12   14   10   13   16
##
## 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##   3   8   8   3   3   5   12   13   10   13   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 = 9.2, df = 10, p-value = 0.5
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.6, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 1.36426160182137"
## [1] "Male first author team size 2018 geometric mean: 1.782602457966"

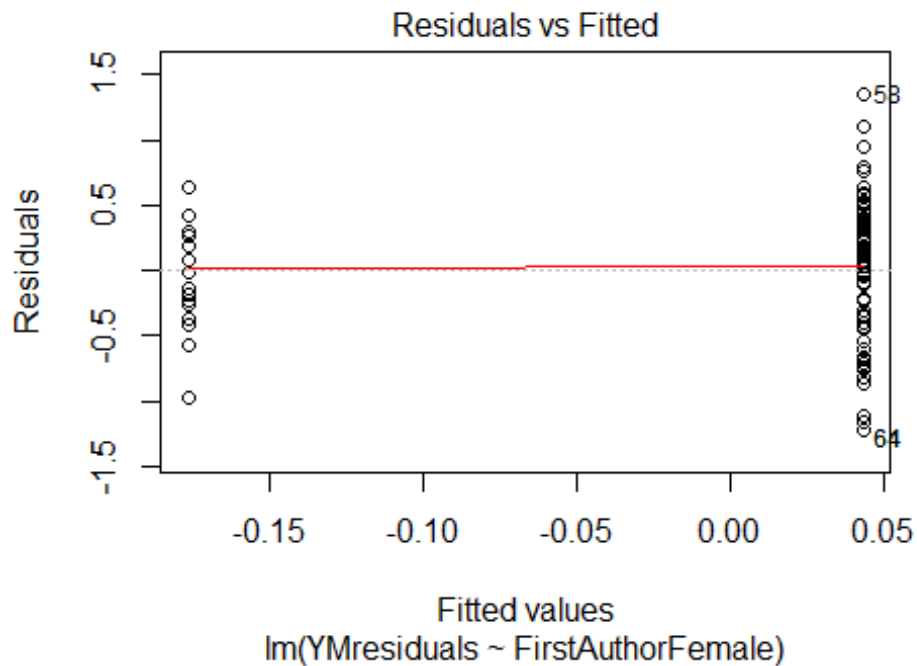
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
```



```
## [1] "Female last author team size 2018 geometric mean: 1.53367468683856"
## [1] "Male last author team size 2018 geometric mean: 1.44224957030741"

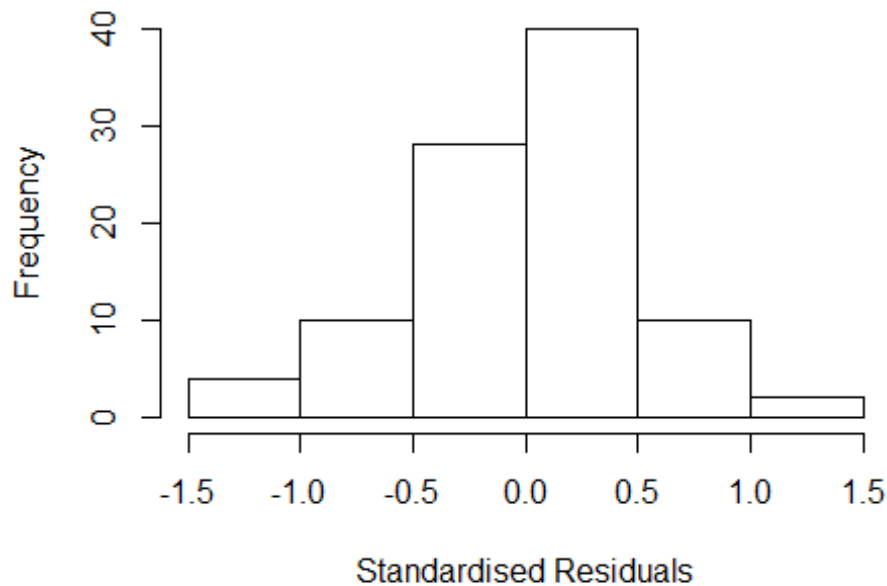
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



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

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	3.260	1	1.806
LastAuthorFemale	2.961	1	1.721
UniqueAuthors	12.384	4	1.370
Year	30.596	10	1.187

Residuals from first and last author and team size



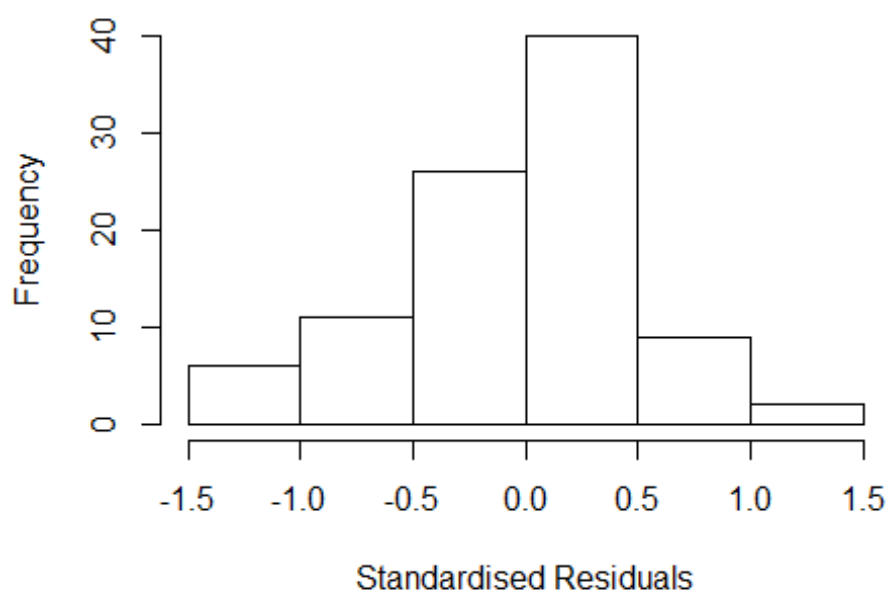
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1944 -0.2679 0.0257 0.3579 1.2379
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.926920 0.270021 3.43 0.00096 ***
## FirstAuthorFemale1 0.250580 0.147668 1.70 0.09375 .
## LastAuthorFemale1 0.000969 0.133117 0.01 0.99421
## UniqueAuthors2 0.148025 0.172841 0.86 0.39442
## UniqueAuthors3 0.358160 0.113073 3.17 0.00220 **
## UniqueAuthors4 0.085022 0.351703 0.24 0.80962
## UniqueAuthors5 0.474569 0.202518 2.34 0.02169 *
## Year2003 -0.062498 0.278195 -0.22 0.82284
## Year2004 0.215363 0.264819 0.81 0.41858
## Year2005 -0.255157 0.306171 -0.83 0.40721
```

```

## Year2006      -0.490428    0.418422   -1.17  0.24477
## Year2007      0.057712    0.247689    0.23  0.81638
## Year2008     -0.201398    0.327984   -0.61  0.54099
## Year2009     -0.080059    0.273642   -0.29  0.77064
## Year2010      0.163735    0.297485    0.55  0.58364
## Year2011     -0.132096    0.301130   -0.44  0.66213
## Year2012     -0.074742    0.264787   -0.28  0.77849
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.502
## Multiple R-squared:  0.176, Adjusted R-squared:  0.00443
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 87 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.522  0.880   0.957   0.911   0.989   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.06e-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 2.060 1      1.435
## LastAuthorFemale 2.316 1      1.522
## Year      2.655 10      1.050

```

Residuals from first and last author



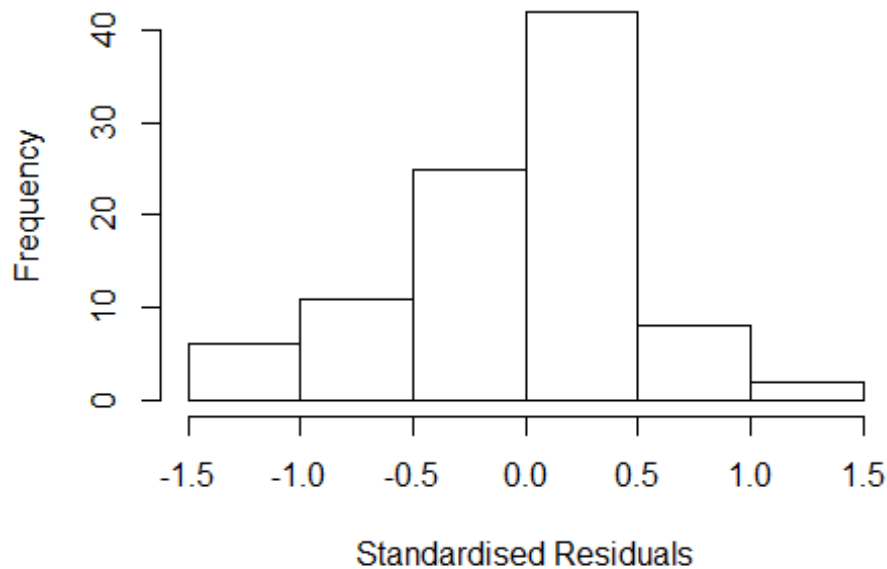
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2791 -0.3379 0.0475 0.3321 1.2859
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0061 0.2464 4.08 0.0001 ***
## FirstAuthorFemale1 0.3193 0.1494 2.14 0.0356 *
## LastAuthorFemale1 -0.0640 0.1392 -0.46 0.6466
## Year2003 -0.0697 0.2870 -0.24 0.8088
## Year2004 0.2021 0.2852 0.71 0.4805
## Year2005 -0.3374 0.2862 -1.18 0.2419
## Year2006 -0.4164 0.4319 -0.96 0.3378
## Year2007 0.0851 0.2536 0.34 0.7381
## Year2008 -0.1568 0.3078 -0.51 0.6118
## Year2009 0.0177 0.2777 0.06 0.9494
## Year2010 0.1936 0.2830 0.68 0.4958
## Year2011 -0.0795 0.2778 -0.29 0.7754
```

```

## Year2012          0.0258      0.2603      0.10      0.9212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.536
## Multiple R-squared:  0.0921, Adjusted R-squared:  -0.0424
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 84 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.545  0.882   0.957   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      1.06e-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 1.316  1      1.147
## Year              1.316 10      1.014

```

Residuals from first author



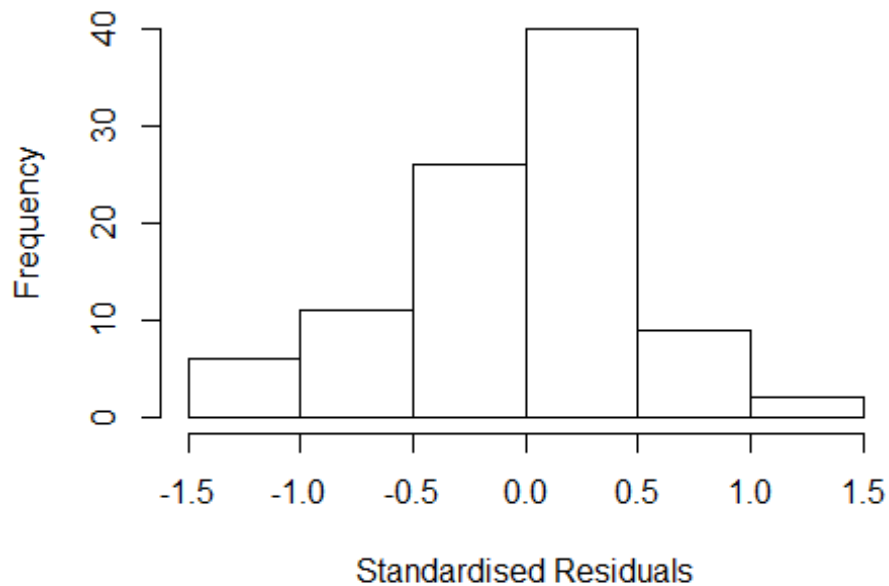
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2943 -0.3457 0.0484 0.3259 1.2707
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0328 0.2202 4.69 1.1e-05 ***
## FirstAuthorFemale1 0.2728 0.1195 2.28 0.025 *
## Year2003 -0.1111 0.2516 -0.44 0.660
## Year2004 0.1598 0.2541 0.63 0.531
## Year2005 -0.3757 0.2564 -1.47 0.147
## Year2006 -0.4336 0.4409 -0.98 0.328
## Year2007 0.0713 0.2402 0.30 0.767
## Year2008 -0.1734 0.3107 -0.56 0.578
## Year2009 -0.0113 0.2631 -0.04 0.966
## Year2010 0.1532 0.2510 0.61 0.543
## Year2011 -0.1202 0.2408 -0.50 0.619
## Year2012 -0.0183 0.2195 -0.08 0.934
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.53
## Multiple R-squared:  0.0907, Adjusted R-squared:  -0.0313
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 83 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.531  0.885   0.956   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      1.06e-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 1.351  1      1.162
## Year            1.351 10      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
## -1.297 -0.367 0.107 0.319 1.268
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2717 0.1577 8.06 5.2e-12 ***
## LastAuthorFemale1 0.1071 0.1164 0.92 0.36
## Year2003 -0.2066 0.2483 -0.83 0.41
## Year2004 -0.0181 0.2076 -0.09 0.93
## Year2005 -0.5048 0.2135 -2.36 0.02 *
## Year2006 -0.4682 0.4555 -1.03 0.31
## Year2007 0.0135 0.2470 0.05 0.96
## Year2008 -0.2081 0.2921 -0.71 0.48
## Year2009 -0.0817 0.2519 -0.32 0.75
## Year2010 0.0409 0.2252 0.18 0.86
## Year2011 -0.2204 0.2320 -0.95 0.34
## Year2012 -0.1764 0.1984 -0.89 0.38
```

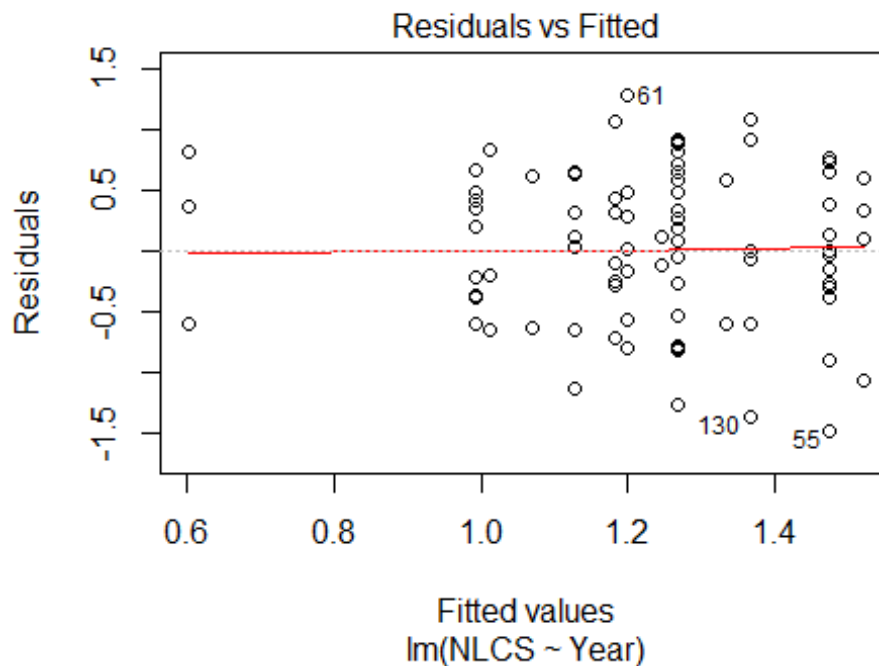


```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.0621, Adjusted R-squared:  -0.0637
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 86 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.543  0.880  0.956  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           1.06e-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:  94"
## [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]"
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    2   11   10    7   10   15   10   10    4    4    3   14   10   10    6
## 2012
##    8
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    2    2    0    2   14    8   10    4    4    3   14    9   10    6
## 2012
##    7
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    2    2    0    2   11    8   10    3    4    3   14    7    8    5
## 2012
##    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 = 6.6, df = 13, 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

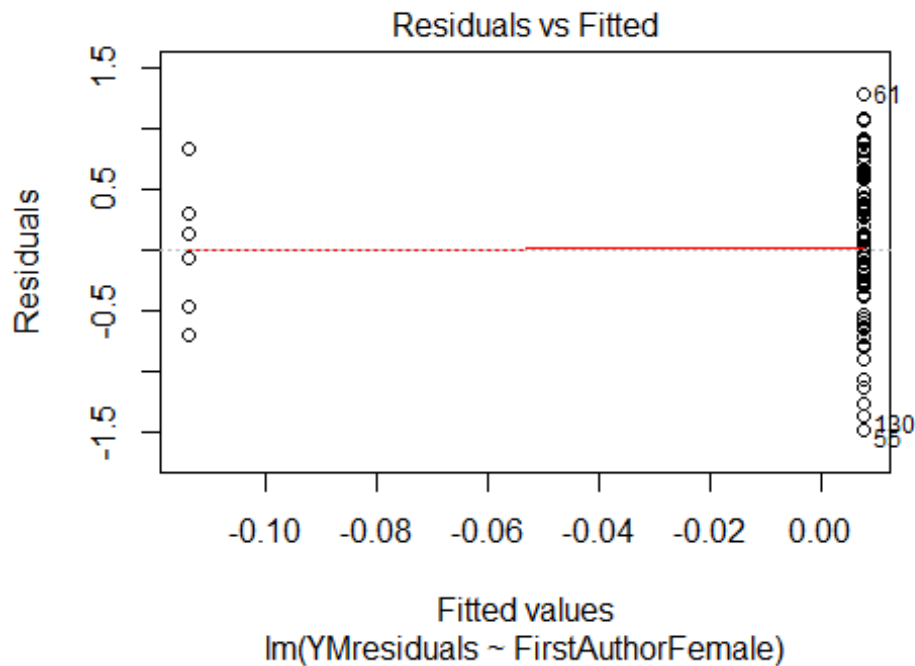
## [1] "Female first author team size 2018 geometric mean: 2.99379516552391"
## [1] "Male first author team size 2018 geometric mean: 2"

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

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4.5, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
```

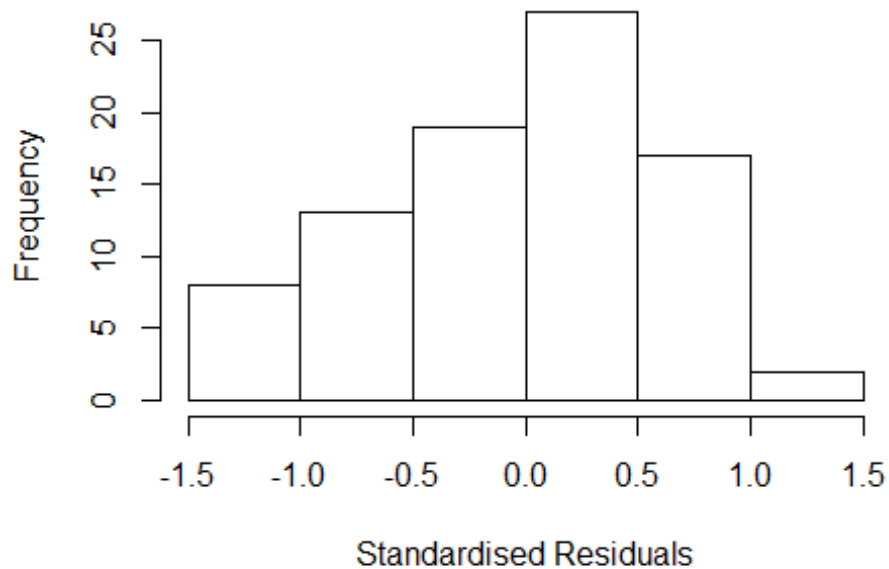
```
## [1] "Female last author team size 2018 geometric mean: 2.66716827534"
## [1] "Male last author team size 2018 geometric mean: 4"

## 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 = 2, 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 2.553 1      1.598
## LastAuthorFemale  1.483 1      1.218
## UniqueAuthors    25.296 4      1.498
## Year              54.754 13     1.166
```

Residuals from first and last author and team size



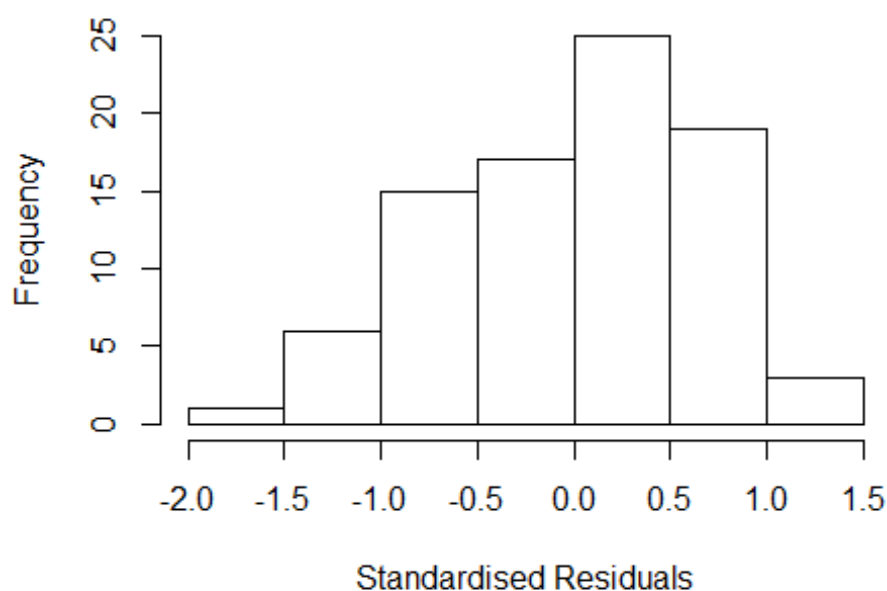
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.338 -0.484 0.122 0.392 1.372
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.77864 0.70277 1.11 0.27
## FirstAuthorFemale1 0.00302 0.33852 0.01 0.99
## LastAuthorFemale1 0.26021 0.27807 0.94 0.35
## UniqueAuthors2 -0.01704 0.31354 -0.05 0.96
## UniqueAuthors3 0.37770 0.26894 1.40 0.16
## UniqueAuthors4 0.18405 0.28914 0.64 0.53
## UniqueAuthors5 0.29414 0.42295 0.70 0.49
## Year1999 0.21166 0.64246 0.33 0.74
## Year2001 0.02714 0.83471 0.03 0.97
## Year2002 0.29589 0.62290 0.48 0.64
```

```

## Year2003      0.08911      0.69461      0.13      0.90
## Year2004     -0.12116      0.66314     -0.18      0.86
## Year2005     -0.83297      0.66241     -1.26      0.21
## Year2006      0.32179      0.69354      0.46      0.64
## Year2007     -0.03957      0.71786     -0.06      0.96
## Year2008      0.16629      0.67830      0.25      0.81
## Year2009     -0.03616      0.67060     -0.05      0.96
## Year2010      0.03889      0.60793      0.06      0.95
## Year2011      0.05487      0.73380      0.07      0.94
## Year2012      0.05723      0.69789      0.08      0.93
##
## Robust residual standard error: 0.694
## Multiple R-squared:  0.16,   Adjusted R-squared:  -0.082
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## one weight is ~= 1. The remaining 85 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.675  0.894   0.961   0.928   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.16e-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 2.188 1      1.479
## LastAuthorFemale  1.512 1      1.230
## Year              2.391 13      1.034

```

Residuals from first and last author



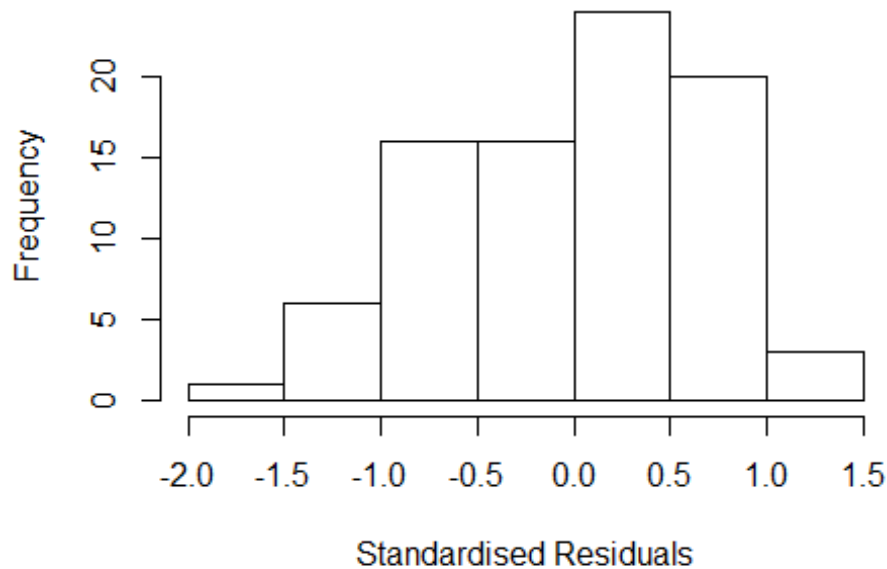
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5347 -0.5419 0.0957 0.5488 1.2852
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12291 0.51294 2.19 0.032 *
## FirstAuthorFemale1 0.03258 0.28262 0.12 0.909
## LastAuthorFemale1 0.19530 0.25825 0.76 0.452
## Year1999 -0.10579 0.48329 -0.22 0.827
## Year2001 -0.28179 0.69603 -0.40 0.687
## Year2002 0.18386 0.52774 0.35 0.729
## Year2003 -0.15000 0.52388 -0.29 0.775
## Year2004 -0.33644 0.50350 -0.67 0.506
## Year2005 -1.03679 0.55375 -1.87 0.065 .
## Year2006 0.21326 0.58213 0.37 0.715
## Year2007 -0.34555 0.59573 -0.58 0.564
## Year2008 0.00551 0.54910 0.01 0.992
```

```

## Year2009          -0.10703      0.52223    -0.20      0.838
## Year2010          -0.18287      0.50037    -0.37      0.716
## Year2011          -0.12803      0.60138    -0.21      0.832
## Year2012          -0.16089      0.54764    -0.29      0.770
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.748
## Multiple R-squared:  0.121, Adjusted R-squared:  -0.0676
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 4 weights are ~= 1. The remaining 82 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.653  0.918  0.948  0.932  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      1.16e-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 1.803 1          1.343
## Year              1.803 13          1.023

```

Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5367 -0.5388 0.0788 0.5319 1.3062
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2855 0.4476 2.87 0.0054 **
## FirstAuthorFemale1 0.0980 0.2820 0.35 0.7293
## Year1999 -0.1385 0.4547 -0.30 0.7616
## Year2001 -0.3145 0.6786 -0.46 0.6445
## Year2002 0.1532 0.5081 0.30 0.7639
## Year2003 -0.2037 0.4838 -0.42 0.6750
## Year2004 -0.3899 0.4720 -0.83 0.4115
## Year2005 -1.0700 0.5280 -2.03 0.0464 *
## Year2006 0.1827 0.5645 0.32 0.7472
## Year2007 -0.3526 0.5702 -0.62 0.5383
## Year2008 -0.0646 0.5254 -0.12 0.9024
## Year2009 -0.1403 0.4944 -0.28 0.7774
```

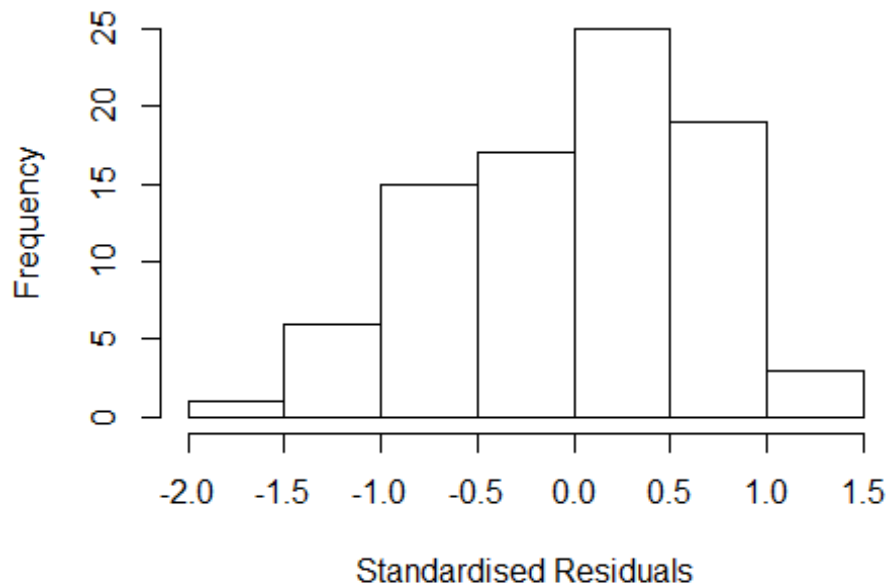


```

## Year2010          -0.2221      0.4808   -0.46   0.6455
## Year2011          -0.2350      0.5798   -0.41   0.6865
## Year2012          -0.2258      0.5132   -0.44   0.6613
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.733
## Multiple R-squared:  0.116, Adjusted R-squared:  -0.0582
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 4 weights are ~= 1. The remaining 82 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.640  0.902  0.951  0.929  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      1.16e-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 1.262 1          1.123
## Year            1.262 13          1.009

```

Residuals from last author



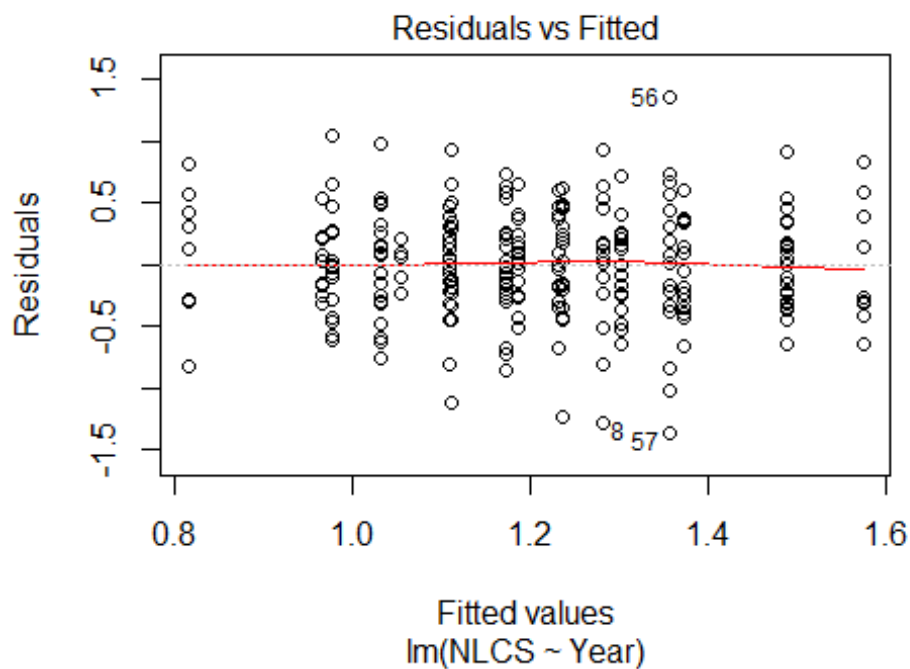
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5361 -0.5428 0.0949 0.5488 1.2898
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1315 0.5298 2.14 0.036 *
## LastAuthorFemale1 0.2030 0.2416 0.84 0.404
## Year1999 -0.0895 0.4800 -0.19 0.853
## Year2001 -0.2655 0.6954 -0.38 0.704
## Year2002 0.2016 0.5245 0.38 0.702
## Year2003 -0.1383 0.5347 -0.26 0.797
## Year2004 -0.3193 0.4961 -0.64 0.522
## Year2005 -1.0209 0.5512 -1.85 0.068 .
## Year2006 0.2311 0.5791 0.40 0.691
## Year2007 -0.3430 0.6179 -0.56 0.581
## Year2008 0.0225 0.5504 0.04 0.968
## Year2009 -0.0911 0.5194 -0.18 0.861
```

```

## Year2010          -0.1739      0.5149   -0.34    0.737
## Year2011          -0.1089      0.5904   -0.18    0.854
## Year2012          -0.1421      0.5366   -0.26    0.792
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.737
## Multiple R-squared:  0.122, Adjusted R-squared:  -0.0516
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 4 weights are ~= 1. The remaining 82 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.643  0.916  0.946  0.931  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      1.16e-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: 86"
## [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
##   26   22   27   29   18   25   23   23   28   29   27   19   22   16   17
## 2011 2012
##   19   26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   16    9   16    5   11   18   14   21   25   17   17   16    9   13
## 2011 2012
##   17   23
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

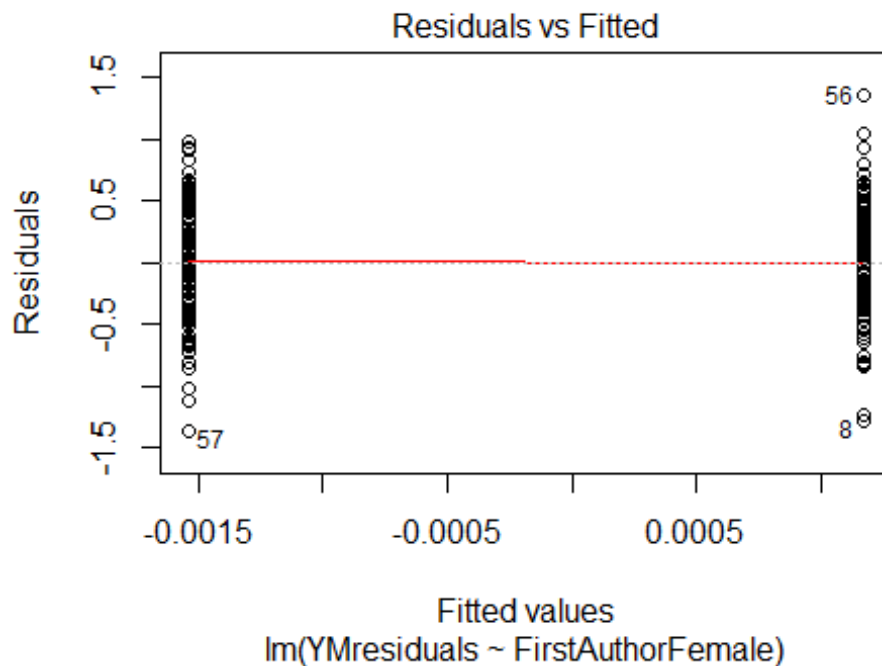
```
## 15 14 6 12 4 11 18 12 15 23 14 16 15 7 13
## 2011 2012
## 15 19
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 31, df = 16, p-value = 0.01
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2, df = 1, p-value = 0.2
## [1] "Female first author team size 2018 geometric mean: 2.70586577923534"
## [1] "Male first author team size 2018 geometric mean: 2.61070249390159"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
```

```
## W = 42, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.76288939685185"
## [1] "Male last author team size 2018 geometric mean: 2.54768217138607"

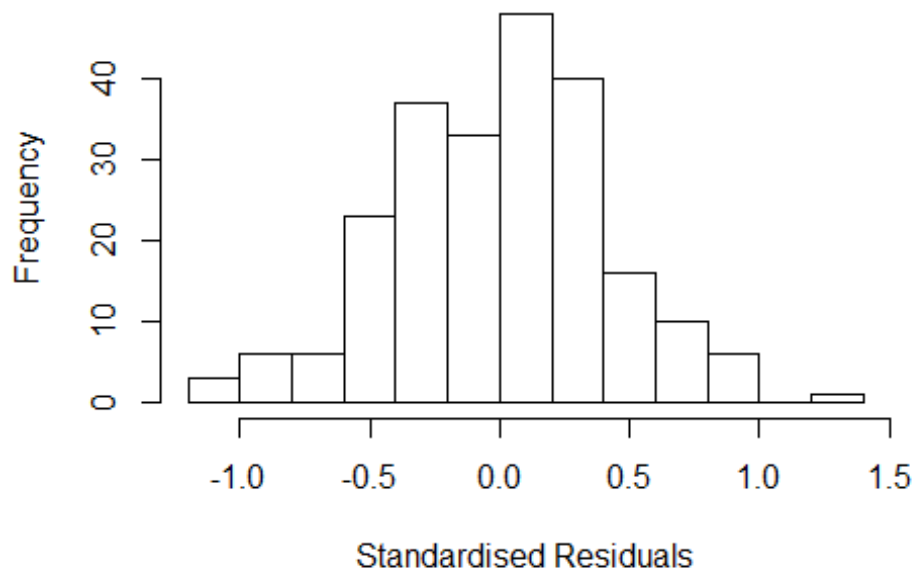
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 51, p-value = 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	1.511	1	1.229
LastAuthorFemale	1.460	1	1.208
UniqueAuthors	4.089	4	1.192
Year	6.373	16	1.060

Residuals from first and last author and team size



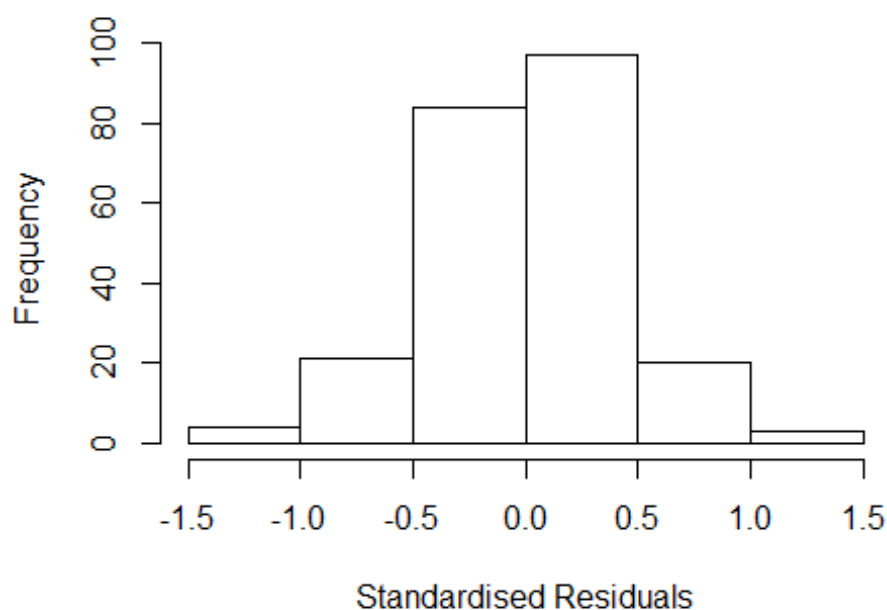
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.171 -0.284 0.028 0.259 1.329
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09399 0.20415 5.36 2.2e-07 ***
## FirstAuthorFemale1 -0.00451 0.06212 -0.07 0.942
## LastAuthorFemale1 0.07344 0.07771 0.95 0.346
## UniqueAuthors2 0.07703 0.15443 0.50 0.618
## UniqueAuthors3 0.22493 0.14938 1.51 0.134
## UniqueAuthors4 0.23414 0.13993 1.67 0.096 .
## UniqueAuthors5 0.24003 0.14410 1.67 0.097 .
## Year1997 0.04451 0.22109 0.20 0.841
## Year1998 0.53674 0.37036 1.45 0.149
## Year1999 0.14373 0.19164 0.75 0.454
```

```

## Year2000      -0.18819    0.17447   -1.08    0.282
## Year2001      -0.35748    0.16204   -2.21    0.028 *
## Year2002       0.02810    0.17189    0.16    0.870
## Year2003      -0.07456    0.17990   -0.41    0.679
## Year2004      -0.31267    0.22586   -1.38    0.168
## Year2005      -0.16750    0.15786   -1.06    0.290
## Year2006      -0.18689    0.17197   -1.09    0.278
## Year2007      -0.10857    0.16311   -0.67    0.506
## Year2008      -0.00270    0.17790   -0.02    0.988
## Year2009      -0.50031    0.26450   -1.89    0.060 .
## Year2010      -0.15782    0.20178   -0.78    0.435
## Year2011      -0.30298    0.18684   -1.62    0.106
## Year2012       0.20128    0.17202    1.17    0.243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.417
## Multiple R-squared:  0.21, Adjusted R-squared:  0.126
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 211 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.289  0.882  0.957  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      4.37e-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 1.245 1      1.116
## LastAuthorFemale  1.284 1      1.133
## Year              1.597 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
## -1.32219 -0.26667 0.00951 0.27697 1.34759
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.32219 0.13681 9.66 <2e-16 ***
## FirstAuthorFemale1 0.00749 0.06139 0.12 0.903
## LastAuthorFemale1 0.06121 0.07620 0.80 0.423
## Year1997 0.03822 0.22485 0.17 0.865
## Year1998 0.42430 0.31398 1.35 0.178
## Year1999 0.09828 0.18781 0.52 0.601
## Year2000 -0.31530 0.15671 -2.01 0.045 *
## Year2001 -0.37719 0.15442 -2.44 0.015 *
## Year2002 -0.03316 0.15900 -0.21 0.835
## Year2003 -0.12702 0.17182 -0.74 0.461
## Year2004 -0.40216 0.19830 -2.03 0.044 *
## Year2005 -0.18319 0.15454 -1.19 0.237
```

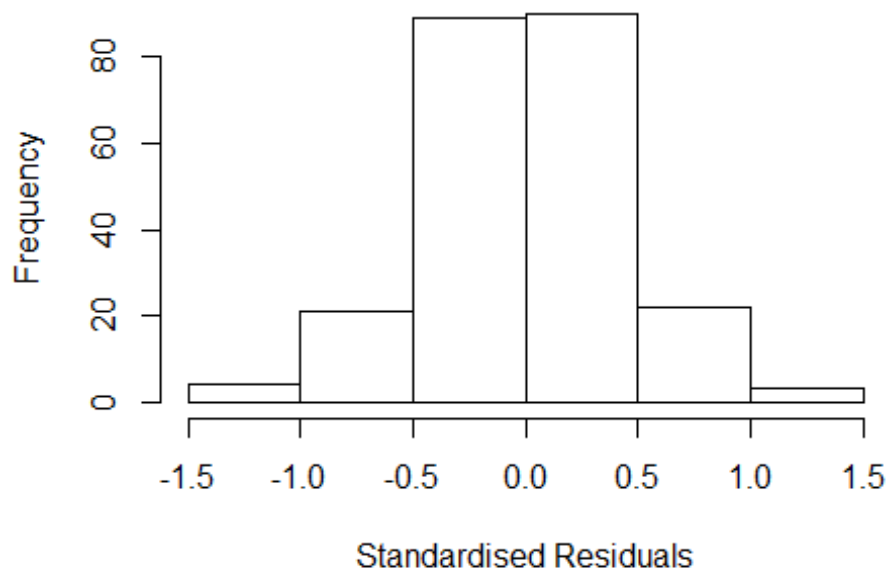


```

## Year2006      -0.22712      0.16092      -1.41      0.160
## Year2007      -0.15522      0.15889      -0.98      0.330
## Year2008      -0.02413      0.16856      -0.14      0.886
## Year2009      -0.53456      0.31680      -1.69      0.093 .
## Year2010      -0.20637      0.20328      -1.02      0.311
## Year2011      -0.34733      0.17849      -1.95      0.053 .
## Year2012      0.16176      0.15809      1.02      0.307
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.415
## Multiple R-squared:  0.186, Adjusted R-squared:  0.117
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 211 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.271 0.880 0.953 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      4.37e-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 1.244 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
## -1.33160 -0.27345  0.00131  0.27455  1.32567
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3316    0.1391   9.57  <2e-16 ***
## FirstAuthorFemale1  0.0188    0.0629   0.30   0.766
## Year1997          0.0507    0.2243   0.23   0.821
## Year1998          0.4152    0.3026   1.37   0.172
## Year1999          0.0898    0.1888   0.48   0.635
## Year2000         -0.2968    0.1617  -1.84   0.068 .
## Year2001         -0.3798    0.1578  -2.41   0.017 *
## Year2002         -0.0367    0.1638  -0.22   0.823
## Year2003         -0.1289    0.1761  -0.73   0.465
## Year2004         -0.4013    0.1975  -2.03   0.043 *
## Year2005         -0.1801    0.1596  -1.13   0.260
## Year2006         -0.2212    0.1654  -1.34   0.183
```

```

## Year2007          -0.1577      0.1621   -0.97    0.332
## Year2008          -0.0152      0.1740   -0.09    0.930
## Year2009          -0.5377      0.3236   -1.66    0.098 .
## Year2010          -0.2176      0.2034   -1.07    0.286
## Year2011          -0.3476      0.1824   -1.91    0.058 .
## Year2012           0.1610      0.1627    0.99    0.324
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.421
## Multiple R-squared:  0.182, Adjusted R-squared:  0.116
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 209 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.297  0.880  0.957  0.905  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.37e-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 1.288 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 ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.3232 -0.2635  0.0148  0.2737  1.3451

```

```

##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3232    0.1356   9.76 <2e-16 ***
## LastAuthorFemale1  0.0626    0.0778   0.80  0.422
## Year1997          0.0398    0.2234   0.18  0.859
## Year1998          0.4252    0.3110   1.37  0.173
## Year1999          0.0991    0.1873   0.53  0.597
## Year2000         -0.3151    0.1575  -2.00  0.047 *
## Year2001         -0.3748    0.1542  -2.43  0.016 *
## Year2002         -0.0309    0.1587  -0.19  0.846
## Year2003         -0.1237    0.1697  -0.73  0.467
## Year2004         -0.3980    0.1914  -2.08  0.039 *
## Year2005         -0.1820    0.1543  -1.18  0.240
## Year2006         -0.2282    0.1606  -1.42  0.157
## Year2007         -0.1548    0.1586  -0.98  0.330
## Year2008         -0.0237    0.1688  -0.14  0.889
## Year2009         -0.5338    0.3135  -1.70  0.090 .
## Year2010         -0.2048    0.2031  -1.01  0.314
## Year2011         -0.3445    0.1826  -1.89  0.061 .
## Year2012          0.1643    0.1583   1.04  0.301
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.422
## Multiple R-squared:  0.185, Adjusted R-squared:  0.119
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 210 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.289  0.882  0.954  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           4.37e-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: 229"
## [1] ""

```

```

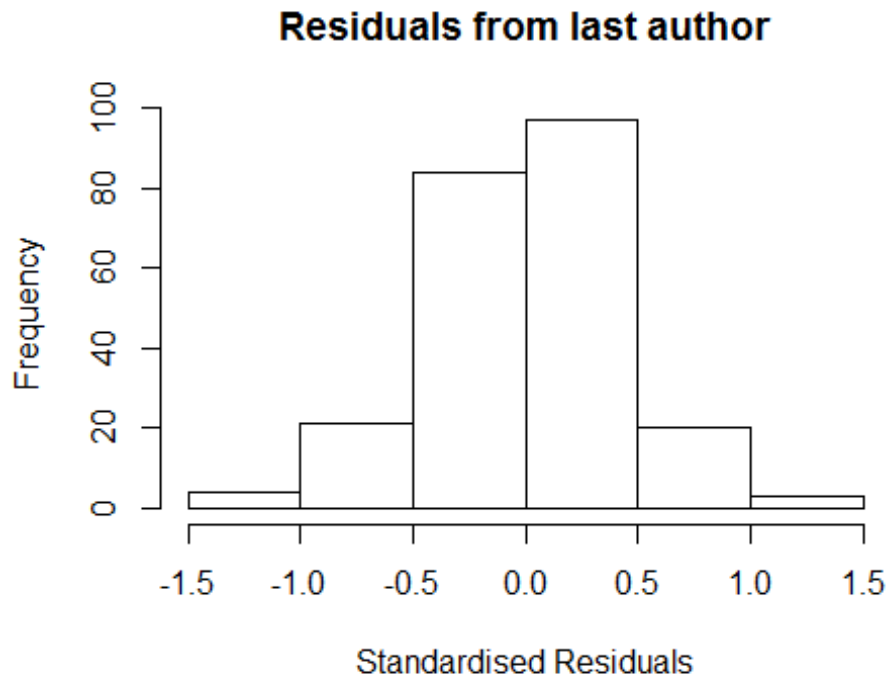
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3001"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2002 2003 2004 2005 2008 2009 2010 2011 2012
##    3    3    3    1    3    4    2    2    2    2    1    3    5    7
##
## 1996 1997 1998 1999 2000 2002 2003 2004 2005 2008 2009 2010 2011 2012
##    1    1    3    1    2    3    2    1    1    1    0    3    4    3
##
## 1996 1997 1998 1999 2000 2002 2003 2004 2005 2008 2009 2010 2011 2012
##    1    1    2    1    1    3    2    1    1    1    0    3    4    3
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 6.789570751021"
## [1] "Male first author team size 2018 geometric mean: 4.57885697021333"

## Warning in wilcox.test.default(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 = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 7.20041148735702"
## [1] "Male last author team size 2018 geometric mean: 5.23582563093831"

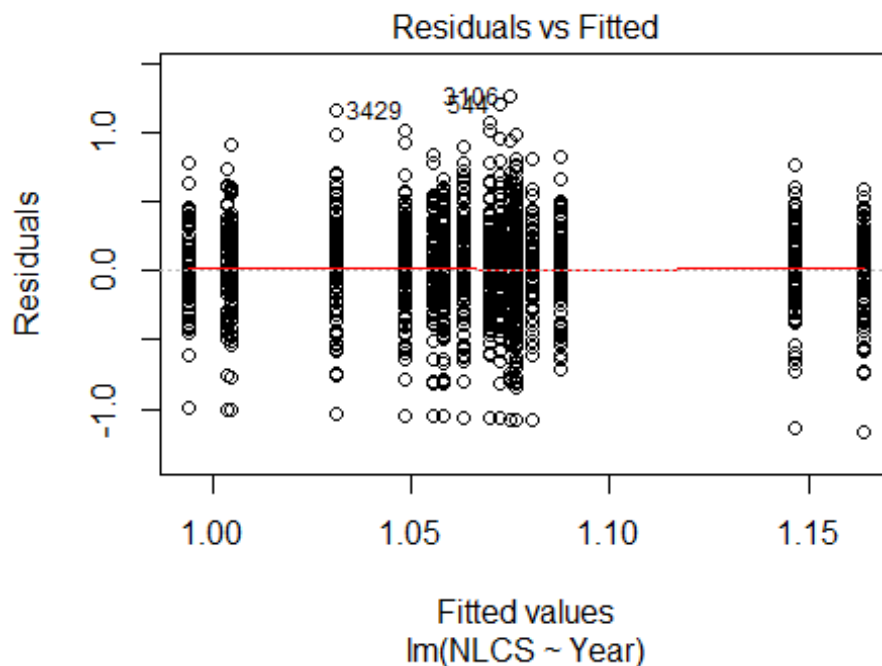
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```

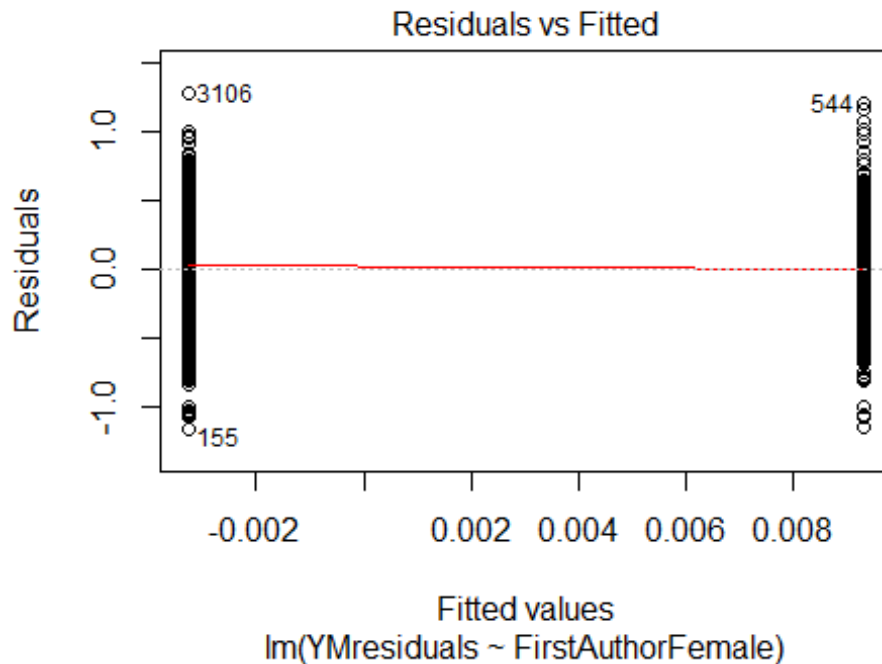


```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## [1] "Regression 2: 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: 24"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3002"
## [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
## 155 148 159 154 129 128 140 125 139 140 138 176 171 195 225
## 2011 2012
## 195 171
##
```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 101 94 115 103 64 69 98 84 104 100 98 119 122 141 154
## 2011 2012
## 147 128
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 74 76 88 79 51 57 76 64 85 79 89 107 103 116 129
## 2011 2012
## 121 103
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 31, df = 16, p-value = 0.01
```

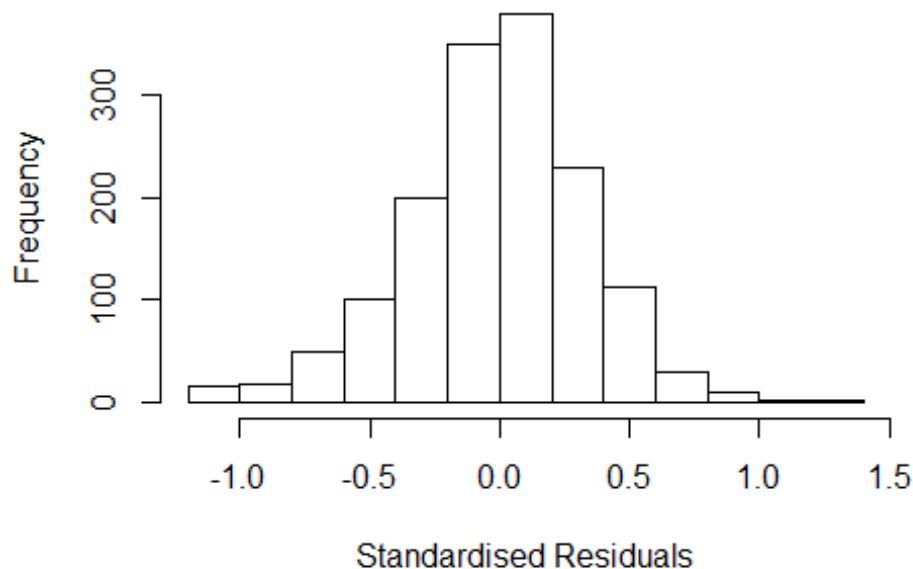


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0011, df = 1, p-value = 1
```



```
## [1] "Female first author team size 2018 geometric mean: 4.38660448262952"
## [1] "Male first author team size 2018 geometric mean: 4.77842808232382"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.89768984142762"
## [1] "Male last author team size 2018 geometric mean: 4.78619494891406"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 540, 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 1.055 1      1.027
## LastAuthorFemale  1.119 1      1.058
## UniqueAuthors     1.531 4      1.055
## Year              1.574 16     1.014
```


Residuals from first and last author and team size



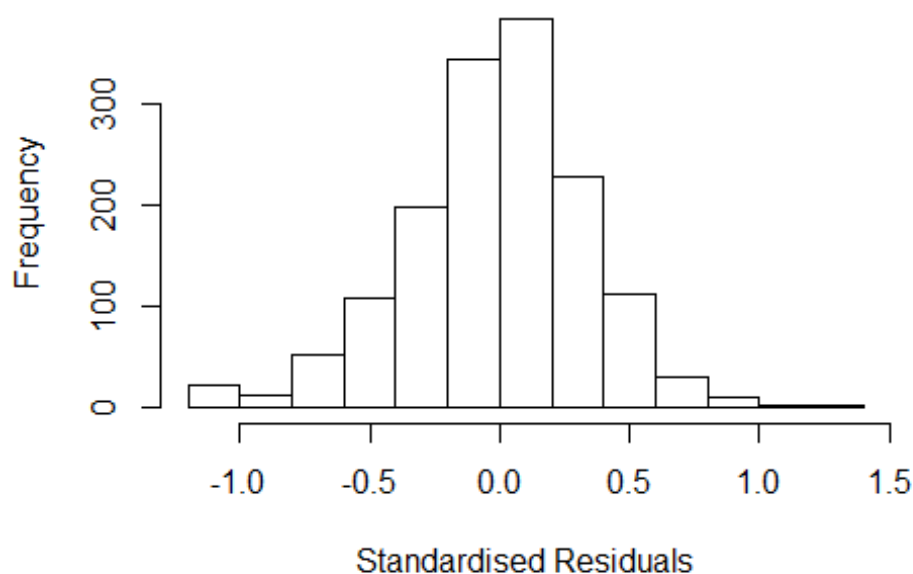
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18574 -0.20517 0.00519 0.20917 1.29849
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.92925 0.07728 12.02 < 2e-16 ***
## FirstAuthorFemale1 -0.00513 0.01987 -0.26 0.79640
## LastAuthorFemale1 0.01699 0.02873 0.59 0.55452
## UniqueAuthors2 0.15802 0.06776 2.33 0.01982 *
## UniqueAuthors3 0.15620 0.06696 2.33 0.01980 *
## UniqueAuthors4 0.19346 0.06835 2.83 0.00471 **
## UniqueAuthors5 0.24795 0.06680 3.71 0.00021 ***
## Year1997 0.04626 0.04916 0.94 0.34685
## Year1998 -0.00844 0.05610 -0.15 0.88043
## Year1999 -0.07040 0.05141 -1.37 0.17112
```

```

## Year2000      -0.11274      0.06495      -1.74      0.08280 .
## Year2001      0.03635      0.05545      0.66      0.51219
## Year2002     -0.07932      0.05368     -1.48      0.13968
## Year2003     -0.11449      0.05250     -2.18      0.02934 *
## Year2004     -0.03301      0.05822     -0.57      0.57079
## Year2005     -0.05122      0.05303     -0.97      0.33423
## Year2006     -0.04274      0.04972     -0.86      0.39019
## Year2007     -0.04046      0.05108     -0.79      0.42846
## Year2008     -0.05495      0.05164     -1.06      0.28748
## Year2009     -0.06838      0.04890     -1.40      0.16218
## Year2010     -0.03762      0.04867     -0.77      0.43963
## Year2011     -0.04176      0.04988     -0.84      0.40263
## Year2012     -0.06850      0.06155     -1.11      0.26595
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.316
## Multiple R-squared:  0.0413, Adjusted R-squared:  0.027
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 146 weights are ~= 1. The remaining 1351 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0525 0.8590 0.9500 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.68e-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 1.050 1      1.025
## LastAuthorFemale  1.052 1      1.026
## Year              1.103 16      1.003

```

Residuals from first and last author



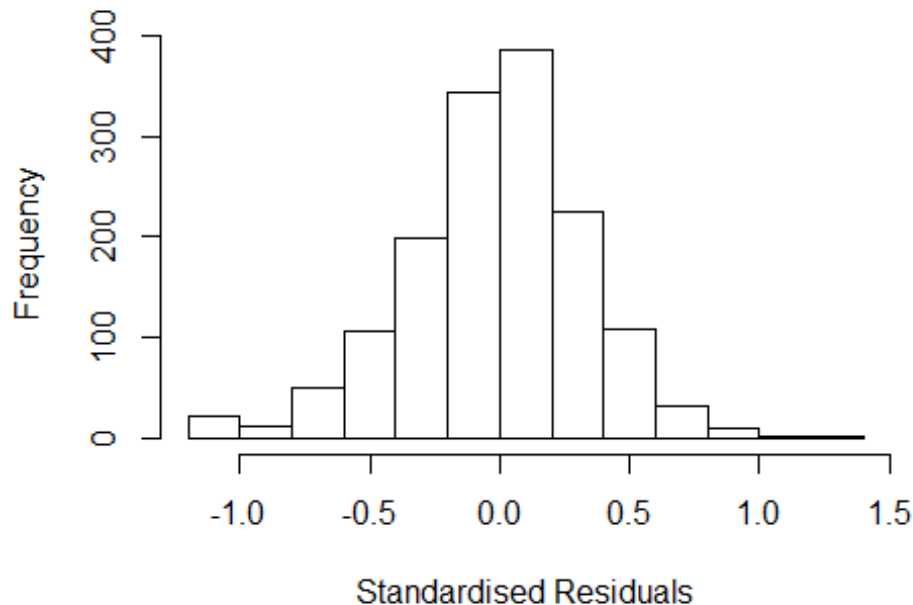
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.14805 -0.21237 0.00457 0.20326 1.26293
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11859 0.03813 29.34 <2e-16 ***
## FirstAuthorFemale1 -0.00125 0.02002 -0.06 0.950
## LastAuthorFemale1 0.02369 0.02873 0.82 0.410
## Year1997 0.03072 0.04929 0.62 0.533
## Year1998 -0.02848 0.05471 -0.52 0.603
## Year1999 -0.08651 0.05024 -1.72 0.085 .
## Year2000 -0.12380 0.06443 -1.92 0.055 .
## Year2001 0.02104 0.05498 0.38 0.702
## Year2002 -0.10759 0.05244 -2.05 0.040 *
## Year2003 -0.10985 0.05199 -2.11 0.035 *
## Year2004 -0.02748 0.05767 -0.48 0.634
## Year2005 -0.04221 0.05221 -0.81 0.419
```

```

## Year2006      -0.03981    0.04950   -0.80    0.421
## Year2007      -0.03068    0.05087   -0.60    0.547
## Year2008      -0.04677    0.05086   -0.92    0.358
## Year2009      -0.05820    0.04841   -1.20    0.229
## Year2010      -0.01816    0.04823   -0.38    0.707
## Year2011      -0.03751    0.05041   -0.74    0.457
## Year2012      -0.05684    0.06064   -0.94    0.349
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.0131, Adjusted R-squared:  0.00112
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1368 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0813 0.8610 0.9520 0.8910 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.68e-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



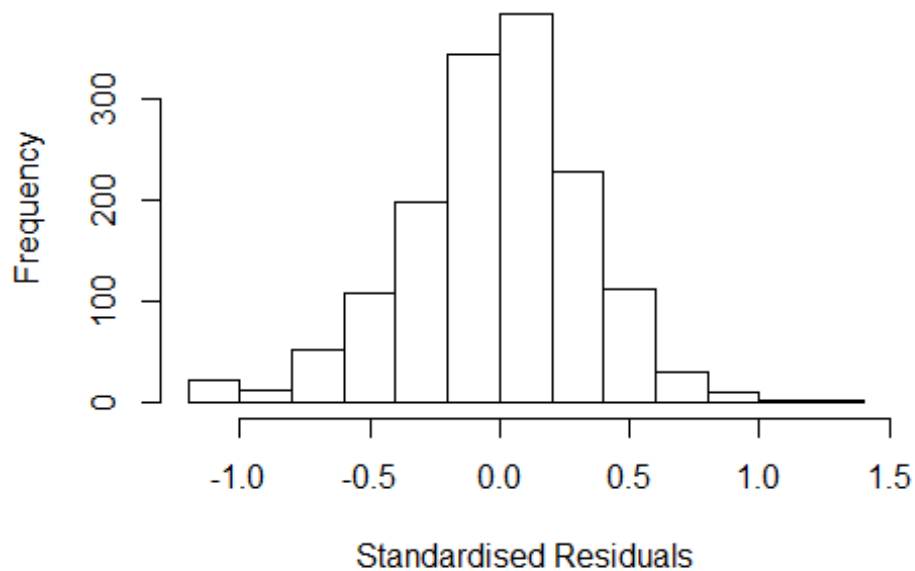
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15286 -0.21447 0.00569 0.20149 1.26000
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.120242 0.037946 29.52 <2e-16 ***
## FirstAuthorFemale1 0.000467 0.020118 0.02 0.981
## Year1997 0.032147 0.049292 0.65 0.514
## Year1998 -0.029718 0.054646 -0.54 0.587
## Year1999 -0.085749 0.050272 -1.71 0.088 .
## Year2000 -0.120999 0.064406 -1.88 0.060 .
## Year2001 0.021270 0.054972 0.39 0.699
## Year2002 -0.107228 0.052310 -2.05 0.041 *
## Year2003 -0.110636 0.051935 -2.13 0.033 *
## Year2004 -0.025656 0.057761 -0.44 0.657
## Year2005 -0.042111 0.052120 -0.81 0.419
## Year2006 -0.038927 0.049446 -0.79 0.431
```

```

## Year2007          -0.028637    0.050829   -0.56    0.573
## Year2008          -0.046921    0.050808   -0.92    0.356
## Year2009          -0.057095    0.048332   -1.18    0.238
## Year2010          -0.018720    0.048106   -0.39    0.697
## Year2011          -0.036241    0.050437   -0.72    0.473
## Year2012          -0.055964    0.060639   -0.92    0.356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.0126, Adjusted R-squared:  0.00126
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1368 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0824 0.8610 0.9530 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      6.68e-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 1.052 1          1.026
## Year            1.052 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
## -1.1491 -0.2124 0.0049 0.2029 1.2633
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1184 0.0379 29.51 <2e-16 ***
## LastAuthorFemale1 0.0235 0.0288 0.82 0.414
## Year1997 0.0307 0.0493 0.62 0.534
## Year1998 -0.0285 0.0547 -0.52 0.603
## Year1999 -0.0866 0.0503 -1.72 0.085 .
## Year2000 -0.1238 0.0645 -1.92 0.055 .
## Year2001 0.0210 0.0550 0.38 0.703
## Year2002 -0.1077 0.0525 -2.05 0.040 *
## Year2003 -0.1100 0.0520 -2.12 0.034 *
## Year2004 -0.0276 0.0577 -0.48 0.633
## Year2005 -0.0423 0.0523 -0.81 0.418
## Year2006 -0.0399 0.0495 -0.81 0.420
```

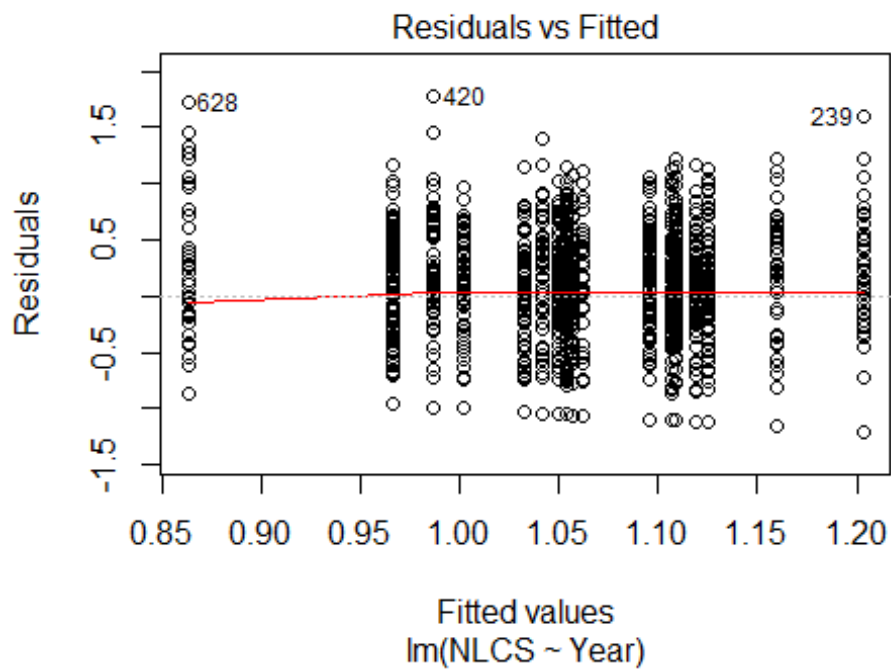
```

## Year2007          -0.0309      0.0508   -0.61    0.543
## Year2008          -0.0469      0.0509   -0.92    0.356
## Year2009          -0.0583      0.0483   -1.21    0.228
## Year2010          -0.0183      0.0483   -0.38    0.704
## Year2011          -0.0377      0.0503   -0.75    0.453
## Year2012          -0.0570      0.0604   -0.94    0.345
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.0131, Adjusted R-squared:  0.0018
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 128 weights are ~= 1. The remaining 1369 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0804 0.8610 0.9520 0.8910 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.68e-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: 1497"
## [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
## 107 89 107 123 99 87 103 114 141 137 132 173 197 137 201
## 2011 2012
## 191 165
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 58 47 49 80 54 43 68 73 99 98 84 123 146 95 136
## 2011 2012

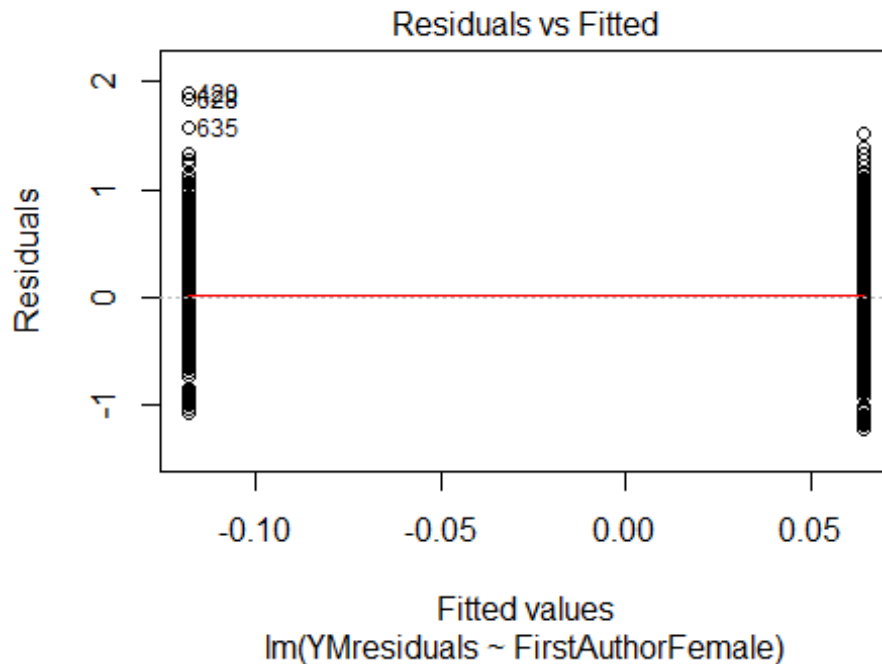
```



```
## 142 129
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 48 40 41 65 48 33 61 58 82 82 69 107 131 81 114
## 2011 2012
## 117 102
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 62, df = 16, p-value = 2e-07
```

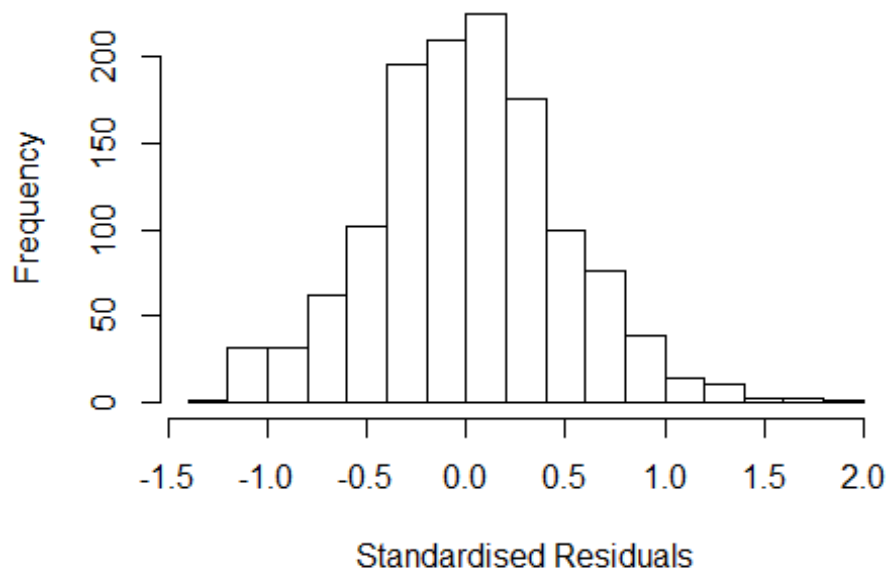


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 19, df = 1, p-value = 2e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.48030415111504"
## [1] "Male first author team size 2018 geometric mean: 3.89826797895952"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.28192209115952"
## [1] "Male last author team size 2018 geometric mean: 4.06028068271424"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.253 1      1.119
## LastAuthorFemale  1.299 1      1.140
## UniqueAuthors    1.928 4      1.085
## Year             1.685 16      1.016
```

Residuals from first and last author and team size



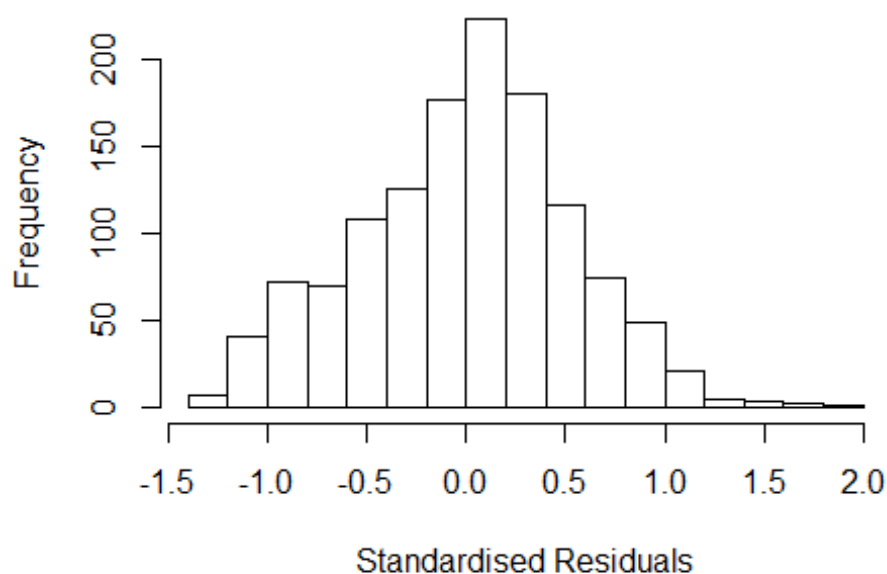
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20035 -0.29280  0.00333  0.29695  1.81769
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.38479    0.08838   4.35 1.4e-05 ***
## FirstAuthorFemale1 -0.12679    0.03145  -4.03 5.9e-05 ***
## LastAuthorFemale1 -0.01368    0.03557  -0.38  0.70
## UniqueAuthors2    0.67108    0.06322  10.61 < 2e-16 ***
## UniqueAuthors3    0.73096    0.06106  11.97 < 2e-16 ***
## UniqueAuthors4    0.74387    0.05866  12.68 < 2e-16 ***
## UniqueAuthors5    0.82414    0.04867  16.93 < 2e-16 ***
## Year1997          0.20209    0.14529   1.39  0.16
## Year1998          0.15058    0.12911   1.17  0.24
## Year1999          0.01323    0.10128   0.13  0.90
```

```

## Year2000      -0.20425    0.12491   -1.64    0.10
## Year2001      0.02719    0.12255    0.22    0.82
## Year2002      0.00223    0.10060    0.02    0.98
## Year2003     -0.05474    0.10404   -0.53    0.60
## Year2004      0.09829    0.09871    1.00    0.32
## Year2005      0.15463    0.09767    1.58    0.11
## Year2006      0.04102    0.09480    0.43    0.67
## Year2007     -0.02323    0.09205   -0.25    0.80
## Year2008      0.02470    0.09299    0.27    0.79
## Year2009      0.03211    0.09174    0.35    0.73
## Year2010     -0.02948    0.08860   -0.33    0.74
## Year2011      0.05545    0.09035    0.61    0.54
## Year2012      0.02974    0.08860    0.34    0.74
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.262, Adjusted R-squared:  0.249
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 114 weights are ~= 1. The remaining 1165 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0444 0.8490 0.9480 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      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 1.038 1      1.019
## LastAuthorFemale  1.028 1      1.014
## Year              1.059 16      1.002

```

Residuals from first and last author



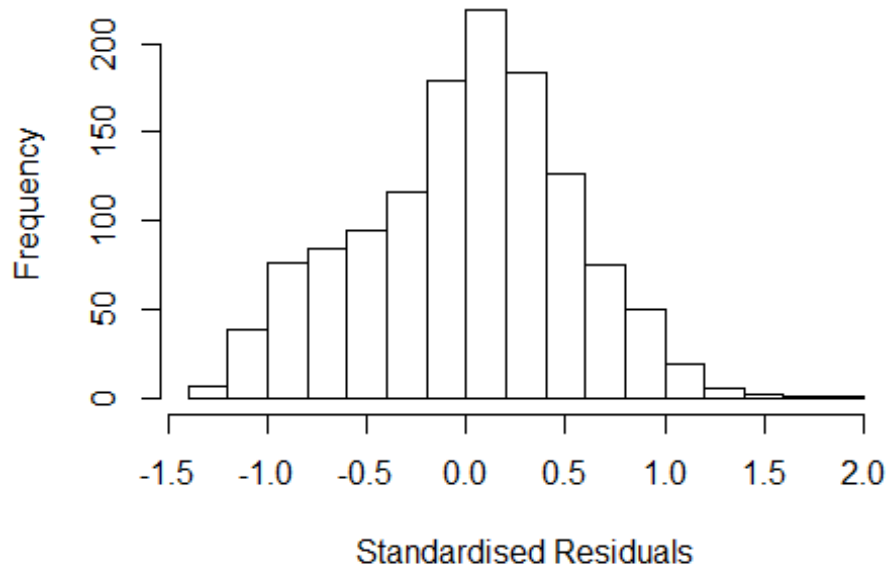
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2508 -0.3583 0.0237 0.3410 1.8739
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0160 0.1038 9.79 <2e-16 ***
## FirstAuthorFemale1 -0.1681 0.0342 -4.91 1e-06 ***
## LastAuthorFemale1 -0.0937 0.0401 -2.34 0.019 *
## Year1997 0.2627 0.1533 1.71 0.087 .
## Year1998 0.2058 0.1703 1.21 0.227
## Year1999 0.0382 0.1350 0.28 0.777
## Year2000 -0.2038 0.1588 -1.28 0.200
## Year2001 0.0869 0.1708 0.51 0.611
## Year2002 0.0754 0.1227 0.61 0.539
## Year2003 0.0724 0.1191 0.61 0.543
## Year2004 0.1548 0.1188 1.30 0.193
## Year2005 0.2347 0.1190 1.97 0.049 *
```

```

## Year2006          0.1369      0.1139      1.20      0.230
## Year2007          0.0653      0.1176      0.56      0.579
## Year2008          0.1408      0.1142      1.23      0.218
## Year2009          0.1396      0.1157      1.21      0.228
## Year2010          0.1152      0.1110      1.04      0.299
## Year2011          0.1747      0.1128      1.55      0.122
## Year2012          0.1344      0.1124      1.20      0.232
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.505
## Multiple R-squared:  0.0587, Adjusted R-squared:  0.0452
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1145 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.139  0.846  0.944  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      7.82e-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 1.028 1          1.014
## Year              1.028 16          1.001

```

Residuals from first author



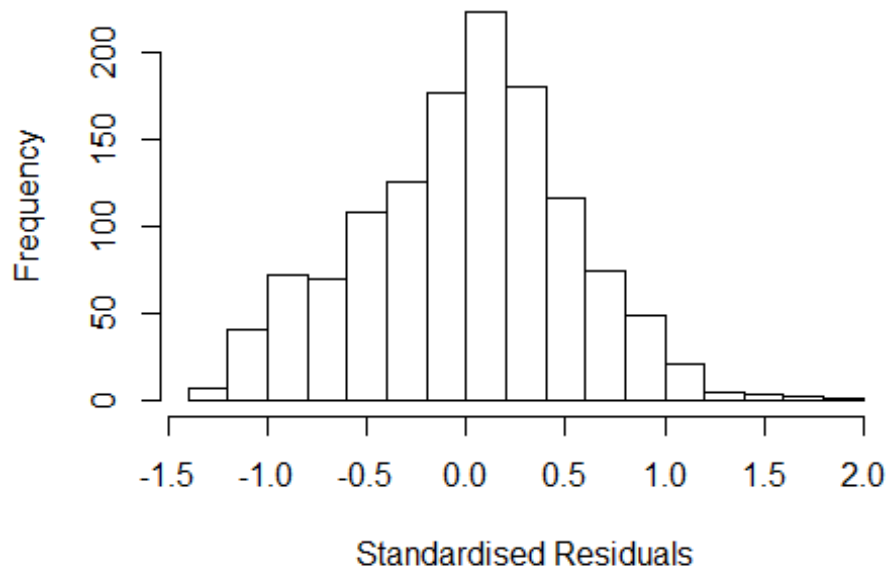
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2393 -0.3647 0.0372 0.3409 1.9079
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9957 0.1027 9.69 < 2e-16 ***
## FirstAuthorFemale1 -0.1851 0.0352 -5.26 1.7e-07 ***
## Year1997 0.2713 0.1541 1.76 0.078 .
## Year1998 0.2117 0.1725 1.23 0.220
## Year1999 0.0414 0.1357 0.31 0.760
## Year2000 -0.1986 0.1608 -1.23 0.217
## Year2001 0.0754 0.1696 0.44 0.657
## Year2002 0.0804 0.1234 0.65 0.515
## Year2003 0.0833 0.1187 0.70 0.483
## Year2004 0.1564 0.1181 1.32 0.186
## Year2005 0.2436 0.1182 2.06 0.040 *
## Year2006 0.1398 0.1139 1.23 0.220
```

```

## Year2007          0.0672      0.1166      0.58      0.564
## Year2008          0.1392      0.1142      1.22      0.223
## Year2009          0.1456      0.1148      1.27      0.205
## Year2010          0.1195      0.1105      1.08      0.280
## Year2011          0.1739      0.1125      1.55      0.122
## Year2012          0.1249      0.1123      1.11      0.266
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.512
## Multiple R-squared:  0.0514, Adjusted R-squared:  0.0386
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 123 weights are ~= 1. The remaining 1156 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.134  0.852  0.944  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      7.82e-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 1.02 1      1.010
## Year              1.02 16      1.001

```


Residuals from last author



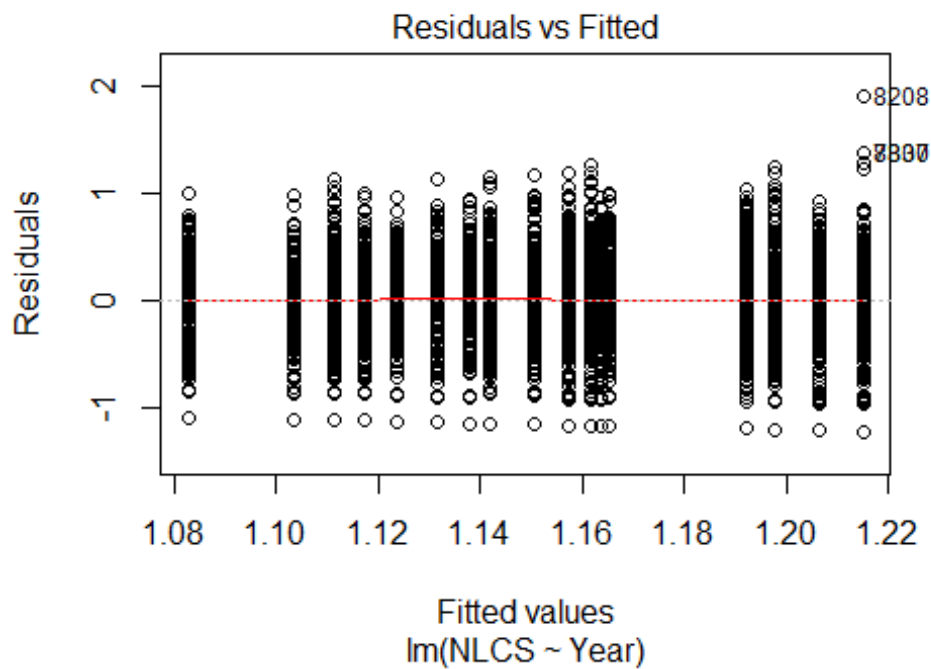
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2319 -0.3631 0.0481 0.3410 1.7611
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9747 0.1036 9.40 <2e-16 ***
## LastAuthorFemale1 -0.1306 0.0424 -3.08 0.0021 **
## Year1997 0.2572 0.1525 1.69 0.0919 .
## Year1998 0.1944 0.1714 1.13 0.2570
## Year1999 0.0242 0.1361 0.18 0.8591
## Year2000 -0.2217 0.1621 -1.37 0.1716
## Year2001 0.0645 0.1775 0.36 0.7162
## Year2002 0.0606 0.1228 0.49 0.6216
## Year2003 0.0558 0.1192 0.47 0.6401
## Year2004 0.1630 0.1192 1.37 0.1718
## Year2005 0.2278 0.1194 1.91 0.0567 .
## Year2006 0.1205 0.1149 1.05 0.2944
```

```

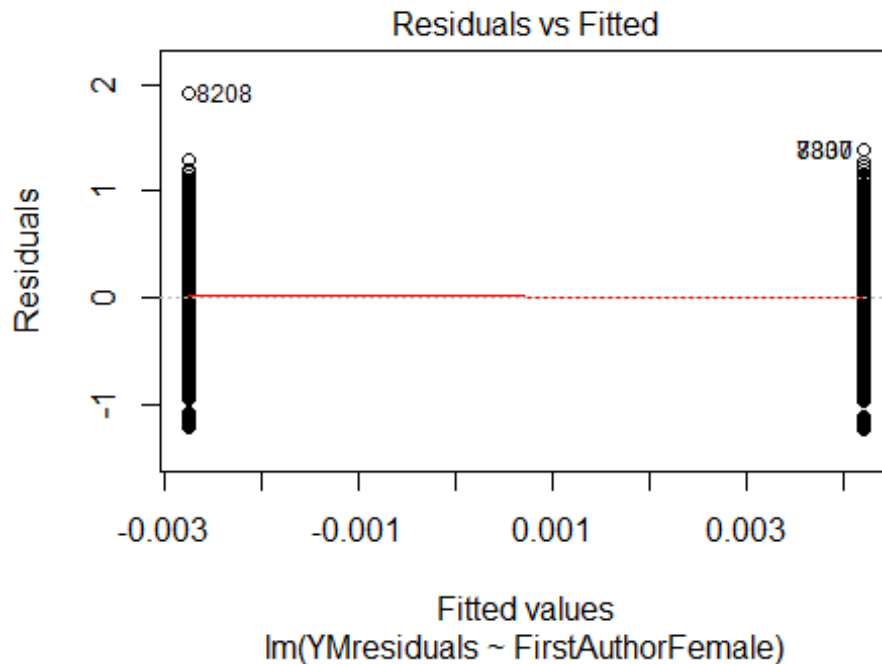
## Year2007          0.0432      0.1183      0.37      0.7148
## Year2008          0.1247      0.1151      1.08      0.2788
## Year2009          0.1468      0.1167      1.26      0.2089
## Year2010          0.1087      0.1114      0.98      0.3296
## Year2011          0.1648      0.1132      1.46      0.1456
## Year2012          0.1432      0.1135      1.26      0.2073
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.505
## Multiple R-squared:  0.0385, Adjusted R-squared:  0.0255
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 1172 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.198  0.848  0.946  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      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: 1279"
## [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
## 471 506 418 441 454 431 422 329 372 330 387 392 387 369 388
## 2011 2012
## 422 427
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 266 246 240 254 227 177 287 211 245 219 258 254 260 245 255
## 2011 2012

```

```
## 302 312
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 220 212 214 218 193 155 249 179 204 189 217 216 221 216 225
## 2011 2012
## 267 283
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 61, df = 16, p-value = 3e-07
```

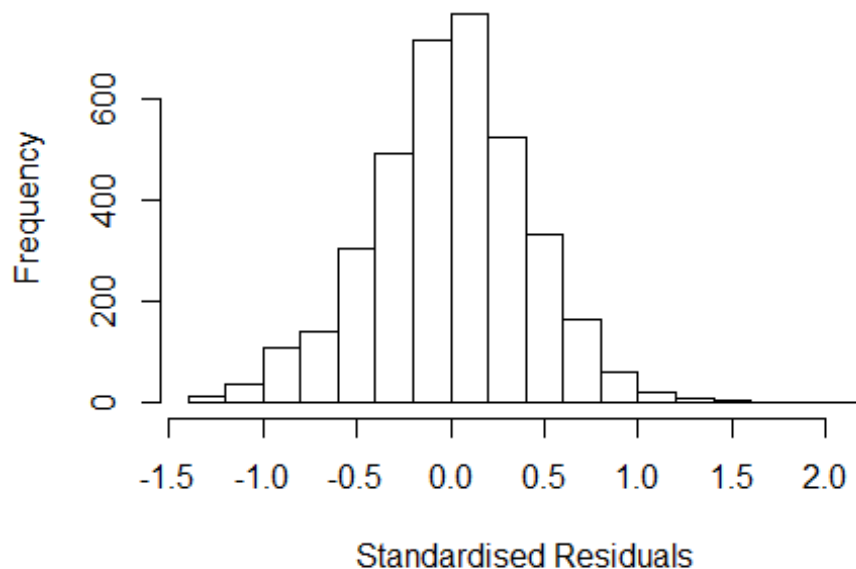


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.28, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 4.58416058302742"
## [1] "Male first author team size 2018 geometric mean: 4.37186283090553"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5900, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.38915065814104"
## [1] "Male last author team size 2018 geometric mean: 4.50429807663843"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4900, 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 1.031 1      1.015
## LastAuthorFemale  1.035 1      1.017
## UniqueAuthors    1.212 4      1.024
## Year             1.232 16      1.007
```

Residuals from first and last author and team size



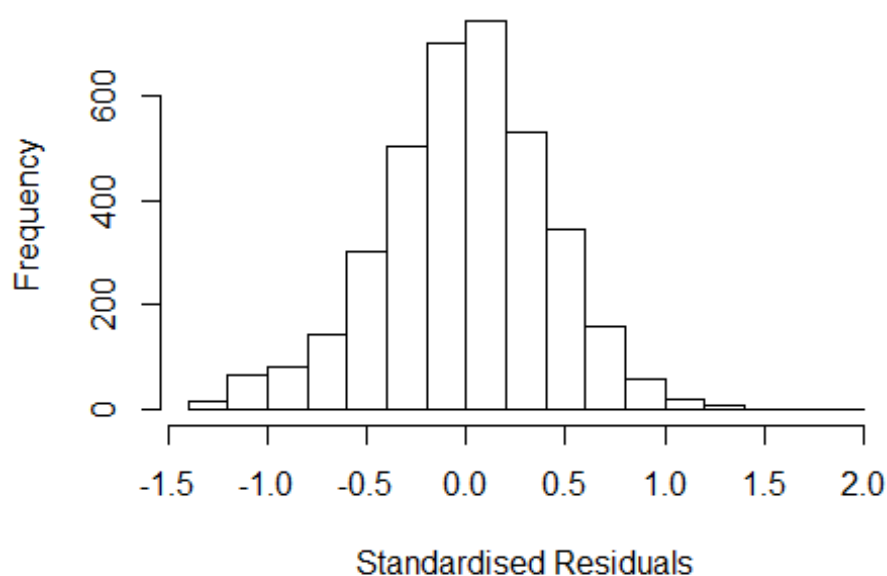
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27173 -0.26865  0.00451  0.26876  2.16149
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9758    0.0519   18.81 < 2e-16 ***
## FirstAuthorFemale1  0.0151    0.0140    1.08  0.2820
## LastAuthorFemale1 -0.0138    0.0172   -0.80  0.4210
## UniqueAuthors2    0.1903    0.0428    4.45 8.8e-06 ***
## UniqueAuthors3    0.2235    0.0423    5.28 1.3e-07 ***
## UniqueAuthors4    0.2536    0.0425    5.97 2.6e-09 ***
## UniqueAuthors5    0.3069    0.0415    7.39 1.8e-13 ***
## Year1997          0.0119    0.0486    0.25  0.8061
## Year1998         -0.0261    0.0453   -0.58  0.5651
## Year1999         -0.0674    0.0474   -1.42  0.1554
```

```

## Year2000          -0.0283      0.0465   -0.61    0.5431
## Year2001          -0.0471      0.0481   -0.98    0.3276
## Year2002          -0.0739      0.0419   -1.76    0.0781 .
## Year2003          -0.1331      0.0450   -2.96    0.0031 **
## Year2004          -0.0982      0.0423   -2.32    0.0203 *
## Year2005          -0.1063      0.0431   -2.47    0.0137 *
## Year2006          -0.1131      0.0440   -2.57    0.0103 *
## Year2007          -0.0946      0.0437   -2.17    0.0304 *
## Year2008           0.0185      0.0441    0.42    0.6753
## Year2009          -0.0686      0.0437   -1.57    0.1168
## Year2010          -0.0576      0.0443   -1.30    0.1930
## Year2011          -0.0113      0.0428   -0.26    0.7912
## Year2012          -0.0608      0.0442   -1.37    0.1693
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.0442, Adjusted R-squared:  0.0384
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 3335 is an outlier with |weight| = 0 ( < 2.7e-05);
## 321 weights are ~= 1. The remaining 3356 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.120  0.864  0.949  0.897  0.986  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           2.72e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1 1.011
## LastAuthorFemale 1.021 1 1.011
## Year 1.044 16 1.001

```

Residuals from first and last author



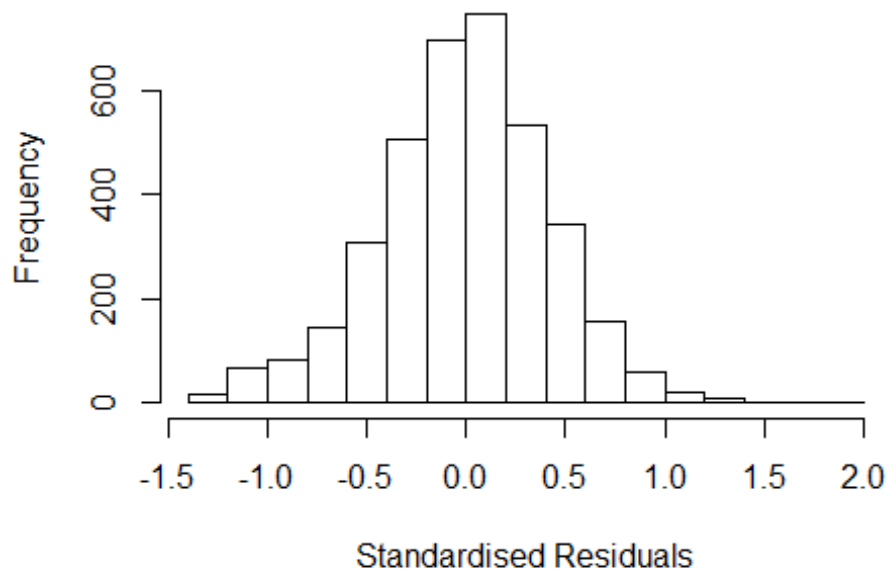
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26042 -0.27228 0.00515 0.26418 1.91391
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1966 0.0345 34.72 <2e-16 ***
## FirstAuthorFemale1 0.0204 0.0142 1.44 0.150
## LastAuthorFemale1 -0.0235 0.0174 -1.35 0.178
## Year1997 0.0153 0.0486 0.31 0.753
## Year1998 -0.0397 0.0456 -0.87 0.384
## Year1999 -0.0659 0.0477 -1.38 0.167
## Year2000 -0.0229 0.0462 -0.50 0.620
## Year2001 -0.0327 0.0475 -0.69 0.492
## Year2002 -0.0606 0.0420 -1.44 0.149
## Year2003 -0.1128 0.0445 -2.53 0.011 *
## Year2004 -0.0745 0.0426 -1.75 0.080 .
## Year2005 -0.0753 0.0425 -1.77 0.077 .
```

```

## Year2006          -0.0918      0.0440    -2.08      0.037 *
## Year2007          -0.0611      0.0431    -1.42      0.157
## Year2008           0.0434      0.0444      0.98      0.328
## Year2009          -0.0375      0.0435    -0.86      0.389
## Year2010          -0.0288      0.0436    -0.66      0.509
## Year2011           0.0154      0.0427      0.36      0.718
## Year2012          -0.0180      0.0437    -0.41      0.680
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.396
## Multiple R-squared:  0.0105, Adjusted R-squared:  0.00566
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 3335 is an outlier with |weight| = 0 ( < 2.7e-05);
## 324 weights are ~= 1. The remaining 3353 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.186  0.864  0.949   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      2.72e-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 1.023 1          1.012
## Year              1.023 16          1.001

```


Residuals from first author



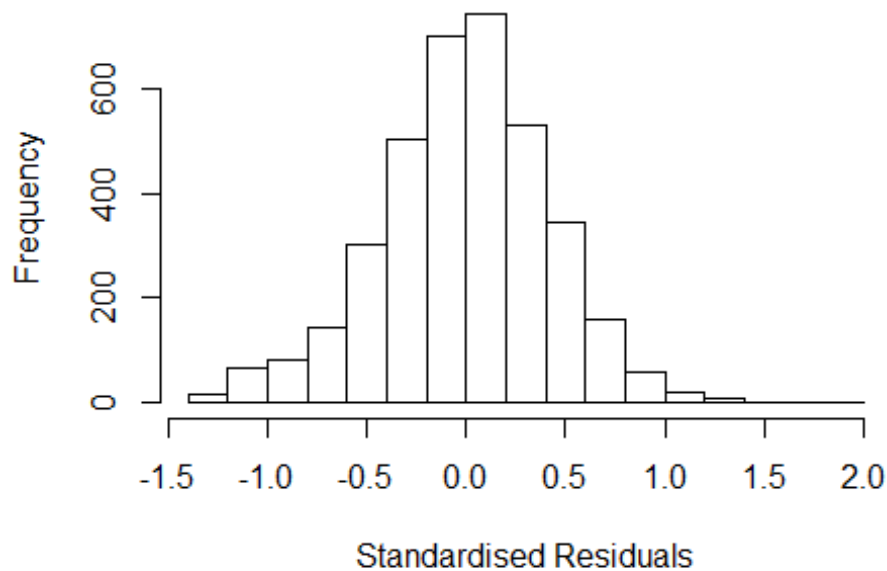
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.25439 -0.27156 0.00673 0.26412 1.91841
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1918 0.0342 34.88 <2e-16 ***
## FirstAuthorFemale1 0.0192 0.0142 1.35 0.177
## Year1997 0.0168 0.0487 0.34 0.730
## Year1998 -0.0392 0.0457 -0.86 0.391
## Year1999 -0.0649 0.0477 -1.36 0.174
## Year2000 -0.0220 0.0461 -0.48 0.633
## Year2001 -0.0334 0.0474 -0.70 0.481
## Year2002 -0.0599 0.0420 -1.43 0.154
## Year2003 -0.1127 0.0446 -2.53 0.011 *
## Year2004 -0.0736 0.0426 -1.73 0.084 .
## Year2005 -0.0757 0.0425 -1.78 0.075 .
## Year2006 -0.0923 0.0440 -2.10 0.036 *
```

```

## Year2007          -0.0610      0.0431   -1.41    0.158
## Year2008           0.0434      0.0444    0.98    0.328
## Year2009          -0.0371      0.0435   -0.85    0.394
## Year2010          -0.0286      0.0436   -0.66    0.513
## Year2011           0.0158      0.0427    0.37    0.711
## Year2012          -0.0171      0.0436   -0.39    0.695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.396
## Multiple R-squared:  0.01,   Adjusted R-squared:  0.00542
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 3335 is an outlier with |weight| = 0 ( < 2.7e-05);
## 322 weights are ~= 1. The remaining 3355 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.199  0.864  0.949  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      2.72e-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 1.021 1          1.011
## Year            1.021 16          1.001

```

Residuals from last author



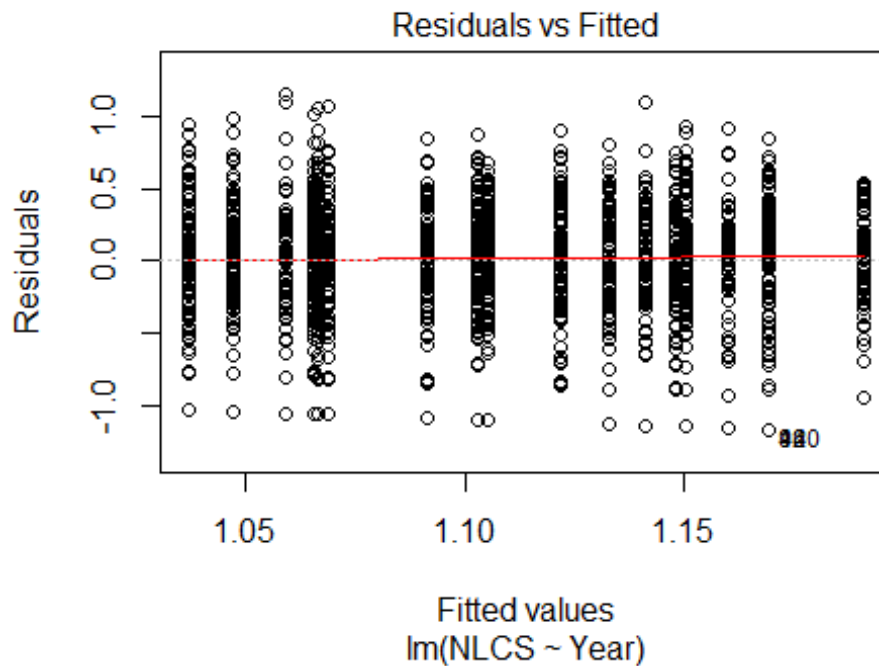
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24928 -0.27096  0.00589  0.26685  1.90534
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2019    0.0344   34.97  <2e-16 ***
## LastAuthorFemale1 -0.0220    0.0174   -1.26    0.208
## Year1997          0.0167    0.0487    0.34    0.732
## Year1998         -0.0380    0.0457   -0.83    0.406
## Year1999         -0.0649    0.0477   -1.36    0.174
## Year2000         -0.0214    0.0462   -0.46    0.644
## Year2001         -0.0309    0.0476   -0.65    0.516
## Year2002         -0.0584    0.0420   -1.39    0.165
## Year2003         -0.1107    0.0446   -2.48    0.013 *
## Year2004         -0.0714    0.0426   -1.68    0.094 .
## Year2005         -0.0720    0.0425   -1.69    0.091 .
## Year2006         -0.0866    0.0439   -1.97    0.049 *
```

```

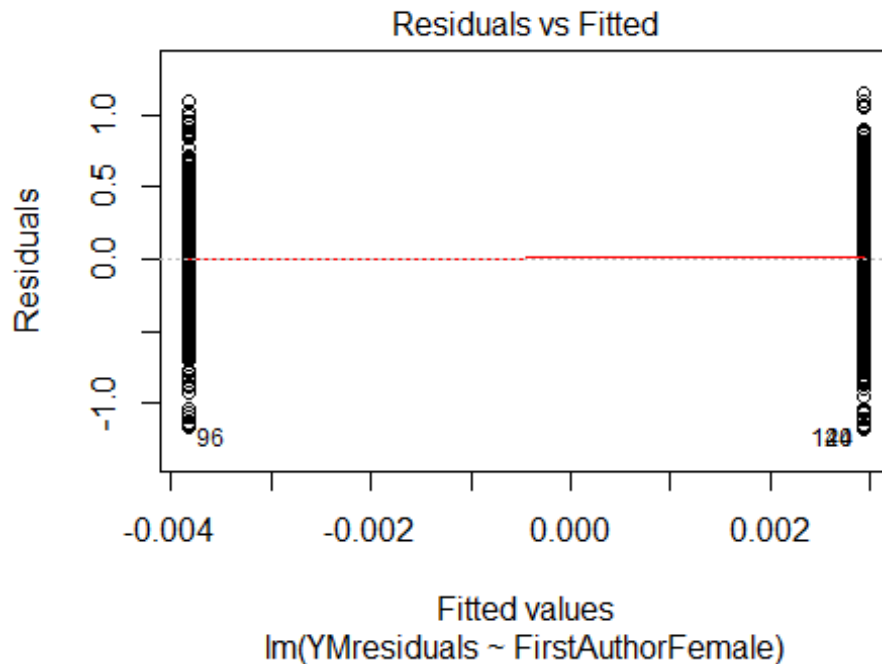
## Year2007          -0.0584      0.0432    -1.35      0.176
## Year2008           0.0473      0.0443      1.07      0.285
## Year2009          -0.0351      0.0436    -0.80      0.421
## Year2010          -0.0252      0.0436    -0.58      0.564
## Year2011           0.0187      0.0427      0.44      0.662
## Year2012          -0.0149      0.0438    -0.34      0.733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.396
## Multiple R-squared:  0.00994,    Adjusted R-squared:  0.00535
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 3335 is an outlier with |weight| = 0 ( < 2.7e-05);
## 325 weights are ~= 1. The remaining 3352 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.179  0.863  0.949  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      2.72e-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: 3678"
## [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
## 196 196 179 169 124 161 162 152 159 165 174 188 203 190 179
## 2011 2012
## 204 161
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 86 97 71 98 52 64 88 92 97 107 121 134 144 123 121

```

```
## 2011 2012
## 131 121
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 76 85 59 76 47 58 79 72 87 93 96 117 133 107 102
## 2011 2012
## 111 98
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 27, df = 16, p-value = 0.04
```

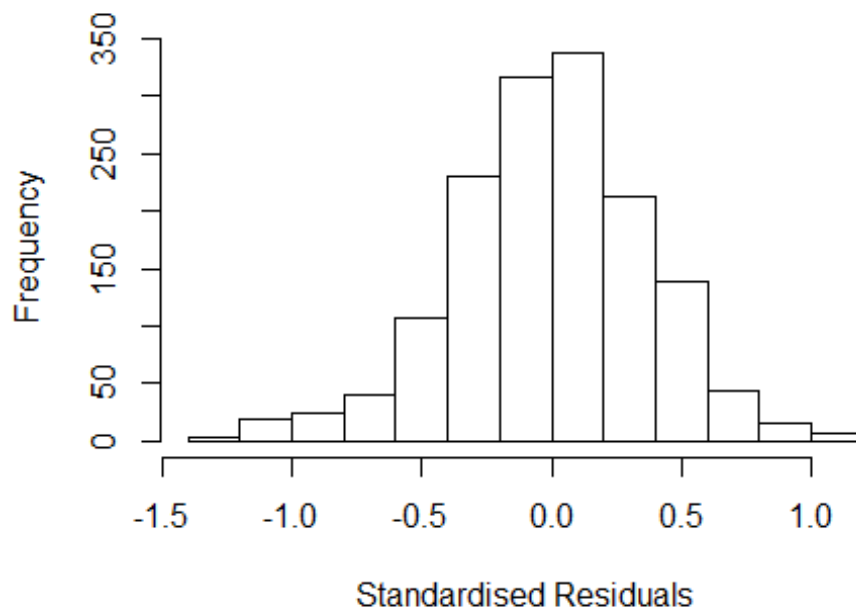


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.6, df = 1, p-value = 0.003
```



```
## [1] "Female first author team size 2018 geometric mean: 4.23746545816691"
## [1] "Male first author team size 2018 geometric mean: 3.58938433175396"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.86518561590414"
## [1] "Male last author team size 2018 geometric mean: 3.92795985962954"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.136 1      1.066
## LastAuthorFemale  1.106 1      1.051
## UniqueAuthors    1.419 4      1.045
## Year             1.578 16      1.014
```

Residuals from first and last author and team size



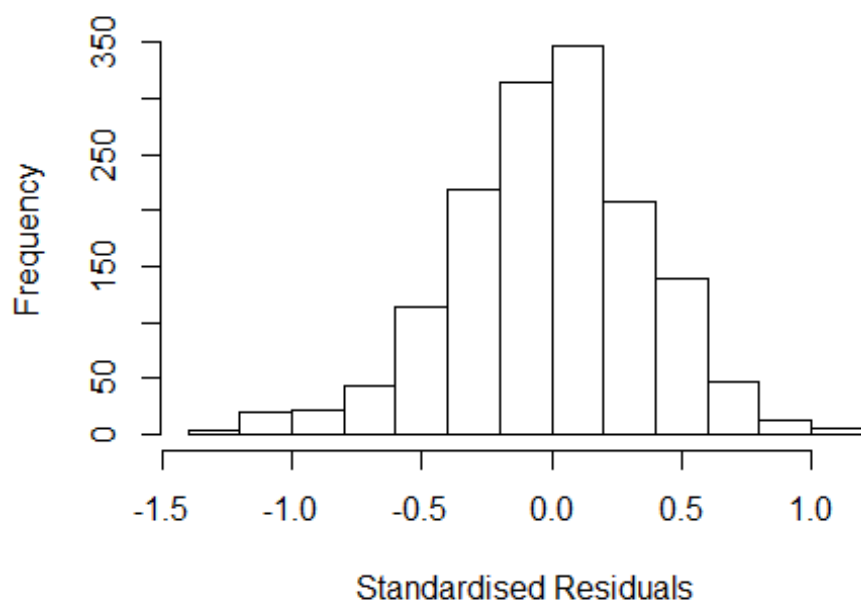
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27587 -0.23588 0.00425 0.22433 1.14602
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1333 0.0743 15.26 < 2e-16 ***
## FirstAuthorFemale1 -0.0208 0.0199 -1.04 0.29791
## LastAuthorFemale1 0.0388 0.0210 1.85 0.06507 .
## UniqueAuthors2 0.0815 0.0551 1.48 0.13969
## UniqueAuthors3 0.0976 0.0553 1.76 0.07782 .
## UniqueAuthors4 0.1426 0.0560 2.55 0.01095 *
## UniqueAuthors5 0.1574 0.0553 2.85 0.00448 **
## Year1997 -0.0798 0.0700 -1.14 0.25454
## Year1998 -0.0270 0.0672 -0.40 0.68775
## Year1999 -0.0248 0.0642 -0.39 0.69903
```

```

## Year2000          -0.0336      0.0775    -0.43    0.66476
## Year2001          -0.1837      0.0717    -2.56    0.01049 *
## Year2002          -0.1463      0.0653    -2.24    0.02533 *
## Year2003          -0.0941      0.0640    -1.47    0.14146
## Year2004          -0.1150      0.0618    -1.86    0.06320 .
## Year2005          -0.1642      0.0623    -2.64    0.00850 **
## Year2006          -0.1507      0.0621    -2.43    0.01539 *
## Year2007          -0.1760      0.0622    -2.83    0.00475 **
## Year2008          -0.2080      0.0593    -3.51    0.00046 ***
## Year2009          -0.1866      0.0617    -3.03    0.00253 **
## Year2010          -0.0982      0.0649    -1.51    0.13069
## Year2011          -0.1783      0.0625    -2.86    0.00436 **
## Year2012          -0.1444      0.0615    -2.35    0.01891 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.354
## Multiple R-squared:  0.0362, Adjusted R-squared:  0.0218
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 143 weights are ~= 1. The remaining 1353 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.167  0.869  0.950  0.897  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.68e-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 1.116 1 1.056
## LastAuthorFemale 1.064 1 1.032
## Year 1.151 16 1.004

```


Residuals from first and last author



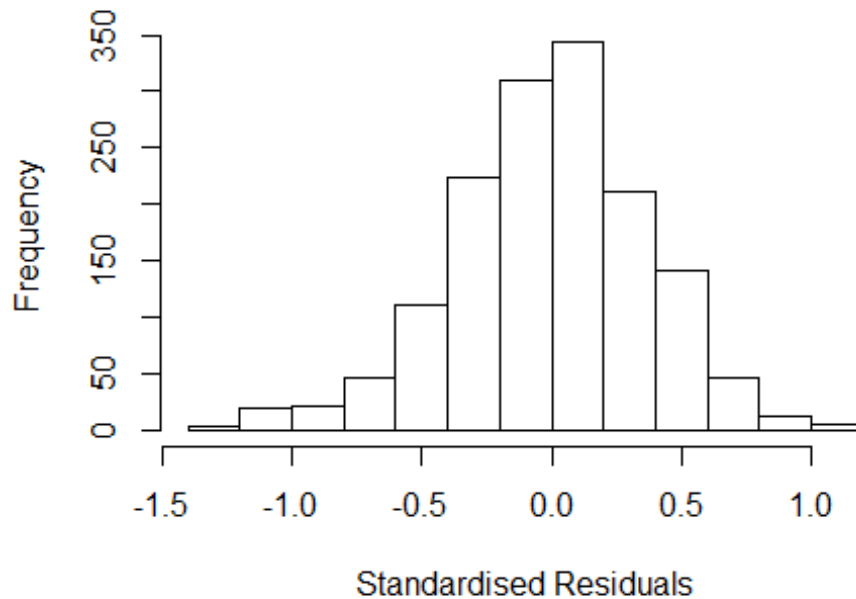
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22755 -0.25058 0.00501 0.22784 1.13623
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2276 0.0518 23.69 <2e-16 ***
## FirstAuthorFemale1 -0.0162 0.0200 -0.81 0.4169
## LastAuthorFemale1 0.0319 0.0207 1.54 0.1230
## Year1997 -0.0694 0.0700 -0.99 0.3212
## Year1998 -0.0259 0.0676 -0.38 0.7013
## Year1999 -0.0269 0.0646 -0.42 0.6776
## Year2000 -0.0344 0.0784 -0.44 0.6611
## Year2001 -0.1548 0.0717 -2.16 0.0311 *
## Year2002 -0.1312 0.0659 -1.99 0.0467 *
## Year2003 -0.0715 0.0639 -1.12 0.2633
## Year2004 -0.0898 0.0619 -1.45 0.1469
## Year2005 -0.1333 0.0616 -2.17 0.0305 *
```

```

## Year2006          -0.1362      0.0630   -2.16   0.0309 *
## Year2007          -0.1485      0.0623   -2.39   0.0172 *
## Year2008          -0.1930      0.0604   -3.19   0.0014 **
## Year2009          -0.1651      0.0620   -2.66   0.0078 **
## Year2010          -0.0720      0.0645   -1.12   0.2646
## Year2011          -0.1478      0.0621   -2.38   0.0174 *
## Year2012          -0.1161      0.0605   -1.92   0.0554 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.0229, Adjusted R-squared:  0.011
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 142 weights are ~= 1. The remaining 1354 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.209  0.867  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      6.68e-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 1.1 1          1.049
## Year              1.1 16          1.003

```

Residuals from first author



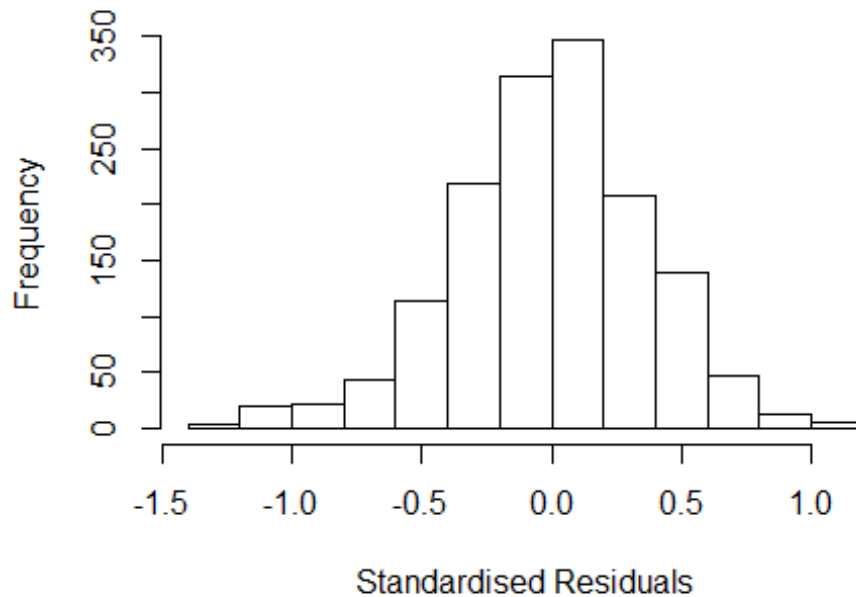
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23201 -0.25002 0.00704 0.23307 1.12781
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2320 0.0517 23.83 <2e-16 ***
## FirstAuthorFemale1 -0.0132 0.0199 -0.66 0.5079
## Year1997 -0.0674 0.0702 -0.96 0.3377
## Year1998 -0.0262 0.0677 -0.39 0.6983
## Year1999 -0.0257 0.0650 -0.40 0.6928
## Year2000 -0.0366 0.0787 -0.46 0.6422
## Year2001 -0.1508 0.0719 -2.10 0.0362 *
## Year2002 -0.1315 0.0659 -1.99 0.0462 *
## Year2003 -0.0674 0.0641 -1.05 0.2931
## Year2004 -0.0889 0.0619 -1.44 0.1511
## Year2005 -0.1310 0.0617 -2.12 0.0341 *
## Year2006 -0.1308 0.0630 -2.08 0.0379 *
```

```

## Year2007          -0.1432      0.0624   -2.30   0.0218 *
## Year2008          -0.1879      0.0605   -3.10   0.0019 **
## Year2009          -0.1600      0.0621   -2.58   0.0101 *
## Year2010          -0.0693      0.0651   -1.06   0.2871
## Year2011          -0.1431      0.0622   -2.30   0.0216 *
## Year2012          -0.1148      0.0606   -1.89   0.0585 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.357
## Multiple R-squared:  0.0213, Adjusted R-squared:  0.01
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 150 weights are ~= 1. The remaining 1346 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.209  0.867  0.946  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      6.68e-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 1.049 1          1.024
## Year            1.049 16          1.001

```

Residuals from last author



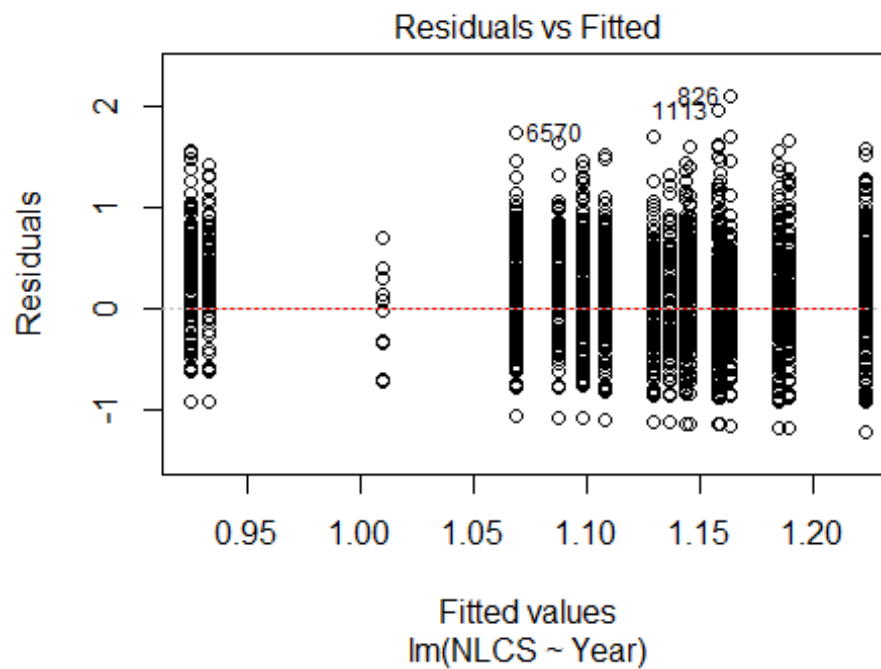
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22159 -0.24928 0.00442 0.22836 1.14434
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2216 0.0512 23.88 <2e-16 ***
## LastAuthorFemale1 0.0301 0.0206 1.46 0.1446
## Year1997 -0.0672 0.0699 -0.96 0.3364
## Year1998 -0.0261 0.0675 -0.39 0.6988
## Year1999 -0.0267 0.0648 -0.41 0.6804
## Year2000 -0.0341 0.0783 -0.44 0.6630
## Year2001 -0.1569 0.0718 -2.19 0.0290 *
## Year2002 -0.1311 0.0658 -1.99 0.0465 *
## Year2003 -0.0738 0.0638 -1.16 0.2477
## Year2004 -0.0922 0.0620 -1.49 0.1368
## Year2005 -0.1322 0.0615 -2.15 0.0318 *
## Year2006 -0.1384 0.0629 -2.20 0.0280 *
```

```

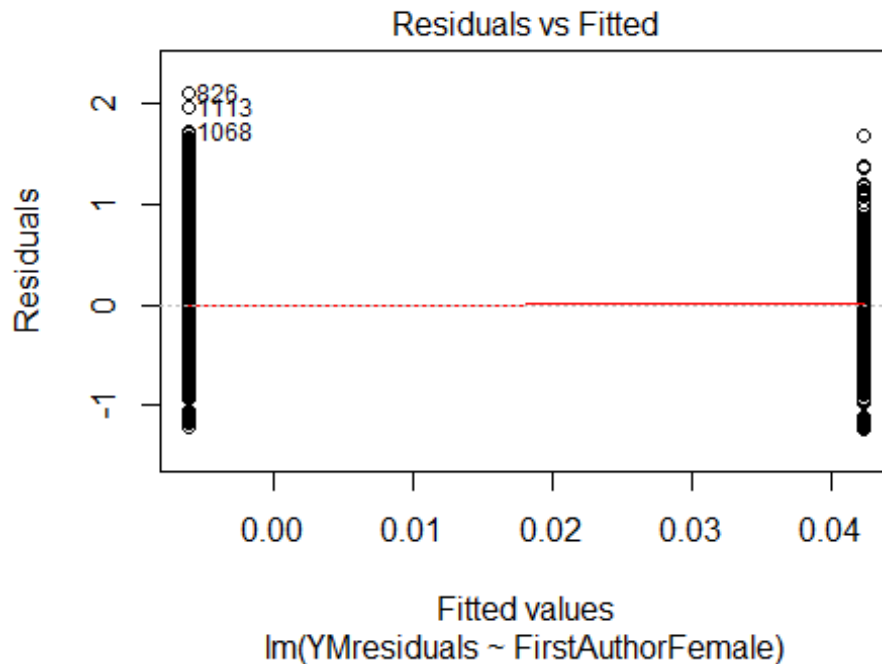
## Year2007          -0.1511      0.0621   -2.43   0.0151 *
## Year2008          -0.1948      0.0605   -3.22   0.0013 **
## Year2009          -0.1664      0.0620   -2.69   0.0073 **
## Year2010          -0.0746      0.0645   -1.16   0.2472
## Year2011          -0.1497      0.0620   -2.41   0.0159 *
## Year2012          -0.1189      0.0603   -1.97   0.0488 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.0223, Adjusted R-squared:  0.0111
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 139 weights are ~= 1. The remaining 1357 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.216  0.866  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      6.68e-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: 1496"
## [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
##   42  412  405  359  339  449  416  468  406  409  477  428  516  481  427
## 2011 2012
##   452  440
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   12  144  136  130  121  120  143  188  171  184  197  184  235  217  205
## 2011 2012

```

```
## 241 251
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 10 110 112 105 101 94 122 161 150 159 174 150 204 182 168
## 2011 2012
## 201 214
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 36, df = 16, p-value = 0.003
```

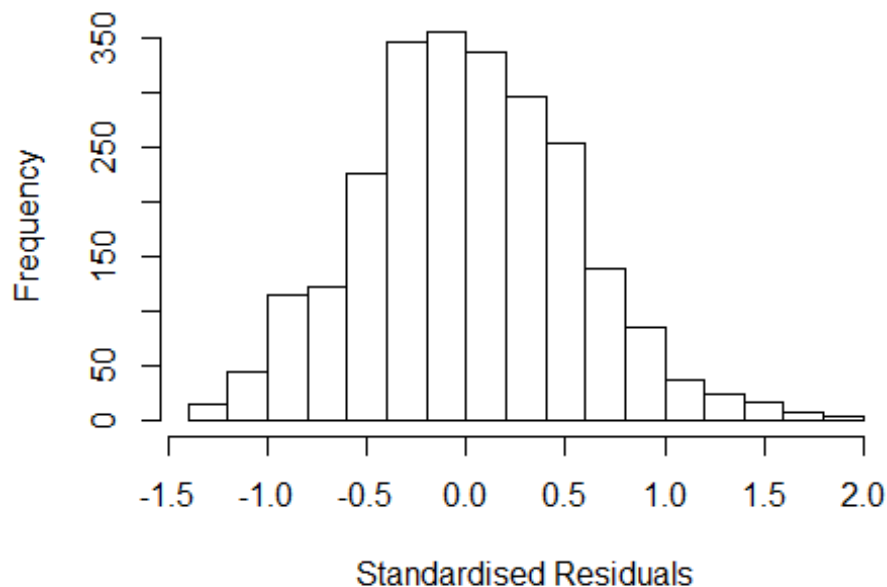


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.6, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 3.79661813769886"
## [1] "Male first author team size 2018 geometric mean: 2.48402741665777"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4400, p-value = 0.003
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.00426154974395"
## [1] "Male last author team size 2018 geometric mean: 2.61284650582829"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.069 1          1.034
## LastAuthorFemale  1.080 1          1.039
## UniqueAuthors    1.198 4          1.023
## Year              1.233 16         1.007
```


Residuals from first and last author and team size



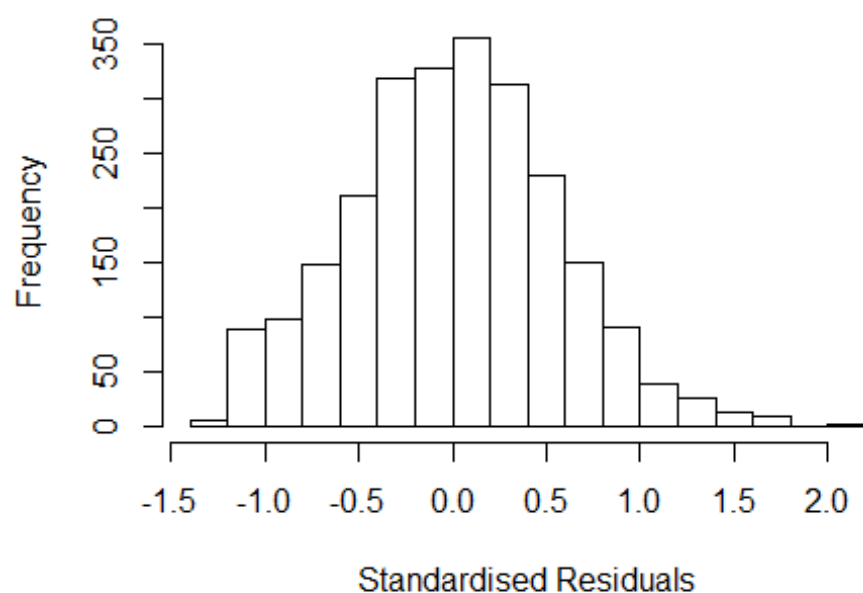
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3537 -0.3548 -0.0106 0.3694 1.9465
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6968 0.1117 6.24 5.1e-10 ***
## FirstAuthorFemale1 0.0281 0.0333 0.84 0.399
## LastAuthorFemale1 0.0607 0.0384 1.58 0.114
## UniqueAuthors2 0.2419 0.0293 8.27 2.2e-16 ***
## UniqueAuthors3 0.2489 0.0335 7.43 1.6e-13 ***
## UniqueAuthors4 0.3804 0.0449 8.47 < 2e-16 ***
## UniqueAuthors5 0.4003 0.0569 7.04 2.5e-12 ***
## Year1997 0.2942 0.1258 2.34 0.019 *
## Year1998 0.2533 0.1193 2.12 0.034 *
## Year1999 0.2228 0.1229 1.81 0.070 .
```

```

## Year2000          0.2316      0.1209      1.92      0.055 .
## Year2001          0.0714      0.1299      0.55      0.583
## Year2002          0.2798      0.1214      2.30      0.021 *
## Year2003          0.0392      0.1230      0.32      0.750
## Year2004          0.1960      0.1203      1.63      0.103
## Year2005          0.3021      0.1179      2.56      0.010 *
## Year2006          0.2482      0.1154      2.15      0.032 *
## Year2007          0.1992      0.1179      1.69      0.091 .
## Year2008          0.1593      0.1152      1.38      0.167
## Year2009          0.1745      0.1165      1.50      0.134
## Year2010          0.1929      0.1169      1.65      0.099 .
## Year2011          0.2562      0.1171      2.19      0.029 *
## Year2012          0.1493      0.1168      1.28      0.201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.529
## Multiple R-squared:  0.0774, Adjusted R-squared:  0.0689
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 178 weights are ~= 1. The remaining 2239 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.147  0.872  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      4.14e-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 1.041 1      1.020
## LastAuthorFemale  1.059 1      1.029
## Year              1.064 16      1.002

```

Residuals from first and last author



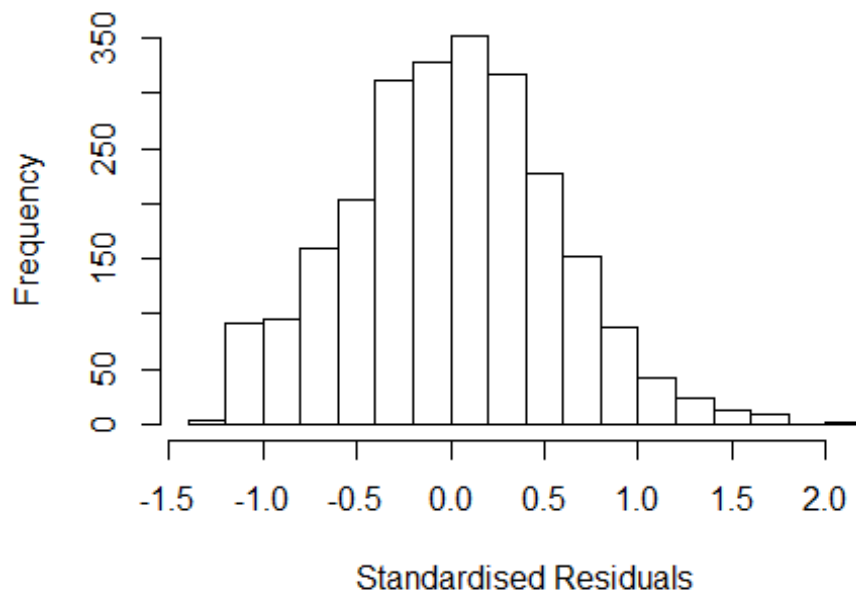
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3346 -0.3591 0.0108 0.3665 2.1396
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9225 0.1450 6.36 2.4e-10 ***
## FirstAuthorFemale1 0.0651 0.0334 1.95 0.051 .
## LastAuthorFemale1 0.0788 0.0387 2.04 0.042 *
## Year1997 0.2683 0.1580 1.70 0.090 .
## Year1998 0.1989 0.1518 1.31 0.190
## Year1999 0.1699 0.1554 1.09 0.274
## Year2000 0.1598 0.1539 1.04 0.299
## Year2001 0.0156 0.1603 0.10 0.922
## Year2002 0.2102 0.1537 1.37 0.171
## Year2003 -0.0337 0.1556 -0.22 0.828
## Year2004 0.1405 0.1527 0.92 0.358
## Year2005 0.2648 0.1515 1.75 0.081 .
```

```

## Year2006          0.1959      0.1494      1.31      0.190
## Year2007          0.1569      0.1512      1.04      0.300
## Year2008          0.1070      0.1491      0.72      0.473
## Year2009          0.1508      0.1506      1.00      0.317
## Year2010          0.1682      0.1513      1.11      0.266
## Year2011          0.2644      0.1508      1.75      0.080 .
## Year2012          0.1441      0.1506      0.96      0.339
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.547
## Multiple R-squared:  0.0239, Adjusted R-squared:  0.0166
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 186 weights are ~= 1. The remaining 2231 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0924 0.8680 0.9530 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      4.14e-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 1.021 1      1.010
## Year              1.021 16      1.001

```

Residuals from first author



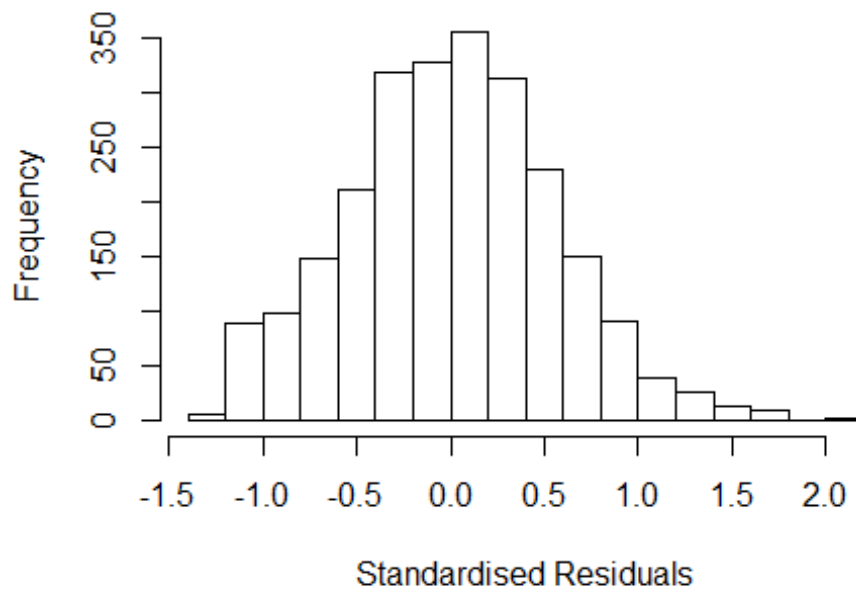
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2702 -0.3624 0.0117 0.3627 2.1375
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9321 0.1378 6.76 1.7e-11 ***
## FirstAuthorFemale1 0.0792 0.0329 2.40 0.016 *
## Year1997 0.2582 0.1514 1.71 0.088 .
## Year1998 0.1914 0.1448 1.32 0.186
## Year1999 0.1678 0.1486 1.13 0.259
## Year2000 0.1558 0.1470 1.06 0.289
## Year2001 0.0113 0.1538 0.07 0.942
## Year2002 0.2024 0.1469 1.38 0.168
## Year2003 -0.0398 0.1490 -0.27 0.789
## Year2004 0.1361 0.1459 0.93 0.351
## Year2005 0.2590 0.1447 1.79 0.074 .
## Year2006 0.1930 0.1424 1.36 0.175
```

```

## Year2007          0.1526      0.1443      1.06      0.290
## Year2008          0.1048      0.1421      0.74      0.461
## Year2009          0.1499      0.1436      1.04      0.297
## Year2010          0.1678      0.1442      1.16      0.245
## Year2011          0.2623      0.1439      1.82      0.068 .
## Year2012          0.1413      0.1435      0.98      0.325
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.548
## Multiple R-squared:  0.0222, Adjusted R-squared:  0.0153
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 2232 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0945 0.8670 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      4.14e-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 1.037 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
## -1.2953 -0.3622 0.0107 0.3647 2.1319
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.93809 0.14646 6.41 1.8e-10 ***
## LastAuthorFemale1 0.09369 0.03810 2.46 0.014 *
## Year1997 0.26349 0.15963 1.65 0.099 .
## Year1998 0.19104 0.15335 1.25 0.213
## Year1999 0.15957 0.15702 1.02 0.310
## Year2000 0.14845 0.15537 0.96 0.339
## Year2001 0.00624 0.16186 0.04 0.969
## Year2002 0.20120 0.15524 1.30 0.195
## Year2003 -0.04637 0.15712 -0.30 0.768
## Year2004 0.13018 0.15426 0.84 0.399
## Year2005 0.25898 0.15314 1.69 0.091 .
## Year2006 0.18871 0.15093 1.25 0.211
```

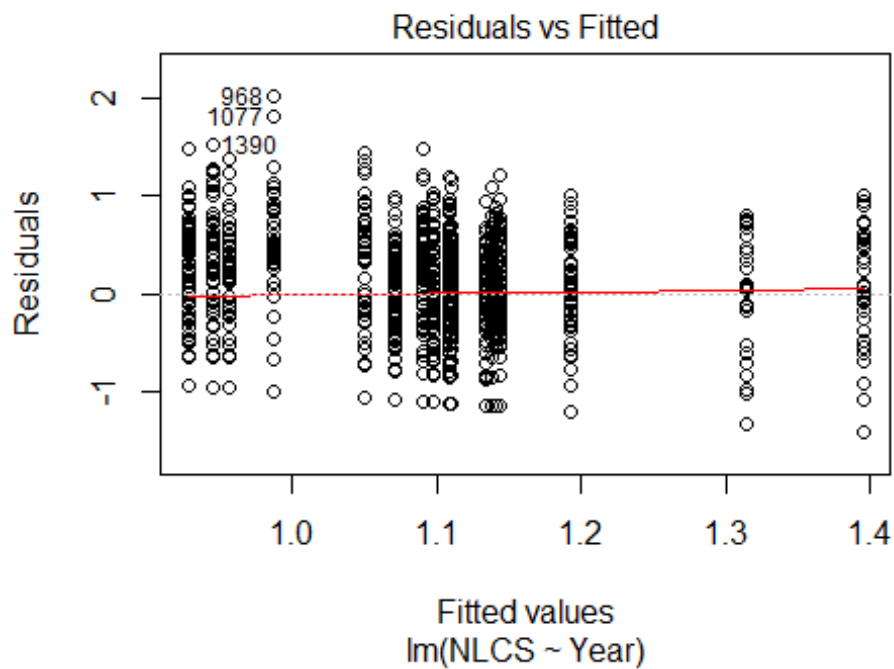
```

## Year2007      0.14840      0.15267      0.97      0.331
## Year2008      0.09887      0.15064      0.66      0.512
## Year2009      0.14221      0.15214      0.93      0.350
## Year2010      0.15858      0.15279      1.04      0.299
## Year2011      0.25608      0.15234      1.68      0.093 .
## Year2012      0.13423      0.15199      0.88      0.377
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.547
## Multiple R-squared:  0.0225, Adjusted R-squared:  0.0156
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 187 weights are ~= 1. The remaining 2230 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0953 0.8670 0.9520 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      4.14e-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: 2417"
## [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
## 157 154 123 180 169 196 204 179 181 195 223 188 180 144 107
## 2011 2012
## 136 145
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 30 39 34 55 54 58 85 72 82 78 89 83 84 65 54
## 2011 2012

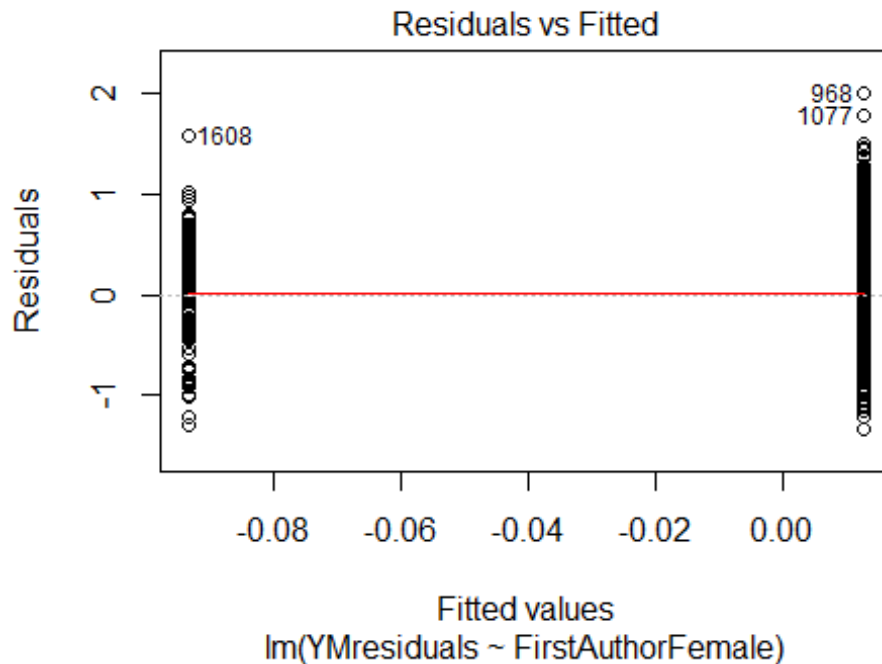
```



```
## 74 64
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 26 32 30 48 50 46 71 51 69 58 73 65 64 55 46
## 2011 2012
## 61 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 = 74, df = 16, p-value = 2e-09
```

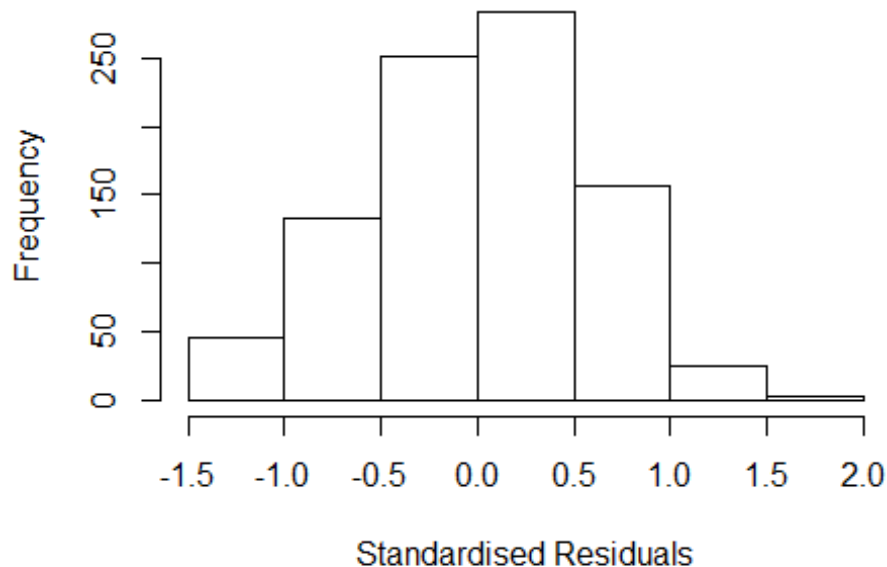


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.57, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 2.44948974278318"
## [1] "Male first author team size 2018 geometric mean: 2.23779111572103"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 210, 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: 2.27508585321903"
##
## 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 1.220 1      1.105
## LastAuthorFemale  1.076 1      1.037
## UniqueAuthors    1.833 4      1.079
## Year              1.839 16     1.019
```

Residuals from first and last author and team size



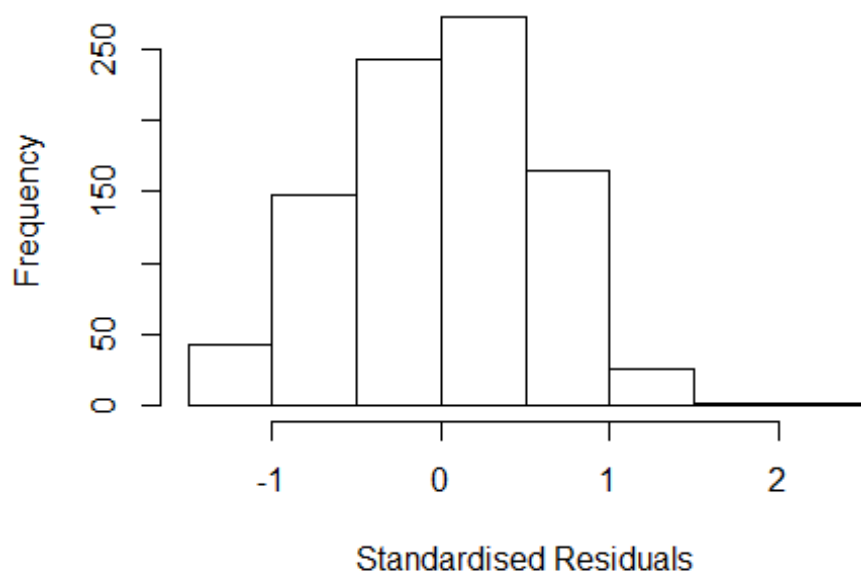
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4247 -0.4027 0.0308 0.4250 1.8986
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3138 0.1293 10.16 < 2e-16 ***
## FirstAuthorFemale1 -0.0905 0.0626 -1.44 0.14899
## LastAuthorFemale1 0.0517 0.0859 0.60 0.54705
## UniqueAuthors2 0.1974 0.0553 3.57 0.00038 ***
## UniqueAuthors3 0.2733 0.0648 4.22 2.7e-05 ***
## UniqueAuthors4 0.2014 0.0755 2.67 0.00777 **
## UniqueAuthors5 0.4381 0.0935 4.69 3.2e-06 ***
## Year1997 -0.1966 0.1567 -1.26 0.20981
## Year1998 -0.0236 0.1736 -0.14 0.89212
## Year1999 -0.5270 0.1601 -3.29 0.00104 **
```

```

## Year2000          -0.4127      0.1849    -2.23   0.02591 *
## Year2001          -0.4888      0.1985    -2.46   0.01401 *
## Year2002          -0.5087      0.1644    -3.09   0.00204 **
## Year2003          -0.4344      0.1557    -2.79   0.00537 **
## Year2004          -0.3613      0.1539    -2.35   0.01915 *
## Year2005          -0.3747      0.1469    -2.55   0.01093 *
## Year2006          -0.4258      0.1436    -2.97   0.00310 **
## Year2007          -0.3991      0.1528    -2.61   0.00917 **
## Year2008          -0.4250      0.1444    -2.94   0.00334 **
## Year2009          -0.2696      0.1476    -1.83   0.06814 .
## Year2010          -0.3559      0.1505    -2.36   0.01828 *
## Year2011          -0.2977      0.1460    -2.04   0.04176 *
## Year2012          -0.3839      0.1547    -2.48   0.01329 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.0703, Adjusted R-squared:  0.047
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 59 weights are ~= 1. The remaining 840 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.309  0.875   0.953   0.916   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.11e-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 1.119 1      1.058
## LastAuthorFemale  1.072 1      1.035
## Year              1.167 16      1.005

```

Residuals from first and last author



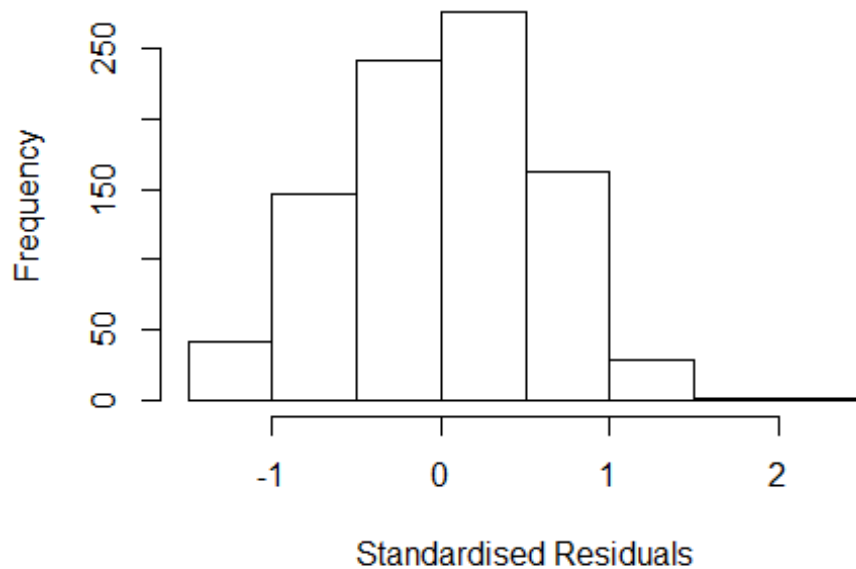
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4337 -0.3943 0.0257 0.4321 2.0472
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37343 0.13284 10.34 <2e-16 ***
## FirstAuthorFemale1 -0.02138 0.06044 -0.35 0.7236
## LastAuthorFemale1 0.07509 0.08789 0.85 0.3932
## Year1997 -0.18843 0.15971 -1.18 0.2384
## Year1998 0.00652 0.17522 0.04 0.9703
## Year1999 -0.48222 0.16390 -2.94 0.0033 **
## Year2000 -0.37629 0.19120 -1.97 0.0494 *
## Year2001 -0.42364 0.20197 -2.10 0.0362 *
## Year2002 -0.43225 0.17208 -2.51 0.0122 *
## Year2003 -0.38511 0.16328 -2.36 0.0186 *
## Year2004 -0.29380 0.16002 -1.84 0.0667 .
## Year2005 -0.22721 0.14694 -1.55 0.1224
```

```

## Year2006      -0.28973    0.14189   -2.04    0.0415 *
## Year2007      -0.30327    0.15362   -1.97    0.0487 *
## Year2008      -0.31051    0.14837   -2.09    0.0367 *
## Year2009      -0.17167    0.14978   -1.15    0.2520
## Year2010      -0.28780    0.15188   -1.89    0.0584 .
## Year2011      -0.19987    0.14713   -1.36    0.1747
## Year2012      -0.23665    0.15030   -1.57    0.1157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.035, Adjusted R-squared:  0.0153
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 833 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.228  0.871  0.948  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      1.11e-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 1.104 1      1.051
## Year      1.104 16      1.003

```

Residuals from first author



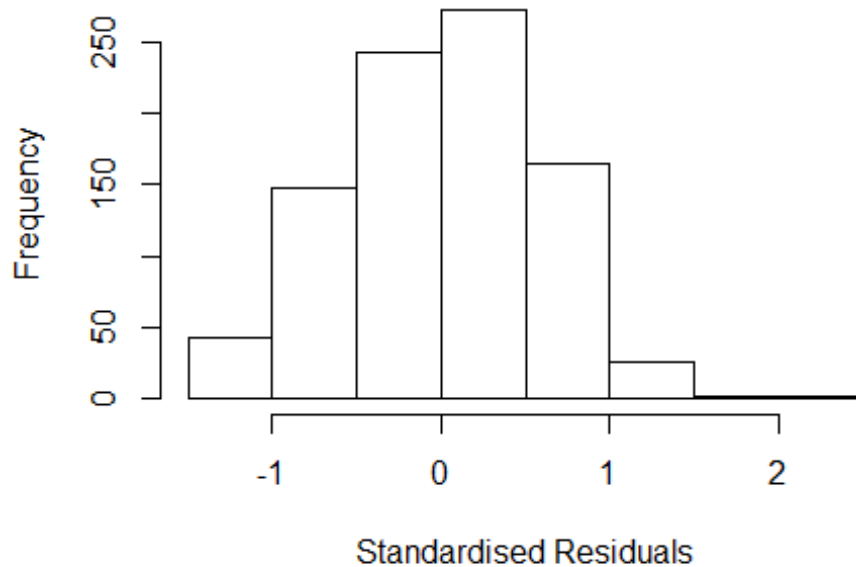
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3780 -0.3992 0.0217 0.4370 2.0461
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.38286 0.13239 10.45 <2e-16 ***
## FirstAuthorFemale1 -0.00927 0.06058 -0.15 0.8784
## Year1997 -0.19942 0.15923 -1.25 0.2108
## Year1998 0.00445 0.17489 0.03 0.9797
## Year1999 -0.48990 0.16366 -2.99 0.0028 **
## Year2000 -0.38368 0.19158 -2.00 0.0455 *
## Year2001 -0.43200 0.20233 -2.14 0.0330 *
## Year2002 -0.43778 0.17240 -2.54 0.0113 *
## Year2003 -0.39213 0.16291 -2.41 0.0163 *
## Year2004 -0.29945 0.15991 -1.87 0.0615 .
## Year2005 -0.23167 0.14658 -1.58 0.1144
## Year2006 -0.29904 0.14139 -2.11 0.0347 *
```

```

## Year2007          -0.30427    0.15383   -1.98    0.0482 *
## Year2008          -0.31506    0.14846   -2.12    0.0341 *
## Year2009          -0.17814    0.14975   -1.19    0.2345
## Year2010          -0.29687    0.15168   -1.96    0.0506 .
## Year2011          -0.20772    0.14689   -1.41    0.1577
## Year2012          -0.23884    0.15020   -1.59    0.1122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.034, Adjusted R-squared:  0.0154
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 833 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.227  0.870   0.947   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      1.11e-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 1.061 1          1.030
## Year            1.061 16          1.002

```


Residuals from last author



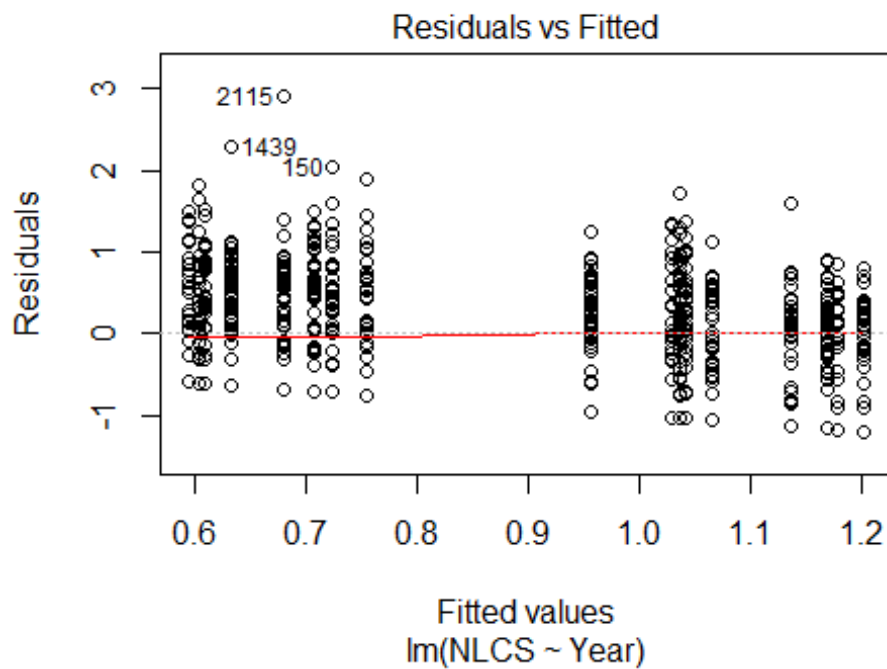
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.447 -0.400 0.024 0.428 2.049
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37076 0.13320 10.29 <2e-16 ***
## LastAuthorFemale1 0.06939 0.08748 0.79 0.4279
## Year1997 -0.18877 0.16004 -1.18 0.2385
## Year1998 0.00673 0.17572 0.04 0.9695
## Year1999 -0.48094 0.16439 -2.93 0.0035 **
## Year2000 -0.37331 0.19141 -1.95 0.0515 .
## Year2001 -0.42276 0.20338 -2.08 0.0379 *
## Year2002 -0.43012 0.17386 -2.47 0.0136 *
## Year2003 -0.38405 0.16348 -2.35 0.0190 *
## Year2004 -0.29211 0.16041 -1.82 0.0689 .
## Year2005 -0.22678 0.14716 -1.54 0.1236
## Year2006 -0.29041 0.14198 -2.05 0.0411 *
```

```

## Year2007          -0.30450      0.15350    -1.98    0.0476 *
## Year2008          -0.30822      0.14861    -2.07    0.0384 *
## Year2009          -0.17050      0.15016    -1.14    0.2565
## Year2010          -0.28788      0.15216    -1.89    0.0588 .
## Year2011          -0.20175      0.14692    -1.37    0.1701
## Year2012          -0.23579      0.15061    -1.57    0.1178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.606
## Multiple R-squared:  0.0349, Adjusted R-squared:  0.0162
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 834 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.229  0.871  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      1.11e-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: 899"
## [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
## 139 133 129 109 126 122 65 90 57 102 70 95 166 91 89
## 2011 2012
## 93 119
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 69 31 58 56 66 23 34 50 33 57 42 61 109 53 51
## 2011 2012

```

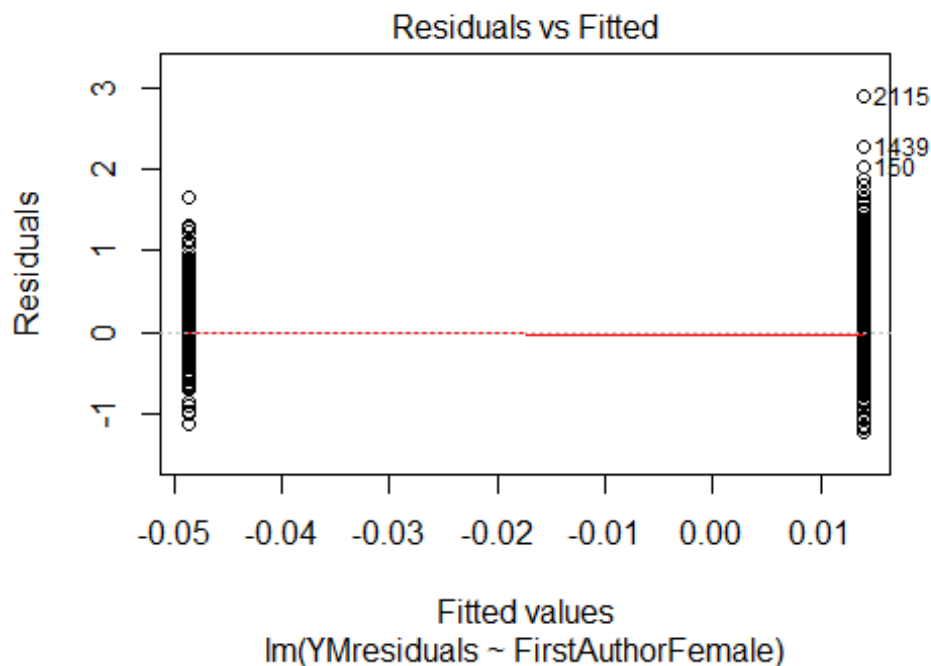
```
## 58 93
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 59 29 49 52 54 20 29 44 28 50 36 55 95 39 42
## 2011 2012
## 48 79
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.03
```



```
##
## 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: 2.79215323394551"
## [1] "Male first author team size 2018 geometric mean: 2.69714285324534"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

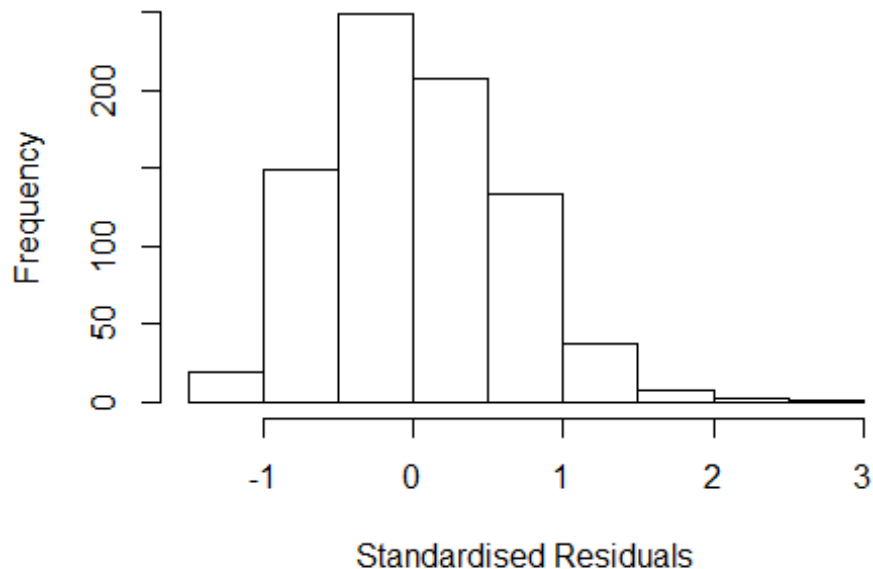
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.03393700979443"
## [1] "Male last author team size 2018 geometric mean: 2.86707225096235"

## Warning in wilcox.test.default(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 = 0.05
## 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 1.471  1      1.213
## LastAuthorFemale  1.433  1      1.197
## UniqueAuthors    1.696  4      1.068
## Year              1.870 16      1.020
```

Residuals from first and last author and team size



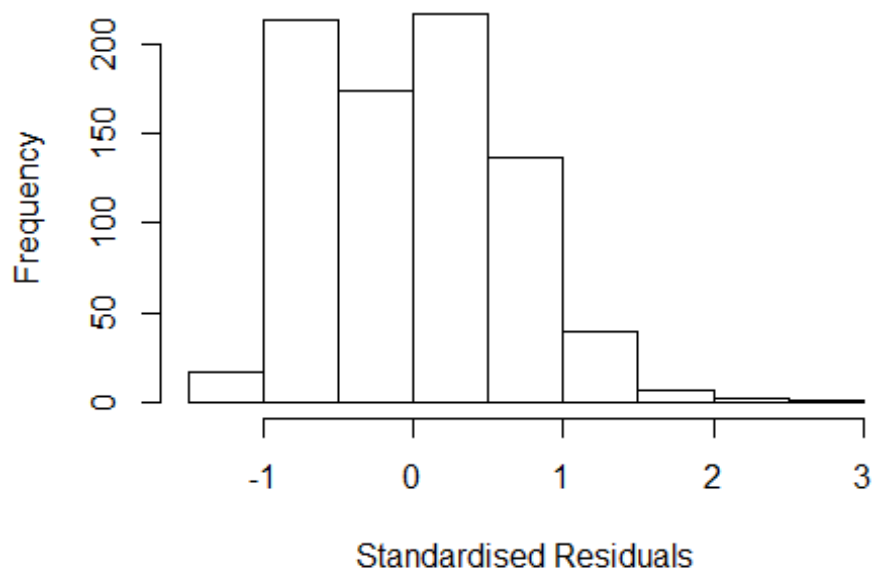
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2115 84055211743 3.581 2012      2208      2      2.841
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3016 -0.4093 -0.0205  0.4240  2.8414
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4093    0.0992   4.13 4.1e-05 ***
## FirstAuthorFemale1 -0.0474    0.0580  -0.82 0.41444
## LastAuthorFemale1 -0.0523    0.0653  -0.80 0.42326
## UniqueAuthors2     0.3518    0.0628   5.61 2.9e-08 ***
## UniqueAuthors3     0.3900    0.0719   5.43 7.6e-08 ***
## UniqueAuthors4     0.3814    0.0975   3.91 0.00010 ***
## UniqueAuthors5     0.4557    0.0983   4.64 4.1e-06 ***
## Year1997          0.3033    0.1774   1.71 0.08775 .
## Year1998         -0.1176    0.1248  -0.94 0.34636
## Year1999         -0.0758    0.1288  -0.59 0.55631
```

```

## Year2000          0.0271      0.1362      0.20  0.84263
## Year2001          0.4620      0.1703      2.71  0.00680 **
## Year2002          0.3295      0.1407      2.34  0.01944 *
## Year2003         -0.1582      0.1356     -1.17  0.24383
## Year2004          0.4480      0.1299      3.45  0.00059 ***
## Year2005         -0.0301      0.1333     -0.23  0.82114
## Year2006          0.4211      0.1283      3.28  0.00108 **
## Year2007          0.2787      0.1351      2.06  0.03945 *
## Year2008         -0.1142      0.1077     -1.06  0.28959
## Year2009          0.4888      0.1129      4.33  1.7e-05 ***
## Year2010          0.2197      0.1369      1.60  0.10893
## Year2011          0.2161      0.1373      1.57  0.11593
## Year2012         -0.0598      0.1120     -0.53  0.59354
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.606
## Multiple R-squared:  0.185, Adjusted R-squared:  0.162
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 803 is an outlier with |weight| = 0 ( < 0.00012);
## 59 weights are ~= 1. The remaining 748 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.179  0.877  0.947  0.910  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          1.24e-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 1.345 1          1.160
## LastAuthorFemale  1.337 1          1.156
## Year              1.212 16          1.006

```

Residuals from first and last author



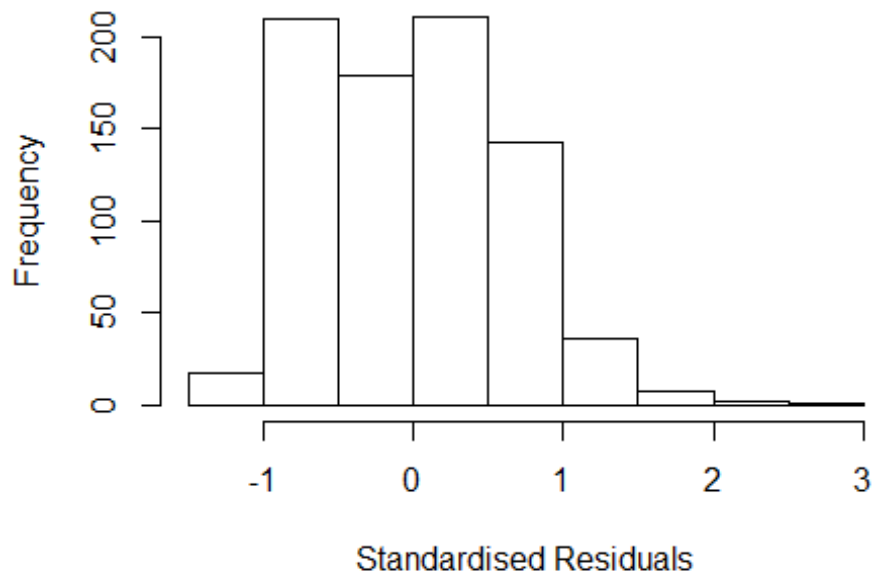
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2115 84055211743 3.581 2012      2208      2      2.946
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2095 -0.5598 -0.0115  0.4679  2.9458
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.63141    0.10053   6.28 5.6e-10 ***
## FirstAuthorFemale1 0.00220    0.05749   0.04 0.96944
## LastAuthorFemale1 -0.05584    0.06698  -0.83 0.40470
## Year1997         0.35568    0.18347   1.94 0.05290 .
## Year1998        -0.07166    0.13165  -0.54 0.58637
## Year1999        -0.07106    0.13768  -0.52 0.60592
## Year2000         0.04615    0.14693   0.31 0.75353
## Year2001         0.57589    0.16266   3.54 0.00042 ***
## Year2002         0.40389    0.15162   2.66 0.00788 **
## Year2003        -0.15400    0.13771  -1.12 0.26379
## Year2004         0.56233    0.13967   4.03 6.2e-05 ***
## Year2005         0.00901    0.13857   0.06 0.94820
```

```

## Year2006          0.52197      0.13753      3.80  0.00016 ***
## Year2007          0.39865      0.14518      2.75  0.00617 **
## Year2008         -0.04594      0.11549     -0.40  0.69087
## Year2009          0.58225      0.12027      4.84  1.6e-06 ***
## Year2010          0.34298      0.14948      2.29  0.02202 *
## Year2011          0.36541      0.13837      2.64  0.00844 **
## Year2012          0.00376      0.12124      0.03  0.97526
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.137, Adjusted R-squared:  0.118
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 803 is an outlier with |weight| = 0 ( < 0.00012);
## 50 weights are ~= 1. The remaining 757 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.112  0.890   0.924   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      1.24e-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 1.115 1          1.056
## Year              1.115 16          1.003

```


Residuals from first author



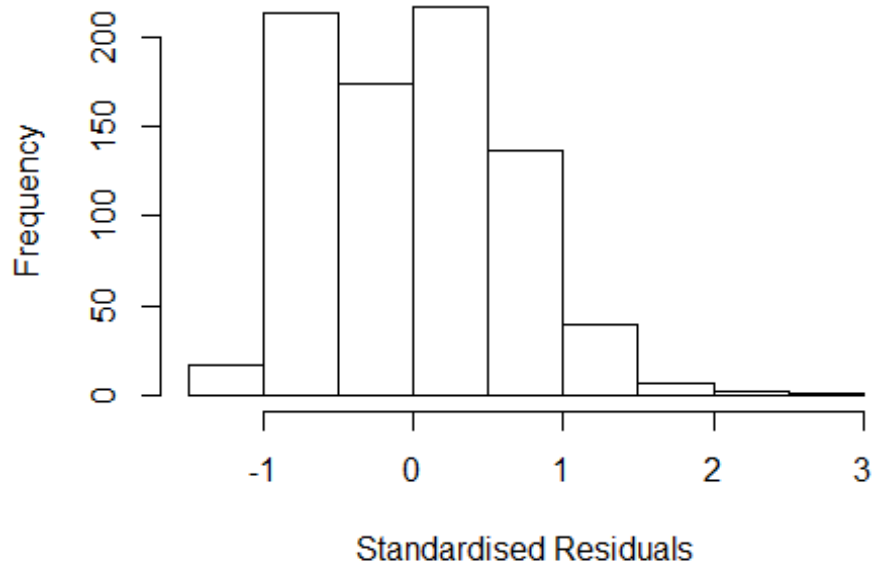
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2115 84055211743 3.581 2012      2208      2      2.946
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2089 -0.5545 -0.0295  0.4648  2.9515
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.62774    0.10019   6.27 6.1e-10 ***
## FirstAuthorFemale1 -0.01683    0.05256  -0.32  0.74897
## Year1997        0.35430    0.18419   1.92  0.05477 .
## Year1998       -0.07323    0.13133  -0.56  0.57725
## Year1999       -0.07297    0.13709  -0.53  0.59468
## Year2000        0.05000    0.14720   0.34  0.73420
## Year2001        0.57767    0.16017   3.61  0.00033 ***
## Year2002        0.40051    0.15112   2.65  0.00820 **
## Year2003       -0.15901    0.13696  -1.16  0.24599
## Year2004        0.56664    0.13880   4.08  4.9e-05 ***
## Year2005        0.00346    0.13763   0.03  0.97993
## Year2006        0.52143    0.13774   3.79  0.00017 ***
```

```

## Year2007          0.40104      0.14475      2.77  0.00573 **
## Year2008         -0.05107      0.11460     -0.45  0.65596
## Year2009          0.58116      0.11989      4.85  1.5e-06 ***
## Year2010          0.33909      0.14993      2.26  0.02399 *
## Year2011          0.36735      0.13831      2.66  0.00807 **
## Year2012          0.00175      0.12063      0.01  0.98843
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.609
## Multiple R-squared:  0.137, Adjusted R-squared:  0.118
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 803 is an outlier with |weight| = 0 ( < 0.00012);
## 55 weights are ~= 1. The remaining 752 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.111  0.890  0.926  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      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 1.111 1          1.054
## Year              1.111 16          1.003

```

Residuals from last author



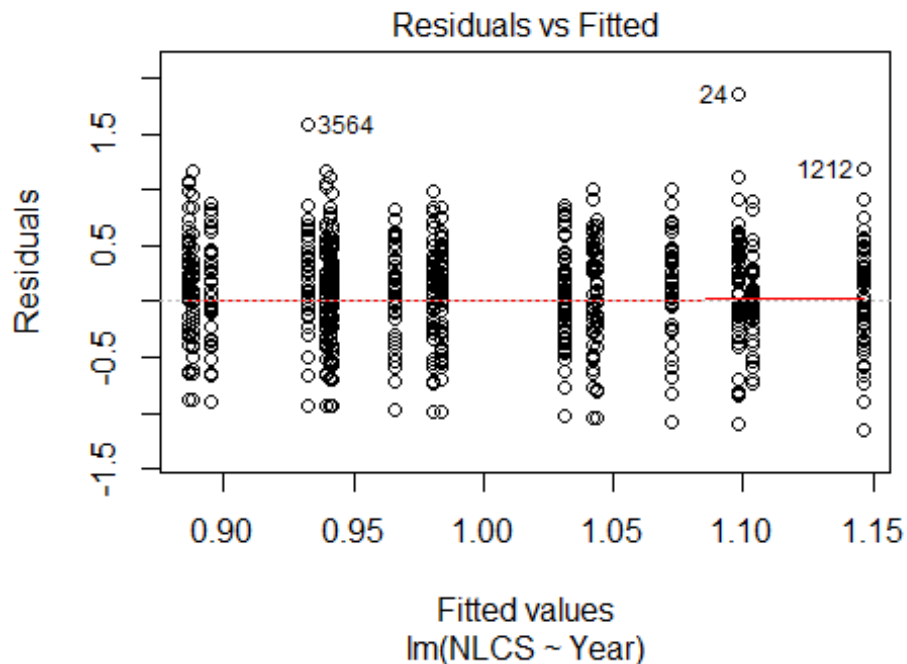
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2115 84055211743 3.581 2012      2208      2      2.946
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2076 -0.5600 -0.0122  0.4683  2.9454
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.63165    0.09961   6.34 3.8e-10 ***
## LastAuthorFemale1 -0.05487    0.06115  -0.90 0.36988
## Year1997        0.35584    0.18364   1.94 0.05302 .
## Year1998       -0.07169    0.13166  -0.54 0.58627
## Year1999       -0.07084    0.13775  -0.51 0.60720
## Year2000        0.04629    0.14720   0.31 0.75323
## Year2001        0.57591    0.16218   3.55 0.00041 ***
## Year2002        0.40392    0.15156   2.67 0.00785 **
## Year2003       -0.15389    0.13764  -1.12 0.26386
## Year2004        0.56236    0.13960   4.03 6.2e-05 ***
## Year2005        0.00910    0.13849   0.07 0.94763
## Year2006        0.52206    0.13756   3.80 0.00016 ***
```

```

## Year2007          0.39869      0.14519      2.75  0.00617 **
## Year2008         -0.04604      0.11521     -0.40  0.68953
## Year2009          0.58209      0.11989      4.86  1.5e-06 ***
## Year2010          0.34310      0.14956      2.29  0.02205 *
## Year2011          0.36574      0.13819      2.65  0.00829 **
## Year2012          0.00394      0.12134      0.03  0.97409
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.137, Adjusted R-squared:  0.119
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 803 is an outlier with |weight| = 0 ( < 0.00012);
## 50 weights are ~= 1. The remaining 757 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.111  0.890  0.924  0.907  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.24e-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: 808"
## [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
## 124 107 126 114 110 89 102 108 100 124 94 98 120 134 154
## 2011 2012
## 127 139
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 58 51 64 64 55 33 53 49 46 50 38 42 56 69 80

```

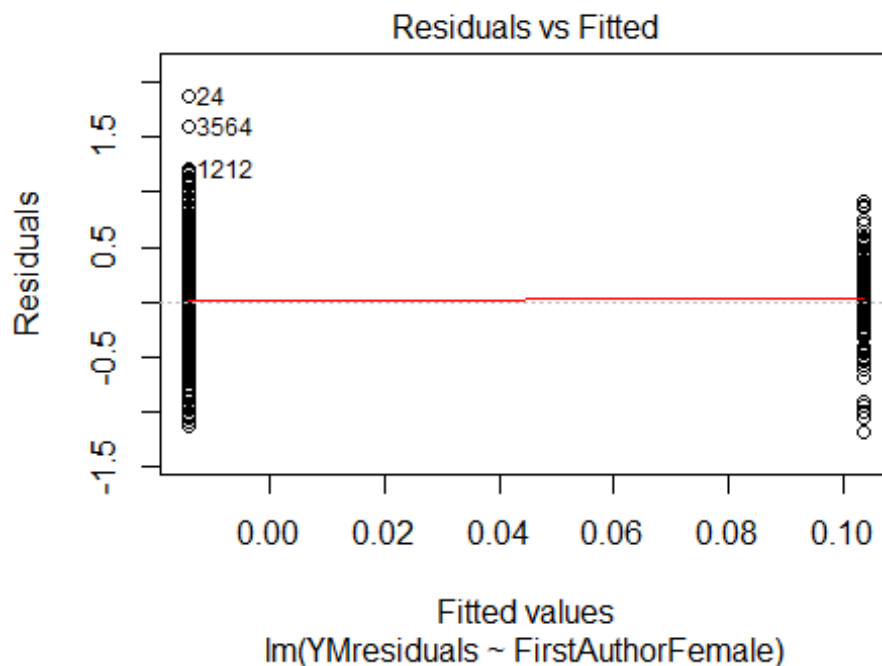
```
## 2011 2012
## 70 70
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 53 48 53 57 48 29 46 43 42 44 33 39 52 62 72
## 2011 2012
## 66 61
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.06
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.6, df = 1, p-value = 0.1
## [1] "Female first author team size 2018 geometric mean: 1.76172958987204"
## [1] "Male first author team size 2018 geometric mean: 1.95584827140533"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

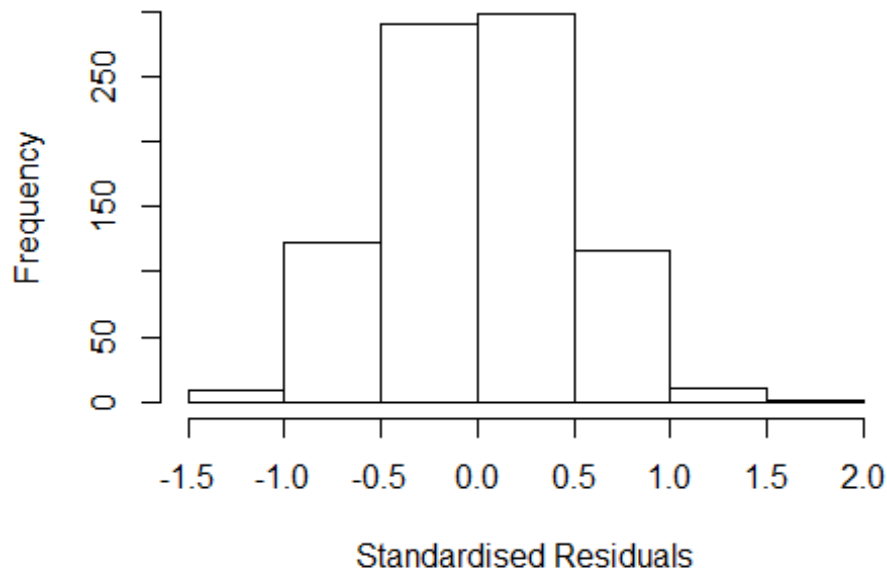
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.81712059283214"
## [1] "Male last author team size 2018 geometric mean: 1.93257017248007"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "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  1.234 1      1.111
## UniqueAuthors    1.571 4      1.058
## Year              1.838 16     1.019
```

Residuals from first and last author and team size



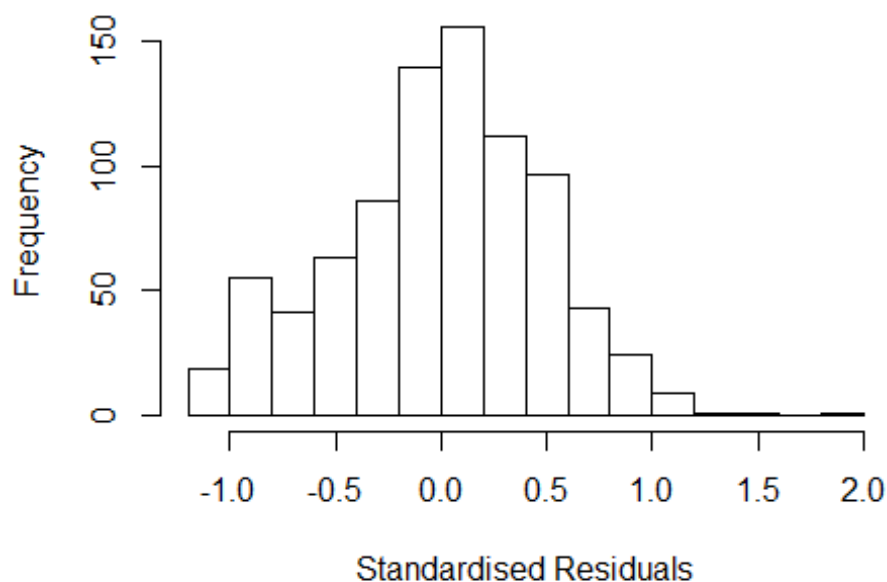
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.15269 -0.30578  0.00421  0.32111  1.99273
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.96127    0.09983   9.63 < 2e-16 ***
## FirstAuthorFemale1 0.08727    0.04728   1.85  0.06529 .
## LastAuthorFemale1 -0.11384    0.06059  -1.88  0.06062 .
## UniqueAuthors2    0.17507    0.04412   3.97  7.9e-05 ***
## UniqueAuthors3    0.24508    0.05094   4.81  1.8e-06 ***
## UniqueAuthors4    0.29063    0.07910   3.67  0.00025 ***
## UniqueAuthors5    0.34215    0.10046   3.41  0.00069 ***
## Year1997         -0.04425    0.12798  -0.35  0.72963
## Year1998         -0.21434    0.12738  -1.68  0.09282 .
## Year1999         -0.17993    0.11278  -1.60  0.11100
```

```

## Year2000          0.01593      0.12423      0.13  0.89800
## Year2001          0.00819      0.12210      0.07  0.94656
## Year2002          0.05595      0.12588      0.44  0.65682
## Year2003         -0.15987      0.12909     -1.24  0.21592
## Year2004         -0.16208      0.11305     -1.43  0.15206
## Year2005         -0.09921      0.12706     -0.78  0.43512
## Year2006         -0.02656      0.11824     -0.22  0.82232
## Year2007         -0.22062      0.11401     -1.94  0.05331 .
## Year2008         -0.12119      0.11859     -1.02  0.30710
## Year2009         -0.15007      0.11351     -1.32  0.18650
## Year2010         -0.10190      0.10541     -0.97  0.33399
## Year2011         -0.15056      0.11253     -1.34  0.18127
## Year2012         -0.17528      0.11282     -1.55  0.12067
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0766, Adjusted R-squared:  0.052
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 83 weights are ~= 1. The remaining 765 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0363 0.8650 0.9480 0.9030 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.18e-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 1.099 1      1.048
## LastAuthorFemale  1.225 1      1.107
## Year              1.334 16      1.009

```


Residuals from first and last author



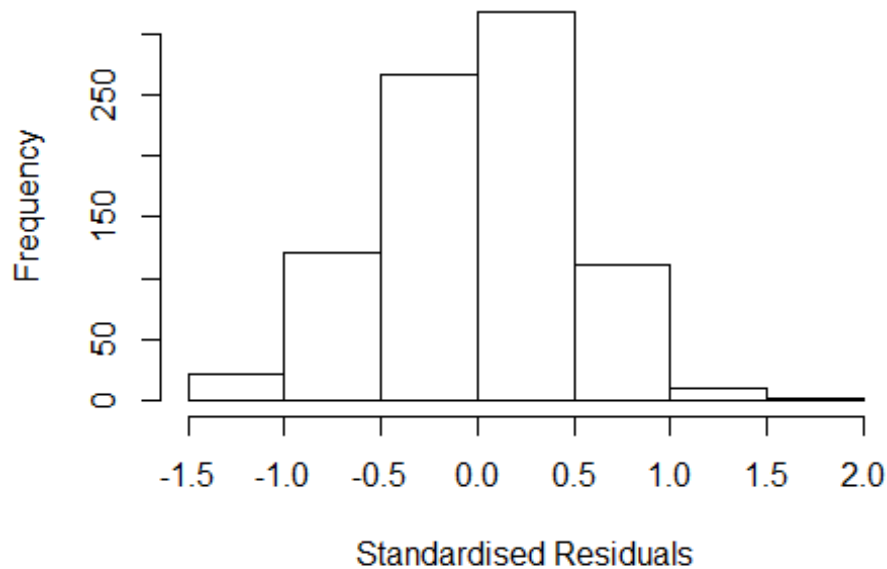
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1912 -0.3056 0.0131 0.3187 1.8693
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0847 0.0957 11.34 <2e-16 ***
## FirstAuthorFemale1 0.1484 0.0468 3.17 0.0016 **
## LastAuthorFemale1 -0.0645 0.0614 -1.05 0.2940
## Year1997 -0.0261 0.1313 -0.20 0.8424
## Year1998 -0.2132 0.1319 -1.62 0.1062
## Year1999 -0.1877 0.1167 -1.61 0.1081
## Year2000 0.0226 0.1252 0.18 0.8568
## Year2001 -0.0146 0.1260 -0.12 0.9080
## Year2002 0.0409 0.1239 0.33 0.7416
## Year2003 -0.1584 0.1299 -1.22 0.2228
## Year2004 -0.1482 0.1176 -1.26 0.2078
## Year2005 -0.1297 0.1263 -1.03 0.3048
```

```

## Year2006          -0.0192      0.1178   -0.16   0.8705
## Year2007          -0.2099      0.1171   -1.79   0.0735 .
## Year2008          -0.1327      0.1195   -1.11   0.2669
## Year2009          -0.1438      0.1144   -1.26   0.2092
## Year2010          -0.0888      0.1066   -0.83   0.4054
## Year2011          -0.1714      0.1115   -1.54   0.1245
## Year2012          -0.1481      0.1162   -1.27   0.2029
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.471
## Multiple R-squared:  0.034, Adjusted R-squared:  0.013
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 87 weights are ~= 1. The remaining 761 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0804 0.8540 0.9450 0.8980 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.104 1      1.051
## Year              1.104 16      1.003

```

Residuals from first author



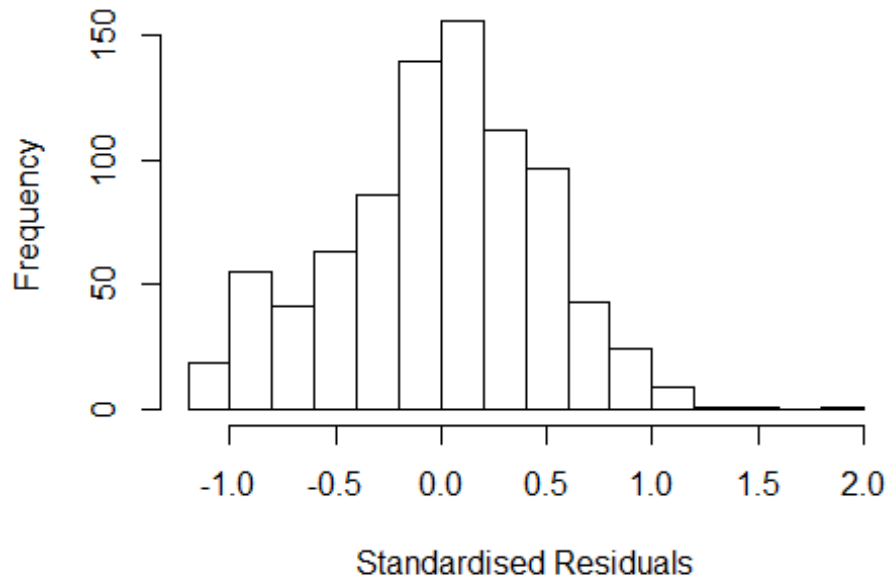
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2417 -0.2996 0.0152 0.3242 1.8707
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0833 0.0954 11.35 <2e-16 ***
## FirstAuthorFemale1 0.1410 0.0481 2.93 0.0035 **
## Year1997 -0.0293 0.1314 -0.22 0.8237
## Year1998 -0.2145 0.1317 -1.63 0.1037
## Year1999 -0.2000 0.1148 -1.74 0.0818 .
## Year2000 0.0174 0.1258 0.14 0.8899
## Year2001 -0.0262 0.1243 -0.21 0.8328
## Year2002 0.0338 0.1232 0.27 0.7840
## Year2003 -0.1637 0.1293 -1.27 0.2060
## Year2004 -0.1488 0.1173 -1.27 0.2050
## Year2005 -0.1382 0.1254 -1.10 0.2706
## Year2006 -0.0208 0.1177 -0.18 0.8595
```

```

## Year2007          -0.2084      0.1168   -1.79   0.0746 .
## Year2008          -0.1336      0.1193   -1.12   0.2633
## Year2009          -0.1412      0.1142   -1.24   0.2170
## Year2010          -0.0930      0.1060   -0.88   0.3807
## Year2011          -0.1745      0.1112   -1.57   0.1168
## Year2012          -0.1544      0.1158   -1.33   0.1828
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0329, Adjusted R-squared:  0.0131
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 773 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0821 0.8560 0.9480 0.9000 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      1.18e-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 1.24 1          1.114
## Year            1.24 16          1.007

```

Residuals from last author



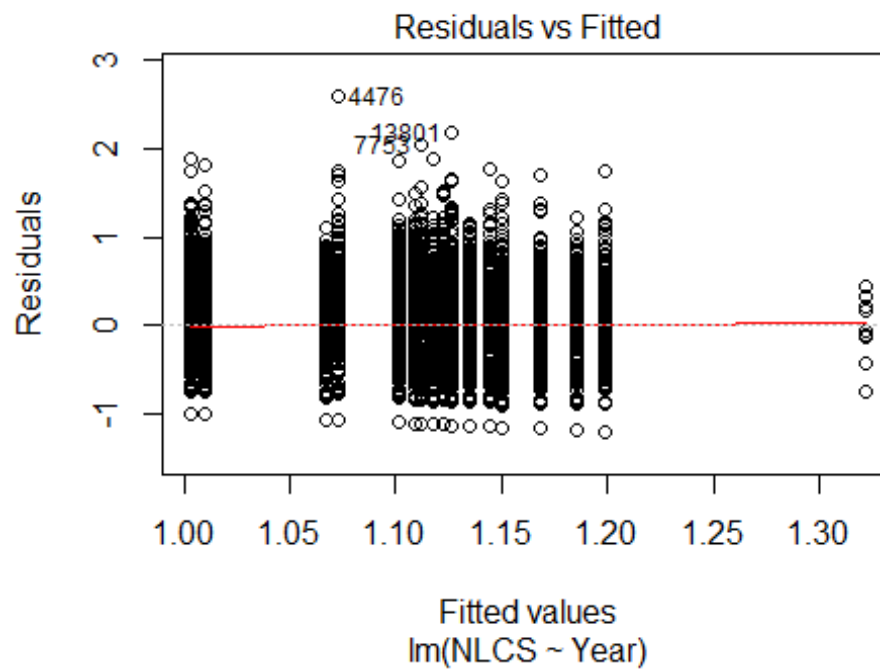
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1443 -0.3192 0.0189 0.3231 1.8599
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09415 0.09262 11.81 <2e-16 ***
## LastAuthorFemale1 -0.04128 0.06241 -0.66 0.509
## Year1997 -0.02128 0.12802 -0.17 0.868
## Year1998 -0.20488 0.12992 -1.58 0.115
## Year1999 -0.18209 0.11445 -1.59 0.112
## Year2000 0.01853 0.12282 0.15 0.880
## Year2001 -0.01859 0.12311 -0.15 0.880
## Year2002 0.05016 0.12067 0.42 0.678
## Year2003 -0.16580 0.12742 -1.30 0.194
## Year2004 -0.14263 0.11564 -1.23 0.218
## Year2005 -0.13476 0.12445 -1.08 0.279
## Year2006 -0.00181 0.11567 -0.02 0.988
```

```

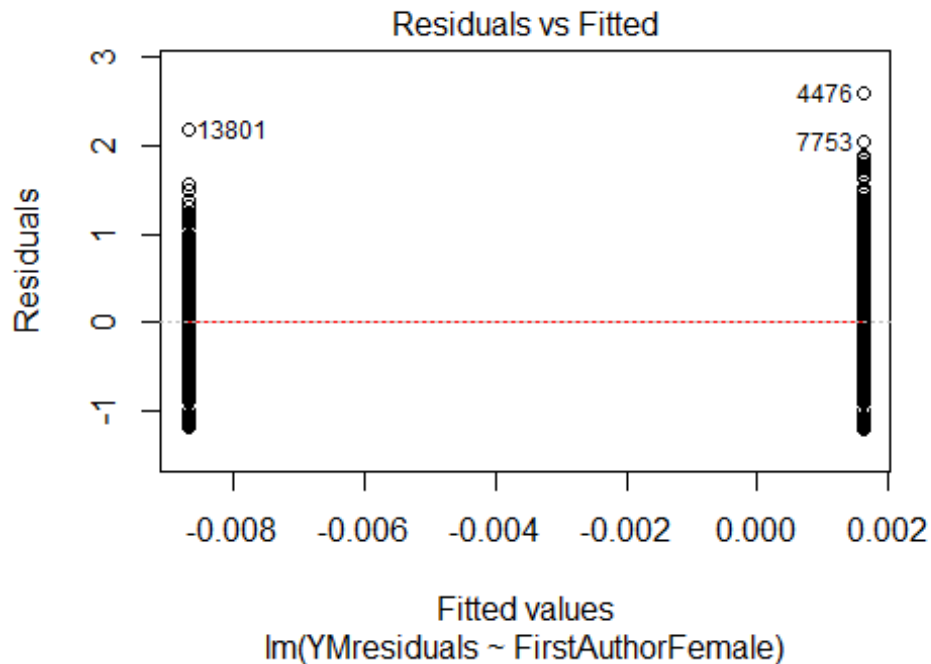
## Year2007          -0.19107      0.11547    -1.65      0.098 .
## Year2008          -0.12432      0.11837    -1.05      0.294
## Year2009          -0.12698      0.11092    -1.14      0.253
## Year2010          -0.07273      0.10370    -0.70      0.483
## Year2011          -0.16349      0.10939    -1.49      0.135
## Year2012          -0.13952      0.11390    -1.22      0.221
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.474
## Multiple R-squared:  0.0248, Adjusted R-squared:  0.00478
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 768 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0899 0.8560 0.9450 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      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: 848"
## [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
## 793 821 740 34 727 746 806 741 717 780 890 821 949 1090 1016
## 2011 2012
## 1129 1056
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 184 202 191 10 210 179 264 268 269 301 336 343 370 473 440
## 2011 2012

```

```
## 536 503
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 144 164 156 6 168 136 215 211 222 241 260 278 280 361 335
## 2011 2012
## 401 376
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 97, df = 16, p-value = 1e-13
```

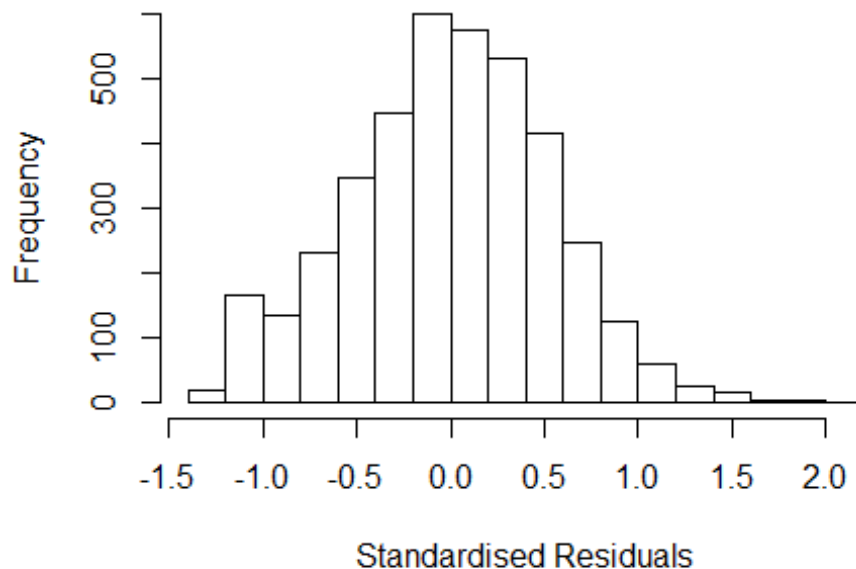


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.8, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 2.85698855084374"
## [1] "Male first author team size 2018 geometric mean: 2.69729298411934"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8200, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.68789407792665"
## [1] "Male last author team size 2018 geometric mean: 2.72636651381558"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6600, 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 1.056 1          1.027
## LastAuthorFemale  1.026 1          1.013
## UniqueAuthors    1.167 4          1.019
## Year             1.172 16          1.005
```


Residuals from first and last author and team size



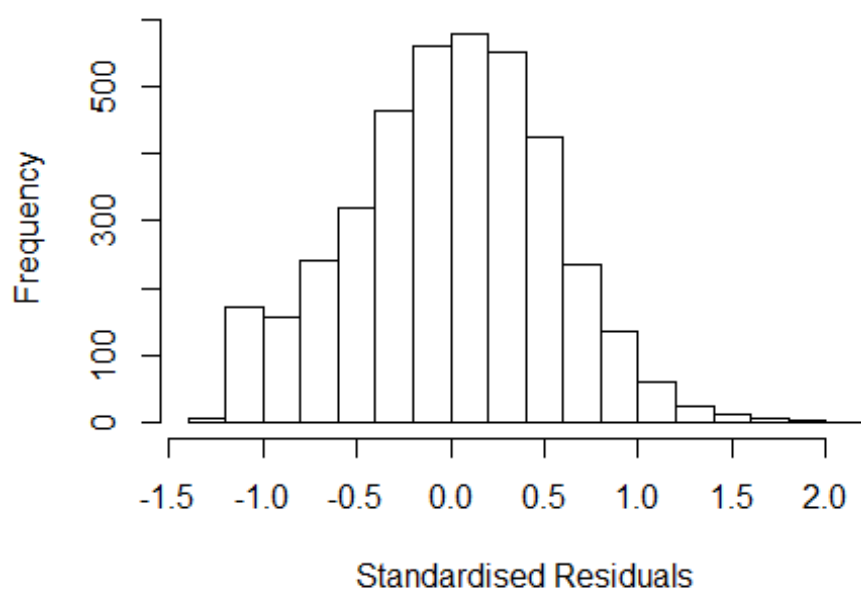
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2596 -0.3530  0.0104  0.3674  2.1103
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.14913    0.04698   24.46  < 2e-16 ***
## FirstAuthorFemale1 -0.01511    0.02312   -0.65  0.51340
## LastAuthorFemale1  0.04319    0.02946    1.47  0.14274
## UniqueAuthors2     0.08300    0.02833    2.93  0.00341 **
## UniqueAuthors3     0.05285    0.02979    1.77  0.07611 .
## UniqueAuthors4     0.08813    0.03370    2.62  0.00894 **
## UniqueAuthors5     0.18613    0.03631    5.13  3.1e-07 ***
## Year1997          -0.03767    0.06279   -0.60  0.54855
## Year1998          -0.02567    0.05884   -0.44  0.66265
## Year1999           0.00771    0.19683    0.04  0.96877
```

```

## Year2000      -0.01575    0.06187   -0.25  0.79900
## Year2001      -0.20858    0.07669   -2.72  0.00656 **
## Year2002      -0.23119    0.06450   -3.58  0.00034 ***
## Year2003      -0.23687    0.06097   -3.89  0.00010 ***
## Year2004      -0.15429    0.06112   -2.52  0.01162 *
## Year2005      -0.09803    0.05375   -1.82  0.06829 .
## Year2006      -0.08052    0.05509   -1.46  0.14389
## Year2007      -0.09927    0.05603   -1.77  0.07651 .
## Year2008      -0.13400    0.05166   -2.59  0.00953 **
## Year2009      -0.08762    0.05226   -1.68  0.09368 .
## Year2010      -0.10746    0.05118   -2.10  0.03581 *
## Year2011      -0.11953    0.05043   -2.37  0.01782 *
## Year2012      -0.13478    0.05099   -2.64  0.00824 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.53
## Multiple R-squared:  0.0193, Adjusted R-squared:  0.0138
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 342 weights are ~= 1. The remaining 3612 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0771 0.8690 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          2.53e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1 1.016
## LastAuthorFemale 1.020 1 1.010
## Year 1.037 16 1.001

```

Residuals from first and last author



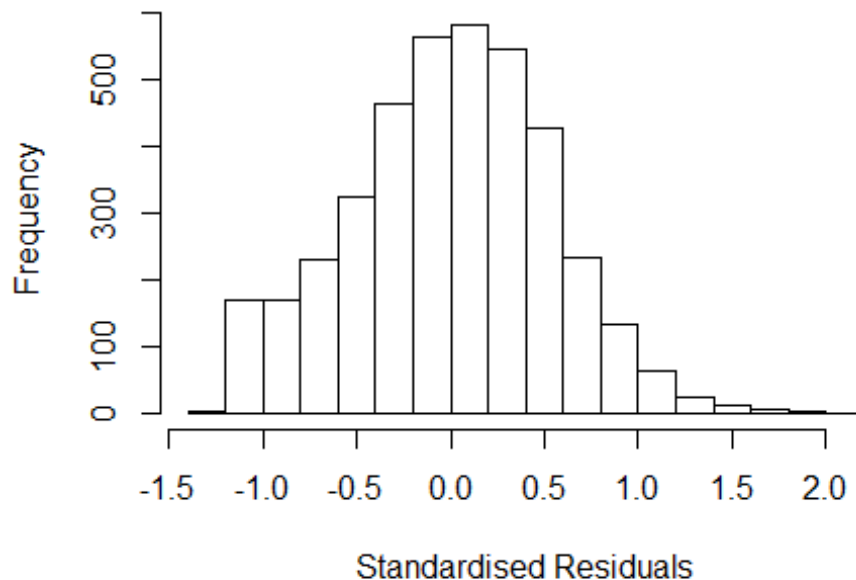
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2355 -0.3584 0.0195 0.3642 2.0905
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.204023 0.043571 27.63 < 2e-16 ***
## FirstAuthorFemale1 -0.000844 0.023070 -0.04 0.97080
## LastAuthorFemale1 0.048935 0.029624 1.65 0.09864 .
## Year1997 -0.032954 0.062891 -0.52 0.60032
## Year1998 -0.024315 0.059590 -0.41 0.68327
## Year1999 0.028651 0.184620 0.16 0.87668
## Year2000 -0.017466 0.062033 -0.28 0.77830
## Year2001 -0.196474 0.076887 -2.56 0.01065 *
## Year2002 -0.215994 0.064459 -3.35 0.00081 ***
## Year2003 -0.228934 0.061180 -3.74 0.00019 ***
## Year2004 -0.136476 0.061573 -2.22 0.02671 *
## Year2005 -0.080643 0.054123 -1.49 0.13631
```

```

## Year2006      -0.066972    0.055435   -1.21  0.22707
## Year2007      -0.083988    0.056095   -1.50  0.13441
## Year2008      -0.119264    0.051957   -2.30  0.02176 *
## Year2009      -0.069180    0.052468   -1.32  0.18741
## Year2010      -0.086531    0.051365   -1.68  0.09214 .
## Year2011      -0.097508    0.050636   -1.93  0.05422 .
## Year2012      -0.108607    0.051030   -2.13  0.03337 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.533
## Multiple R-squared:  0.0119, Adjusted R-squared:  0.00734
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 350 weights are ~= 1. The remaining 3604 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0888 0.8700 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      2.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.025 1      1.012
## Year      1.025 16      1.001

```

Residuals from first author



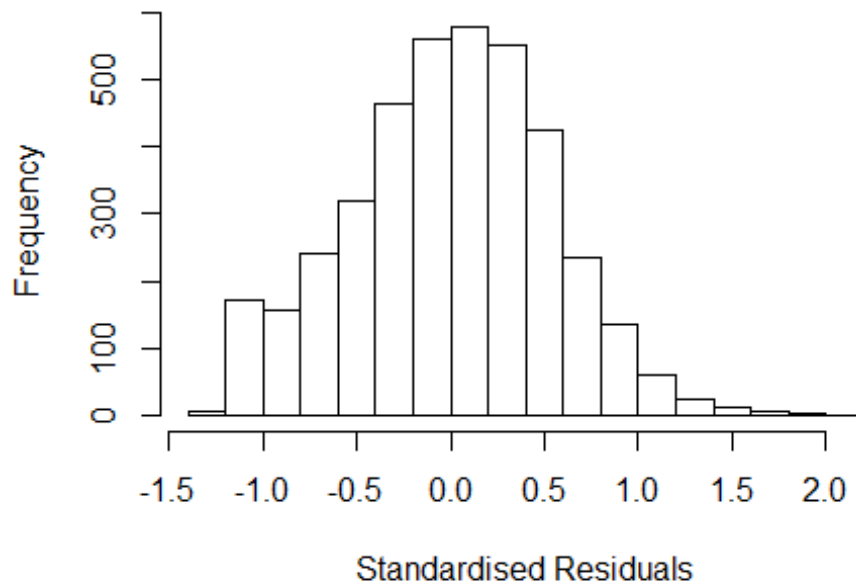
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2051 -0.3594 0.0197 0.3624 2.0857
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20510 0.04347 27.72 <2e-16 ***
## FirstAuthorFemale1 0.00343 0.02303 0.15 0.8815
## Year1997 -0.03003 0.06264 -0.48 0.6316
## Year1998 -0.02246 0.05945 -0.38 0.7056
## Year1999 0.02673 0.18455 0.14 0.8849
## Year2000 -0.01533 0.06197 -0.25 0.8046
## Year2001 -0.19264 0.07686 -2.51 0.0122 *
## Year2002 -0.21365 0.06429 -3.32 0.0009 ***
## Year2003 -0.22767 0.06107 -3.73 0.0002 ***
## Year2004 -0.13277 0.06140 -2.16 0.0306 *
## Year2005 -0.07828 0.05403 -1.45 0.1475
## Year2006 -0.06519 0.05531 -1.18 0.2386
```

```

## Year2007          -0.07967      0.05595      -1.42      0.1546
## Year2008          -0.11649      0.05183      -2.25      0.0247 *
## Year2009          -0.06719      0.05239      -1.28      0.1997
## Year2010          -0.08370      0.05120      -1.63      0.1022
## Year2011          -0.09452      0.05054      -1.87      0.0615 .
## Year2012          -0.10390      0.05083      -2.04      0.0410 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.533
## Multiple R-squared:  0.0112, Adjusted R-squared:  0.00693
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 352 weights are ~= 1. The remaining 3602 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0912 0.8690 0.9500 0.9050 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.013 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
## -1.2353 -0.3583  0.0195  0.3643  2.0906
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2039     0.0435  27.69  < 2e-16 ***
## LastAuthorFemale1  0.0488     0.0295   1.65  0.09834 .
## Year1997         -0.0329     0.0629  -0.52  0.60062
## Year1998         -0.0243     0.0596  -0.41  0.68324
## Year1999          0.0286     0.1846   0.15  0.87701
## Year2000         -0.0175     0.0620  -0.28  0.77797
## Year2001         -0.1965     0.0769  -2.56  0.01064 *
## Year2002         -0.2160     0.0645  -3.35  0.00081 ***
## Year2003         -0.2290     0.0611  -3.75  0.00018 ***
## Year2004         -0.1365     0.0616  -2.22  0.02668 *
## Year2005         -0.0807     0.0541  -1.49  0.13568
## Year2006         -0.0671     0.0553  -1.21  0.22564
```

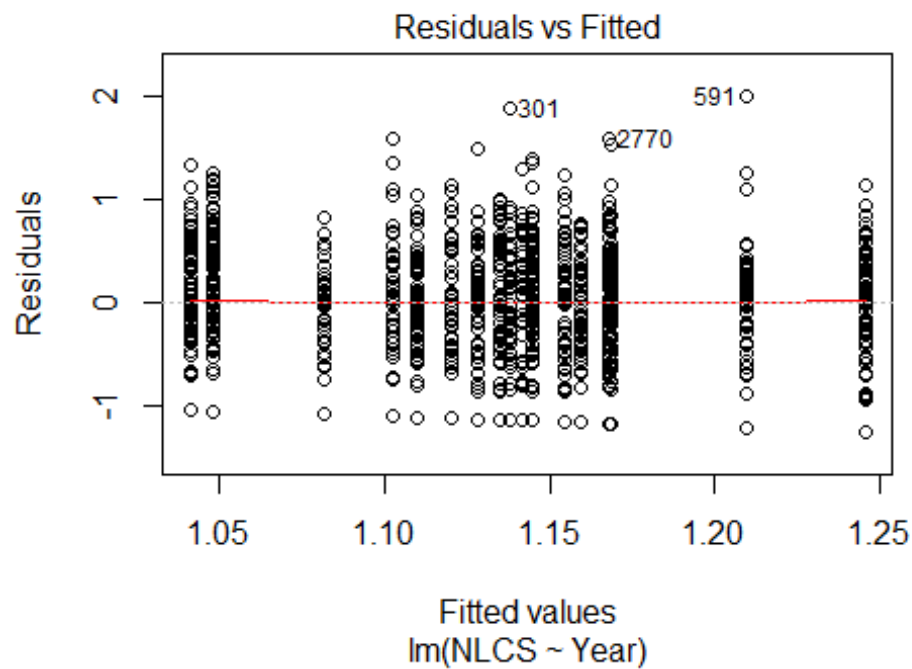
```

## Year2007          -0.0841      0.0560    -1.50   0.13342
## Year2008          -0.1193      0.0519    -2.30   0.02166 *
## Year2009          -0.0692      0.0525    -1.32   0.18696
## Year2010          -0.0866      0.0513    -1.69   0.09183 .
## Year2011          -0.0975      0.0506    -1.93   0.05414 .
## Year2012          -0.1086      0.0510    -2.13   0.03326 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.532
## Multiple R-squared:  0.0119, Adjusted R-squared:  0.00759
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 352 weights are ~= 1. The remaining 3602 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0885 0.8700 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      2.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3954"
## [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
## 154 154 136 171 139 116 133 84 95 137 154 205 174 189 146
## 2011 2012
## 178 206
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 47 46 56 53 56 41 56 40 46 59 72 92 85 92 71
## 2011 2012

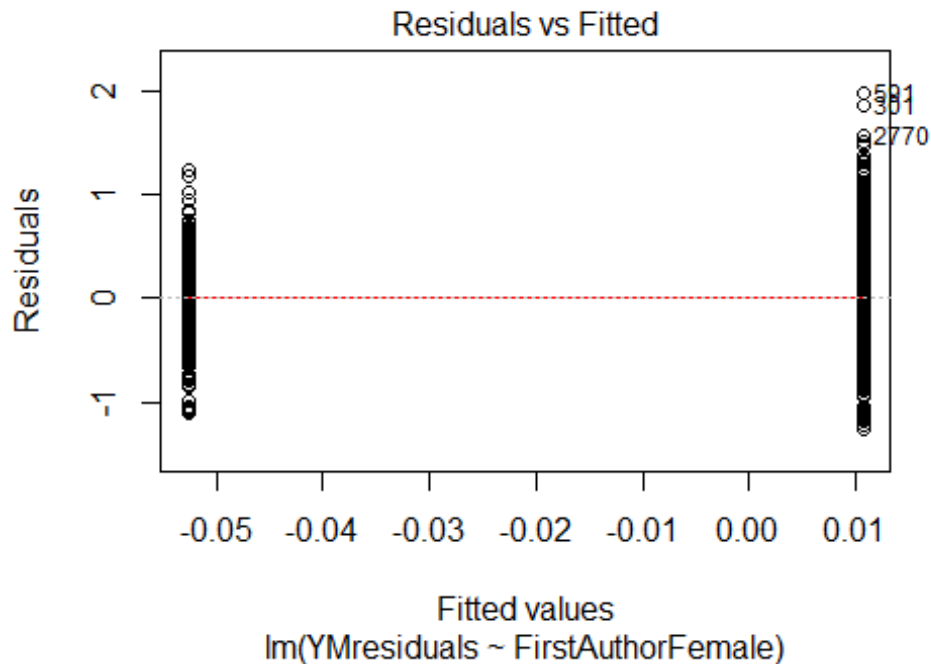
```



```
## 92 113
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 34 38 43 43 44 32 42 30 43 45 59 73 68 73 49
## 2011 2012
## 71 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 = 28, df = 16, p-value = 0.03
```

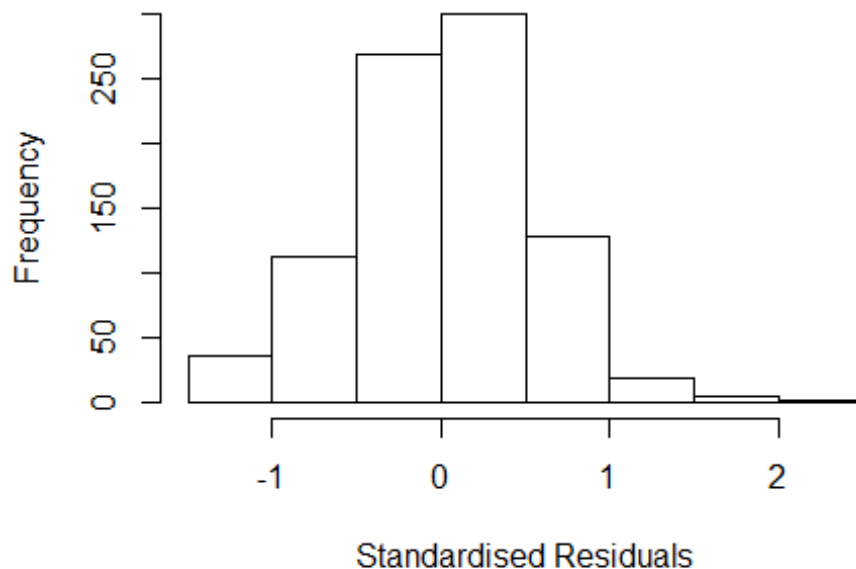


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.7, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 3.41041649874016"
## [1] "Male first author team size 2018 geometric mean: 3.0968967520866"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.71173542276274"
## [1] "Male last author team size 2018 geometric mean: 3.05440910251146"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.152 1      1.073
## LastAuthorFemale  1.095 1      1.046
## UniqueAuthors    1.472 4      1.050
## Year              1.563 16     1.014
```

Residuals from first and last author and team size



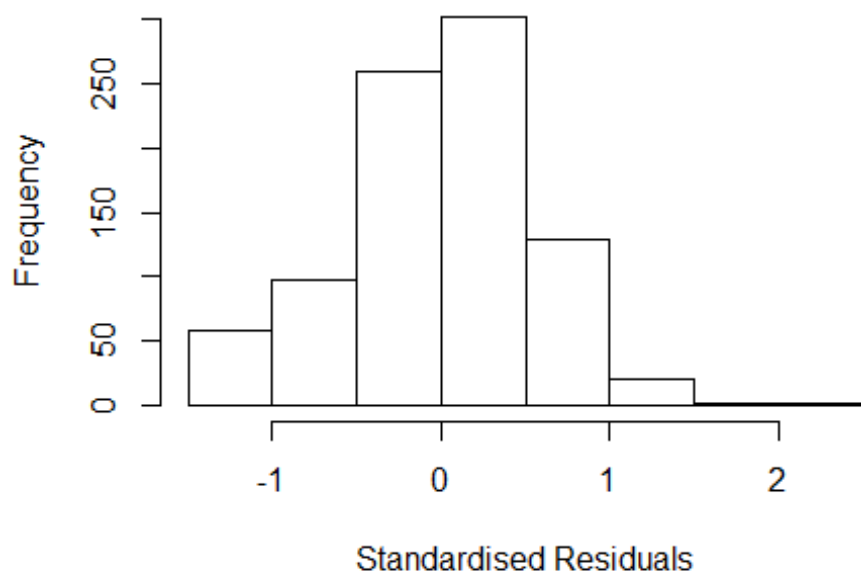
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3367 -0.3543 0.0207 0.3554 2.2357
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8143 0.1297 6.28 5.4e-10 ***
## FirstAuthorFemale1 -0.1018 0.0468 -2.18 0.02966 *
## LastAuthorFemale1 0.0322 0.0702 0.46 0.64709
## UniqueAuthors2 0.2998 0.0673 4.45 9.5e-06 ***
## UniqueAuthors3 0.2699 0.0732 3.69 0.00024 ***
## UniqueAuthors4 0.3024 0.0761 3.98 7.6e-05 ***
## UniqueAuthors5 0.2880 0.0729 3.95 8.4e-05 ***
## Year1997 0.0889 0.1542 0.58 0.56416
## Year1998 0.0786 0.1389 0.57 0.57172
## Year1999 0.1489 0.1500 0.99 0.32095
```

```

## Year2000          0.1297      0.1460      0.89  0.37457
## Year2001          0.1018      0.1678      0.61  0.54425
## Year2002          0.0626      0.1583      0.40  0.69266
## Year2003         -0.0108      0.1383     -0.08  0.93802
## Year2004          0.0709      0.1499      0.47  0.63620
## Year2005          0.0317      0.1465      0.22  0.82894
## Year2006          0.0367      0.1303      0.28  0.77854
## Year2007          0.0749      0.1335      0.56  0.57497
## Year2008          0.0752      0.1293      0.58  0.56116
## Year2009          0.0684      0.1407      0.49  0.62683
## Year2010          0.2200      0.1364      1.61  0.10720
## Year2011          0.0933      0.1300      0.72  0.47302
## Year2012         -0.0664      0.1400     -0.47  0.63556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.529
## Multiple R-squared:  0.0527, Adjusted R-squared:  0.0281
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 797 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0349 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      1.15e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.114 1      1.055
## LastAuthorFemale  1.038 1      1.019
## Year              1.150 16      1.004

```

Residuals from first and last author



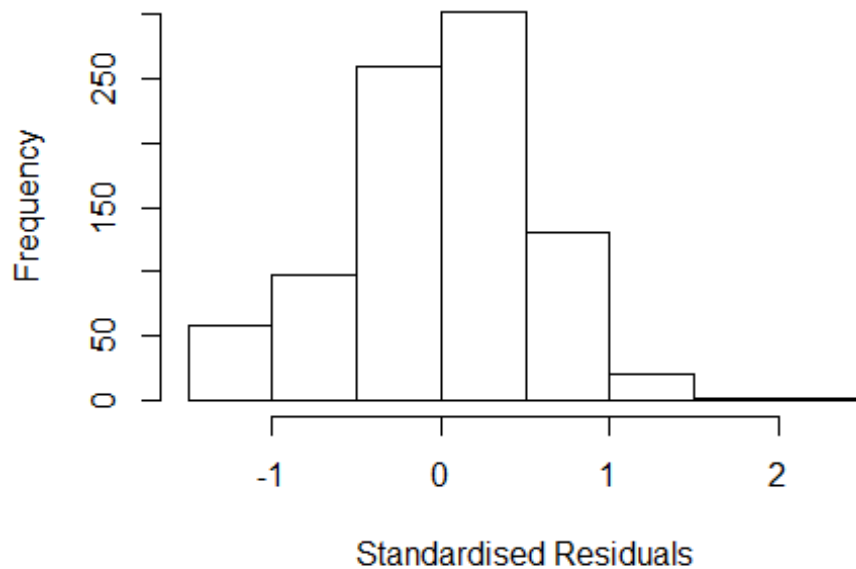
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2883 -0.3579 0.0333 0.3510 2.0097
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0214 0.1114 9.17 <2e-16 ***
## FirstAuthorFemale1 -0.0668 0.0468 -1.43 0.154
## LastAuthorFemale1 0.0480 0.0709 0.68 0.499
## Year1997 0.1214 0.1482 0.82 0.413
## Year1998 0.1063 0.1326 0.80 0.423
## Year1999 0.1679 0.1425 1.18 0.239
## Year2000 0.1261 0.1408 0.90 0.371
## Year2001 0.1405 0.1581 0.89 0.374
## Year2002 0.0734 0.1538 0.48 0.633
## Year2003 0.0504 0.1343 0.38 0.707
## Year2004 0.0928 0.1463 0.63 0.526
## Year2005 0.0702 0.1430 0.49 0.624
```

```

## Year2006          0.0703      0.1259      0.56      0.576
## Year2007          0.1206      0.1291      0.93      0.350
## Year2008          0.1155      0.1237      0.93      0.351
## Year2009          0.0883      0.1362      0.65      0.517
## Year2010          0.2669      0.1336      2.00      0.046 *
## Year2011          0.1462      0.1262      1.16      0.247
## Year2012          -0.0180      0.1381     -0.13      0.896
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.0167, Adjusted R-squared:  -0.00416
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 800 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##   0.120  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      1.15e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.113 1          1.055
## Year              1.113 16          1.003

```

Residuals from first author



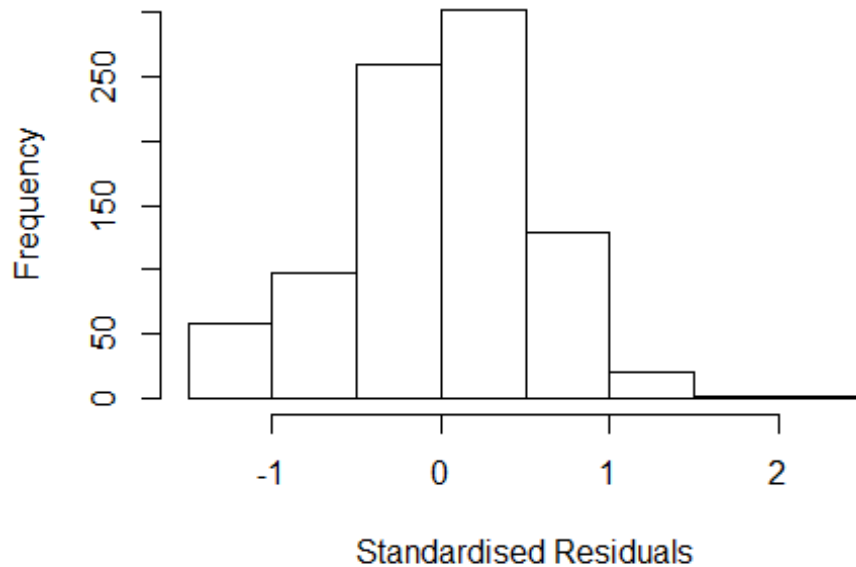
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2914 -0.3534 0.0375 0.3510 2.0040
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0216 0.1110 9.20 <2e-16 ***
## FirstAuthorFemale1 -0.0649 0.0472 -1.38 0.169
## Year1997 0.1241 0.1477 0.84 0.401
## Year1998 0.1112 0.1323 0.84 0.401
## Year1999 0.1733 0.1417 1.22 0.222
## Year2000 0.1308 0.1400 0.93 0.350
## Year2001 0.1412 0.1581 0.89 0.372
## Year2002 0.0760 0.1540 0.49 0.622
## Year2003 0.0534 0.1336 0.40 0.689
## Year2004 0.0968 0.1457 0.66 0.507
## Year2005 0.0721 0.1423 0.51 0.612
## Year2006 0.0723 0.1252 0.58 0.563
```

```

## Year2007          0.1239      0.1284      0.96      0.335
## Year2008          0.1181      0.1231      0.96      0.338
## Year2009          0.0884      0.1358      0.65      0.515
## Year2010          0.2697      0.1328      2.03      0.043 *
## Year2011          0.1517      0.1251      1.21      0.225
## Year2012          -0.0130      0.1372     -0.09      0.925
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.0162, Adjusted R-squared:  -0.00343
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 804 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.124  0.869  0.950  0.904  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.15e-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 1.034 1      1.017
## Year              1.034 16      1.001

```


Residuals from last author



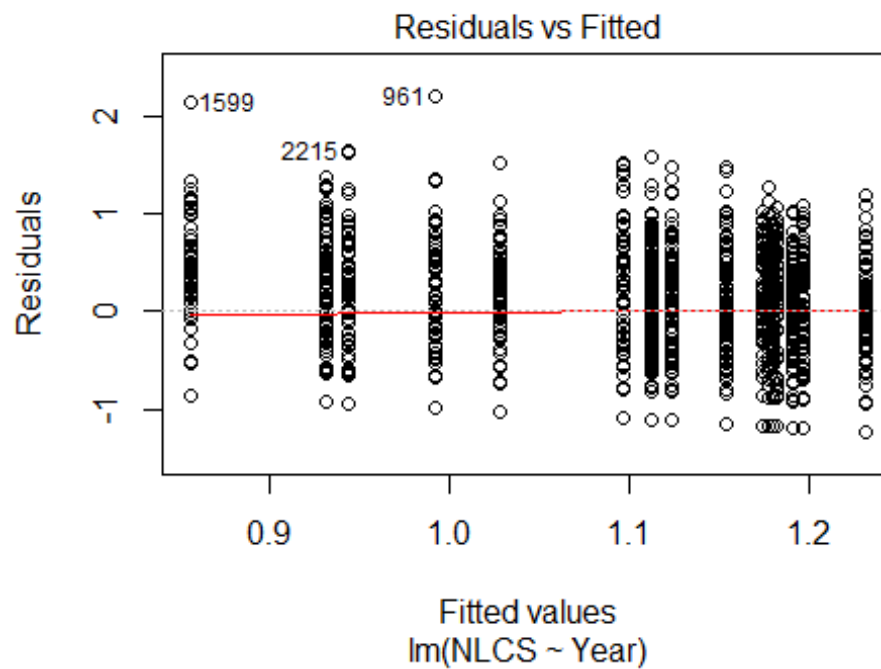
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2762 -0.3657 0.0357 0.3488 2.0153
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0192 0.1119 9.11 <2e-16 ***
## LastAuthorFemale1 0.0426 0.0711 0.60 0.550
## Year1997 0.1166 0.1488 0.78 0.433
## Year1998 0.1021 0.1330 0.77 0.443
## Year1999 0.1645 0.1432 1.15 0.251
## Year2000 0.1182 0.1411 0.84 0.403
## Year2001 0.1289 0.1576 0.82 0.413
## Year2002 0.0627 0.1537 0.41 0.684
## Year2003 0.0335 0.1341 0.25 0.803
## Year2004 0.0823 0.1469 0.56 0.576
## Year2005 0.0560 0.1430 0.39 0.695
## Year2006 0.0594 0.1262 0.47 0.638
```

```

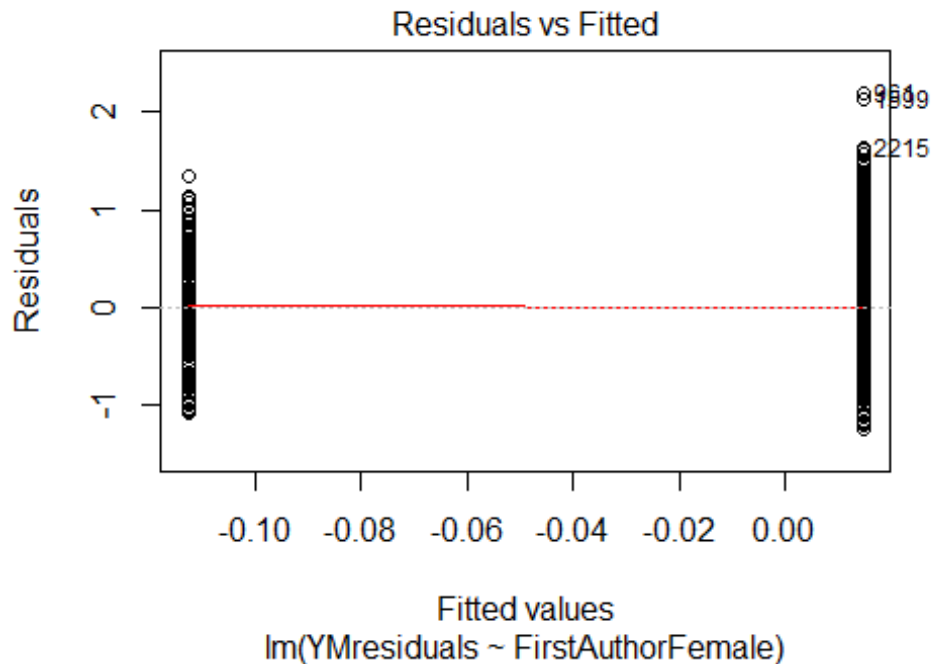
## Year2007          0.1056      0.1293      0.82      0.414
## Year2008          0.1143      0.1243      0.92      0.358
## Year2009          0.0730      0.1353      0.54      0.589
## Year2010          0.2570      0.1339      1.92      0.055 .
## Year2011          0.1369      0.1266      1.08      0.280
## Year2012         -0.0333      0.1375     -0.24      0.809
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.0147, Adjusted R-squared:  -0.005
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 806 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.119  0.870   0.952   0.904   0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.15e-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: 869"
## [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
## 193 247 215 261 233 222 240 160 158 193 179 219 210 215 187
## 2011 2012
## 237 199
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 50 60 66 88 68 91 114 86 85 93 99 106 120 116 101
## 2011 2012

```

```
## 145 108
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 45 47 54 81 59 69 95 67 78 75 85 90 101 91 80
## 2011 2012
## 122 85
## [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
```

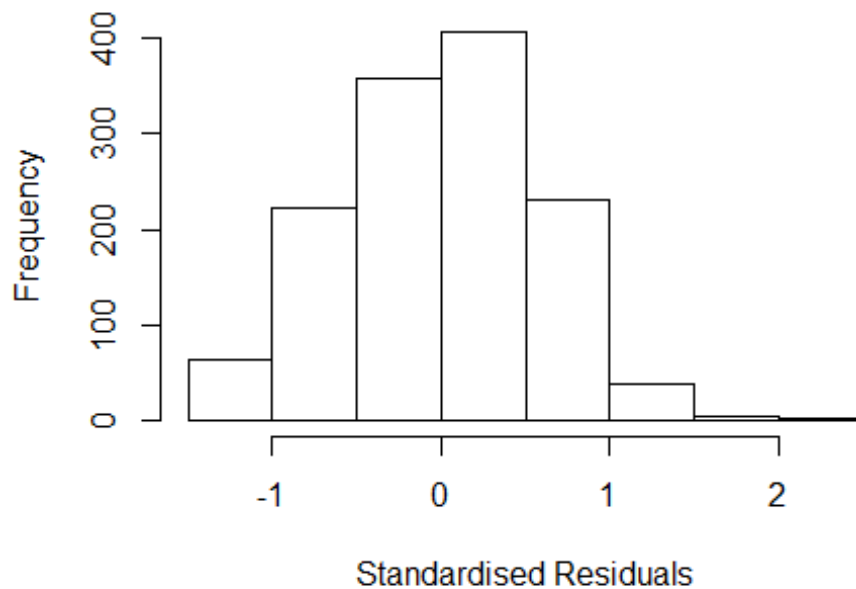


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.19, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 2.63214802590498"
## [1] "Male first author team size 2018 geometric mean: 2.15049869949851"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 130, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.70192007704123"
## [1] "Male last author team size 2018 geometric mean: 2.1367748325645"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, 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  1.079 1      1.039
## UniqueAuthors     1.288 4      1.032
## Year               1.301 16     1.008
```

Residuals from first and last author and team size



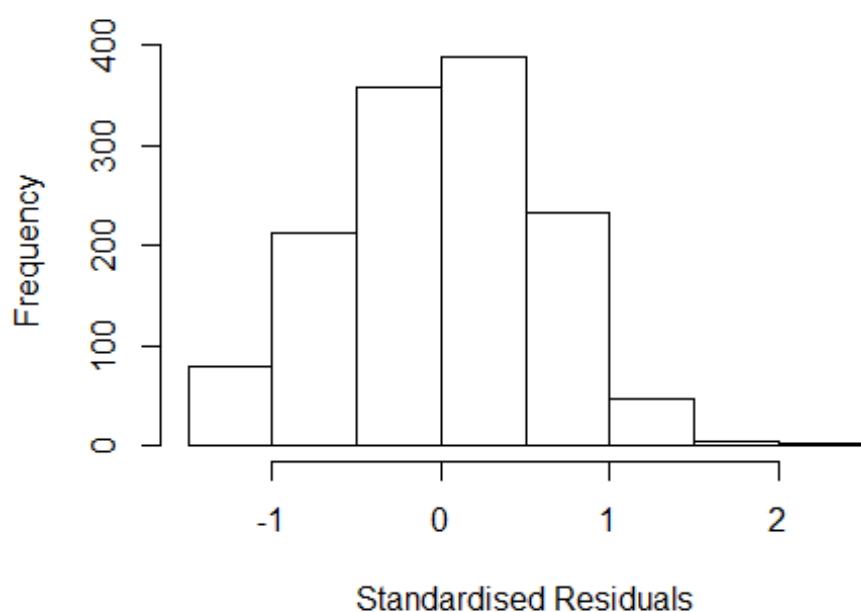
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3931 -0.4131 0.0213 0.4195 2.3017
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1492 0.1120 10.26 < 2e-16 ***
## FirstAuthorFemale1 -0.0786 0.0550 -1.43 0.15330
## LastAuthorFemale1 0.0163 0.0642 0.25 0.79975
## UniqueAuthors2 0.1265 0.0425 2.98 0.00296 **
## UniqueAuthors3 0.1900 0.0506 3.76 0.00018 ***
## UniqueAuthors4 0.3202 0.0683 4.69 3.1e-06 ***
## UniqueAuthors5 0.0903 0.0774 1.17 0.24379
## Year1997 -0.0717 0.1374 -0.52 0.60200
## Year1998 0.0243 0.1415 0.17 0.86383
## Year1999 -0.2519 0.1364 -1.85 0.06502 .
```

```

## Year2000          -0.1601      0.1635   -0.98   0.32753
## Year2001          -0.4188      0.1510   -2.77   0.00562 **
## Year2002          -0.3002      0.1367   -2.20   0.02828 *
## Year2003          -0.2045      0.1428   -1.43   0.15231
## Year2004          -0.1630      0.1315   -1.24   0.21534
## Year2005          -0.0773      0.1271   -0.61   0.54291
## Year2006          -0.0173      0.1235   -0.14   0.88834
## Year2007          -0.0762      0.1232   -0.62   0.53612
## Year2008          -0.0281      0.1193   -0.24   0.81372
## Year2009          -0.1074      0.1253   -0.86   0.39166
## Year2010          -0.0569      0.1286   -0.44   0.65834
## Year2011          -0.1401      0.1223   -1.15   0.25219
## Year2012          -0.1521      0.1353   -1.12   0.26114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.609
## Multiple R-squared:  0.0529, Adjusted R-squared:  0.0369
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 105 weights are ~= 1. The remaining 1219 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.122  0.873  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          7.55e-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 1.075 1 1.037
## LastAuthorFemale 1.076 1 1.037
## Year 1.069 16 1.002

```

Residuals from first and last author

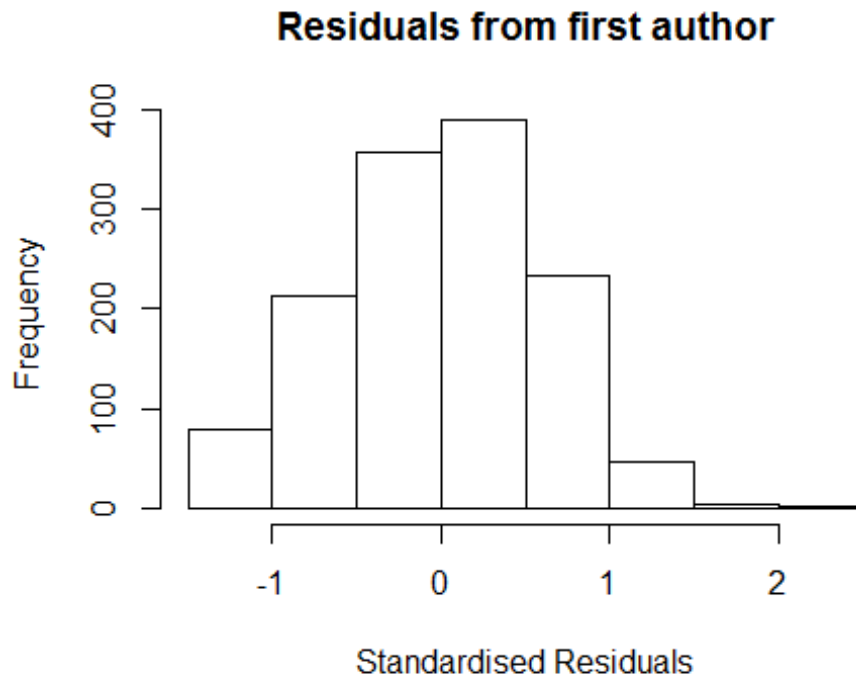


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2379 -0.4272 0.0212 0.4279 2.2261
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21519 0.10816 11.24 <2e-16 ***
## FirstAuthorFemale1 -0.04941 0.05429 -0.91 0.3629
## LastAuthorFemale1 0.02208 0.06396 0.35 0.7300
## Year1997 -0.05739 0.13426 -0.43 0.6691
## Year1998 0.01699 0.13898 0.12 0.9027
## Year1999 -0.24233 0.13507 -1.79 0.0730 .
## Year2000 -0.17716 0.16581 -1.07 0.2855
## Year2001 -0.41270 0.15131 -2.73 0.0065 **
## Year2002 -0.28162 0.13637 -2.07 0.0391 *
## Year2003 -0.20176 0.14301 -1.41 0.1585
## Year2004 -0.14129 0.12882 -1.10 0.2729
## Year2005 -0.03705 0.12571 -0.29 0.7683
```

```

## Year2006          0.02275    0.11971    0.19    0.8493
## Year2007          -0.04836    0.12003   -0.40    0.6871
## Year2008          -0.00733    0.11635   -0.06    0.9498
## Year2009          -0.07160    0.12286   -0.58    0.5601
## Year2010          -0.00383    0.12438   -0.03    0.9755
## Year2011          -0.10258    0.11864   -0.86    0.3874
## Year2012          -0.10075    0.13168   -0.77    0.4444
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.034, Adjusted R-squared:  0.0207
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 89 weights are ~= 1. The remaining 1235 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.151  0.865  0.946  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      7.55e-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 1.035 1         1.017
## Year              1.035 16         1.001

```

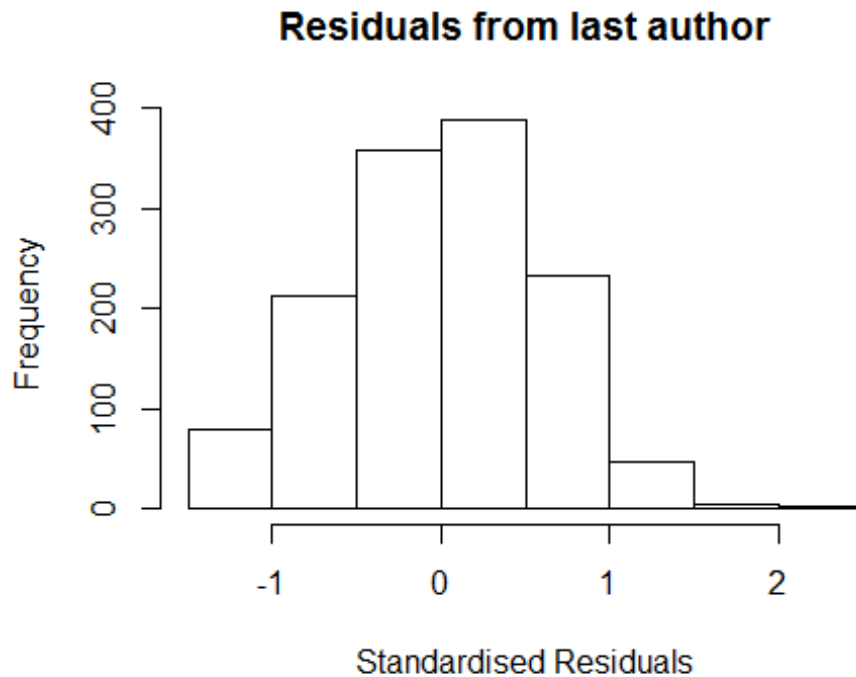



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2384 -0.4283  0.0201  0.4269  2.2254
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.21680    0.10790   11.28  <2e-16 ***
## FirstAuthorFemale1 -0.04314    0.05382   -0.80   0.4230
## Year1997        -0.05861    0.13407   -0.44   0.6621
## Year1998         0.01800    0.13908    0.13   0.8970
## Year1999        -0.24320    0.13501   -1.80   0.0719 .
## Year2000        -0.17720    0.16590   -1.07   0.2857
## Year2001        -0.41409    0.15132   -2.74   0.0063 **
## Year2002        -0.28257    0.13635   -2.07   0.0384 *
## Year2003        -0.20262    0.14283   -1.42   0.1563
## Year2004        -0.14150    0.12880   -1.10   0.2721
## Year2005        -0.03671    0.12579   -0.29   0.7704
## Year2006         0.02155    0.11956    0.18   0.8570
```

```

## Year2007      -0.04877    0.12006   -0.41    0.6846
## Year2008      -0.00785    0.11634   -0.07    0.9462
## Year2009      -0.07229    0.12283   -0.59    0.5563
## Year2010      -0.00522    0.12420   -0.04    0.9665
## Year2011      -0.10265    0.11870   -0.86    0.3873
## Year2012      -0.10106    0.13171   -0.77    0.4431
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.0339, Adjusted R-squared:  0.0214
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 86 weights are ~= 1. The remaining 1238 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.151  0.865  0.947  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      7.55e-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 1.036 1          1.018
## Year            1.036 16          1.001

```



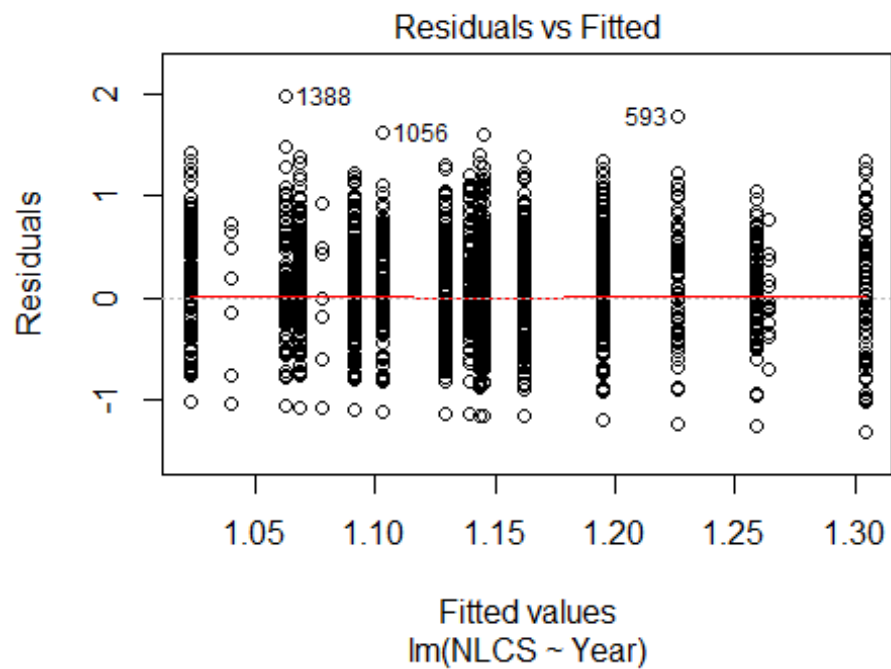
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2338 -0.4235 0.0245 0.4301 2.2285
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21113 0.10852 11.16 <2e-16 ***
## LastAuthorFemale1 0.00344 0.06319 0.05 0.9566
## Year1997 -0.05547 0.13486 -0.41 0.6809
## Year1998 0.01713 0.13946 0.12 0.9023
## Year1999 -0.24058 0.13551 -1.78 0.0761 .
## Year2000 -0.17603 0.16672 -1.06 0.2913
## Year2001 -0.41129 0.15129 -2.72 0.0066 **
## Year2002 -0.28197 0.13730 -2.05 0.0402 *
## Year2003 -0.20361 0.14300 -1.42 0.1547
## Year2004 -0.13926 0.12929 -1.08 0.2816
## Year2005 -0.03434 0.12602 -0.27 0.7853
## Year2006 0.02266 0.12006 0.19 0.8503
```

```

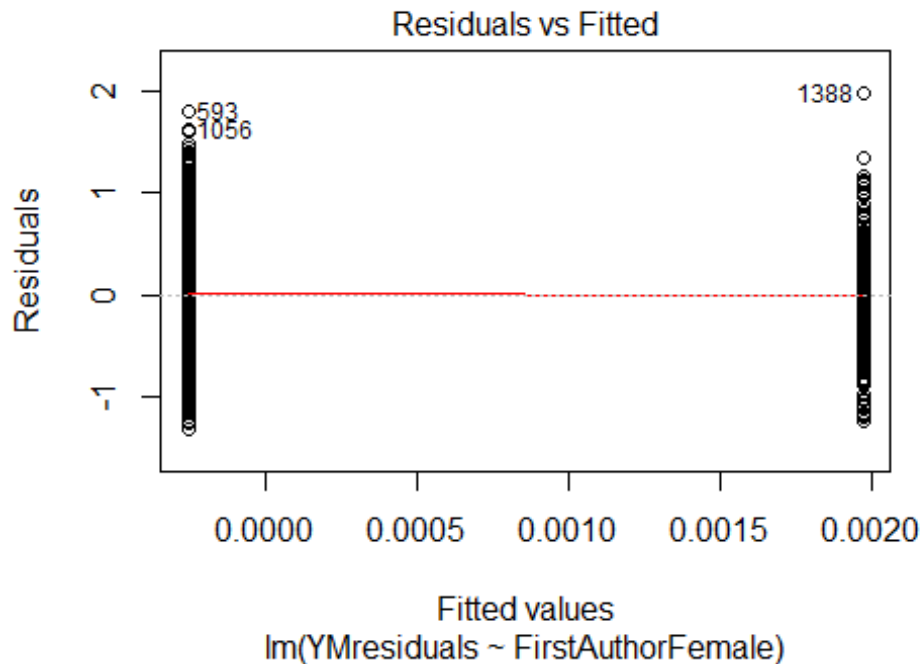
## Year2007          -0.04815      0.12057    -0.40    0.6897
## Year2008          -0.00645      0.11680    -0.06    0.9559
## Year2009          -0.07086      0.12327    -0.57    0.5655
## Year2010          -0.00558      0.12468    -0.04    0.9643
## Year2011          -0.10431      0.11900    -0.88    0.3809
## Year2012          -0.10350      0.13182    -0.79    0.4325
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.61
## Multiple R-squared:  0.0334, Adjusted R-squared:  0.0208
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 91 weights are ~= 1. The remaining 1233 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.153  0.868  0.948  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      7.55e-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: 1324"
## [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
## 241 223 223 21 280 263 273 319 20 435 36 488 500 500 481
## 2011 2012
## 519 530
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 88 81 95 8 120 100 103 169 7 194 14 223 257 263 245
## 2011 2012

```

```
## 261 299
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 60 78 7 100 75 82 132 6 161 10 166 188 191 185
## 2011 2012
## 198 237
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 54, df = 16, p-value = 4e-06
```

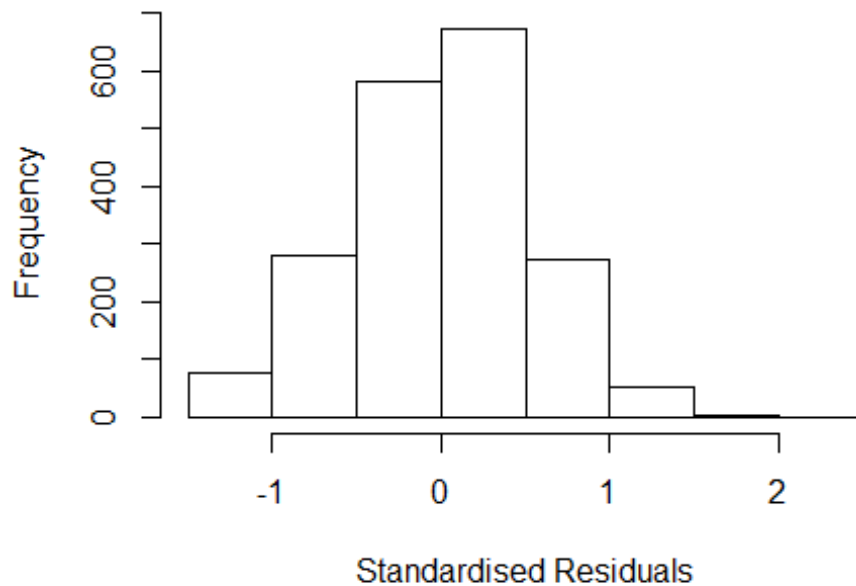


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.7, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.32008604948992"
## [1] "Male first author team size 2018 geometric mean: 3.07022638468696"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2900, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.16475758394742"
## [1] "Male last author team size 2018 geometric mean: 3.09700231889752"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2400, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.039 1          1.019
## LastAuthorFemale  1.036 1          1.018
## UniqueAuthors    1.207 4          1.024
## Year             1.272 16          1.008
```

Residuals from first and last author and team size



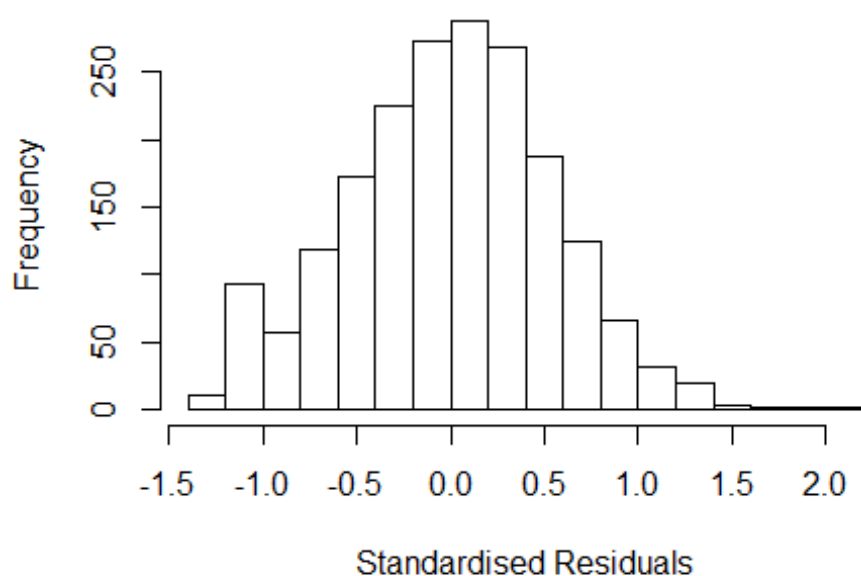
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4100 -0.3483 0.0199 0.3570 2.1600
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16318 0.08892 13.08 < 2e-16 ***
## FirstAuthorFemale1 -0.02881 0.03843 -0.75 0.4535
## LastAuthorFemale1 -0.06689 0.04145 -1.61 0.1067
## UniqueAuthors2 0.10300 0.04256 2.42 0.0156 *
## UniqueAuthors3 0.13502 0.04431 3.05 0.0023 **
## UniqueAuthors4 0.26790 0.04888 5.48 4.8e-08 ***
## UniqueAuthors5 0.35216 0.04962 7.10 1.8e-12 ***
## Year1997 -0.10535 0.12371 -0.85 0.3946
## Year1998 -0.02833 0.10622 -0.27 0.7897
## Year1999 -0.20033 0.35877 -0.56 0.5767
```

```

## Year2000      -0.19414      0.10669      -1.82      0.0690 .
## Year2001      -0.18750      0.11356      -1.65      0.0989 .
## Year2002      -0.18832      0.10477      -1.80      0.0724 .
## Year2003      -0.28808      0.10394      -2.77      0.0056 **
## Year2004      -0.00618      0.22758      -0.03      0.9783
## Year2005      -0.16807      0.09542      -1.76      0.0783 .
## Year2006      -0.06019      0.14749      -0.41      0.6833
## Year2007      -0.19438      0.09431      -2.06      0.0394 *
## Year2008      -0.19311      0.09202      -2.10      0.0360 *
## Year2009      -0.15512      0.09373      -1.65      0.0981 .
## Year2010      -0.18248      0.09313      -1.96      0.0502 .
## Year2011      -0.24705      0.09299      -2.66      0.0080 **
## Year2012      -0.23205      0.09155      -2.53      0.0113 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.529
## Multiple R-squared:  0.0499, Adjusted R-squared:  0.039
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 150 weights are ~= 1. The remaining 1793 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0584 0.8690 0.9520 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.15e-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 1.035 1 1.017
## LastAuthorFemale 1.038 1 1.019
## Year 1.060 16 1.002

```


Residuals from first and last author



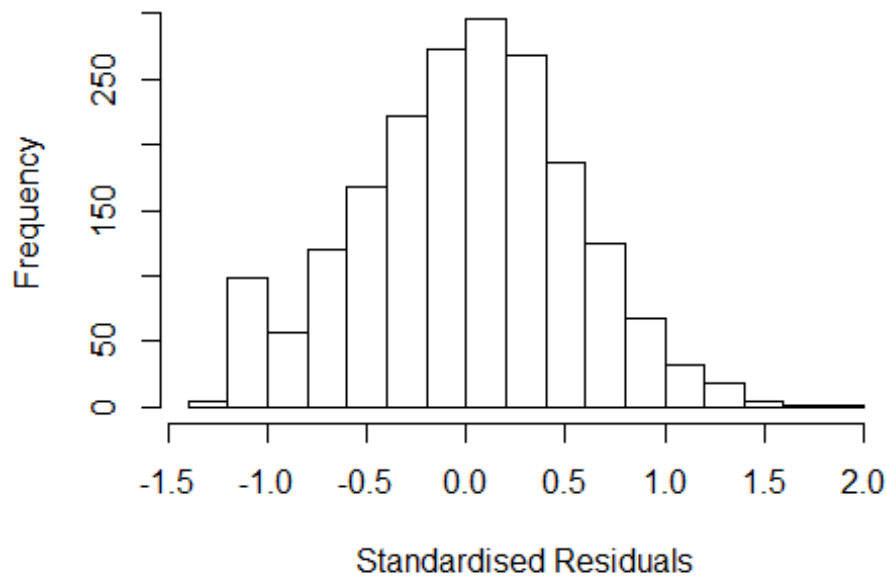
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2704 -0.3602 0.0121 0.3602 2.0053
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2567 0.0831 15.12 <2e-16 ***
## FirstAuthorFemale1 -0.0159 0.0386 -0.41 0.680
## LastAuthorFemale1 -0.0493 0.0413 -1.19 0.233
## Year1997 -0.0508 0.1235 -0.41 0.681
## Year1998 0.0137 0.1060 0.13 0.897
## Year1999 -0.2063 0.3066 -0.67 0.501
## Year2000 -0.1760 0.1049 -1.68 0.094 .
## Year2001 -0.1568 0.1148 -1.37 0.172
## Year2002 -0.1689 0.1046 -1.62 0.106
## Year2003 -0.2648 0.1048 -2.53 0.012 *
## Year2004 0.0143 0.2371 0.06 0.952
## Year2005 -0.1248 0.0934 -1.34 0.182
```

```

## Year2006          -0.0495      0.1364   -0.36    0.717
## Year2007          -0.1366      0.0925   -1.48    0.140
## Year2008          -0.1364      0.0905   -1.51    0.132
## Year2009          -0.0772      0.0922   -0.84    0.402
## Year2010          -0.1195      0.0916   -1.30    0.192
## Year2011          -0.1796      0.0916   -1.96    0.050 .
## Year2012          -0.1589      0.0892   -1.78    0.075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.0131, Adjusted R-squared:  0.00385
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 160 weights are ~= 1. The remaining 1783 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.129  0.866  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      5.15e-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 1.028 1      1.014
## Year              1.028 16      1.001

```

Residuals from first author



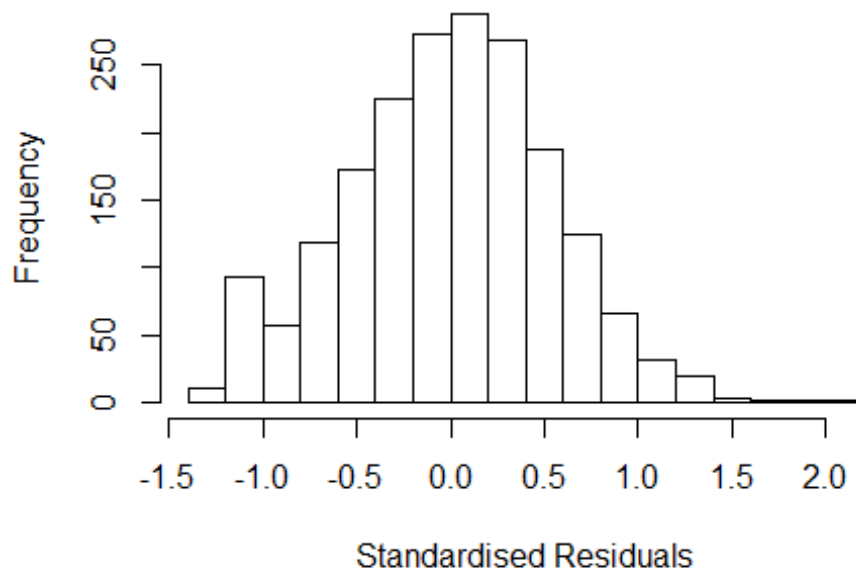
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2665 -0.3593 0.0151 0.3587 1.9645
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2533 0.0824 15.20 <2e-16 ***
## FirstAuthorFemale1 -0.0203 0.0385 -0.53 0.599
## Year1997 -0.0538 0.1233 -0.44 0.662
## Year1998 0.0132 0.1057 0.12 0.901
## Year1999 -0.2087 0.3127 -0.67 0.505
## Year2000 -0.1728 0.1042 -1.66 0.097 .
## Year2001 -0.1576 0.1146 -1.38 0.169
## Year2002 -0.1690 0.1041 -1.62 0.105
## Year2003 -0.2643 0.1042 -2.54 0.011 *
## Year2004 0.0074 0.2389 0.03 0.975
## Year2005 -0.1277 0.0927 -1.38 0.169
## Year2006 -0.0508 0.1372 -0.37 0.711
```

```

## Year2007          -0.1367      0.0919   -1.49    0.137
## Year2008          -0.1385      0.0899   -1.54    0.124
## Year2009          -0.0793      0.0916   -0.87    0.387
## Year2010          -0.1214      0.0910   -1.33    0.182
## Year2011          -0.1822      0.0910   -2.00    0.045 *
## Year2012          -0.1595      0.0887   -1.80    0.072 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.536
## Multiple R-squared:  0.0123, Adjusted R-squared:  0.00358
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 161 weights are ~= 1. The remaining 1782 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.150  0.867  0.950  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.15e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.032 1          1.016
## Year            1.032 16          1.001

```

Residuals from last author



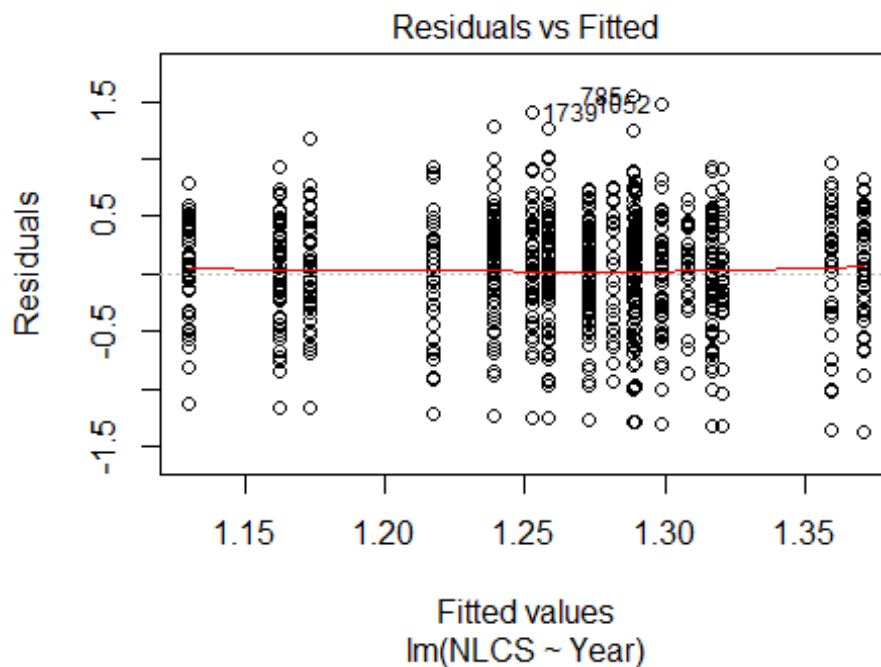
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2683 -0.3575 0.0116 0.3577 1.9921
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2554 0.0832 15.10 <2e-16 ***
## LastAuthorFemale1 -0.0508 0.0412 -1.23 0.217
## Year1997 -0.0520 0.1234 -0.42 0.674
## Year1998 0.0129 0.1059 0.12 0.903
## Year1999 -0.2049 0.3065 -0.67 0.504
## Year2000 -0.1760 0.1050 -1.68 0.094 .
## Year2001 -0.1566 0.1148 -1.36 0.173
## Year2002 -0.1688 0.1046 -1.61 0.107
## Year2003 -0.2646 0.1049 -2.52 0.012 *
## Year2004 0.0159 0.2371 0.07 0.947
## Year2005 -0.1253 0.0934 -1.34 0.180
## Year2006 -0.0498 0.1364 -0.37 0.715
```

```

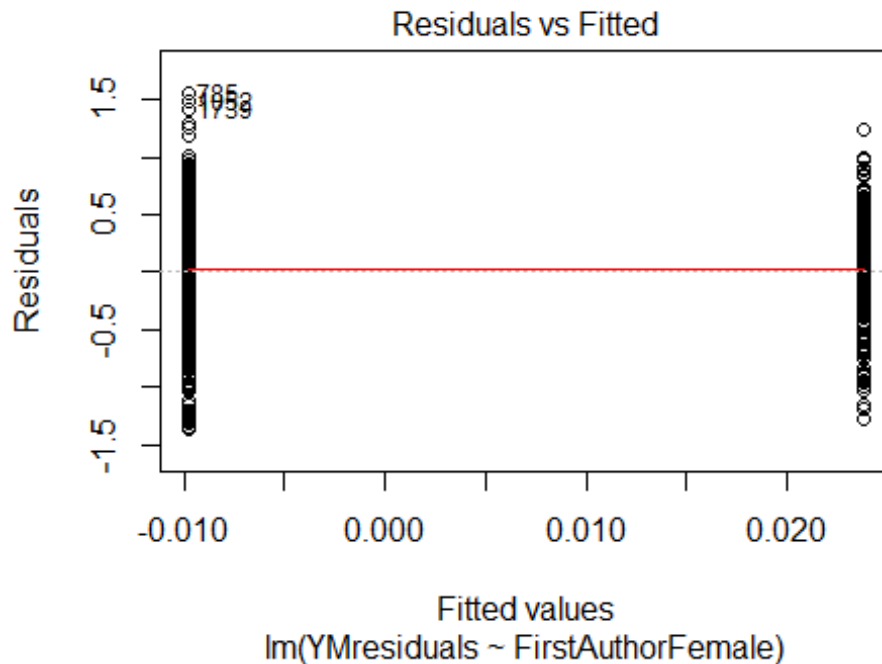
## Year2007          -0.1370      0.0926    -1.48      0.139
## Year2008          -0.1367      0.0906    -1.51      0.132
## Year2009          -0.0773      0.0922    -0.84      0.402
## Year2010          -0.1200      0.0916    -1.31      0.191
## Year2011          -0.1804      0.0916    -1.97      0.049 *
## Year2012          -0.1596      0.0893    -1.79      0.074 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.013, Adjusted R-squared:  0.00426
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1786 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.135  0.867  0.950  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.15e-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: 1943"
## [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
##   66   72   67   80   91   93  100   74   66   95  103  103   94  124  139
## 2011 2012
##  127  145
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   36   37   28   47   40   53   48   45   60   57   72   55   77   91
## 2011 2012

```

```
## 76 85
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 18 33 29 24 35 30 46 35 34 53 47 66 48 63 78
## 2011 2012
## 66 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 = 14, df = 16, p-value = 0.6
```



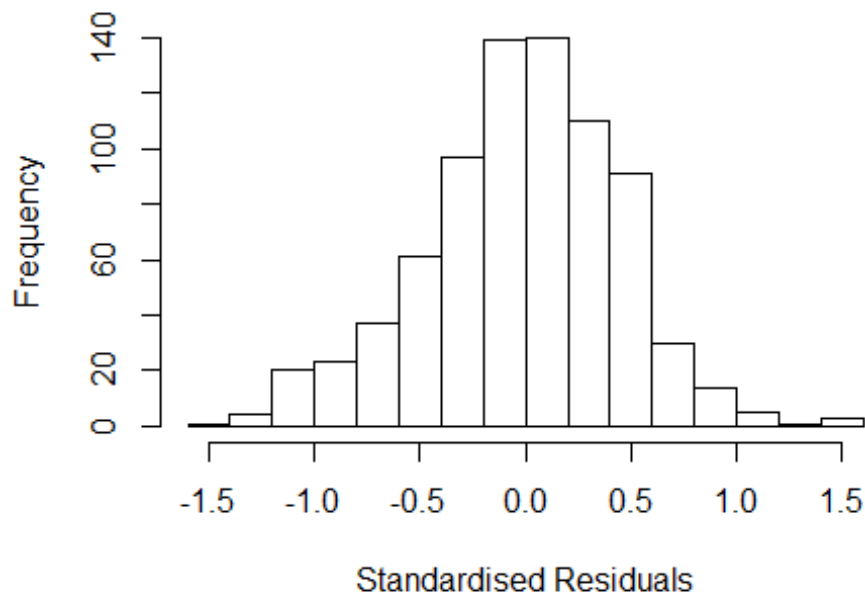
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.9, df = 1, p-value = 0.002
## [1] "Female first author team size 2018 geometric mean: 5.14235751607089"
## [1] "Male first author team size 2018 geometric mean: 3.61866284398956"
## Warning in wilcox.test.default(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 = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.76777487342011"
## [1] "Male last author team size 2018 geometric mean: 3.8553059282983"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.125	1	1.061
LastAuthorFemale	1.074	1	1.036
UniqueAuthors	1.470	4	1.049
Year	1.640	16	1.016

Residuals from first and last author and team size



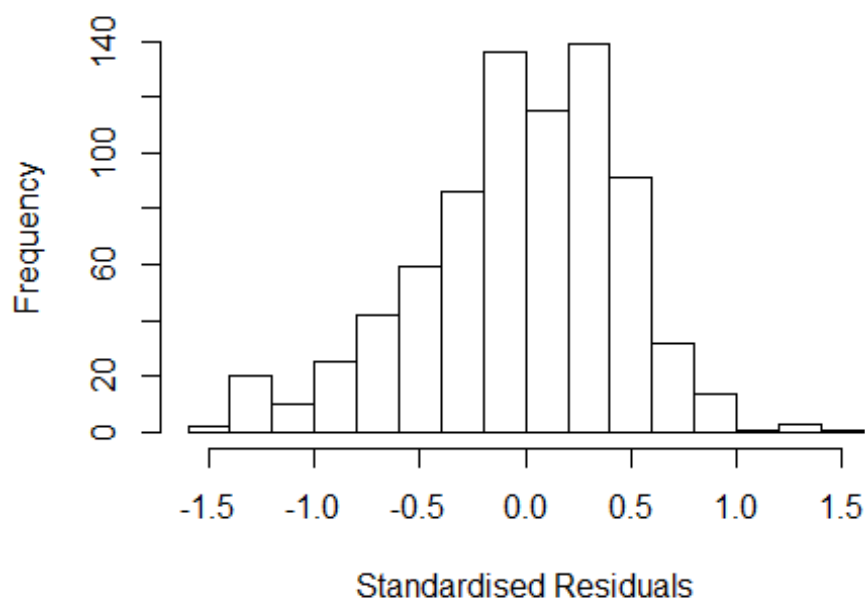
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.43410 -0.29243  0.00971  0.30349  1.45062
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9406    0.1412    6.66 5.3e-11 ***
## FirstAuthorFemale1 -0.0316    0.0352   -0.90 0.37010
## LastAuthorFemale1  0.0188    0.0393    0.48 0.63273
## UniqueAuthors2    0.2356    0.1051    2.24 0.02528 *
## UniqueAuthors3    0.3911    0.1036    3.78 0.00017 ***
## UniqueAuthors4    0.3365    0.1086    3.10 0.00202 **
## UniqueAuthors5    0.5250    0.0948    5.54 4.3e-08 ***
## Year1997         -0.0590    0.1648   -0.36 0.72063
## Year1998          0.1107    0.1446    0.77 0.44387
## Year1999          0.0163    0.1413    0.12 0.90841
```

```

## Year2000      -0.1507      0.1492      -1.01      0.31296
## Year2001      0.0548      0.1735      0.32      0.75219
## Year2002     -0.1670      0.1352     -1.23      0.21722
## Year2003     -0.0315      0.1464     -0.22      0.82943
## Year2004      0.1746      0.1419      1.23      0.21884
## Year2005     -0.0405      0.1313     -0.31      0.75789
## Year2006     -0.1008      0.1341     -0.75      0.45241
## Year2007      0.0221      0.1278      0.17      0.86276
## Year2008     -0.1089      0.1367     -0.80      0.42584
## Year2009      0.0243      0.1261      0.19      0.84754
## Year2010     -0.0659      0.1267     -0.52      0.60314
## Year2011     -0.0948      0.1266     -0.75      0.45400
## Year2012     -0.1018      0.1272     -0.80      0.42362
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.123, Adjusted R-squared:  0.097
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 713 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.249  0.872  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      1.29e-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 1.096 1      1.047
## LastAuthorFemale  1.070 1      1.034
## Year              1.173 16      1.005

```

Residuals from first and last author



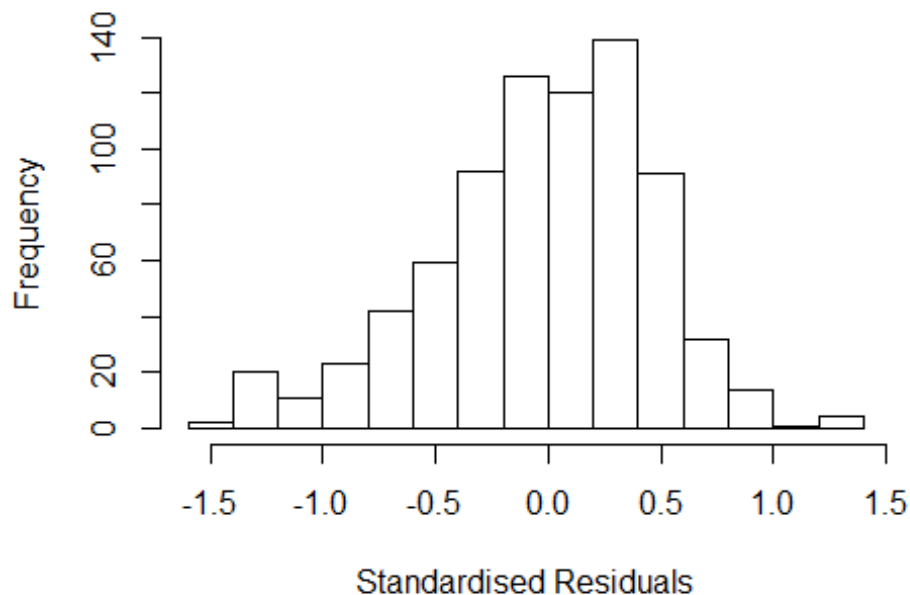
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.44019 -0.30131  0.00934  0.32146  1.40660
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.27947    0.12732   10.05  <2e-16 ***
## FirstAuthorFemale1  0.01531    0.03678    0.42    0.68
## LastAuthorFemale1  0.05176    0.04079    1.27    0.20
## Year1997        -0.05732    0.16432   -0.35    0.73
## Year1998         0.15122    0.15155    1.00    0.32
## Year1999         0.06501    0.15049    0.43    0.67
## Year2000        -0.15858    0.15863   -1.00    0.32
## Year2001         0.09344    0.18624    0.50    0.62
## Year2002        -0.10050    0.14691   -0.68    0.49
## Year2003        -0.01048    0.15903   -0.07    0.95
## Year2004         0.16073    0.15303    1.05    0.29
## Year2005        -0.00497    0.14416   -0.03    0.97
```

```

## Year2006      -0.03852    0.14387   -0.27    0.79
## Year2007      0.08214    0.14087    0.58    0.56
## Year2008     -0.07428    0.14622   -0.51    0.61
## Year2009      0.03208    0.13753    0.23    0.82
## Year2010     -0.01460    0.13874   -0.11    0.92
## Year2011     -0.03407    0.13935   -0.24    0.81
## Year2012     -0.01026    0.13741   -0.07    0.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.0261, Adjusted R-squared:  0.0029
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 696 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.300  0.874  0.946  0.894  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      1.29e-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 1.096 1      1.047
## Year              1.096 16      1.003

```

Residuals from first author



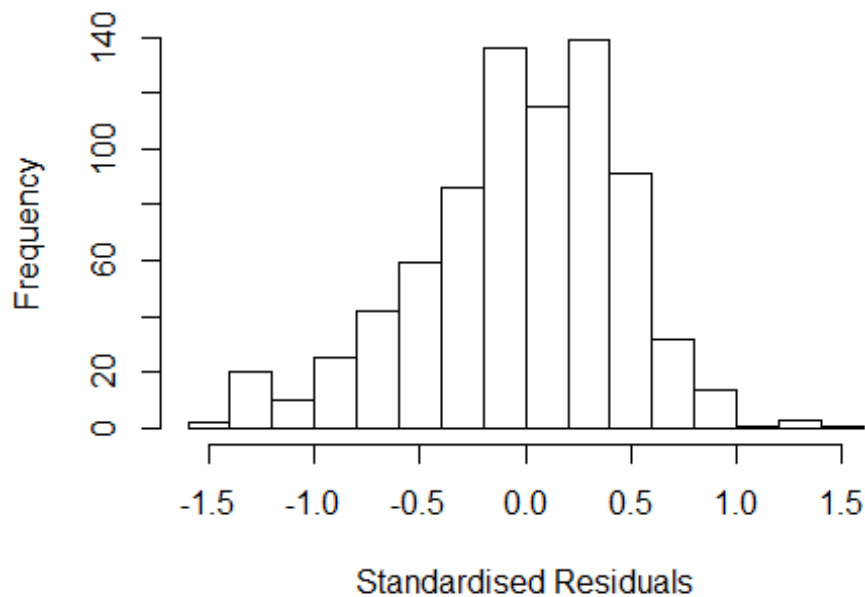
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4496 -0.3092 0.0177 0.3275 1.3971
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28759 0.12876 10.00 <2e-16 ***
## FirstAuthorFemale1 0.02165 0.03712 0.58 0.56
## Year1997 -0.05759 0.16647 -0.35 0.73
## Year1998 0.14928 0.15372 0.97 0.33
## Year1999 0.06954 0.15202 0.46 0.65
## Year2000 -0.15775 0.16145 -0.98 0.33
## Year2001 0.09086 0.19037 0.48 0.63
## Year2002 -0.10382 0.14871 -0.70 0.49
## Year2003 -0.01167 0.16032 -0.07 0.94
## Year2004 0.16200 0.15447 1.05 0.29
## Year2005 -0.00548 0.14627 -0.04 0.97
## Year2006 -0.03432 0.14572 -0.24 0.81
```

```

## Year2007          0.08012    0.14301    0.56    0.58
## Year2008          -0.07037    0.14846   -0.47    0.64
## Year2009          0.02968    0.13937    0.21    0.83
## Year2010          -0.01494    0.14110   -0.11    0.92
## Year2011          -0.03273    0.14118   -0.23    0.82
## Year2012          -0.00691    0.13937   -0.05    0.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.0242, Adjusted R-squared:  0.0023
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 704 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.293  0.871  0.946  0.895  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      1.29e-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 1.068 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
## -1.4439 -0.3045  0.0114  0.3247  1.4012
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.280679   0.127069   10.08  <2e-16 ***
## LastAuthorFemale1 0.054221   0.041136    1.32    0.19
## Year1997       -0.056441   0.163992   -0.34    0.73
## Year1998        0.153120   0.151433    1.01    0.31
## Year1999        0.065540   0.150234    0.44    0.66
## Year2000       -0.157361   0.158283   -0.99    0.32
## Year2001        0.096095   0.185745    0.52    0.61
## Year2002       -0.096492   0.146197   -0.66    0.51
## Year2003       -0.006986   0.158563   -0.04    0.96
## Year2004        0.163188   0.152772    1.07    0.29
## Year2005       -0.000998   0.143140   -0.01    0.99
## Year2006       -0.035603   0.143660   -0.25    0.80
```

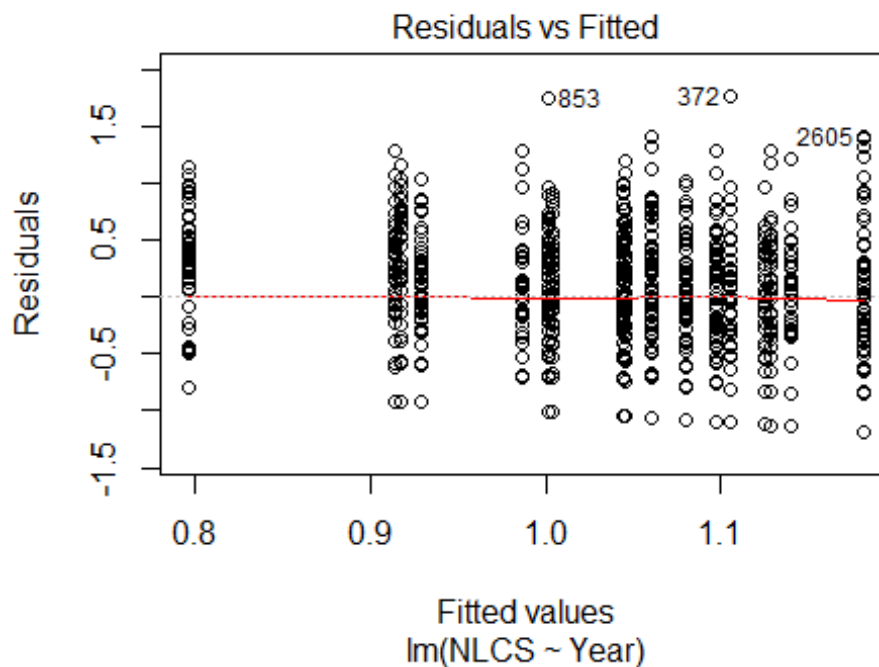
```

## Year2007      0.085950   0.140362   0.61   0.54
## Year2008     -0.072227   0.145985  -0.49   0.62
## Year2009      0.035783   0.136742   0.26   0.79
## Year2010     -0.011762   0.138512  -0.08   0.93
## Year2011     -0.029902   0.138425  -0.22   0.83
## Year2012     -0.006411   0.136483  -0.05   0.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.0259, Adjusted R-squared:  0.004
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 700 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.298  0.872  0.946  0.895  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 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
##  100  105  92  90  99  145  137  162  132  131  125  118  133  127  119
## 2011 2012
##  122  144
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   38  48  42  43  44  52  69  78  59  66  52  64  74  70  66
## 2011 2012

```



```
## 72 97
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 35 38 37 34 41 43 57 68 53 59 47 59 60 59 60
## 2011 2012
## 59 85
## [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
```



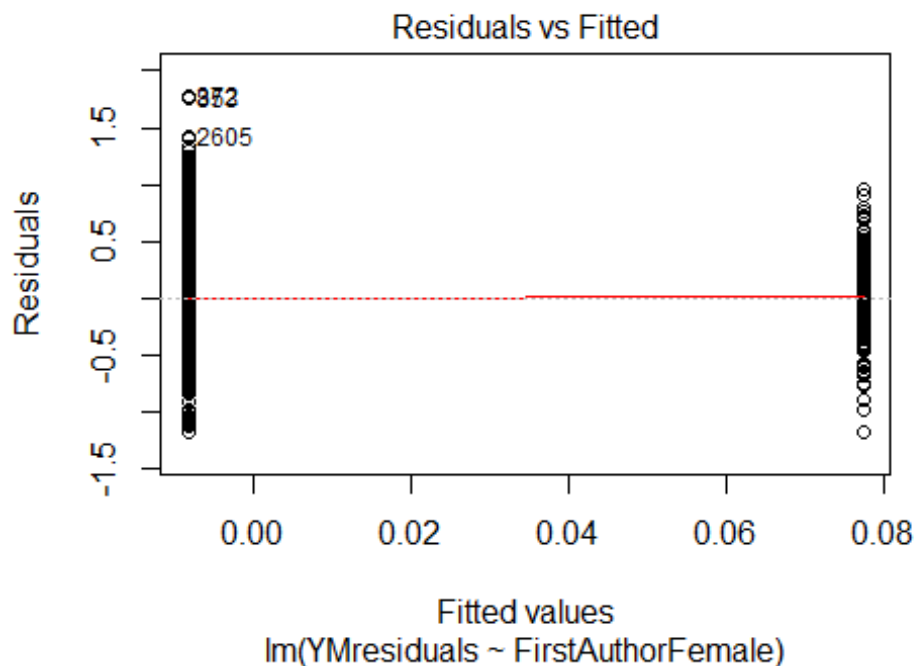
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 1.81712059283214"
## [1] "Male first author team size 2018 geometric mean: 1.86021870372604"

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

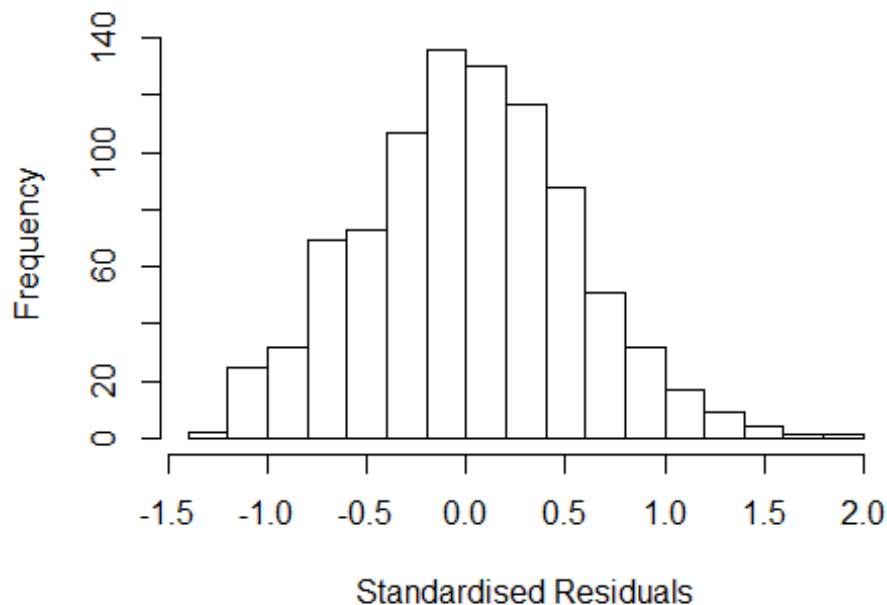
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 67, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.41421356237309"
## [1] "Male last author team size 2018 geometric mean: 1.8805341170524"

## Warning in wilcox.test.default(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 = 0.6
## 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 1.173  1      1.083
## LastAuthorFemale  1.281  1      1.132
## UniqueAuthors    1.623  4      1.062
## Year              1.797 16      1.018
```

Residuals from first and last author and team size



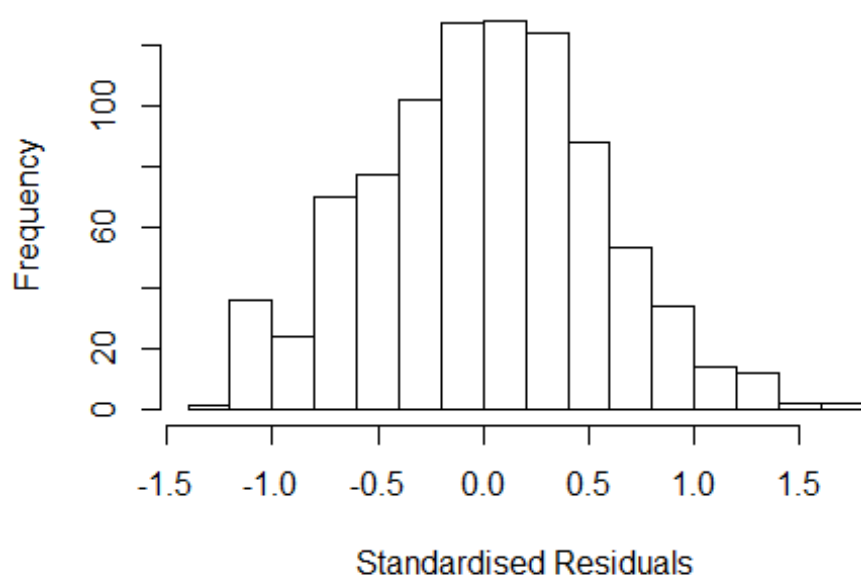
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24800 -0.34941 0.00715 0.36149 1.82272
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08427 0.07657 14.16 <2e-16 ***
## FirstAuthorFemale1 0.13414 0.06370 2.11 0.0355 *
## LastAuthorFemale1 -0.06535 0.06795 -0.96 0.3364
## UniqueAuthors2 0.10515 0.04729 2.22 0.0265 *
## UniqueAuthors3 0.11348 0.05334 2.13 0.0336 *
## UniqueAuthors4 0.20519 0.08072 2.54 0.0112 *
## UniqueAuthors5 0.29655 0.09439 3.14 0.0017 **
## Year1997 0.02663 0.11406 0.23 0.8154
## Year1998 -0.03816 0.10837 -0.35 0.7248
## Year1999 -0.22358 0.10760 -2.08 0.0380 *
```

```

## Year2000      -0.00448    0.10244   -0.04    0.9651
## Year2001      -0.13902    0.13547   -1.03    0.3051
## Year2002      -0.15499    0.10581   -1.46    0.1433
## Year2003      -0.39277    0.11892   -3.30    0.0010 ***
## Year2004      -0.29104    0.12642   -2.30    0.0216 *
## Year2005      -0.11325    0.09613   -1.18    0.2391
## Year2006      -0.12343    0.10193   -1.21    0.2263
## Year2007      -0.14547    0.10122   -1.44    0.1511
## Year2008      -0.25692    0.09416   -2.73    0.0065 **
## Year2009      -0.15266    0.10337   -1.48    0.1401
## Year2010      -0.07556    0.10236   -0.74    0.4606
## Year2011      -0.02210    0.10399   -0.21    0.8318
## Year2012      -0.15799    0.09091   -1.74    0.0826 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0621, Adjusted R-squared:  0.0384
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 810 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.221  0.861  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      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 1.132 1      1.064
## LastAuthorFemale  1.253 1      1.119
## Year              1.173 16      1.005

```

Residuals from first and last author



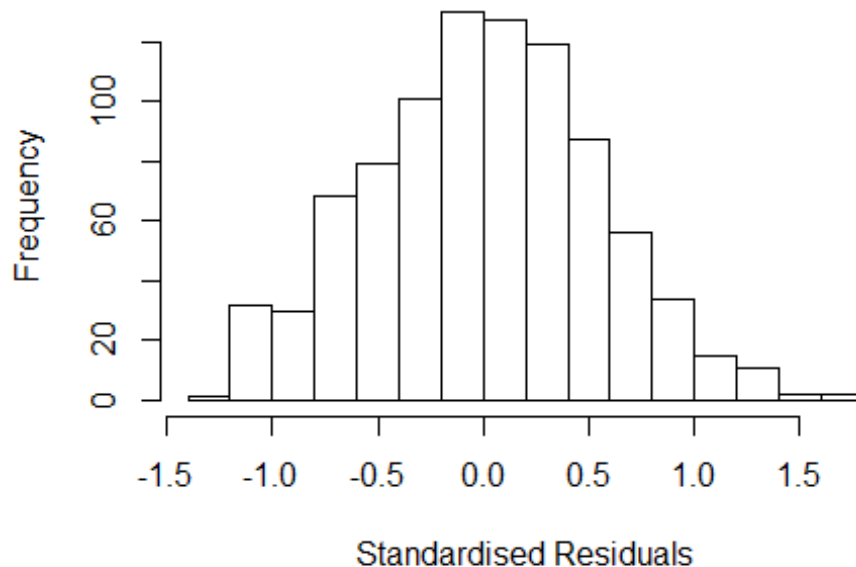
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2699 -0.3572 0.0149 0.3682 1.7714
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1545 0.0719 16.05 <2e-16 ***
## FirstAuthorFemale1 0.1615 0.0638 2.53 0.0115 *
## LastAuthorFemale1 -0.0587 0.0686 -0.86 0.3918
## Year1997 0.0319 0.1157 0.28 0.7828
## Year1998 -0.0538 0.1082 -0.50 0.6188
## Year1999 -0.2425 0.1097 -2.21 0.0273 *
## Year2000 -0.0298 0.1022 -0.29 0.7708
## Year2001 -0.1282 0.1357 -0.94 0.3450
## Year2002 -0.1467 0.1065 -1.38 0.1690
## Year2003 -0.3896 0.1192 -3.27 0.0011 **
## Year2004 -0.2740 0.1237 -2.22 0.0270 *
## Year2005 -0.0937 0.0972 -0.96 0.3354
```

```

## Year2006          -0.1459      0.1052   -1.39   0.1661
## Year2007          -0.1395      0.1017   -1.37   0.1707
## Year2008          -0.2736      0.0948   -2.89   0.0040 **
## Year2009          -0.1526      0.1033   -1.48   0.1400
## Year2010          -0.0461      0.1037   -0.44   0.6570
## Year2011           0.0240      0.1042    0.23   0.8177
## Year2012          -0.1147      0.0896   -1.28   0.2010
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.047, Adjusted R-squared:  0.0274
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 83 weights are ~= 1. The remaining 811 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.249  0.856  0.948  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      1.12e-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 1.022 1      1.011
## Year              1.022 16      1.001

```

Residuals from first author



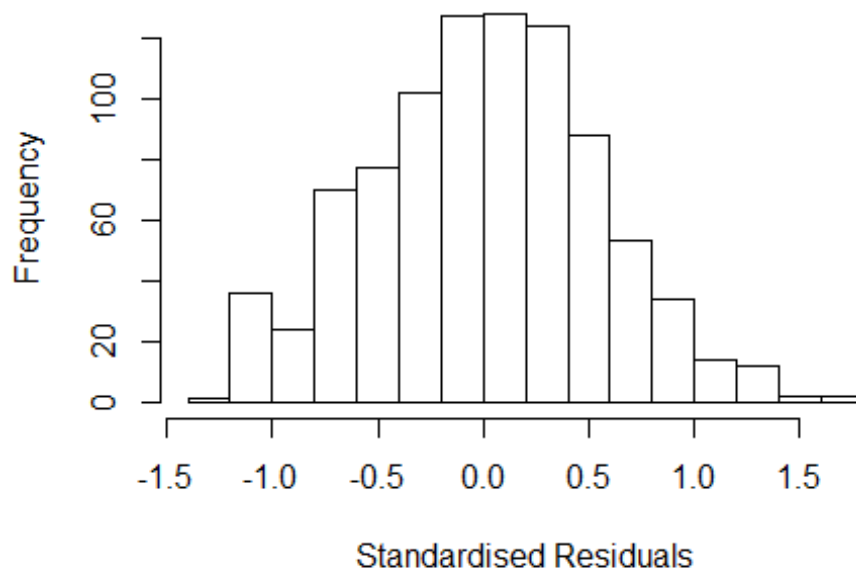
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2523 -0.3589 0.0145 0.3691 1.7763
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1559 0.0721 16.02 < 2e-16 ***
## FirstAuthorFemale1 0.1458 0.0612 2.38 0.01739 *
## Year1997 0.0298 0.1160 0.26 0.79766
## Year1998 -0.0602 0.1074 -0.56 0.57509
## Year1999 -0.2508 0.1082 -2.32 0.02069 *
## Year2000 -0.0326 0.1021 -0.32 0.74998
## Year2001 -0.1391 0.1335 -1.04 0.29799
## Year2002 -0.1534 0.1058 -1.45 0.14770
## Year2003 -0.3920 0.1182 -3.32 0.00095 ***
## Year2004 -0.2798 0.1230 -2.27 0.02317 *
## Year2005 -0.1010 0.0965 -1.05 0.29541
## Year2006 -0.1513 0.1049 -1.44 0.14968
```

```

## Year2007          -0.1427      0.1015   -1.41  0.15981
## Year2008          -0.2793      0.0943   -2.96  0.00313 **
## Year2009          -0.1568      0.1031   -1.52  0.12883
## Year2010          -0.0493      0.1043   -0.47  0.63617
## Year2011           0.0185      0.1034    0.18  0.85788
## Year2012          -0.1161      0.0898   -1.29  0.19607
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.536
## Multiple R-squared:  0.0463, Adjusted R-squared:  0.0278
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 810 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.250  0.857   0.947   0.905   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.12e-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 1.135 1          1.065
## Year            1.135 16          1.004

```


Residuals from last author



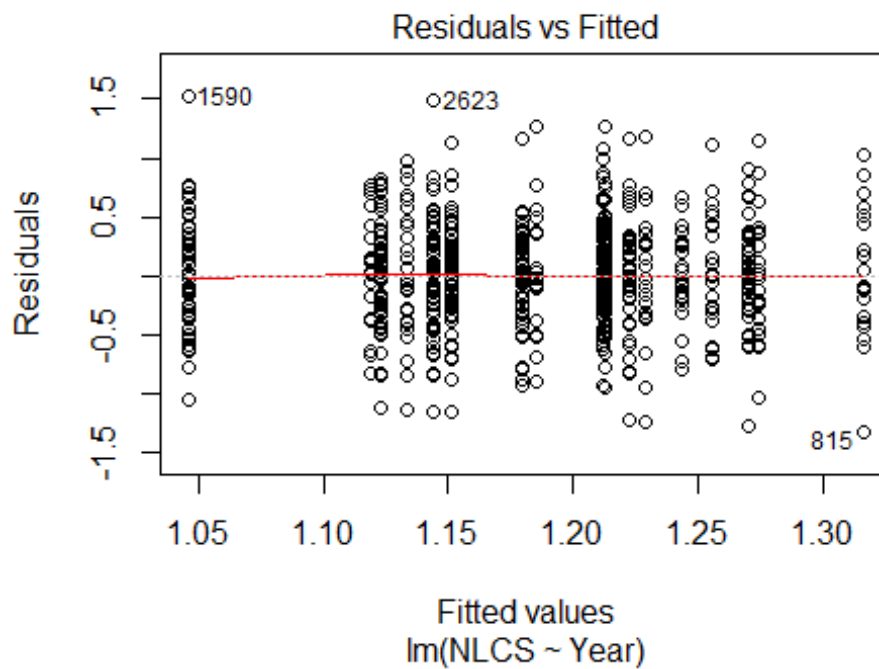
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19524 -0.35956 0.00725 0.36200 1.74890
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1695 0.0738 15.85 <2e-16 ***
## LastAuthorFemale1 -0.0102 0.0646 -0.16 0.8747
## Year1997 0.0257 0.1166 0.22 0.8256
## Year1998 -0.0464 0.1115 -0.42 0.6772
## Year1999 -0.2510 0.1112 -2.26 0.0242 *
## Year2000 -0.0415 0.1036 -0.40 0.6886
## Year2001 -0.1434 0.1356 -1.06 0.2905
## Year2002 -0.1571 0.1079 -1.46 0.1455
## Year2003 -0.3941 0.1226 -3.21 0.0014 **
## Year2004 -0.2805 0.1253 -2.24 0.0254 *
## Year2005 -0.1048 0.0991 -1.06 0.2906
## Year2006 -0.1507 0.1060 -1.42 0.1556
```

```

## Year2007          -0.1409      0.1042    -1.35    0.1770
## Year2008          -0.2726      0.0969    -2.81    0.0050 **
## Year2009          -0.1562      0.1051    -1.49    0.1375
## Year2010          -0.0464      0.1062    -0.44    0.6621
## Year2011           0.0236      0.1051     0.22    0.8223
## Year2012          -0.1198      0.0918    -1.30    0.1923
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.537
## Multiple R-squared:  0.0413, Adjusted R-squared:  0.0227
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 809 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.267  0.856  0.946  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      1.12e-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 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
## 108 107 123 109 132 80 114 117 113 130 185 172 163 157 155
## 2011 2012
## 139 150
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 24 22 21 26 28 20 27 28 42 49 57 74 63 62 74
## 2011 2012

```

```
## 63 80
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 22 15 18 17 21 14 19 23 31 39 42 57 48 37 52
## 2011 2012
## 50 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 = 27, df = 16, p-value = 0.04
```



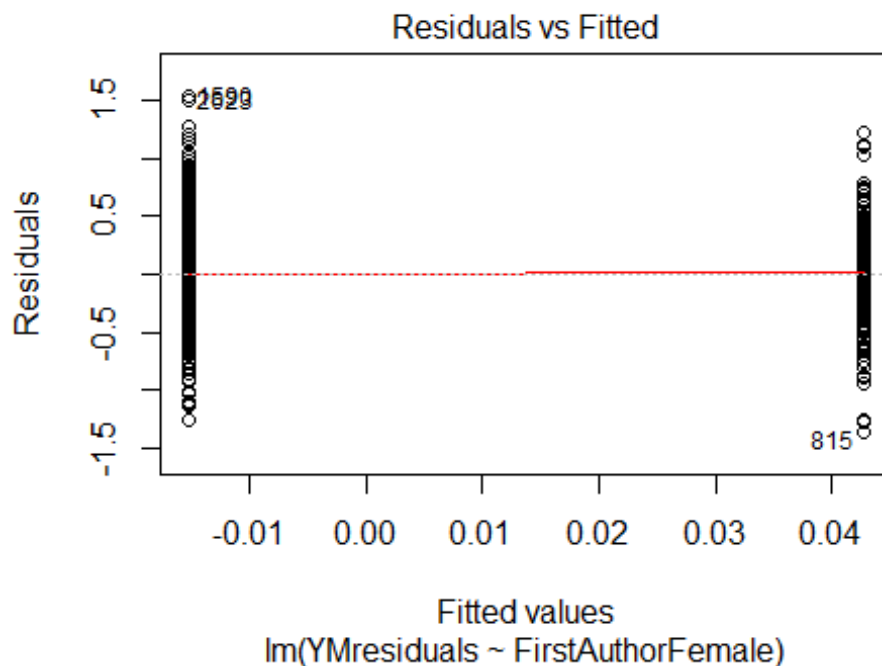
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.15, df = 1, p-value = 0.7

## [1] "Female first author team size 2018 geometric mean: 3.86073926338072"
## [1] "Male first author team size 2018 geometric mean: 3.13621696812939"

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

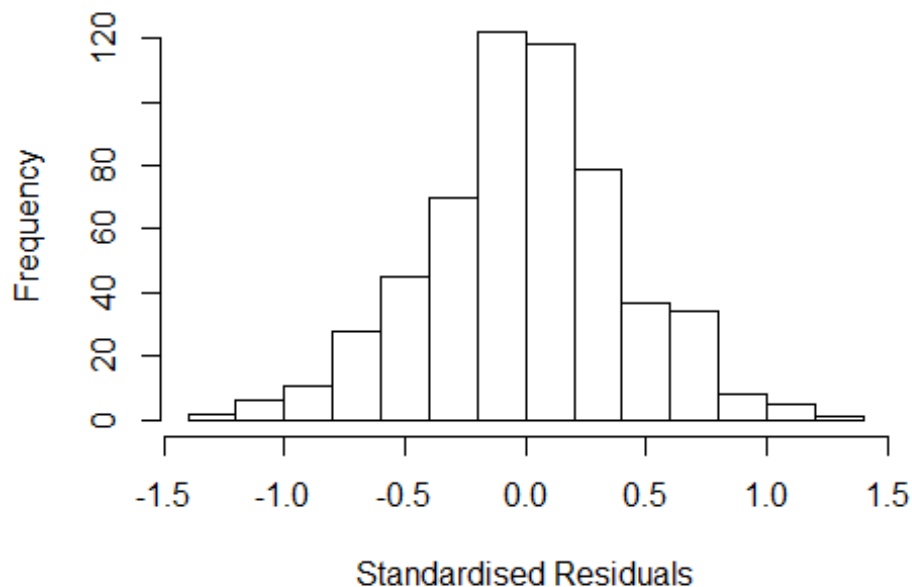
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 320, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.79375509213377"
## [1] "Male last author team size 2018 geometric mean: 3.53737589270767"

## Warning in wilcox.test.default(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 = 0.09
## 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 1.164 1      1.079
## LastAuthorFemale  1.115 1      1.056
## UniqueAuthors    1.655 4      1.065
## Year              1.898 16     1.020
```

Residuals from first and last author and team size



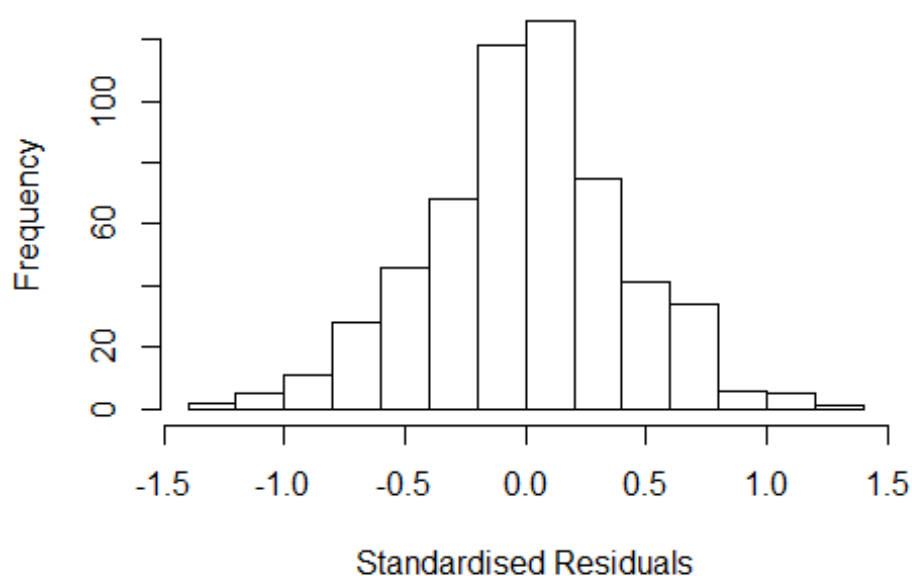
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27441 -0.24794 -0.00664  0.25422  1.25512
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1160    0.1104   10.11  <2e-16 ***
## FirstAuthorFemale1  0.0689    0.0396    1.74   0.083 .
## LastAuthorFemale1  0.0434    0.0434    1.00   0.318
## UniqueAuthors2     0.1161    0.0855    1.36   0.175
## UniqueAuthors3     0.0591    0.0854    0.69   0.489
## UniqueAuthors4     0.0895    0.0883    1.01   0.311
## UniqueAuthors5     0.1603    0.0896    1.79   0.074 .
## Year1997           0.0246    0.2021    0.12   0.903
## Year1998          -0.1486    0.1491   -1.00   0.319
## Year1999          -0.0168    0.1380   -0.12   0.903
```

```

## Year2000          -0.0327      0.1507   -0.22    0.828
## Year2001           0.0987      0.1738    0.57    0.570
## Year2002          -0.0542      0.1560   -0.35    0.728
## Year2003          -0.0268      0.1101   -0.24    0.808
## Year2004           0.0096      0.1158    0.08    0.934
## Year2005           0.0251      0.1031    0.24    0.808
## Year2006          -0.2120      0.1202   -1.76    0.078 .
## Year2007          -0.0112      0.1022   -0.11    0.913
## Year2008          -0.0596      0.1112   -0.54    0.592
## Year2009          -0.0432      0.1123   -0.38    0.701
## Year2010          -0.0996      0.0991   -1.00    0.315
## Year2011          -0.0621      0.1006   -0.62    0.537
## Year2012          -0.1260      0.1055   -1.19    0.233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.375
## Multiple R-squared:  0.0473, Adjusted R-squared:  0.00867
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 516 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.224  0.838  0.950  0.888  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          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 1.139 1          1.067
## LastAuthorFemale  1.078 1          1.038
## Year              1.228 16          1.006

```

Residuals from first and last author



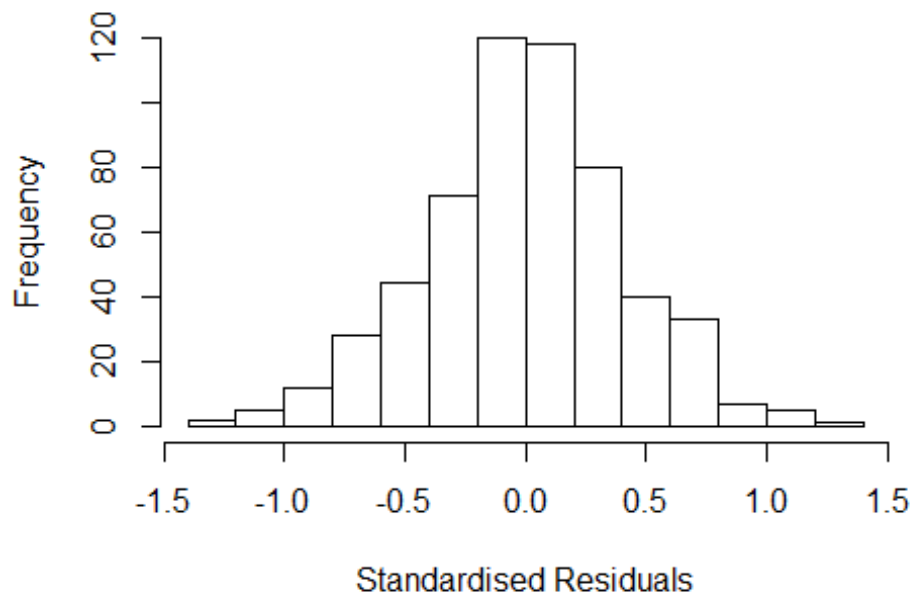
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26787 -0.25868  0.00581  0.25188  1.29068
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18807    0.09239   12.86  <2e-16 ***
## FirstAuthorFemale1  0.07980    0.03960    2.02   0.044 *
## LastAuthorFemale1  0.04795    0.04242    1.13   0.259
## Year1997         0.02974    0.20247    0.15   0.883
## Year1998        -0.13408    0.15534   -0.86   0.388
## Year1999         0.00745    0.13504    0.06   0.956
## Year2000        -0.01351    0.14682   -0.09   0.927
## Year2001         0.11100    0.16418    0.68   0.499
## Year2002        -0.04489    0.15683   -0.29   0.775
## Year2003        -0.01097    0.11207   -0.10   0.922
## Year2004         0.03854    0.11926    0.32   0.747
## Year2005         0.03740    0.10705    0.35   0.727
```

```

## Year2006      -0.20613    0.12261   -1.68    0.093 .
## Year2007      -0.00276    0.10557   -0.03    0.979
## Year2008      -0.04415    0.11467   -0.39    0.700
## Year2009      -0.01093    0.11374   -0.10    0.923
## Year2010      -0.07845    0.10315   -0.76    0.447
## Year2011      -0.04916    0.10330   -0.48    0.634
## Year2012      -0.08983    0.10705   -0.84    0.402
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0369, Adjusted R-squared:  0.00524
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 516 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.225  0.838  0.952  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      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 1.135 1      1.065
## Year      1.135 16      1.004

```


Residuals from first author



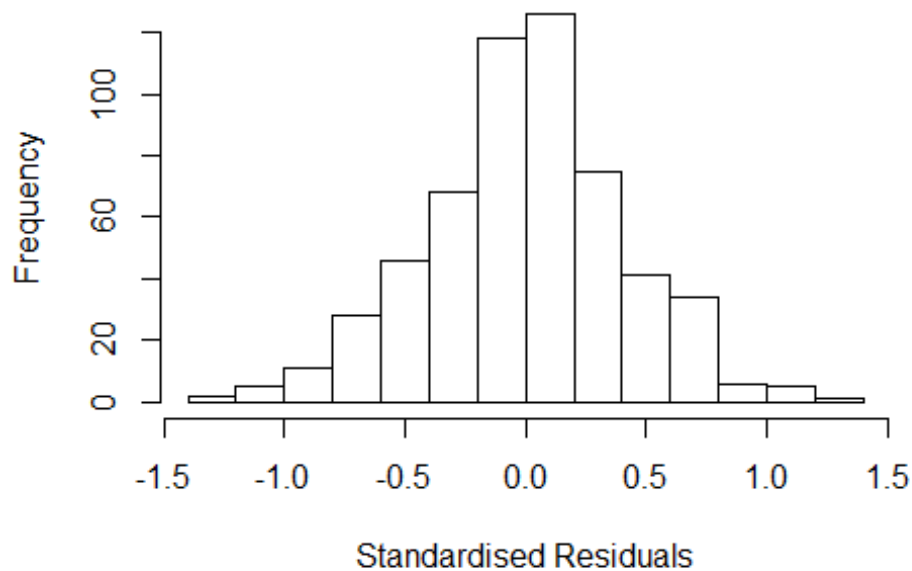
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27058 -0.25772 0.00349 0.24672 1.27945
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.190642 0.091633 12.99 <2e-16 ***
## FirstAuthorFemale1 0.079935 0.039719 2.01 0.045 *
## Year1997 0.039647 0.198259 0.20 0.842
## Year1998 -0.132679 0.154300 -0.86 0.390
## Year1999 0.019837 0.135036 0.15 0.883
## Year2000 -0.000384 0.146197 0.00 0.998
## Year2001 0.123113 0.161224 0.76 0.445
## Year2002 -0.044163 0.158208 -0.28 0.780
## Year2003 -0.011294 0.111731 -0.10 0.920
## Year2004 0.045978 0.117934 0.39 0.697
## Year2005 0.042914 0.106181 0.40 0.686
## Year2006 -0.204570 0.122097 -1.68 0.094 .
```

```

## Year2007          0.005905    0.105116    0.06    0.955
## Year2008          -0.034274    0.114067   -0.30    0.764
## Year2009          -0.004555    0.112638   -0.04    0.968
## Year2010          -0.069255    0.102207   -0.68    0.498
## Year2011          -0.043729    0.102951   -0.42    0.671
## Year2012          -0.086472    0.106325   -0.81    0.416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.382
## Multiple R-squared:  0.035, Adjusted R-squared:  0.00507
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 508 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.238  0.835   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      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 1.081 1          1.040
## Year            1.081 16          1.002

```

Residuals from last author



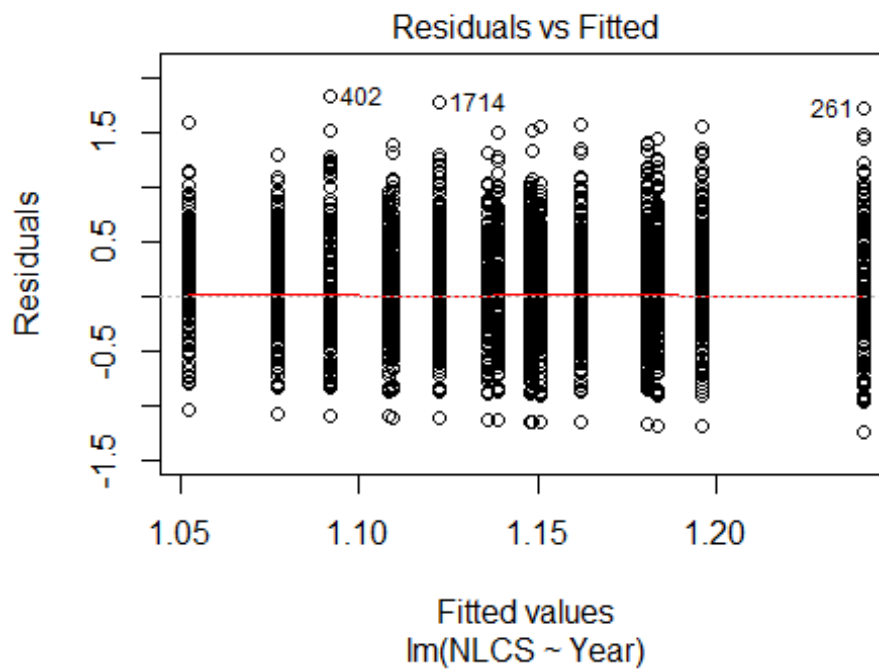
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24614 -0.25669  0.00124  0.24826  1.26097
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.197123   0.093596   12.79  <2e-16 ***
## LastAuthorFemale1  0.047932   0.042966    1.12    0.27
## Year1997         0.017524   0.197233    0.09    0.93
## Year1998        -0.116803   0.156606   -0.75    0.46
## Year1999         0.027634   0.139738    0.20    0.84
## Year2000         0.000906   0.145838    0.01    1.00
## Year2001         0.117346   0.160664    0.73    0.47
## Year2002        -0.050464   0.156241   -0.32    0.75
## Year2003         0.001235   0.114536    0.01    0.99
## Year2004         0.057025   0.119059    0.48    0.63
## Year2005         0.049017   0.108245    0.45    0.65
## Year2006        -0.187294   0.123413   -1.52    0.13
```

```

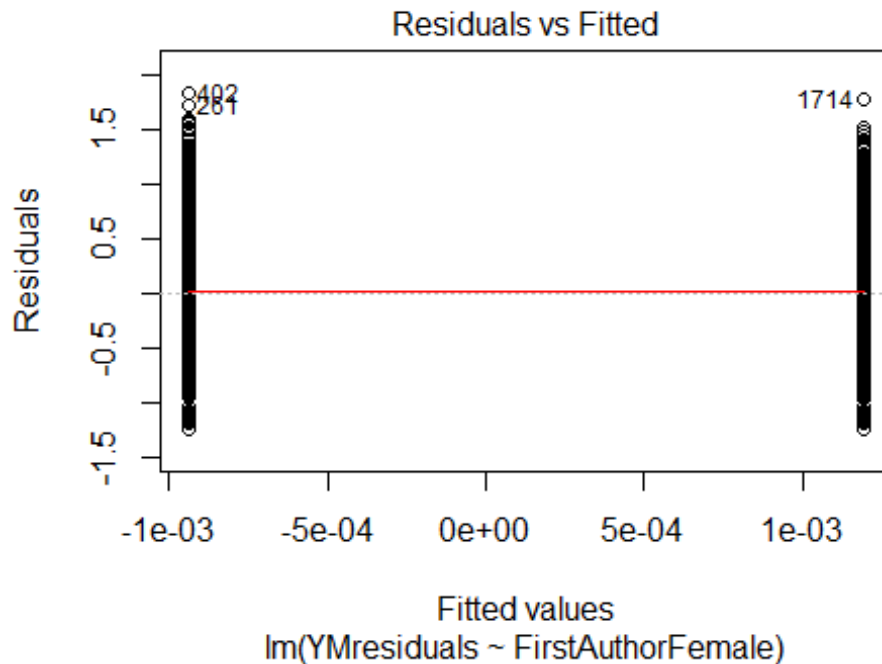
## Year2007      0.017906  0.107113  0.17  0.87
## Year2008     -0.029992  0.115022 -0.26  0.79
## Year2009      0.002952  0.115781  0.03  0.98
## Year2010     -0.070518  0.104470 -0.68  0.50
## Year2011     -0.039641  0.104694 -0.38  0.71
## Year2012     -0.077343  0.108945 -0.71  0.48
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.0291, Adjusted R-squared:  -0.000974
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 520 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.278  0.849  0.951  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      1.77e-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: 566"
## [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
##  215  198  232  205  238  254  250  225  223  234  319  324  318  280  324
## 2011 2012
##  328  353
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  174  158  178  160  176  181  205  189  187  197  266  276  266  242  272
## 2011 2012

```

```
## 278 305
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 155 143 163 147 163 167 190 173 176 178 228 257 235 217 242
## 2011 2012
## 250 272
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 70, df = 16, p-value = 9e-09
```

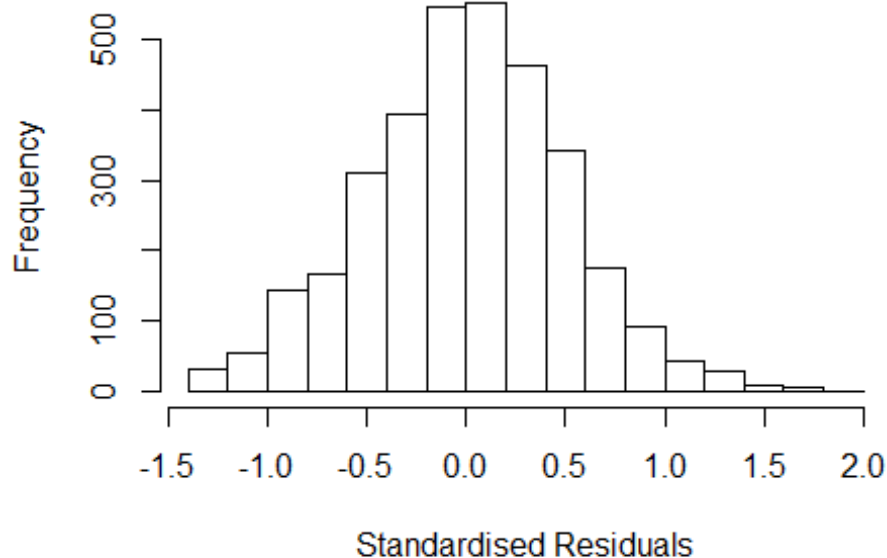


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.5, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.7709168455417"
## [1] "Male first author team size 2018 geometric mean: 2.38859426969158"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14000, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.72227872341041"
## [1] "Male last author team size 2018 geometric mean: 2.48276638029818"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 13000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.121 1 1.059
## LastAuthorFemale 1.091 1 1.045
## UniqueAuthors 1.176 4 1.020
## Year 1.207 16 1.006
```

Residuals from first and last author and team size



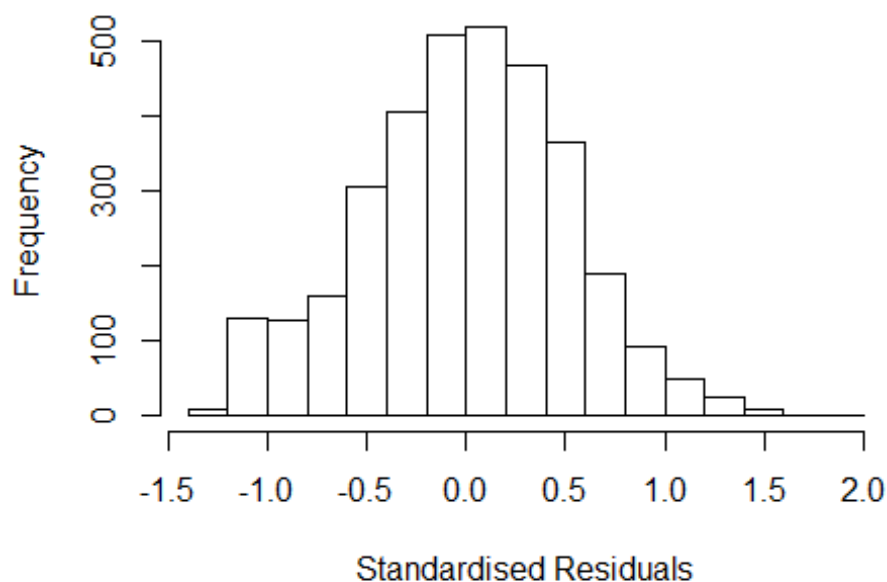
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3737 -0.3371 0.0126 0.3338 1.8930
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0730 0.0578 18.55 <2e-16 ***
## FirstAuthorFemale1 -0.0179 0.0187 -0.96 0.3368
## LastAuthorFemale1 -0.0323 0.0186 -1.74 0.0825 .
## UniqueAuthors2 0.2218 0.0253 8.76 <2e-16 ***
## UniqueAuthors3 0.2685 0.0270 9.94 <2e-16 ***
## UniqueAuthors4 0.2925 0.0326 8.98 <2e-16 ***
## UniqueAuthors5 0.3331 0.0335 9.95 <2e-16 ***
## Year1997 -0.1623 0.0760 -2.13 0.0329 *
## Year1998 -0.1045 0.0654 -1.60 0.1101
## Year1999 -0.1256 0.0697 -1.80 0.0717 .
```

```

## Year2000          -0.0895      0.0674   -1.33   0.1844
## Year2001          -0.1288      0.0674   -1.91   0.0562 .
## Year2002          -0.1755      0.0695   -2.53   0.0116 *
## Year2003          -0.1301      0.0641   -2.03   0.0424 *
## Year2004          -0.0978      0.0671   -1.46   0.1450
## Year2005          -0.1141      0.0643   -1.77   0.0760 .
## Year2006          -0.1550      0.0617   -2.51   0.0120 *
## Year2007          -0.1701      0.0621   -2.74   0.0062 **
## Year2008          -0.1174      0.0615   -1.91   0.0564 .
## Year2009          -0.1057      0.0646   -1.64   0.1019
## Year2010          -0.0697      0.0638   -1.09   0.2748
## Year2011          -0.0848      0.0639   -1.33   0.1846
## Year2012          -0.0578      0.0642   -0.90   0.3685
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.489
## Multiple R-squared:  0.0586, Adjusted R-squared:  0.0524
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 262 weights are ~= 1. The remaining 3094 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.101  0.872  0.949   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          2.98e-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 1.095 1 1.046
## LastAuthorFemale 1.055 1 1.027
## Year 1.063 16 1.002

```


Residuals from first and last author



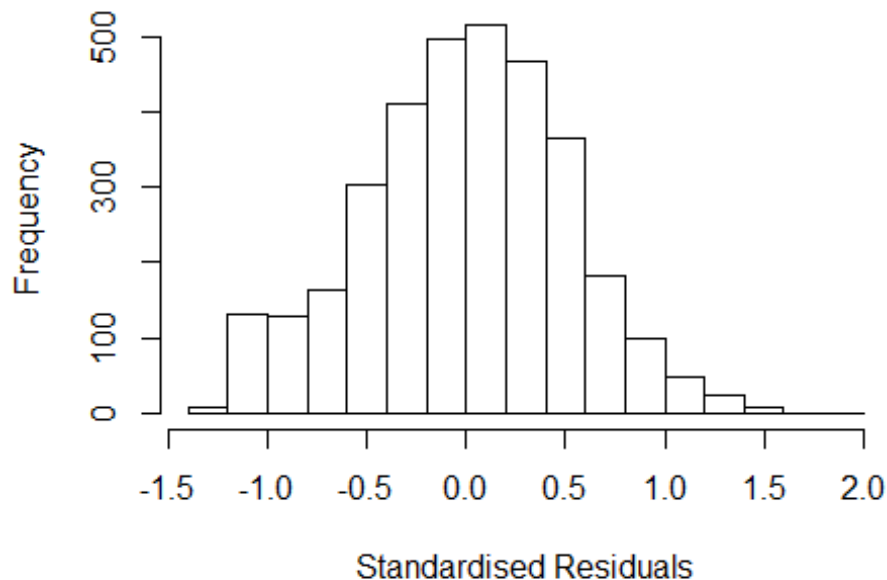
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2569 -0.3441 0.0155 0.3389 1.8072
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.24794 0.05399 23.12 <2e-16 ***
## FirstAuthorFemale1 0.00896 0.01901 0.47 0.6375
## LastAuthorFemale1 -0.01882 0.01895 -0.99 0.3208
## Year1997 -0.18702 0.07840 -2.39 0.0171 *
## Year1998 -0.13023 0.06627 -1.97 0.0495 *
## Year1999 -0.14241 0.06989 -2.04 0.0417 *
## Year2000 -0.10852 0.06845 -1.59 0.1130
## Year2001 -0.15805 0.06815 -2.32 0.0205 *
## Year2002 -0.19161 0.06903 -2.78 0.0055 **
## Year2003 -0.12825 0.06390 -2.01 0.0448 *
## Year2004 -0.08752 0.06783 -1.29 0.1970
## Year2005 -0.10932 0.06427 -1.70 0.0890 .
```

```

## Year2006      -0.13985    0.06161   -2.27   0.0233 *
## Year2007      -0.14826    0.06218   -2.38   0.0172 *
## Year2008      -0.08896    0.06162   -1.44   0.1489
## Year2009      -0.09653    0.06468   -1.49   0.1357
## Year2010      -0.06054    0.06367   -0.95   0.3418
## Year2011      -0.06795    0.06362   -1.07   0.2855
## Year2012      -0.05003    0.06402   -0.78   0.4345
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.505
## Multiple R-squared:  0.00851,    Adjusted R-squared:  0.00316
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 285 weights are ~= 1. The remaining 3071 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.173  0.872  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      2.98e-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 1.051 1      1.025
## Year      1.051 16      1.002

```

Residuals from first author



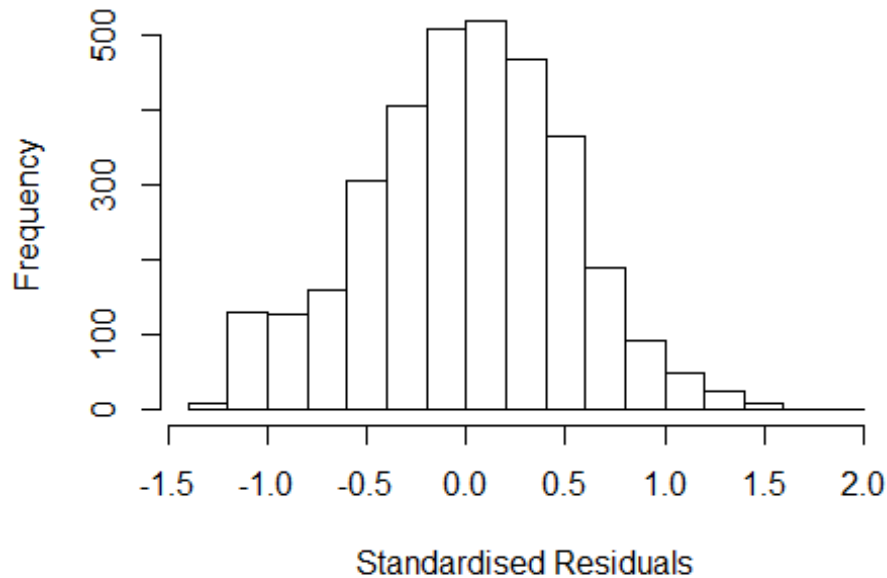
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2478 -0.3446  0.0155  0.3370  1.8176
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.24425    0.05401   23.04  <2e-16 ***
## FirstAuthorFemale1 0.00356    0.01871    0.19  0.8493
## Year1997      -0.18854    0.07868   -2.40  0.0166 *
## Year1998      -0.13121    0.06642   -1.98  0.0483 *
## Year1999      -0.14358    0.07000   -2.05  0.0403 *
## Year2000      -0.10867    0.06856   -1.59  0.1130
## Year2001      -0.15941    0.06832   -2.33  0.0197 *
## Year2002      -0.19225    0.06915   -2.78  0.0055 **
## Year2003      -0.13066    0.06395   -2.04  0.0411 *
## Year2004      -0.08846    0.06797   -1.30  0.1931
## Year2005      -0.11005    0.06442   -1.71  0.0877 .
## Year2006      -0.14155    0.06174   -2.29  0.0219 *
```

```

## Year2007          -0.14963    0.06234   -2.40    0.0164 *
## Year2008          -0.09078    0.06173   -1.47    0.1415
## Year2009          -0.09807    0.06483   -1.51    0.1304
## Year2010          -0.06214    0.06377   -0.97    0.3300
## Year2011          -0.07005    0.06372   -1.10    0.2717
## Year2012          -0.05169    0.06415   -0.81    0.4204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.505
## Multiple R-squared:  0.00824,    Adjusted R-squared:  0.00319
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 279 weights are ~= 1. The remaining 3077 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.168  0.871  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      2.98e-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 1.013 1          1.006
## Year              1.013 16          1.000

```

Residuals from last author



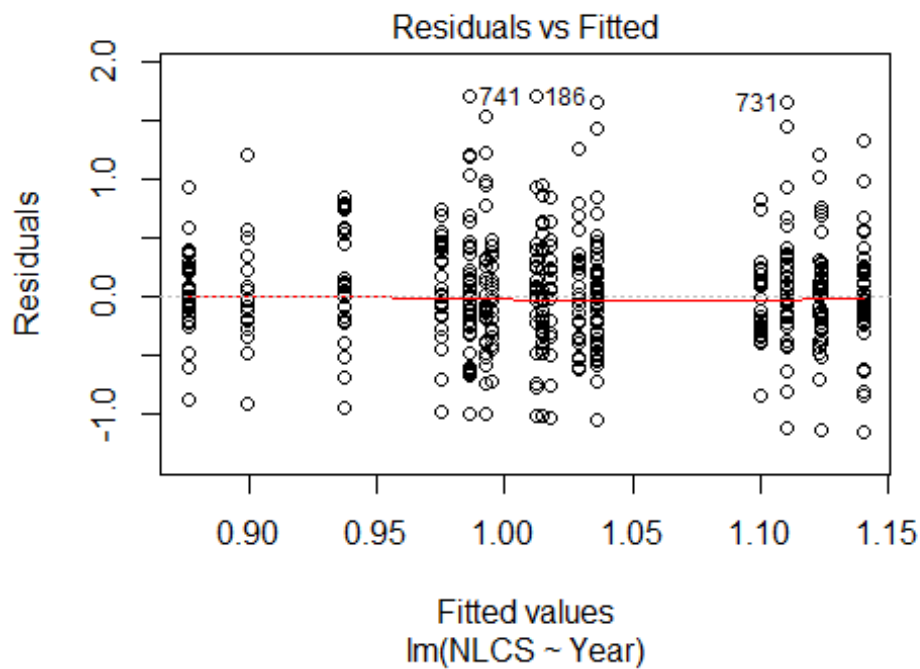
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2500 -0.3438  0.0145  0.3397  1.8134
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2500     0.0538   23.22  <2e-16 ***
## LastAuthorFemale1 -0.0162     0.0186   -0.87   0.3848
## Year1997         -0.1866     0.0784   -2.38   0.0174 *
## Year1998         -0.1304     0.0663   -1.97   0.0494 *
## Year1999         -0.1429     0.0699   -2.04   0.0410 *
## Year2000         -0.1078     0.0684   -1.58   0.1151
## Year2001         -0.1573     0.0682   -2.31   0.0211 *
## Year2002         -0.1909     0.0690   -2.77   0.0057 **
## Year2003         -0.1282     0.0639   -2.01   0.0449 *
## Year2004         -0.0865     0.0678   -1.28   0.2020
## Year2005         -0.1080     0.0641   -1.68   0.0923 .
## Year2006         -0.1384     0.0615   -2.25   0.0245 *
```

```

## Year2007          -0.1473      0.0621   -2.37   0.0178 *
## Year2008          -0.0882      0.0616   -1.43   0.1526
## Year2009          -0.0951      0.0646   -1.47   0.1410
## Year2010          -0.0595      0.0637   -0.94   0.3498
## Year2011          -0.0660      0.0634   -1.04   0.2982
## Year2012          -0.0484      0.0638   -0.76   0.4488
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.506
## Multiple R-squared:  0.00845,    Adjusted R-squared:  0.0034
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 280 weights are ~= 1. The remaining 3076 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.171  0.872  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      2.98e-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: 3356"
## [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
##   45   31   38   35   40   25   37   25   32   20   38   41   44   68   62
## 2011 2012
##   42   68
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   38   29   28   23   24   19   29   20   26   18   33   31   35   53   47
## 2011 2012

```

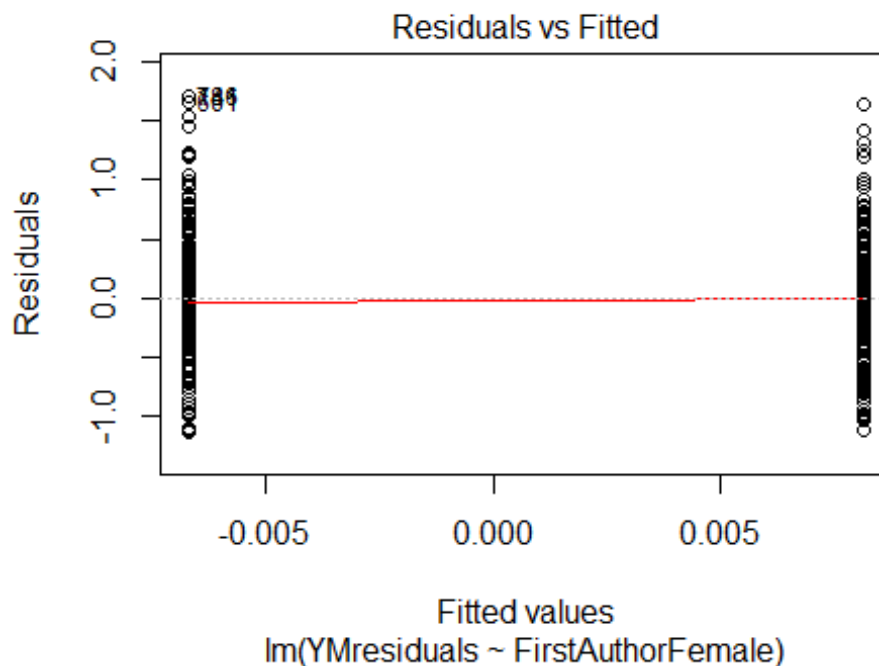
```
## 37 43
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 34 28 26 21 23 18 27 17 25 18 26 31 31 45 43
## 2011 2012
## 35 41
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 39, df = 16, p-value = 0.001
```



```
##
## 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: 2.34312421092553"
## [1] "Male first author team size 2018 geometric mean: 2.41460584598154"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

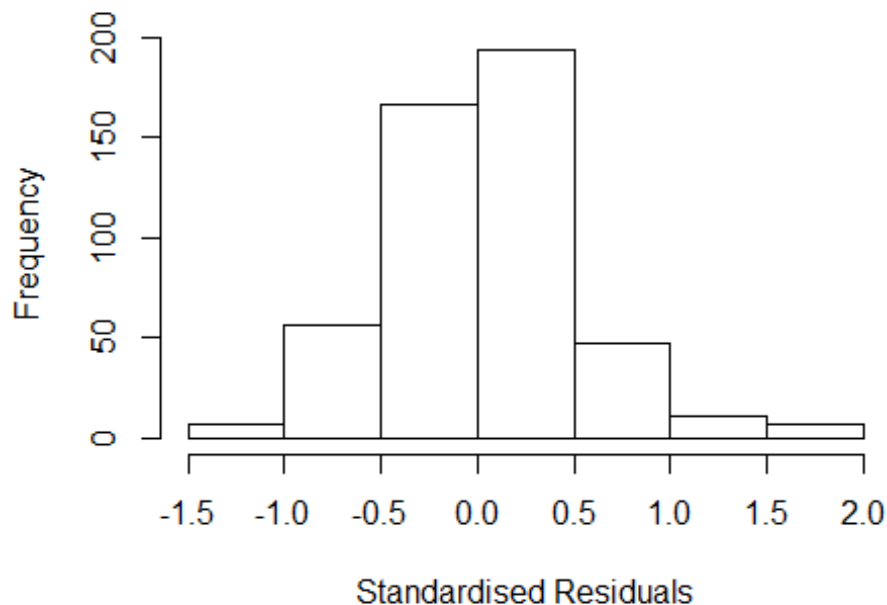
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.59985909370602"
## [1] "Male last author team size 2018 geometric mean: 2.10203647935734"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.851  1      1.360
## LastAuthorFemale  1.801  1      1.342
## UniqueAuthors    2.156  4      1.101
## Year              2.769 16      1.032
```


Residuals from first and last author and team size



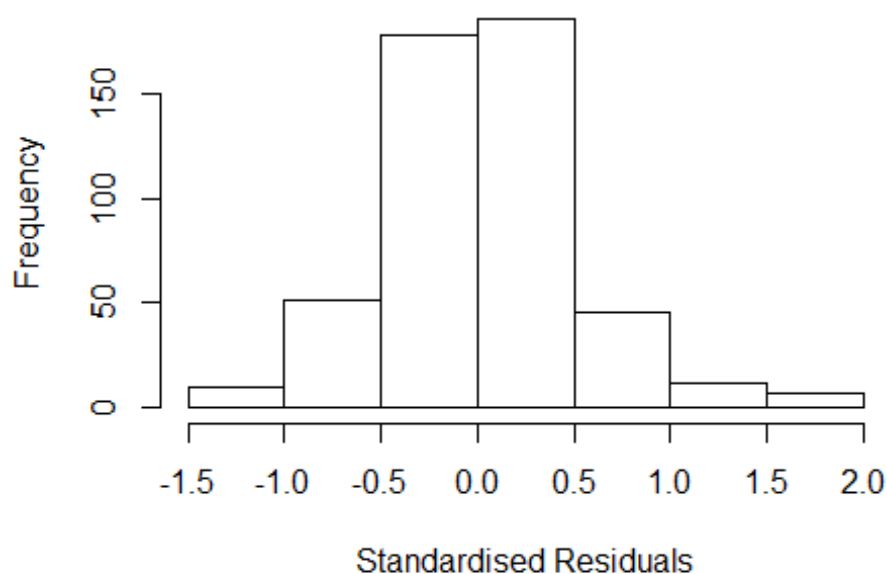
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.150 -0.284 0.015 0.286 1.782
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03096 0.07843 13.15 <2e-16 ***
## FirstAuthorFemale1 0.02995 0.05431 0.55 0.582
## LastAuthorFemale1 -0.05417 0.05262 -1.03 0.304
## UniqueAuthors2 -0.03801 0.06070 -0.63 0.532
## UniqueAuthors3 0.11616 0.06854 1.69 0.091 .
## UniqueAuthors4 -0.02334 0.07291 -0.32 0.749
## UniqueAuthors5 0.03984 0.08462 0.47 0.638
## Year1997 -0.16304 0.15021 -1.09 0.278
## Year1998 -0.08697 0.11841 -0.73 0.463
## Year1999 -0.04530 0.10458 -0.43 0.665
```

```

## Year2000      -0.05833    0.14724   -0.40    0.692
## Year2001      -0.12225    0.12657   -0.97    0.335
## Year2002      -0.06742    0.16196   -0.42    0.677
## Year2003       0.09390    0.14963    0.63    0.531
## Year2004       0.09127    0.10638    0.86    0.391
## Year2005       0.02461    0.12071    0.20    0.839
## Year2006      -0.05228    0.12568   -0.42    0.678
## Year2007      -0.15091    0.09819   -1.54    0.125
## Year2008       0.11904    0.10594    1.12    0.262
## Year2009      -0.04288    0.10209   -0.42    0.675
## Year2010       0.09171    0.10054    0.91    0.362
## Year2011       0.00366    0.14835    0.02    0.980
## Year2012      -0.12424    0.11596   -1.07    0.285
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.0497, Adjusted R-squared:  0.00482
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 436 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0395 0.8510 0.9480 0.8780 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          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-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.672 1      1.293
## LastAuthorFemale  1.598 1      1.264
## Year              1.611 16      1.015

```

Residuals from first and last author



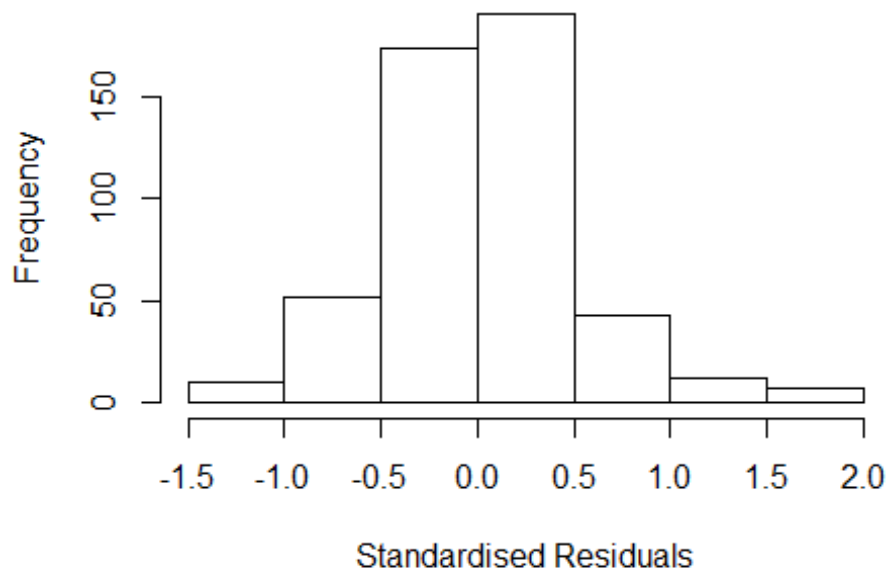
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1652 -0.2775 0.0111 0.3025 1.7692
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03491 0.08014 12.91 <2e-16 ***
## FirstAuthorFemale1 0.03892 0.05239 0.74 0.46
## LastAuthorFemale1 -0.05980 0.05093 -1.17 0.24
## Year1997 -0.13380 0.14875 -0.90 0.37
## Year1998 -0.09034 0.12360 -0.73 0.47
## Year1999 -0.04243 0.10693 -0.40 0.69
## Year2000 -0.04892 0.15007 -0.33 0.74
## Year2001 -0.11160 0.12269 -0.91 0.36
## Year2002 -0.06939 0.16113 -0.43 0.67
## Year2003 0.09654 0.14851 0.65 0.52
## Year2004 0.08515 0.10512 0.81 0.42
## Year2005 0.03441 0.12108 0.28 0.78
```

```

## Year2006      -0.05383    0.12578   -0.43    0.67
## Year2007      -0.15098    0.10183   -1.48    0.14
## Year2008       0.13032    0.10685    1.22    0.22
## Year2009      -0.03709    0.10405   -0.36    0.72
## Year2010       0.09450    0.10269    0.92    0.36
## Year2011      -0.00279    0.14351   -0.02    0.98
## Year2012      -0.11511    0.11922   -0.97    0.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.427
## Multiple R-squared:  0.0356, Adjusted R-squared:  -0.00132
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 441 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0469 0.8610 0.9480 0.8780 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          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-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.277 1      1.130
## Year      1.277 16      1.008

```

Residuals from first author

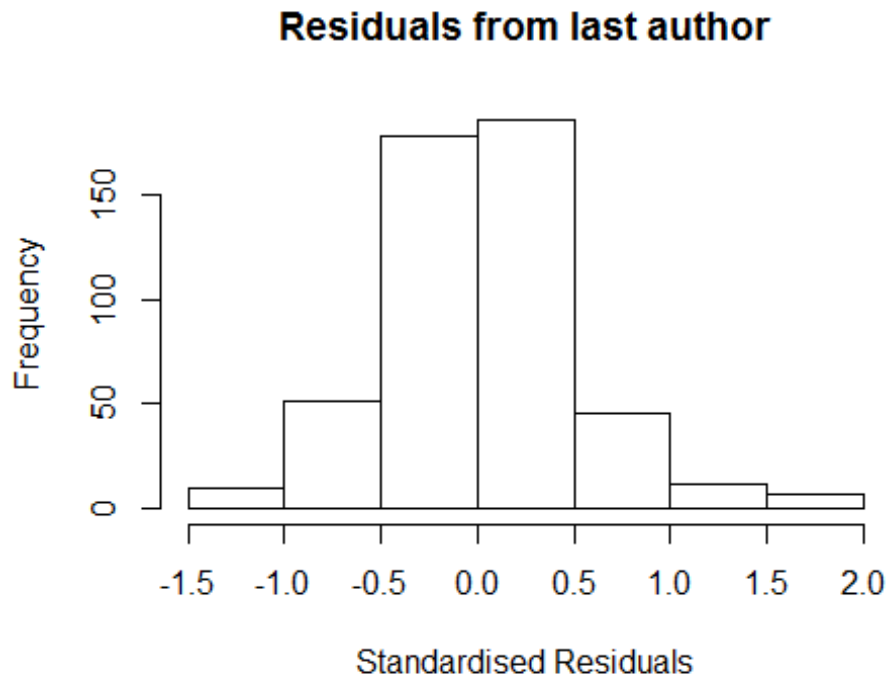


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1388 -0.2815 0.0158 0.2873 1.7803
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02172 0.08028 12.73 <2e-16 ***
## FirstAuthorFemale1 0.01255 0.04606 0.27 0.79
## Year1997 -0.13261 0.14885 -0.89 0.37
## Year1998 -0.08997 0.12352 -0.73 0.47
## Year1999 -0.04060 0.10767 -0.38 0.71
## Year2000 -0.04808 0.14855 -0.32 0.75
## Year2001 -0.12617 0.12255 -1.03 0.30
## Year2002 -0.07112 0.16091 -0.44 0.66
## Year2003 0.08305 0.14965 0.55 0.58
## Year2004 0.08286 0.10365 0.80 0.42
## Year2005 0.03013 0.12046 0.25 0.80
## Year2006 -0.04517 0.12644 -0.36 0.72
```

```

## Year2007      -0.14357    0.10116   -1.42    0.16
## Year2008      0.11709    0.10497    1.12    0.27
## Year2009     -0.03705    0.10400   -0.36    0.72
## Year2010      0.08957    0.10234    0.88    0.38
## Year2011     -0.00717    0.14151   -0.05    0.96
## Year2012     -0.11302    0.11821   -0.96    0.34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.0323, Adjusted R-squared:  -0.00267
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 444 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0409 0.8580 0.9520 0.8780 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-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.229 1          1.108
## Year            1.229 16          1.006

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1722 -0.2803 0.0117 0.2907 1.7550
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04332 0.07972 13.09 <2e-16 ***
## LastAuthorFemale1 -0.04269 0.04487 -0.95 0.34
## Year1997 -0.12968 0.14878 -0.87 0.38
## Year1998 -0.08641 0.12434 -0.69 0.49
## Year1999 -0.04603 0.10803 -0.43 0.67
## Year2000 -0.05582 0.14749 -0.38 0.71
## Year2001 -0.11785 0.12336 -0.96 0.34
## Year2002 -0.06569 0.15954 -0.41 0.68
## Year2003 0.08932 0.14710 0.61 0.54
## Year2004 0.09232 0.10500 0.88 0.38
## Year2005 0.03569 0.12096 0.30 0.77
## Year2006 -0.05006 0.12629 -0.40 0.69
```

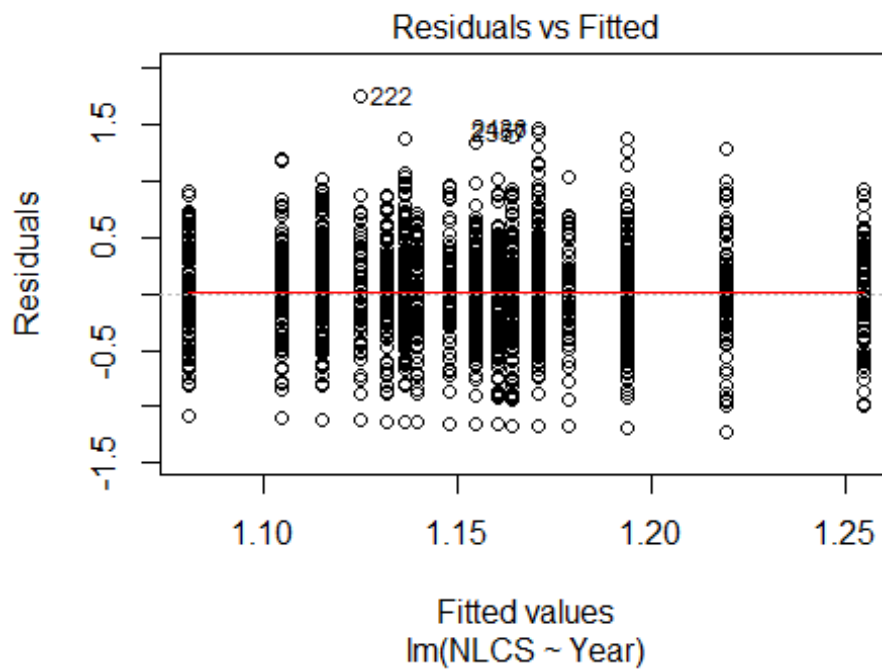
```

## Year2007          -0.14471      0.10161      -1.42      0.16
## Year2008           0.12891      0.10801       1.19      0.23
## Year2009          -0.02954      0.10362      -0.29      0.78
## Year2010           0.09669      0.10311       0.94      0.35
## Year2011          -0.00437      0.14361      -0.03      0.98
## Year2012          -0.10936      0.11891      -0.92      0.36
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.427
## Multiple R-squared:  0.0334, Adjusted R-squared:  -0.00147
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 444 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0524 0.8640 0.9500 0.8790 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.04e-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: 489"
## [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
## 106 84 93 81 97 113 116 110 103 121 148 179 171 186 244
## 2011 2012
## 245 221
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 85 63 71 64 55 78 84 95 86 100 122 141 144 152 199
## 2011 2012

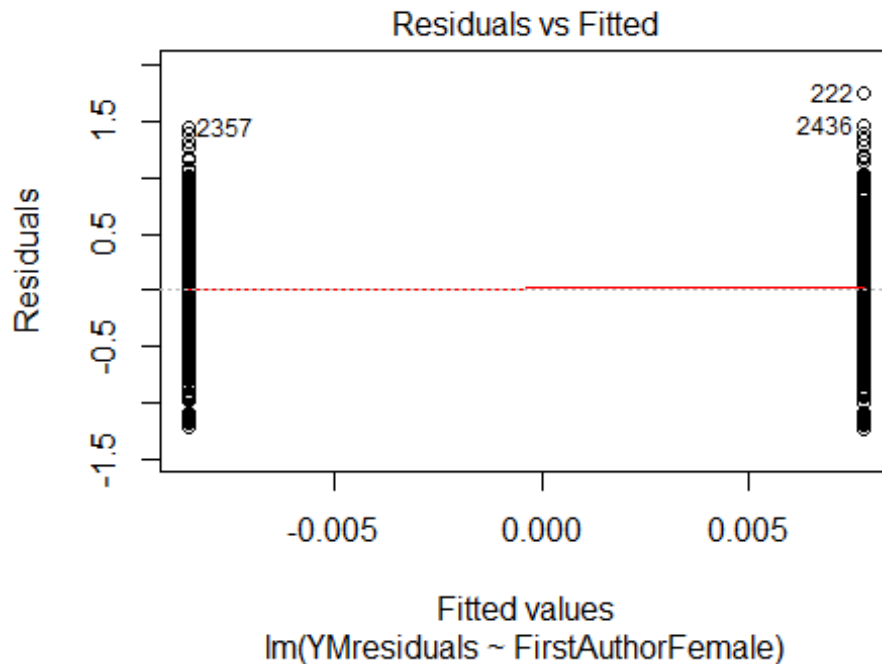
```



```
## 192 187
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 77 59 63 57 47 71 74 79 76 92 108 127 123 136 177
## 2011 2012
## 173 175
## [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
```

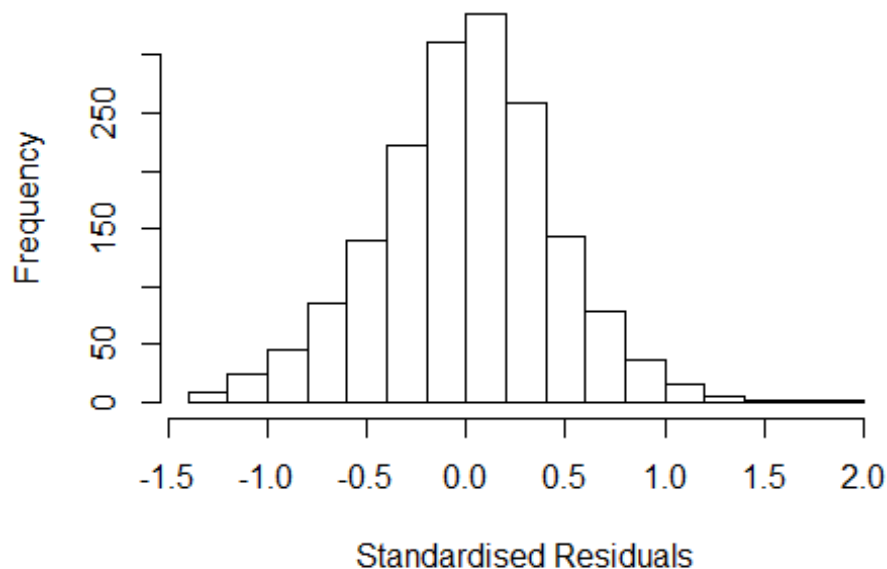


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.55, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 3.00967095967759"
## [1] "Male first author team size 2018 geometric mean: 2.17895775827836"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5200, p-value = 4e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.80428730789598"
## [1] "Male last author team size 2018 geometric mean: 2.63965184658661"
##
## 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 1.124 1      1.060
## LastAuthorFemale  1.144 1      1.070
## UniqueAuthors    1.255 4      1.029
## Year             1.318 16      1.009
```

Residuals from first and last author and team size



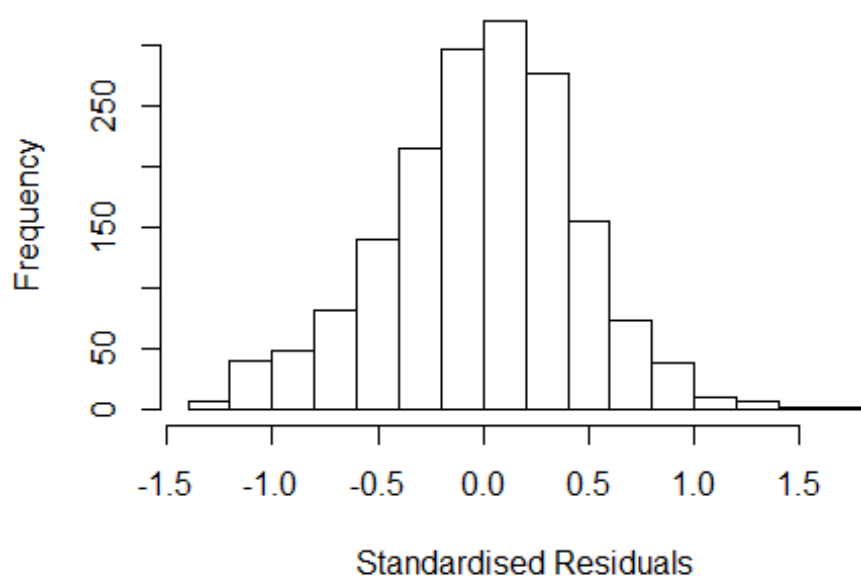
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2842 -0.2854 0.0143 0.2810 1.9180
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03895 0.05815 17.87 < 2e-16 ***
## FirstAuthorFemale1 -0.02415 0.02277 -1.06 0.289
## LastAuthorFemale1 -0.02246 0.02316 -0.97 0.332
## UniqueAuthors2 0.20612 0.03397 6.07 1.6e-09 ***
## UniqueAuthors3 0.26367 0.03608 7.31 4.1e-13 ***
## UniqueAuthors4 0.21406 0.03990 5.36 9.2e-08 ***
## UniqueAuthors5 0.26435 0.04135 6.39 2.1e-10 ***
## Year1997 -0.07796 0.08008 -0.97 0.330
## Year1998 -0.02075 0.08680 -0.24 0.811
## Year1999 -0.00797 0.09191 -0.09 0.931
```

```

## Year2000      0.01601    0.08583    0.19    0.852
## Year2001      0.03913    0.07878    0.50    0.619
## Year2002      0.01418    0.08082    0.18    0.861
## Year2003     -0.06021    0.07431   -0.81    0.418
## Year2004      0.03282    0.07959    0.41    0.680
## Year2005     -0.06125    0.07024   -0.87    0.383
## Year2006     -0.02832    0.06885   -0.41    0.681
## Year2007      0.01033    0.06631    0.16    0.876
## Year2008     -0.11066    0.06619   -1.67    0.095 .
## Year2009     -0.00874    0.06805   -0.13    0.898
## Year2010     -0.04795    0.06649   -0.72    0.471
## Year2011     -0.09114    0.06568   -1.39    0.165
## Year2012     -0.09169    0.06921   -1.32    0.185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.422
## Multiple R-squared:  0.0566, Adjusted R-squared:  0.0444
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 128 weights are ~= 1. The remaining 1586 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0034 0.8690 0.9530 0.8980 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "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  1.118 1      1.057
## Year              1.068 16      1.002

```

Residuals from first and last author



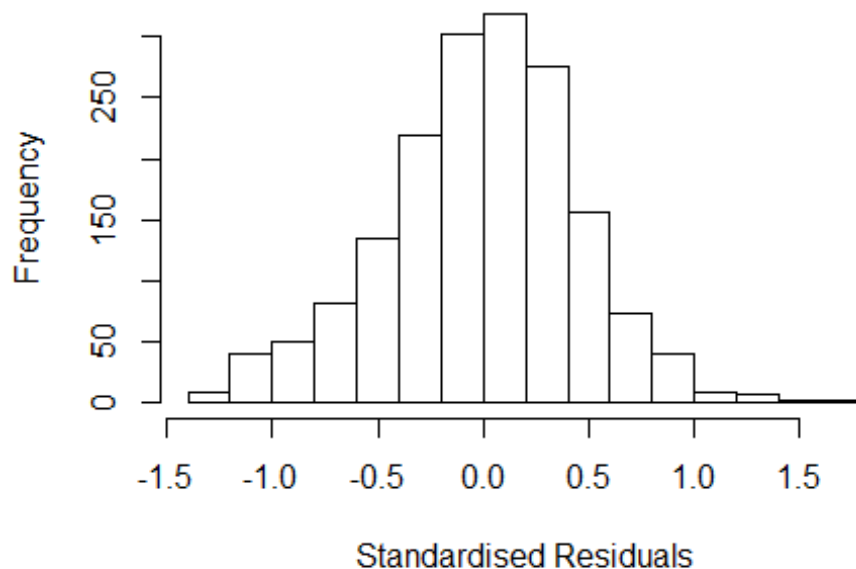
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2234 -0.2923 0.0155 0.2906 1.7783
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15482 0.05981 19.31 <2e-16 ***
## FirstAuthorFemale1 -0.00841 0.02307 -0.36 0.72
## LastAuthorFemale1 -0.01817 0.02357 -0.77 0.44
## Year1997 -0.05412 0.08384 -0.65 0.52
## Year1998 0.00319 0.08649 0.04 0.97
## Year1999 0.00391 0.09336 0.04 0.97
## Year2000 0.02629 0.08999 0.29 0.77
## Year2001 0.06122 0.08084 0.76 0.45
## Year2002 0.06857 0.08259 0.83 0.41
## Year2003 -0.01623 0.07661 -0.21 0.83
## Year2004 0.08872 0.08117 1.09 0.27
## Year2005 -0.01181 0.07319 -0.16 0.87
```

```

## Year2006      0.03556    0.07232    0.49    0.62
## Year2007      0.05463    0.07046    0.78    0.44
## Year2008     -0.04041    0.06989   -0.58    0.56
## Year2009      0.05617    0.07217    0.78    0.44
## Year2010      0.02752    0.06855    0.40    0.69
## Year2011     -0.01550    0.06815   -0.23    0.82
## Year2012     -0.00783    0.07111   -0.11    0.91
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.00755,    Adjusted R-squared:  -0.00299
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 128 weights are ~= 1. The remaining 1586 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0478 0.8650 0.9520 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



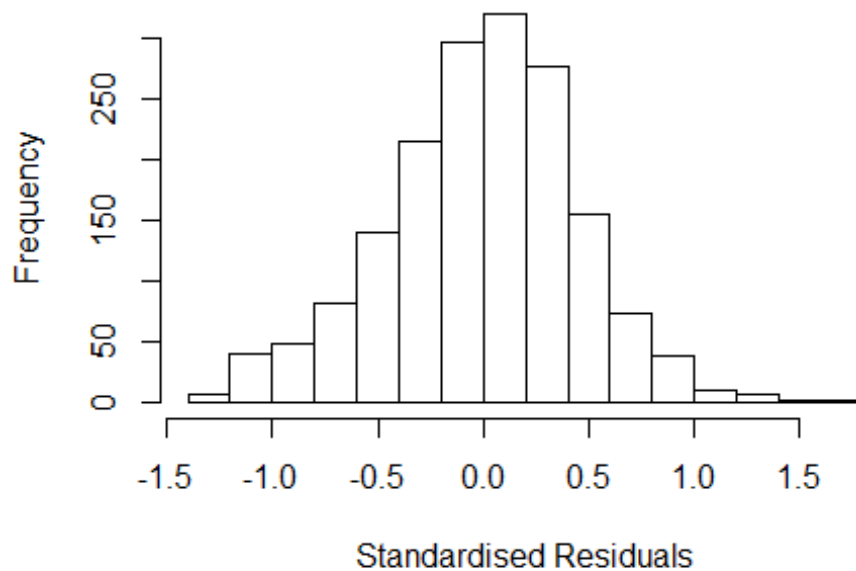
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2183 -0.2877 0.0152 0.2911 1.7835
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15133 0.05961 19.32 <2e-16 ***
## FirstAuthorFemale1 -0.01412 0.02236 -0.63 0.53
## Year1997 -0.05587 0.08372 -0.67 0.50
## Year1998 0.00219 0.08637 0.03 0.98
## Year1999 0.00397 0.09356 0.04 0.97
## Year2000 0.02375 0.08960 0.27 0.79
## Year2001 0.06155 0.08078 0.76 0.45
## Year2002 0.06698 0.08245 0.81 0.42
## Year2003 -0.01845 0.07647 -0.24 0.81
## Year2004 0.08662 0.08106 1.07 0.29
## Year2005 -0.01385 0.07313 -0.19 0.85
## Year2006 0.03382 0.07222 0.47 0.64
```

```

## Year2007          0.05382    0.07033    0.77    0.44
## Year2008          -0.04189    0.06978   -0.60    0.55
## Year2009          0.05356    0.07188    0.75    0.46
## Year2010          0.02461    0.06825    0.36    0.72
## Year2011          -0.01564    0.06813   -0.23    0.82
## Year2012          -0.01066    0.07067   -0.15    0.88
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.0072, Adjusted R-squared:  -0.00275
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1590 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0455 0.8640 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      5.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.045 1          1.022
## Year            1.045 16          1.001

```


Residuals from last author



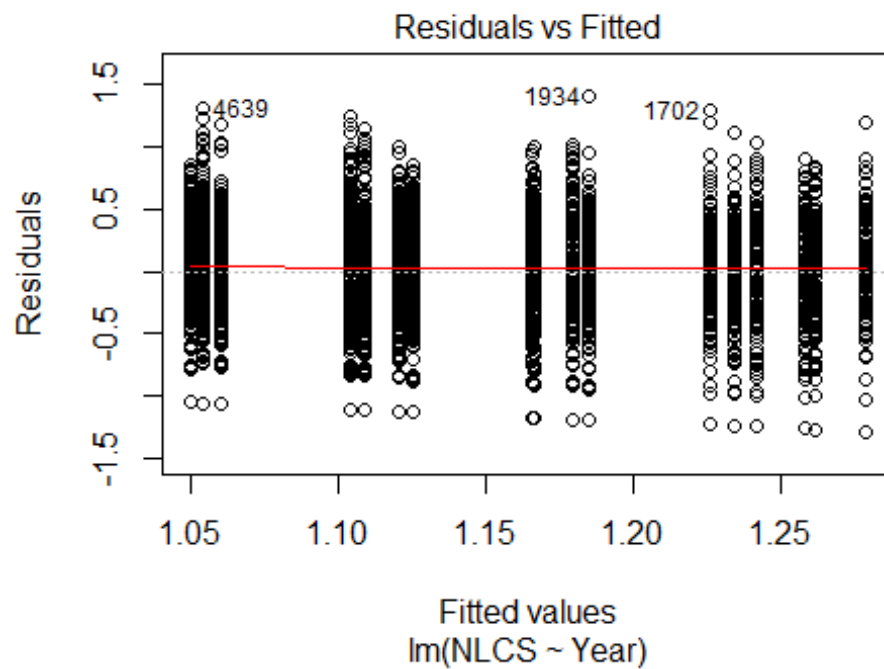
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2205 -0.2918 0.0141 0.2900 1.7800
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15313 0.05960 19.35 <2e-16 ***
## LastAuthorFemale1 -0.02083 0.02284 -0.91 0.36
## Year1997 -0.05414 0.08390 -0.65 0.52
## Year1998 0.00244 0.08653 0.03 0.98
## Year1999 0.00263 0.09328 0.03 0.98
## Year2000 0.02534 0.08995 0.28 0.78
## Year2001 0.06066 0.08079 0.75 0.45
## Year2002 0.06739 0.08247 0.82 0.41
## Year2003 -0.01689 0.07671 -0.22 0.83
## Year2004 0.08788 0.08112 1.08 0.28
## Year2005 -0.01305 0.07316 -0.18 0.86
## Year2006 0.03468 0.07226 0.48 0.63
```

```

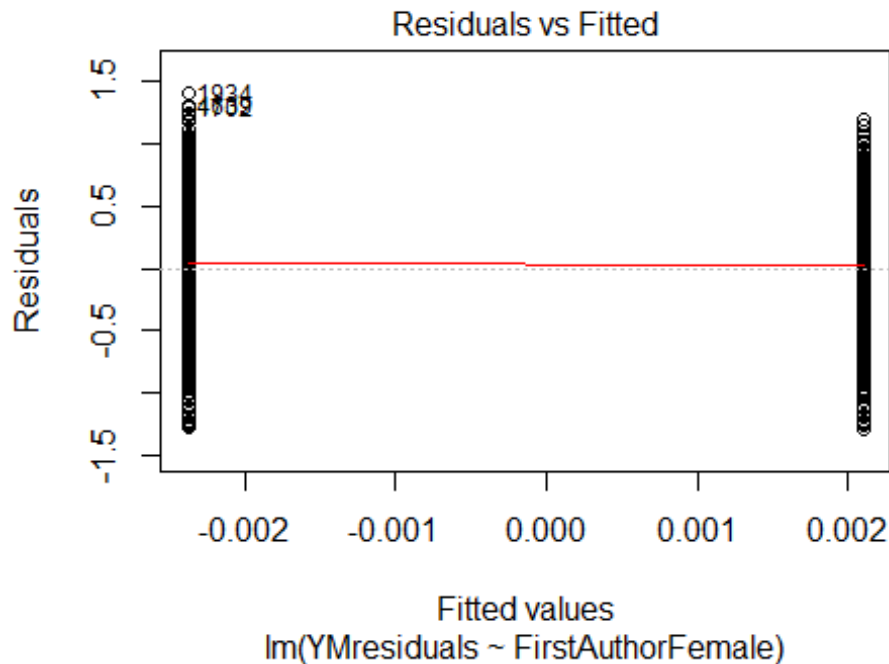
## Year2007      0.05361      0.07043      0.76      0.45
## Year2008     -0.04151      0.06980     -0.59      0.55
## Year2009      0.05429      0.07199      0.75      0.45
## Year2010      0.02627      0.06849      0.38      0.70
## Year2011     -0.01764      0.06793     -0.26      0.80
## Year2012     -0.00938      0.07107     -0.13      0.90
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.00747,    Adjusted R-squared:  -0.00248
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 130 weights are ~= 1. The remaining 1584 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0466 0.8650 0.9520 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.83e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1714"
## [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
## 167 156 157 168 174 213 196 176 199 178 246 270 255 310 326
## 2011 2012
## 320 360
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 135 117 119 133 128 151 158 142 152 150 212 234 220 274 279
## 2011 2012

```

```
## 263 310
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 123 101 108 112 110 130 142 123 140 129 190 213 199 235 247
## 2011 2012
## 240 278
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```

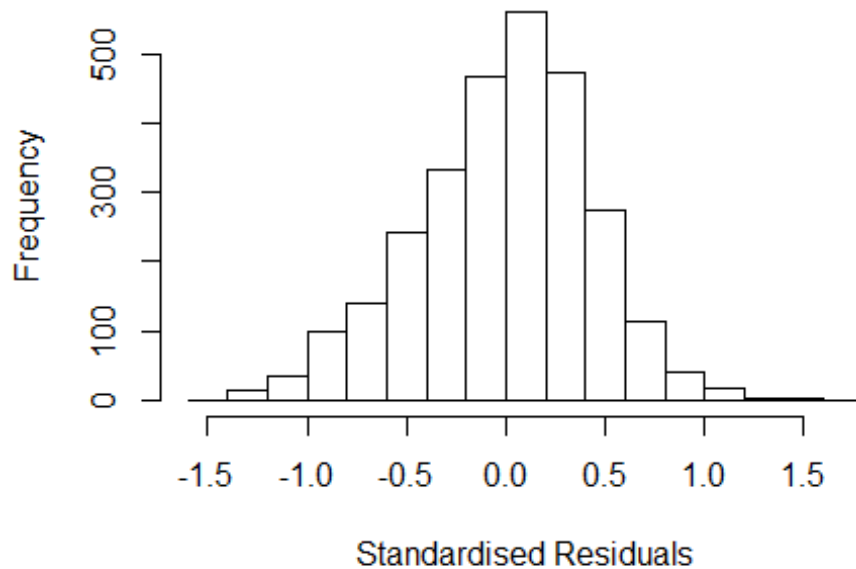


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 18, df = 1, p-value = 2e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 3.36926383546452"
## [1] "Male first author team size 2018 geometric mean: 3.88411141013961"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7500, p-value = 0.009
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.32696354087939"
## [1] "Male last author team size 2018 geometric mean: 3.76537667692571"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8400, 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 1.074 1          1.037
## LastAuthorFemale  1.057 1          1.028
## UniqueAuthors    1.177 4          1.021
## Year             1.210 16          1.006
```

Residuals from first and last author and team size



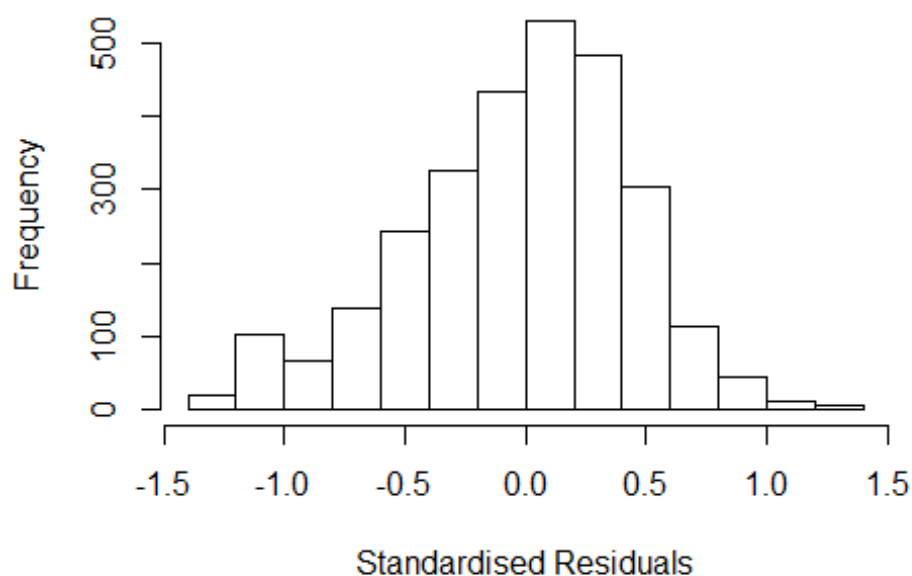
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4017 -0.2928  0.0263  0.2891  1.7043
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.92379    0.05160   17.90  <2e-16 ***
## FirstAuthorFemale1 -0.01841    0.01750   -1.05  0.2929
## LastAuthorFemale1 -0.02098    0.01716   -1.22  0.2217
## UniqueAuthors2     0.33903    0.03202   10.59  <2e-16 ***
## UniqueAuthors3     0.34045    0.03299   10.32  <2e-16 ***
## UniqueAuthors4     0.38757    0.03287   11.79  <2e-16 ***
## UniqueAuthors5     0.42466    0.03271   12.98  <2e-16 ***
## Year1997           0.05860    0.06818    0.86  0.3902
## Year1998           0.09142    0.06149    1.49  0.1372
## Year1999           0.06002    0.06125    0.98  0.3272
```

```

## Year2000      0.07163      0.06081      1.18      0.2389
## Year2001      0.00187      0.06374      0.03      0.9766
## Year2002      0.06394      0.05909      1.08      0.2793
## Year2003      0.04417      0.05723      0.77      0.4403
## Year2004     -0.04709      0.06125     -0.77      0.4421
## Year2005     -0.05210      0.06221     -0.84      0.4024
## Year2006     -0.08266      0.05604     -1.48      0.1403
## Year2007     -0.07165      0.05527     -1.30      0.1950
## Year2008     -0.15424      0.05602     -2.75      0.0059 **
## Year2009     -0.08510      0.05677     -1.50      0.1340
## Year2010     -0.09907      0.05418     -1.83      0.0676 .
## Year2011     -0.14409      0.05448     -2.64      0.0082 **
## Year2012     -0.17129      0.05497     -3.12      0.0019 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.423
## Multiple R-squared:  0.118, Adjusted R-squared:  0.111
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 220 weights are ~= 1. The remaining 2600 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0679 0.8670 0.9490 0.9010 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.55e-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 1.049 1      1.024
## LastAuthorFemale  1.040 1      1.020
## Year              1.072 16      1.002

```

Residuals from first and last author



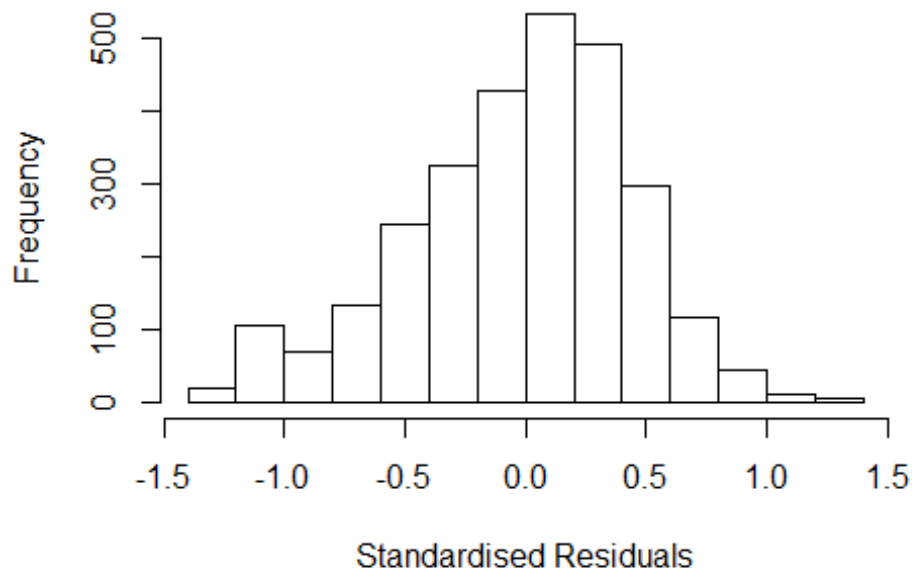
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2875 -0.3048 0.0369 0.2960 1.3951
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17940 0.05023 23.48 <2e-16 ***
## FirstAuthorFemale1 0.01145 0.01811 0.63 0.527
## LastAuthorFemale1 -0.01371 0.01772 -0.77 0.439
## Year1997 0.06820 0.07085 0.96 0.336
## Year1998 0.10806 0.06444 1.68 0.094 .
## Year1999 0.09388 0.06348 1.48 0.139
## Year2000 0.09087 0.06586 1.38 0.168
## Year2001 0.03561 0.06615 0.54 0.590
## Year2002 0.08161 0.06079 1.34 0.180
## Year2003 0.07496 0.06251 1.20 0.231
## Year2004 0.00646 0.06227 0.10 0.917
## Year2005 0.00526 0.06470 0.08 0.935
```

```

## Year2006      -0.02850    0.05937   -0.48    0.631
## Year2007      -0.03610    0.05791   -0.62    0.533
## Year2008      -0.10829    0.05971   -1.81    0.070 .
## Year2009      -0.05146    0.06018   -0.86    0.393
## Year2010      -0.05714    0.05756   -0.99    0.321
## Year2011      -0.09981    0.05706   -1.75    0.080 .
## Year2012      -0.11504    0.05831   -1.97    0.049 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0254, Adjusted R-squared:  0.0191
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 211 weights are ~= 1. The remaining 2609 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.281  0.871  0.950  0.898  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      3.55e-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 1.044 1      1.022
## Year      1.044 16      1.001

```


Residuals from first author



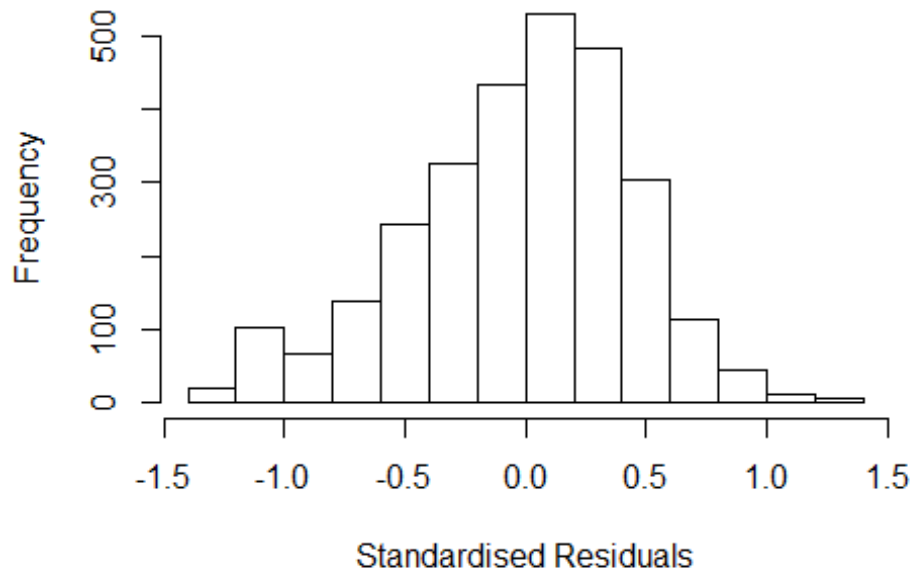
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2918 -0.3016 0.0354 0.2982 1.3995
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17494 0.04975 23.62 <2e-16 ***
## FirstAuthorFemale1 0.00854 0.01818 0.47 0.639
## Year1997 0.06813 0.07090 0.96 0.337
## Year1998 0.10834 0.06448 1.68 0.093 .
## Year1999 0.09581 0.06340 1.51 0.131
## Year2000 0.09065 0.06584 1.38 0.169
## Year2001 0.03565 0.06612 0.54 0.590
## Year2002 0.08126 0.06069 1.34 0.181
## Year2003 0.07517 0.06248 1.20 0.229
## Year2004 0.00659 0.06223 0.11 0.916
## Year2005 0.00495 0.06467 0.08 0.939
## Year2006 -0.02878 0.05936 -0.48 0.628
```

```

## Year2007          -0.03691    0.05785   -0.64    0.524
## Year2008          -0.10871    0.05965   -1.82    0.068 .
## Year2009          -0.05168    0.06020   -0.86    0.391
## Year2010          -0.05772    0.05753   -1.00    0.316
## Year2011          -0.09993    0.05704   -1.75    0.080 .
## Year2012          -0.11642    0.05816   -2.00    0.045 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0252, Adjusted R-squared:  0.0193
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 220 weights are ~= 1. The remaining 2600 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.278  0.870  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      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 1.035 1          1.017
## Year            1.035 16          1.001

```

Residuals from last author



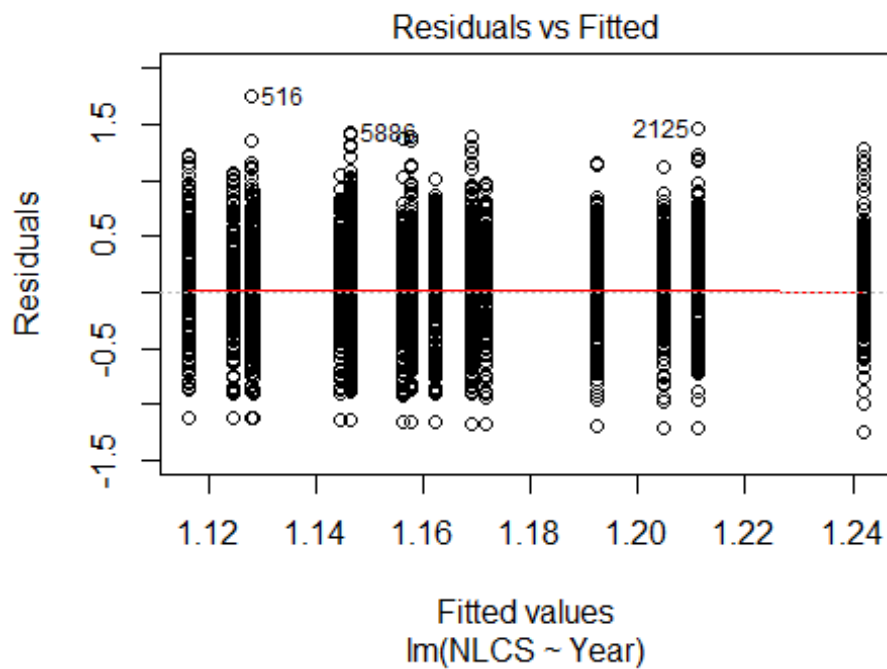
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2925 -0.3054 0.0368 0.2950 1.3903
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18245 0.04988 23.71 <2e-16 ***
## LastAuthorFemale1 -0.01131 0.01781 -0.63 0.526
## Year1997 0.06890 0.07088 0.97 0.331
## Year1998 0.11003 0.06439 1.71 0.088 .
## Year1999 0.09571 0.06343 1.51 0.131
## Year2000 0.09229 0.06585 1.40 0.161
## Year2001 0.03687 0.06614 0.56 0.577
## Year2002 0.08321 0.06072 1.37 0.171
## Year2003 0.07725 0.06229 1.24 0.215
## Year2004 0.00829 0.06213 0.13 0.894
## Year2005 0.00690 0.06464 0.11 0.915
## Year2006 -0.02667 0.05930 -0.45 0.653
```

```

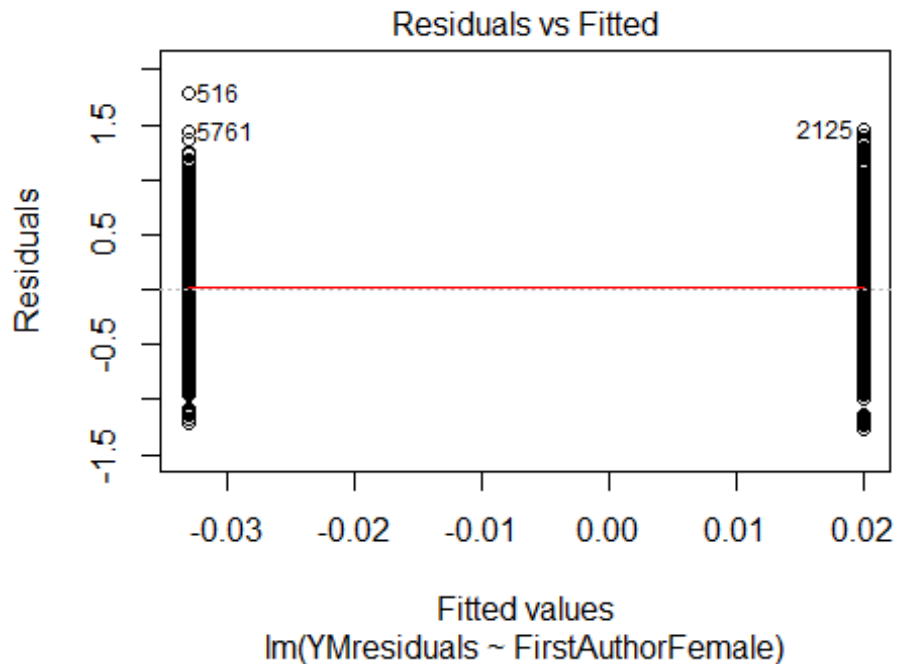
## Year2007          -0.03311      0.05756   -0.58      0.565
## Year2008          -0.10589      0.05956   -1.78      0.076 .
## Year2009          -0.04914      0.06001   -0.82      0.413
## Year2010          -0.05428      0.05730   -0.95      0.344
## Year2011          -0.09703      0.05679   -1.71      0.088 .
## Year2012          -0.11191      0.05799   -1.93      0.054 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0252, Adjusted R-squared:  0.0193
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 219 weights are ~= 1. The remaining 2601 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.284  0.870   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      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: 2820"
## [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
## 249 235 212 182 247 279 244 235 231 230 275 339 315 412 355
## 2011 2012
## 434 455
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 203 181 158 131 193 174 211 197 189 199 250 286 264 347 292
## 2011 2012

```

```
## 374 376
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 181 165 149 122 179 159 201 180 174 173 212 266 238 310 265
## 2011 2012
## 342 335
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 56, df = 16, p-value = 3e-06
```

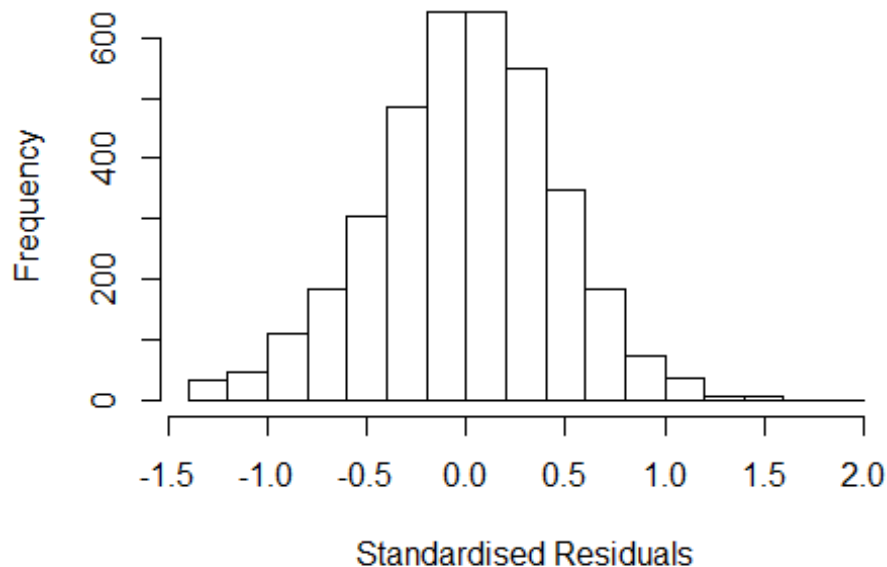


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.3, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.22013398458405"
## [1] "Male first author team size 2018 geometric mean: 2.77079733772588"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.04785159721272"
## [1] "Male last author team size 2018 geometric mean: 3.16072620436508"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.090 1          1.044
## LastAuthorFemale  1.054 1          1.027
## UniqueAuthors    1.112 4          1.013
## Year              1.135 16         1.004
```

Residuals from first and last author and team size



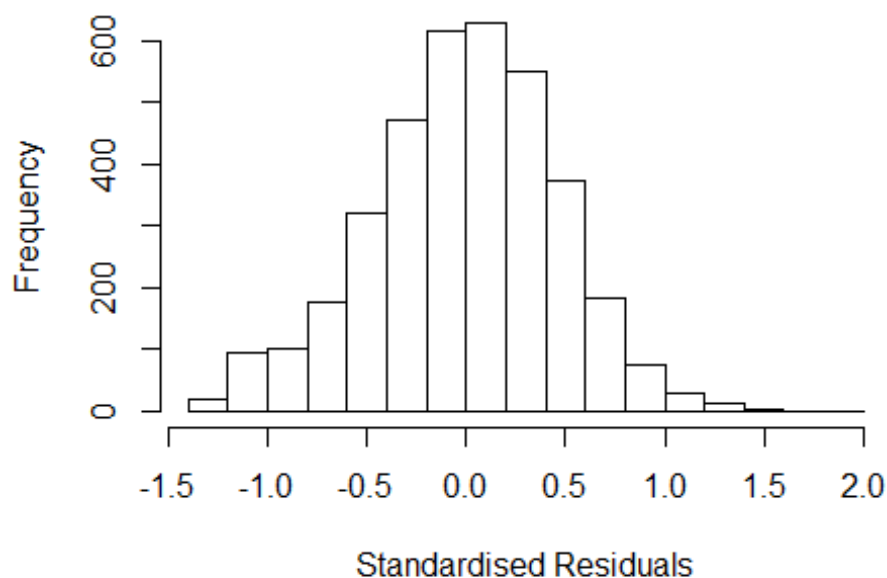
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33755 -0.29563  0.00907  0.29950  1.93362
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9691    0.0503   19.25 < 2e-16 ***
## FirstAuthorFemale1 0.0360    0.0165    2.18  0.029 *
## LastAuthorFemale1 0.0140    0.0157    0.89  0.373
## UniqueAuthors2    0.1889    0.0237    7.97 2.0e-15 ***
## UniqueAuthors3    0.2393    0.0253    9.44 < 2e-16 ***
## UniqueAuthors4    0.2250    0.0294    7.66 2.4e-14 ***
## UniqueAuthors5    0.2868    0.0281   10.20 < 2e-16 ***
## Year1997         -0.0237    0.0605   -0.39  0.696
## Year1998          0.1073    0.0601    1.79  0.074 .
## Year1999          0.0294    0.0631    0.47  0.640
```

```

## Year2000          -0.0249      0.0573   -0.43    0.664
## Year2001           0.0318      0.0633    0.50    0.616
## Year2002           0.0570      0.0582    0.98    0.327
## Year2003          -0.0576      0.0578   -1.00    0.319
## Year2004          -0.0572      0.0579   -0.99    0.323
## Year2005           0.0386      0.0572    0.68    0.499
## Year2006          -0.0183      0.0568   -0.32    0.747
## Year2007          -0.0235      0.0531   -0.44    0.658
## Year2008          -0.0344      0.0562   -0.61    0.540
## Year2009          -0.0275      0.0537   -0.51    0.609
## Year2010          -0.0120      0.0553   -0.22    0.829
## Year2011          -0.0265      0.0531   -0.50    0.618
## Year2012          -0.0477      0.0541   -0.88    0.378
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.442
## Multiple R-squared:  0.0502, Adjusted R-squared:  0.0445
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 282 weights are ~= 1. The remaining 3369 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0167 0.8700 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          2.74e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.067 1 1.033
## LastAuthorFemale 1.046 1 1.023
## Year 1.040 16 1.001

```


Residuals from first and last author



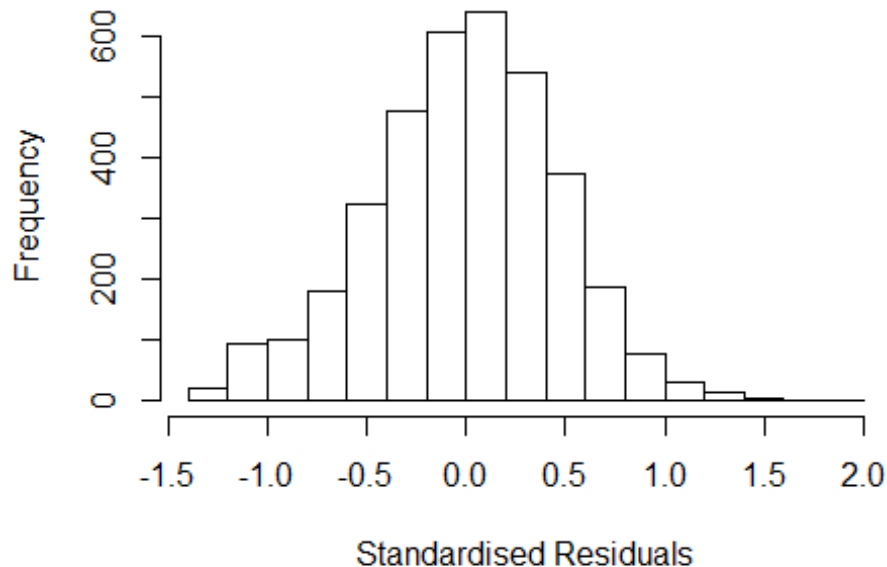
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27263 -0.30056  0.00876  0.30492  1.80389
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.107044   0.050344   21.99  <2e-16 ***
## FirstAuthorFemale1 0.052910   0.016650    3.18  0.0015 **
## LastAuthorFemale1  0.009850   0.015906    0.62  0.5358
## Year1997        -0.031938   0.062249   -0.51  0.6079
## Year1998         0.102828   0.062645    1.64  0.1008
## Year1999         0.045571   0.064226    0.71  0.4780
## Year2000        -0.017444   0.059909   -0.29  0.7709
## Year2001         0.062036   0.064771    0.96  0.3382
## Year2002         0.092451   0.060147    1.54  0.1244
## Year2003        -0.016796   0.059600   -0.28  0.7781
## Year2004        -0.015199   0.059869   -0.25  0.7996
## Year2005         0.072495   0.058399    1.24  0.2145
```

```

## Year2006          0.026856    0.058423    0.46    0.6458
## Year2007          0.015766    0.055242    0.29    0.7754
## Year2008          0.016741    0.057063    0.29    0.7693
## Year2009          0.016517    0.055575    0.30    0.7663
## Year2010          0.028288    0.057258    0.49    0.6213
## Year2011          0.021972    0.054837    0.40    0.6887
## Year2012          0.000656    0.056015    0.01    0.9907
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.00918,    Adjusted R-squared:  0.00427
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 300 weights are ~= 1. The remaining 3351 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0717 0.8680 0.9500 0.9010 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.031 1      1.015
## Year              1.031 16      1.001

```

Residuals from first author



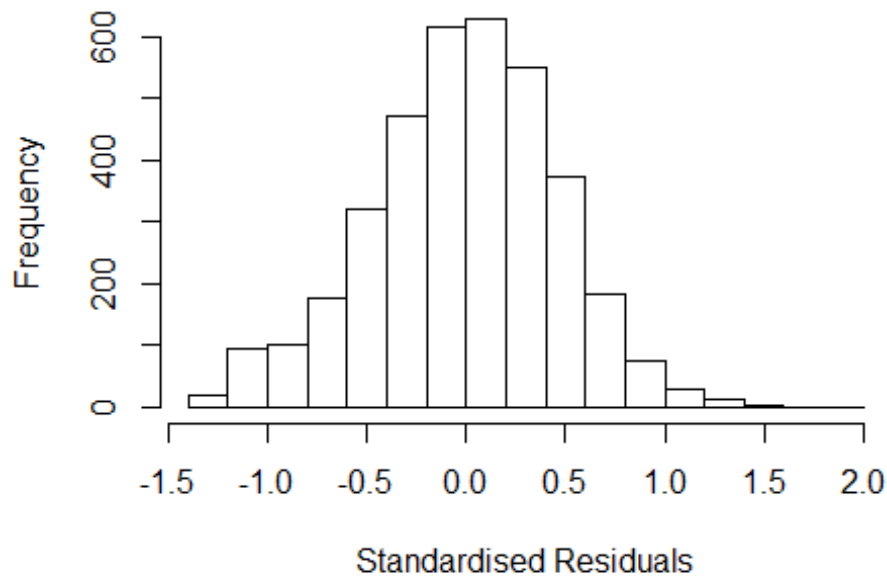
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26940 -0.30065 0.00816 0.30538 1.80082
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10972 0.05005 22.17 < 2e-16 ***
## FirstAuthorFemale1 0.05567 0.01644 3.39 0.00071 ***
## Year1997 -0.03155 0.06226 -0.51 0.61242
## Year1998 0.10401 0.06264 1.66 0.09693 .
## Year1999 0.04615 0.06421 0.72 0.47235
## Year2000 -0.01631 0.05986 -0.27 0.78531
## Year2001 0.06268 0.06484 0.97 0.33375
## Year2002 0.09312 0.06017 1.55 0.12180
## Year2003 -0.01584 0.05957 -0.27 0.79030
## Year2004 -0.01473 0.05993 -0.25 0.80590
## Year2005 0.07354 0.05836 1.26 0.20773
## Year2006 0.02786 0.05838 0.48 0.63325
```

```

## Year2007          0.01666    0.05522    0.30  0.76290
## Year2008          0.01747    0.05704    0.31  0.75938
## Year2009          0.01757    0.05554    0.32  0.75177
## Year2010          0.02935    0.05721    0.51  0.60802
## Year2011          0.02266    0.05483    0.41  0.67934
## Year2012          0.00143    0.05602    0.03  0.97957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.00904,    Adjusted R-squared:  0.00441
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 301 weights are ~= 1. The remaining 3350 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0728 0.8670 0.9510 0.9010 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.012 1          1.006
## Year            1.012 16          1.000

```

Residuals from last author



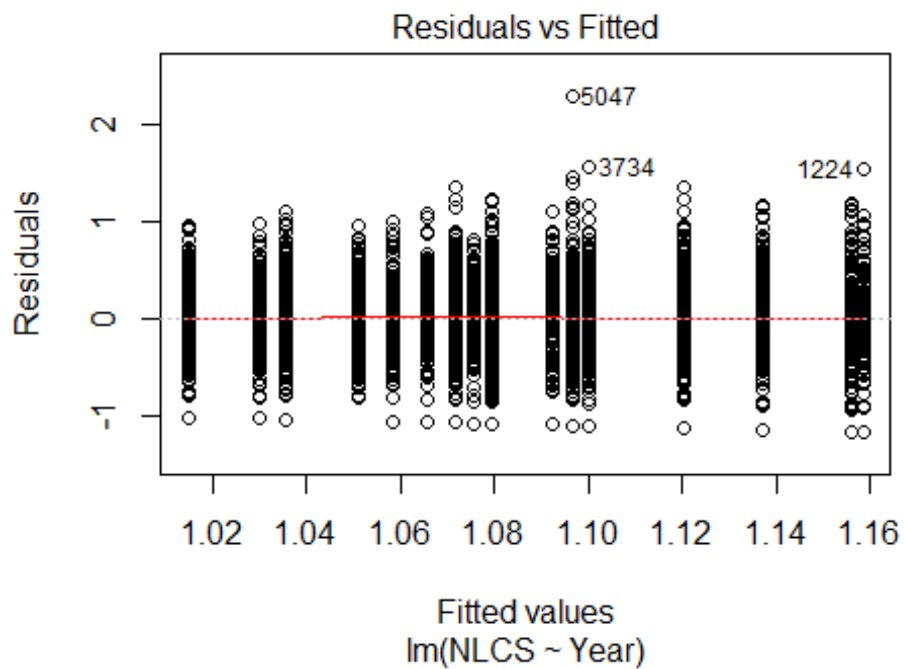
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2540 -0.2974 0.0118 0.3053 1.7823
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12952 0.04992 22.63 <2e-16 ***
## LastAuthorFemale1 0.02364 0.01573 1.50 0.13
## Year1997 -0.03282 0.06231 -0.53 0.60
## Year1998 0.10085 0.06299 1.60 0.11
## Year1999 0.04310 0.06438 0.67 0.50
## Year2000 -0.02176 0.06021 -0.36 0.72
## Year2001 0.06244 0.06507 0.96 0.34
## Year2002 0.09711 0.06051 1.60 0.11
## Year2003 -0.01422 0.05977 -0.24 0.81
## Year2004 -0.01251 0.06019 -0.21 0.84
## Year2005 0.07395 0.05867 1.26 0.21
## Year2006 0.03102 0.05880 0.53 0.60
```

```

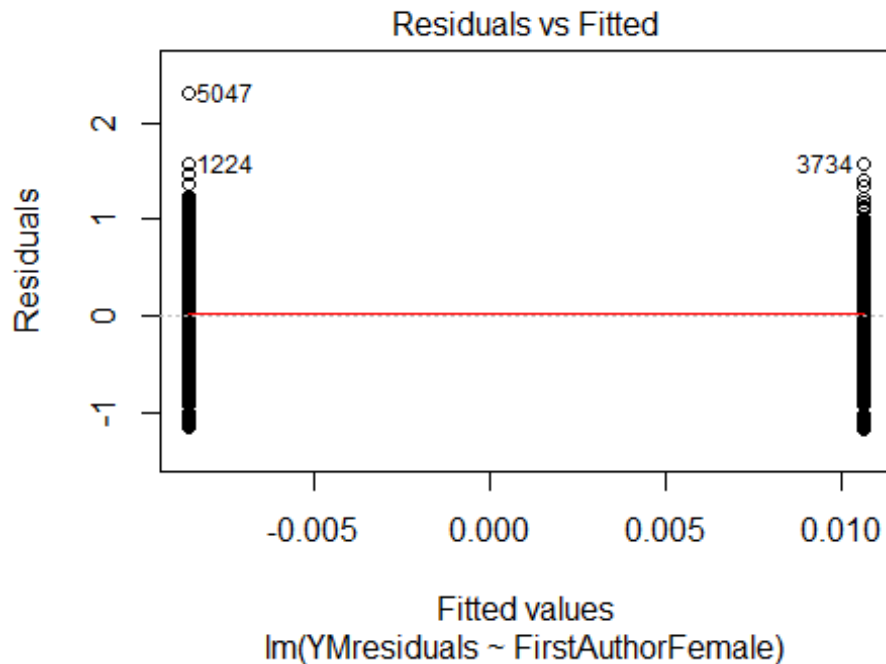
## Year2007      0.01944    0.05545    0.35    0.73
## Year2008      0.02078    0.05730    0.36    0.72
## Year2009      0.02157    0.05591    0.39    0.70
## Year2010      0.03686    0.05753    0.64    0.52
## Year2011      0.02880    0.05512    0.52    0.60
## Year2012      0.00599    0.05625    0.11    0.92
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.00632,    Adjusted R-squared:  0.00167
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 305 weights are ~= 1. The remaining 3346 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0812 0.8680 0.9510 0.9010 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3651"
## [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
## 262 256 210 215 215 274 221 234 234 227 241 238 296 344 304
## 2011 2012
## 346 336
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 220 181 157 157 162 153 180 193 174 186 207 197 254 296 252
## 2011 2012

```

```
## 298 286
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 199 167 142 144 147 136 168 175 153 161 194 187 225 263 231
## 2011 2012
## 271 259
## [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
```

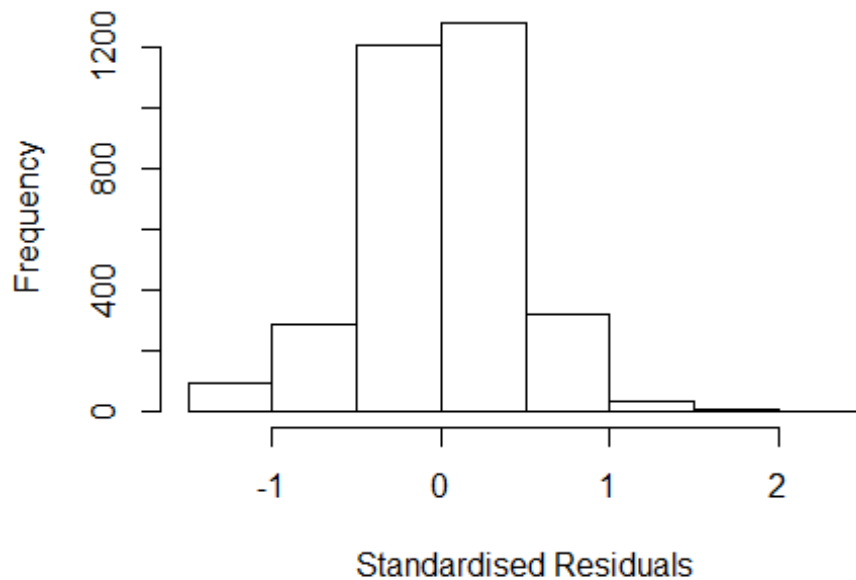


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 10, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 3.23452773619394"
## [1] "Male first author team size 2018 geometric mean: 2.61830744164052"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9200, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.02977147576635"
## [1] "Male last author team size 2018 geometric mean: 2.86751808096685"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7900, 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 1.055 1          1.027
## LastAuthorFemale  1.057 1          1.028
## UniqueAuthors    1.163 4          1.019
## Year             1.225 16          1.006
```


Residuals from first and last author and team size



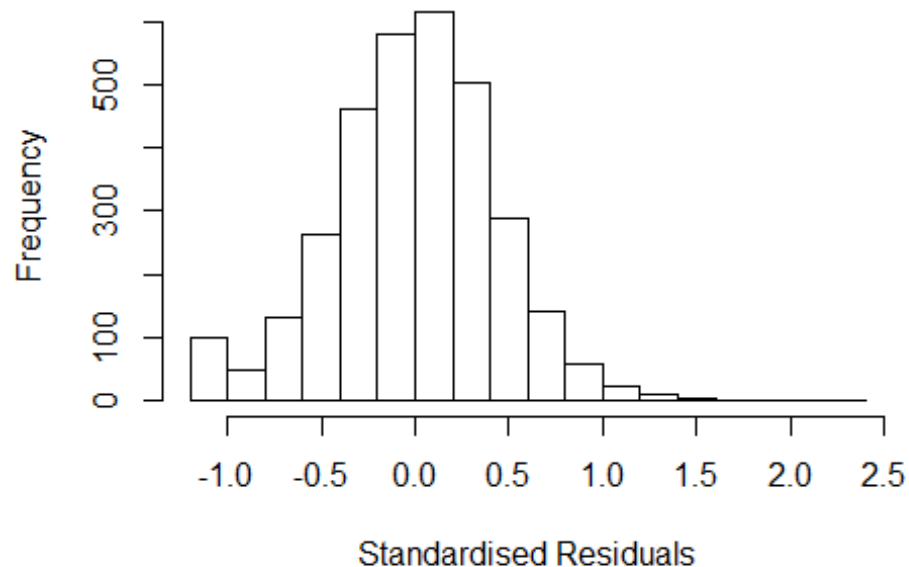
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26374 -0.28141 0.00875 0.27393 2.38192
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03094 0.04263 24.19 <2e-16 ***
## FirstAuthorFemale1 0.02762 0.01553 1.78 0.075 .
## LastAuthorFemale1 0.01461 0.01606 0.91 0.363
## UniqueAuthors2 0.04090 0.02755 1.48 0.138
## UniqueAuthors3 0.06966 0.02908 2.40 0.017 *
## UniqueAuthors4 0.07289 0.03229 2.26 0.024 *
## UniqueAuthors5 0.14420 0.03504 4.12 4e-05 ***
## Year1997 0.07188 0.05471 1.31 0.189
## Year1998 0.07399 0.05631 1.31 0.189
## Year1999 0.00631 0.05076 0.12 0.901
```

```

## Year2000          0.05482    0.05040    1.09    0.277
## Year2001          0.00796    0.05823    0.14    0.891
## Year2002          0.03162    0.04883    0.65    0.517
## Year2003         -0.04042    0.04909   -0.82    0.410
## Year2004         -0.03021    0.05119   -0.59    0.555
## Year2005         -0.05369    0.04949   -1.08    0.278
## Year2006         -0.10481    0.04832   -2.17    0.030 *
## Year2007         -0.06933    0.04782   -1.45    0.147
## Year2008         -0.01646    0.04808   -0.34    0.732
## Year2009         -0.07380    0.04557   -1.62    0.105
## Year2010         -0.01598    0.04859   -0.33    0.742
## Year2011         -0.02386    0.04609   -0.52    0.605
## Year2012          0.00971    0.04934    0.20    0.844
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.0186, Adjusted R-squared:  0.0118
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2863 is an outlier with |weight| = 0 ( < 3.1e-05);
## 253 weights are ~= 1. The remaining 2968 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0848 0.8730 0.9520 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.10e-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 1.049 1          1.024
## LastAuthorFemale 1.050 1          1.025
## Year          1.058 16          1.002

```

Residuals from first and last author



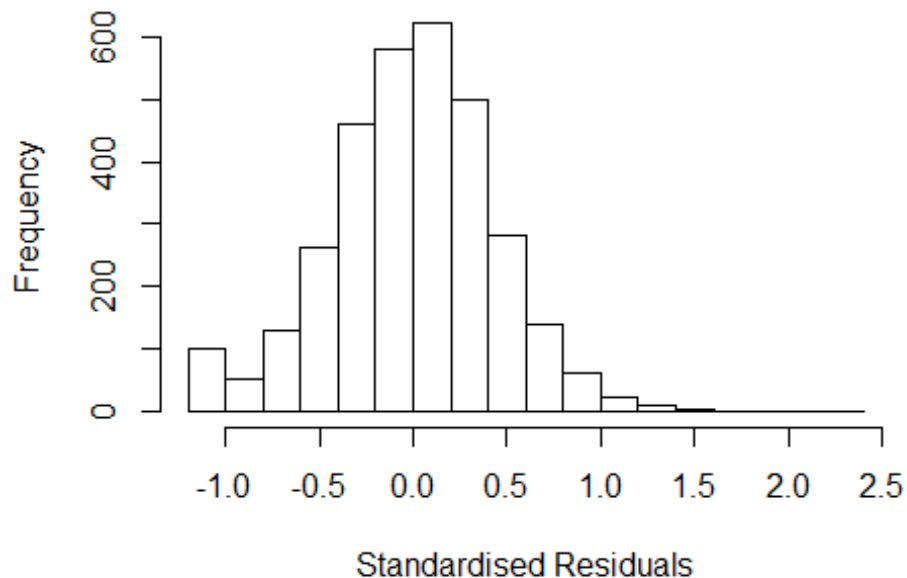
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1781 -0.2852 0.0128 0.2704 2.3175
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.072266 0.038210 28.06 <2e-16 ***
## FirstAuthorFemale1 0.029705 0.015563 1.91 0.056 .
## LastAuthorFemale1 0.012806 0.016076 0.80 0.426
## Year1997 0.072450 0.054641 1.33 0.185
## Year1998 0.076082 0.056268 1.35 0.176
## Year1999 0.011637 0.050774 0.23 0.819
## Year2000 0.063278 0.050449 1.25 0.210
## Year2001 0.017102 0.057563 0.30 0.766
## Year2002 0.041850 0.048515 0.86 0.388
## Year2003 -0.025903 0.048787 -0.53 0.596
## Year2004 -0.013667 0.050984 -0.27 0.789
## Year2005 -0.036114 0.049080 -0.74 0.462
```

```

## Year2006      -0.085628    0.048294   -1.77    0.076 .
## Year2007      -0.052166    0.046913   -1.11    0.266
## Year2008       0.003645    0.047620    0.08    0.939
## Year2009      -0.055826    0.045210   -1.23    0.217
## Year2010       0.001164    0.048431    0.02    0.981
## Year2011      -0.000747    0.045027   -0.02    0.987
## Year2012       0.029893    0.049116    0.61    0.543
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.414
## Multiple R-squared:  0.0114, Adjusted R-squared:  0.00589
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2863 is an outlier with |weight| = 0 ( < 3.1e-05);
## 278 weights are ~= 1. The remaining 2943 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.116  0.872   0.950   0.899   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.10e-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 1.031 1          1.016
## Year              1.031 16          1.001

```

Residuals from first author



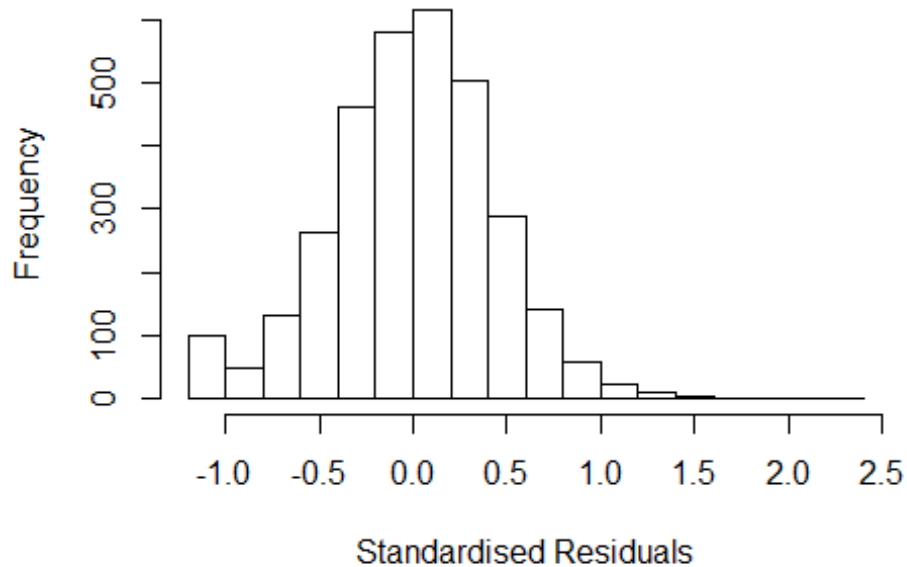
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1798 -0.2860 0.0103 0.2716 2.3138
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.075494 0.038000 28.30 <2e-16 ***
## FirstAuthorFemale1 0.032208 0.015461 2.08 0.037 *
## Year1997 0.072127 0.054631 1.32 0.187
## Year1998 0.076926 0.056162 1.37 0.171
## Year1999 0.012031 0.050789 0.24 0.813
## Year2000 0.064306 0.050456 1.27 0.203
## Year2001 0.018332 0.057447 0.32 0.750
## Year2002 0.042004 0.048557 0.87 0.387
## Year2003 -0.024874 0.048688 -0.51 0.609
## Year2004 -0.013995 0.051061 -0.27 0.784
## Year2005 -0.036758 0.049092 -0.75 0.454
## Year2006 -0.084672 0.048276 -1.75 0.080 .
```

```

## Year2007          -0.052217    0.046914    -1.11    0.266
## Year2008          0.003974    0.047632     0.08    0.934
## Year2009         -0.055475    0.045225    -1.23    0.220
## Year2010          0.002275    0.048422     0.05    0.963
## Year2011         -0.000313    0.045020    -0.01    0.994
## Year2012          0.030307    0.049120     0.62    0.537
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.414
## Multiple R-squared:  0.0112, Adjusted R-squared:  0.00597
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2863 is an outlier with |weight| = 0 ( < 3.1e-05);
## 279 weights are ~= 1. The remaining 2942 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.118  0.872  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      3.10e-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 1.031 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
## -1.1773 -0.2862 0.0104 0.2730 2.3042
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07986 0.03805 28.38 <2e-16 ***
## LastAuthorFemale1 0.01892 0.01598 1.18 0.237
## Year1997 0.07347 0.05462 1.35 0.179
## Year1998 0.07857 0.05620 1.40 0.162
## Year1999 0.01163 0.05074 0.23 0.819
## Year2000 0.06598 0.05032 1.31 0.190
## Year2001 0.01838 0.05766 0.32 0.750
## Year2002 0.04484 0.04856 0.92 0.356
## Year2003 -0.02194 0.04876 -0.45 0.653
## Year2004 -0.00901 0.05071 -0.18 0.859
## Year2005 -0.03215 0.04907 -0.66 0.512
## Year2006 -0.08089 0.04818 -1.68 0.093 .
```

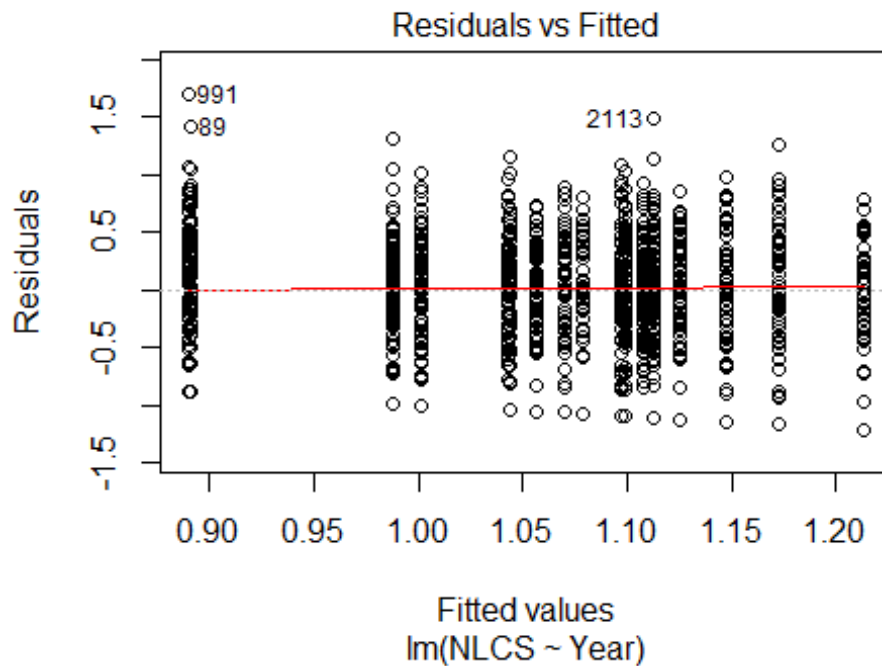
```

## Year2007          -0.04837      0.04695    -1.03      0.303
## Year2008           0.00840      0.04750      0.18      0.860
## Year2009          -0.05071      0.04512    -1.12      0.261
## Year2010           0.00669      0.04820      0.14      0.890
## Year2011           0.00493      0.04488      0.11      0.913
## Year2012           0.03598      0.04886      0.74      0.462
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.414
## Multiple R-squared:  0.0104, Adjusted R-squared:  0.00512
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 2863 is an outlier with |weight| = 0 ( < 3.1e-05);
## 264 weights are ~= 1. The remaining 2957 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.120  0.870  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      3.10e-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: 3222"
## [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
##   97   86   79   71   91  149   89   84  116   91   84   91  102  102  139
## 2011 2012
##  113  133
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   71   56   48   45   51   69   81   73   85   80   67   79   83   75  108

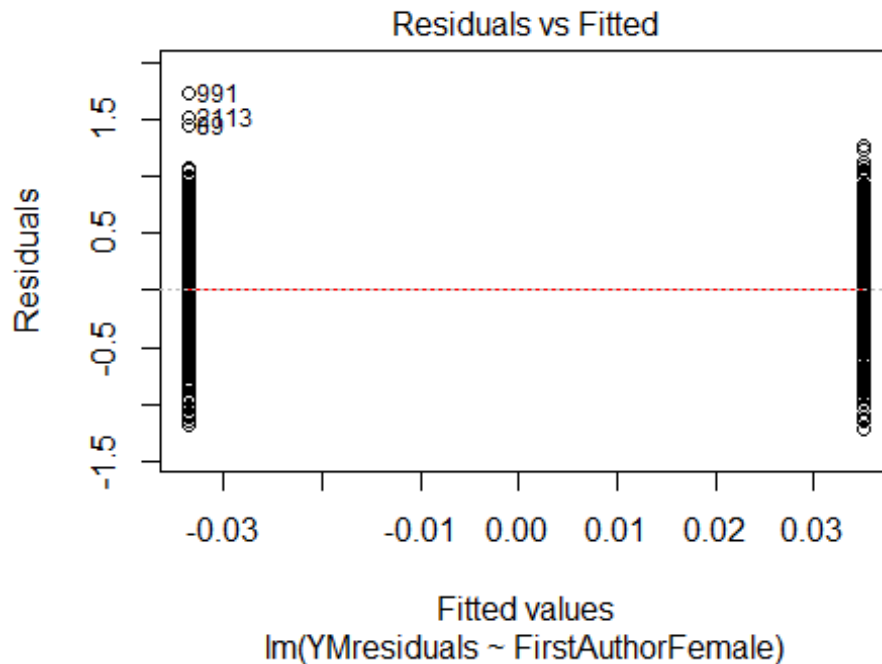
```



```
## 2011 2012
## 93 116
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 52 43 40 47 62 78 63 75 70 65 70 70 66 93
## 2011 2012
## 82 104
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 47, df = 16, p-value = 7e-05
```

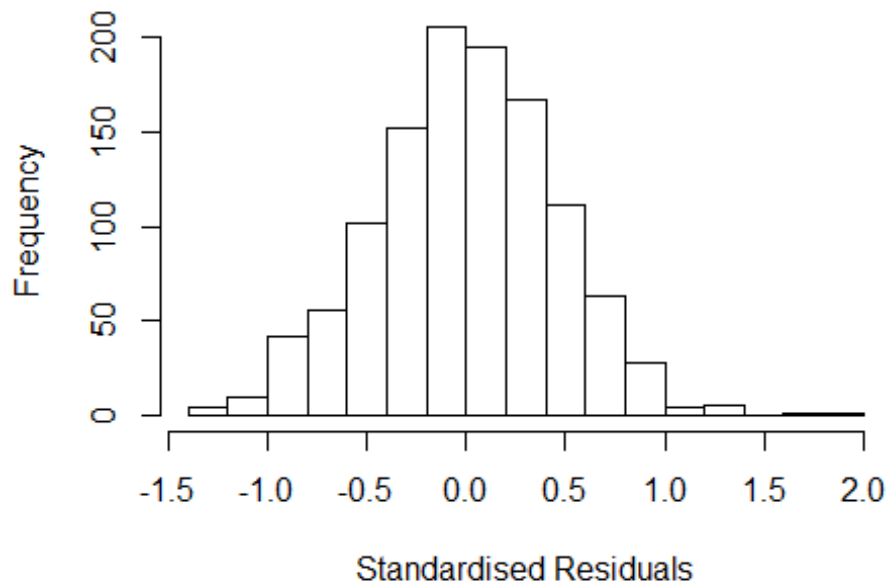


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.5, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 3.9623167021184"
## [1] "Male first author team size 2018 geometric mean: 3.80870197851137"
##
## 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: 4.13919809565097"
## [1] "Male last author team size 2018 geometric mean: 3.73034828018047"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.075 1      1.037
## LastAuthorFemale  1.091 1      1.045
## UniqueAuthors    1.325 4      1.036
## Year             1.369 16      1.010
```

Residuals from first and last author and team size



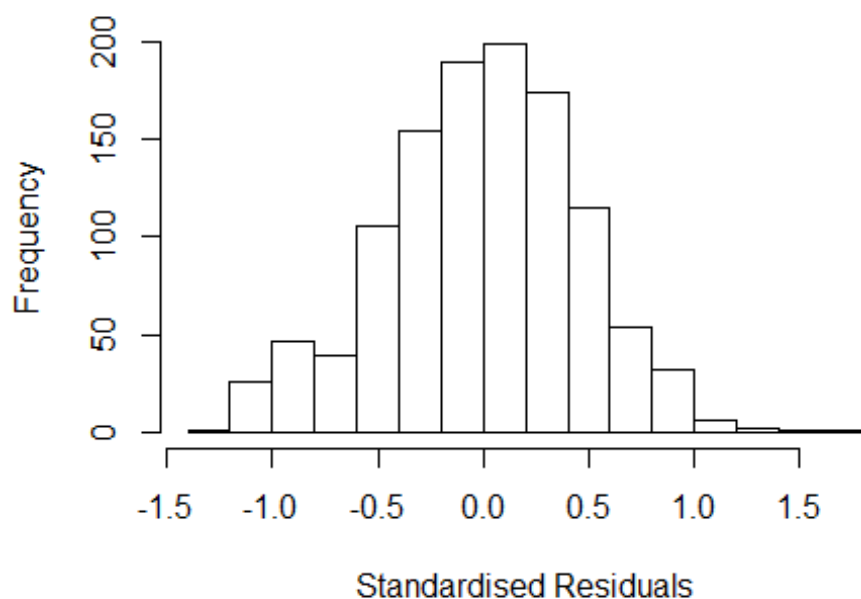
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24349 -0.29523  0.00445  0.29885  1.91722
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6892    0.0987    6.98 4.9e-12 ***
## FirstAuthorFemale1 0.0463    0.0275    1.68 0.09301 .
## LastAuthorFemale1 0.0395    0.0286    1.38 0.16773
## UniqueAuthors2    0.1871    0.0519    3.61 0.00032 ***
## UniqueAuthors3    0.2028    0.0536    3.78 0.00016 ***
## UniqueAuthors4    0.1185    0.0606    1.95 0.05093 .
## UniqueAuthors5    0.2454    0.0579    4.24 2.4e-05 ***
## Year1997          0.2802    0.1222    2.29 0.02196 *
## Year1998          0.4003    0.1132    3.54 0.00042 ***
## Year1999          0.2296    0.1068    2.15 0.03182 *
```

```

## Year2000          0.1702      0.1133      1.50  0.13333
## Year2001          0.2664      0.1160      2.30  0.02179 *
## Year2002          0.2407      0.1008      2.39  0.01713 *
## Year2003          0.1954      0.0995      1.96  0.04980 *
## Year2004         -0.0254      0.1049     -0.24  0.80833
## Year2005          0.1020      0.1055      0.97  0.33393
## Year2006          0.2053      0.1038      1.98  0.04828 *
## Year2007          0.1740      0.0984      1.77  0.07738 .
## Year2008          0.2231      0.1036      2.15  0.03156 *
## Year2009          0.1090      0.1010      1.08  0.28061
## Year2010          0.2006      0.0962      2.09  0.03720 *
## Year2011          0.0760      0.1000      0.76  0.44788
## Year2012          0.1851      0.0981      1.89  0.05959 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.449
## Multiple R-squared:  0.0761, Adjusted R-squared:  0.058
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 94 weights are ~= 1. The remaining 1053 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0285 0.8770 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          8.72e-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 1.062 1          1.030
## LastAuthorFemale  1.054 1          1.027
## Year              1.096 16          1.003

```

Residuals from first and last author



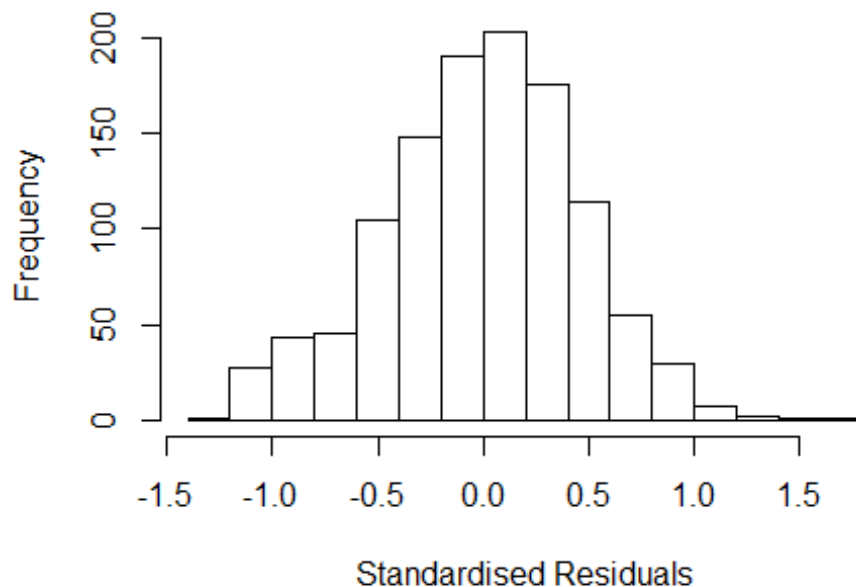
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22072 -0.30179  0.00892  0.30124  1.77084
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8356    0.0880    9.49  <2e-16 ***
## FirstAuthorFemale1 0.0637    0.0281    2.26   0.0238 *
## LastAuthorFemale1 0.0404    0.0287    1.41   0.1595
## Year1997        0.2591    0.1213    2.14   0.0330 *
## Year1998        0.3851    0.1104    3.49   0.0005 ***
## Year1999        0.2322    0.1045    2.22   0.0265 *
## Year2000        0.1652    0.1175    1.41   0.1599
## Year2001        0.2884    0.1148    2.51   0.0121 *
## Year2002        0.2528    0.1013    2.49   0.0128 *
## Year2003        0.2109    0.0993    2.12   0.0338 *
## Year2004       -0.0254    0.1066   -0.24   0.8117
## Year2005        0.1232    0.1061    1.16   0.2459
```

```

## Year2006          0.1894      0.1050      1.80      0.0714 .
## Year2007          0.1755      0.0983      1.79      0.0744 .
## Year2008          0.2371      0.1033      2.29      0.0219 *
## Year2009          0.1335      0.1013      1.32      0.1875
## Year2010          0.2154      0.0962      2.24      0.0253 *
## Year2011          0.0842      0.0992      0.85      0.3963
## Year2012          0.2101      0.0969      2.17      0.0304 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.449
## Multiple R-squared:  0.0487, Adjusted R-squared:  0.0335
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1050 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0855 0.8760 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      8.72e-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 1.056 1      1.028
## Year              1.056 16      1.002

```

Residuals from first author



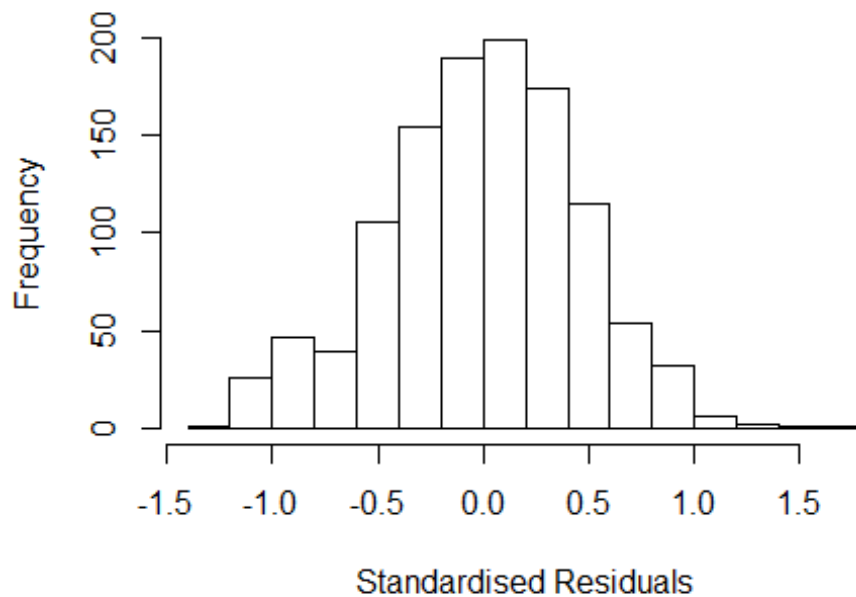
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2354 -0.2972 0.0145 0.3025 1.7636
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8463 0.0878 9.64 < 2e-16 ***
## FirstAuthorFemale1 0.0706 0.0281 2.51 0.01222 *
## Year1997 0.2630 0.1210 2.17 0.03000 *
## Year1998 0.3891 0.1099 3.54 0.00042 ***
## Year1999 0.2293 0.1047 2.19 0.02873 *
## Year2000 0.1670 0.1177 1.42 0.15624
## Year2001 0.2915 0.1143 2.55 0.01092 *
## Year2002 0.2536 0.1014 2.50 0.01256 *
## Year2003 0.2153 0.0989 2.18 0.02974 *
## Year2004 -0.0289 0.1066 -0.27 0.78628
## Year2005 0.1244 0.1062 1.17 0.24190
## Year2006 0.1907 0.1051 1.81 0.07002 .
```

```

## Year2007          0.1760      0.0982      1.79  0.07333 .
## Year2008          0.2384      0.1032      2.31  0.02110 *
## Year2009          0.1330      0.1009      1.32  0.18788
## Year2010          0.2166      0.0961      2.26  0.02432 *
## Year2011          0.0846      0.0990      0.85  0.39327
## Year2012          0.2129      0.0969      2.20  0.02821 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.449
## Multiple R-squared:  0.047, Adjusted R-squared:  0.0326
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1050 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0879 0.8760 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      8.72e-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 1.046 1          1.023
## Year            1.046 16          1.001

```


Residuals from last author



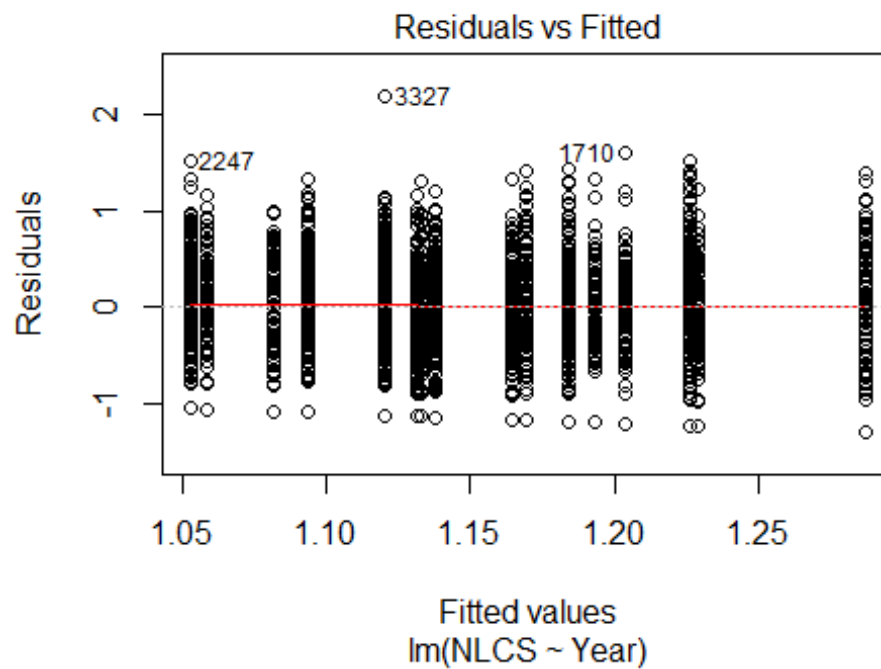
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2332 -0.2914  0.0139  0.2971  1.7434
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8559     0.0862   9.93 < 2e-16 ***
## LastAuthorFemale1 0.0519     0.0287   1.81  0.07060 .
## Year1997        0.2612     0.1209   2.16  0.03091 *
## Year1998        0.3773     0.1089   3.46  0.00055 ***
## Year1999        0.2323     0.1037   2.24  0.02530 *
## Year2000        0.1718     0.1173   1.46  0.14329
## Year2001        0.2908     0.1141   2.55  0.01094 *
## Year2002        0.2605     0.1003   2.60  0.00955 **
## Year2003        0.2257     0.0976   2.31  0.02101 *
## Year2004       -0.0183     0.1057  -0.17  0.86285
## Year2005        0.1293     0.1054   1.23  0.22017
## Year2006        0.1983     0.1038   1.91  0.05647 .
```

```

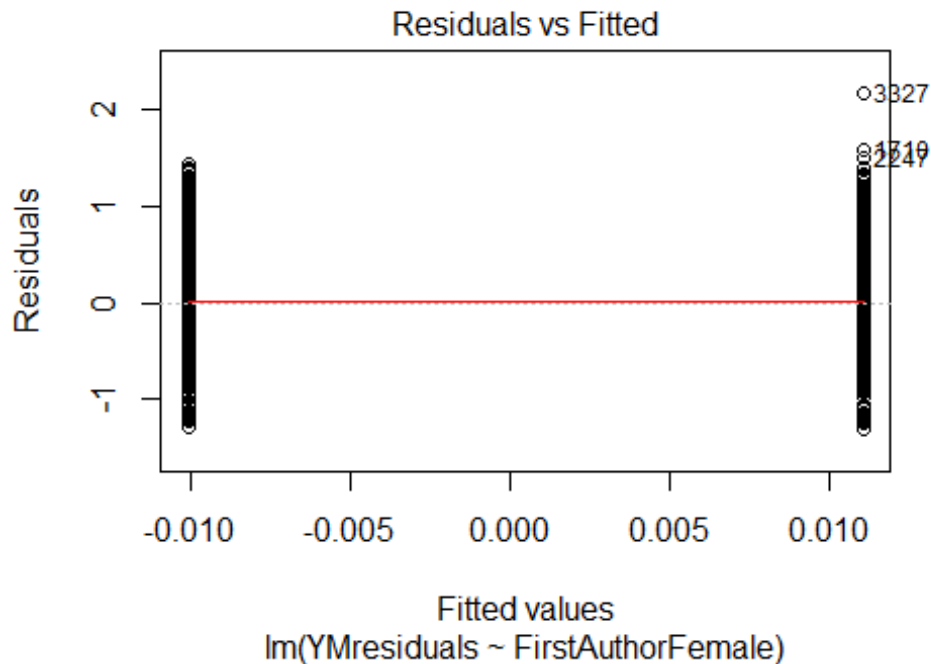
## Year2007          0.1882      0.0977      1.93  0.05427 .
## Year2008          0.2452      0.1020      2.40  0.01635 *
## Year2009          0.1445      0.1000      1.45  0.14873
## Year2010          0.2258      0.0947      2.38  0.01728 *
## Year2011          0.0954      0.0981      0.97  0.33077
## Year2012          0.2180      0.0956      2.28  0.02276 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.0442, Adjusted R-squared:  0.0298
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 98 weights are ~= 1. The remaining 1049 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0994 0.8740 0.9530 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      8.72e-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: 1147"
## [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
## 114 112 101 101 136 141 161 135 124 139 168 193 232 256 296
## 2011 2012
## 314 307
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 95 99 66 75 110 97 140 125 107 129 151 169 205 237 262
## 2011 2012

```

```
## 277 270
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 89 87 61 67 102 90 130 116 101 115 135 145 184 217 244
## 2011 2012
## 263 240
## [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
```

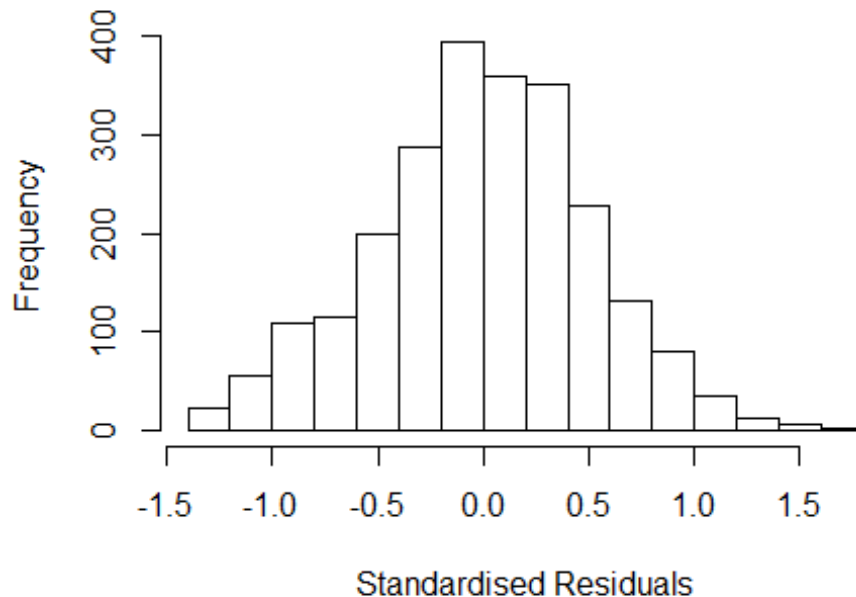


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 12, df = 1, p-value = 6e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 2.6765952218248"
## [1] "Male first author team size 2018 geometric mean: 2.09420720710488"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, p-value = 0.003
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.43045008567346"
## [1] "Male last author team size 2018 geometric mean: 2.44468651742482"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9400, 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 1.132 1      1.064
## LastAuthorFemale  1.116 1      1.056
## UniqueAuthors     1.178 4      1.021
## Year              1.223 16      1.006
```

Residuals from first and last author and team size



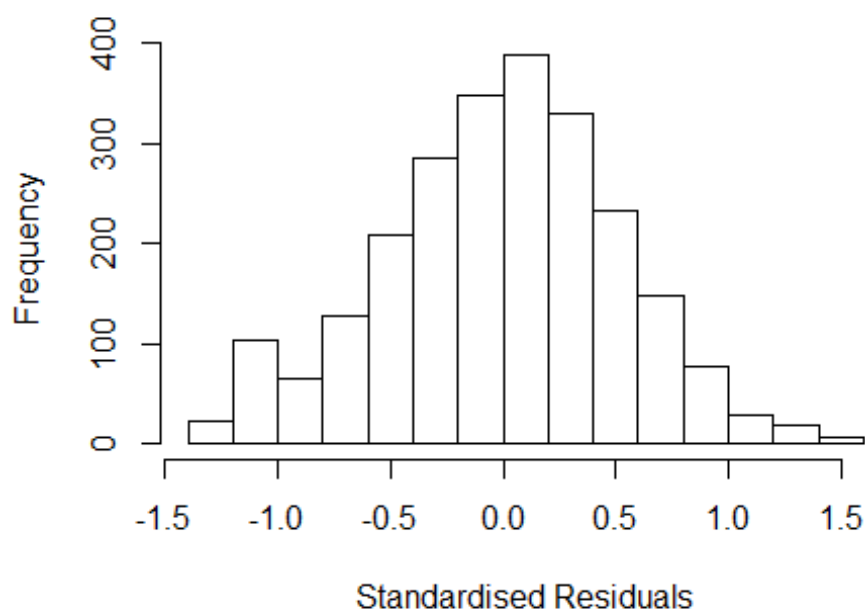
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.33230 -0.33283 0.00321 0.32876 1.75091
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1577 0.0787 14.71 < 2e-16 ***
## FirstAuthorFemale1 -0.0304 0.0228 -1.33 0.1824
## LastAuthorFemale1 -0.0359 0.0225 -1.60 0.1103
## UniqueAuthors2 0.2600 0.0291 8.94 < 2e-16 ***
## UniqueAuthors3 0.2770 0.0322 8.61 < 2e-16 ***
## UniqueAuthors4 0.2958 0.0383 7.73 1.6e-14 ***
## UniqueAuthors5 0.2657 0.0398 6.68 3.0e-11 ***
## Year1997 -0.0800 0.0925 -0.86 0.3871
## Year1998 -0.0954 0.1037 -0.92 0.3576
## Year1999 -0.1646 0.1006 -1.64 0.1020
```

```

## Year2000          -0.1880      0.0920   -2.04   0.0411 *
## Year2001          -0.2298      0.0925   -2.48   0.0131 *
## Year2002          -0.0908      0.0938   -0.97   0.3330
## Year2003          -0.1386      0.0909   -1.52   0.1277
## Year2004          -0.0820      0.0894   -0.92   0.3592
## Year2005          -0.1126      0.0859   -1.31   0.1902
## Year2006          -0.1796      0.0847   -2.12   0.0341 *
## Year2007          -0.1816      0.0858   -2.12   0.0344 *
## Year2008          -0.2557      0.0862   -2.97   0.0031 **
## Year2009          -0.1777      0.0829   -2.14   0.0321 *
## Year2010          -0.1351      0.0827   -1.63   0.1023
## Year2011          -0.1976      0.0827   -2.39   0.0170 *
## Year2012          -0.2483      0.0844   -2.94   0.0033 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.0672, Adjusted R-squared:  0.0586
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 194 weights are ~= 1. The remaining 2192 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.192  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          4.19e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.106 1 1.052
## LastAuthorFemale 1.092 1 1.045
## Year 1.051 16 1.002

```

Residuals from first and last author



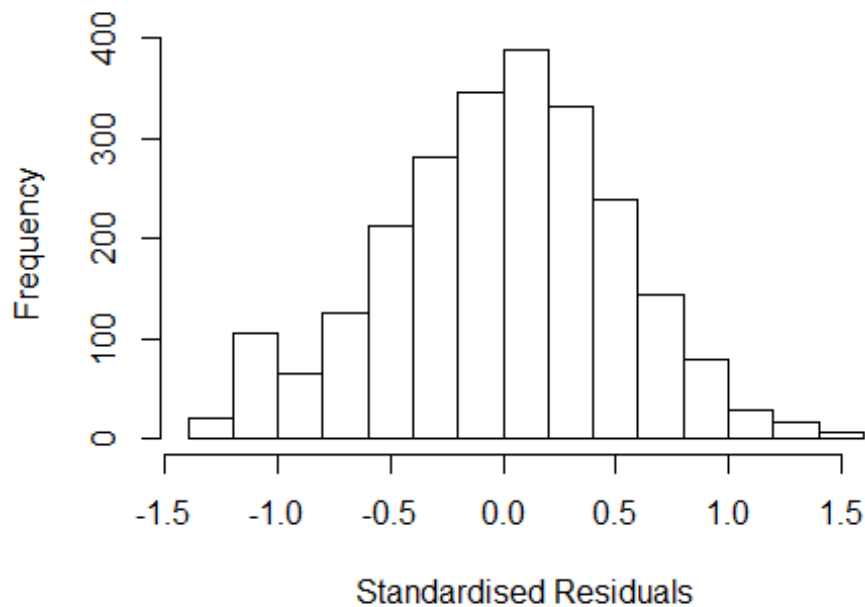
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3284 -0.3445  0.0227  0.3432  1.5768
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.32844    0.07748   17.15  <2e-16 ***
## FirstAuthorFemale1 -0.00501    0.02317   -0.22   0.8288
## LastAuthorFemale1  -0.03641    0.02284   -1.59   0.1110
## Year1997          -0.11475    0.09490   -1.21   0.2267
## Year1998          -0.12449    0.10194   -1.22   0.2221
## Year1999          -0.19586    0.10236   -1.91   0.0558 .
## Year2000          -0.19835    0.09502   -2.09   0.0369 *
## Year2001          -0.22394    0.09357   -2.39   0.0168 *
## Year2002          -0.07202    0.09586   -0.75   0.4526
## Year2003          -0.15431    0.09329   -1.65   0.0982 .
## Year2004          -0.06064    0.09171   -0.66   0.5086
## Year2005          -0.10924    0.08740   -1.25   0.2115
```

```

## Year2006          -0.16351    0.08661   -1.89    0.0592 .
## Year2007          -0.15781    0.08668   -1.82    0.0688 .
## Year2008          -0.25303    0.08768   -2.89    0.0039 **
## Year2009          -0.16167    0.08461   -1.91    0.0562 .
## Year2010          -0.12762    0.08388   -1.52    0.1282
## Year2011          -0.17887    0.08410   -2.13    0.0335 *
## Year2012          -0.21680    0.08608   -2.52    0.0118 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.512
## Multiple R-squared:  0.0138, Adjusted R-squared:  0.00626
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 211 weights are ~= 1. The remaining 2175 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.322  0.868  0.950  0.902  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.19e-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 1.032 1      1.016
## Year              1.032 16      1.001

```


Residuals from first author



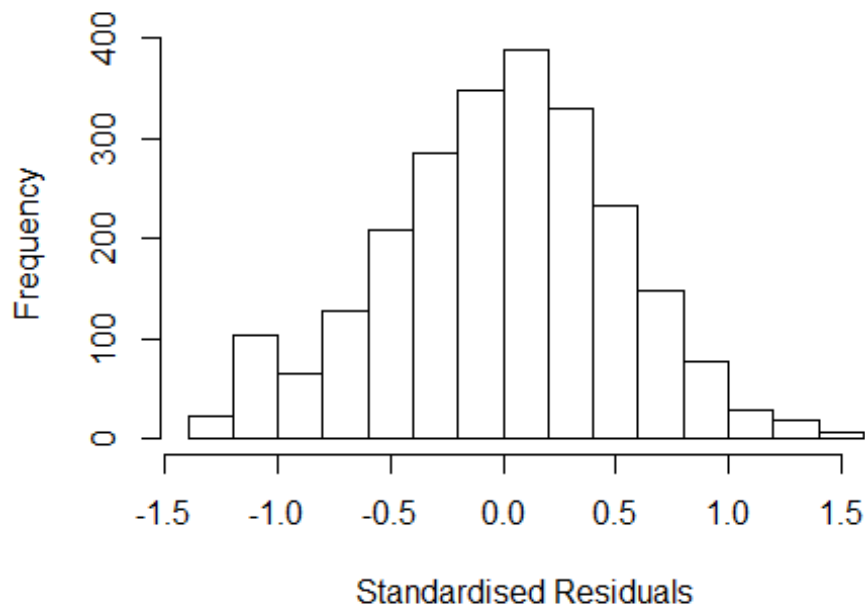
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3213 -0.3425 0.0171 0.3500 1.5877
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3213 0.0772 17.11 <2e-16 ***
## FirstAuthorFemale1 -0.0179 0.0225 -0.80 0.4252
## Year1997 -0.1176 0.0948 -1.24 0.2148
## Year1998 -0.1295 0.1019 -1.27 0.2039
## Year1999 -0.1967 0.1021 -1.93 0.0542 .
## Year2000 -0.1994 0.0948 -2.10 0.0354 *
## Year2001 -0.2276 0.0933 -2.44 0.0148 *
## Year2002 -0.0749 0.0957 -0.78 0.4338
## Year2003 -0.1590 0.0928 -1.71 0.0868 .
## Year2004 -0.0629 0.0913 -0.69 0.4909
## Year2005 -0.1129 0.0870 -1.30 0.1946
## Year2006 -0.1665 0.0862 -1.93 0.0536 .
```

```

## Year2007          -0.1629      0.0863   -1.89   0.0591 .
## Year2008          -0.2561      0.0875   -2.93   0.0034 **
## Year2009          -0.1658      0.0843   -1.97   0.0493 *
## Year2010          -0.1313      0.0836   -1.57   0.1162
## Year2011          -0.1822      0.0840   -2.17   0.0302 *
## Year2012          -0.2201      0.0858   -2.57   0.0103 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.511
## Multiple R-squared:  0.0127, Adjusted R-squared:  0.00562
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 203 weights are ~= 1. The remaining 2183 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.314  0.868  0.950  0.903  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.19e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.02 1          1.010
## Year              1.02 16          1.001

```

Residuals from last author



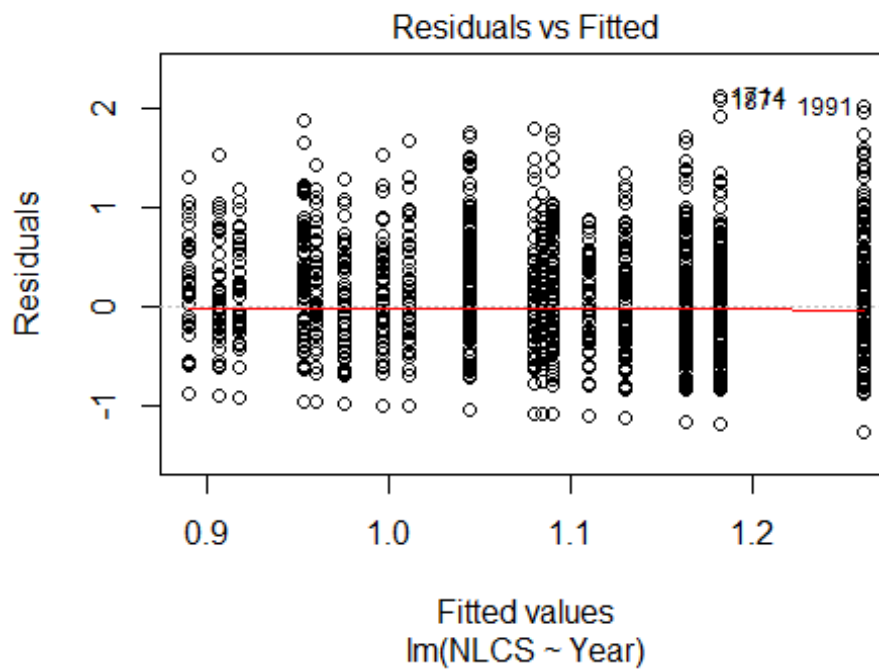
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3272 -0.3451  0.0221  0.3434  1.5786
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3272     0.0771   17.21  <2e-16 ***
## LastAuthorFemale1 -0.0382     0.0222   -1.72  0.0855 .
## Year1997         -0.1150     0.0949   -1.21  0.2259
## Year1998         -0.1250     0.1018   -1.23  0.2198
## Year1999         -0.1964     0.1023   -1.92  0.0551 .
## Year2000         -0.1991     0.0949   -2.10  0.0360 *
## Year2001         -0.2238     0.0936   -2.39  0.0169 *
## Year2002         -0.0726     0.0958   -0.76  0.4490
## Year2003         -0.1547     0.0933   -1.66  0.0973 .
## Year2004         -0.0615     0.0916   -0.67  0.5020
## Year2005         -0.1099     0.0874   -1.26  0.2085
## Year2006         -0.1638     0.0866   -1.89  0.0588 .
```

```

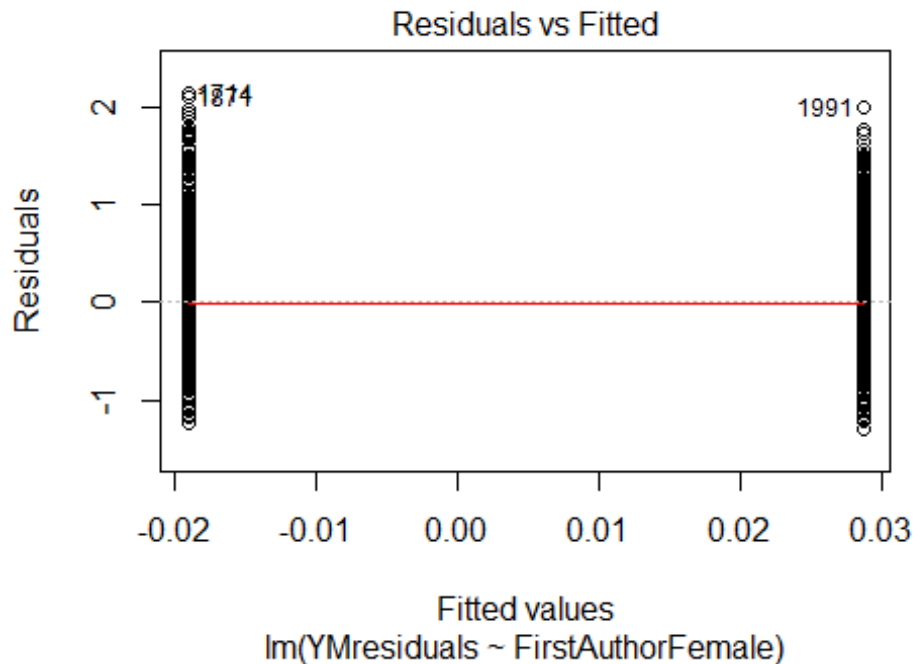
## Year2007          -0.1584      0.0867   -1.83    0.0678 .
## Year2008          -0.2539      0.0876   -2.90    0.0038 **
## Year2009          -0.1621      0.0846   -1.92    0.0556 .
## Year2010          -0.1285      0.0838   -1.53    0.1256
## Year2011          -0.1797      0.0841   -2.14    0.0326 *
## Year2012          -0.2179      0.0859   -2.54    0.0113 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.511
## Multiple R-squared:  0.0137, Adjusted R-squared:  0.00665
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 210 weights are ~= 1. The remaining 2176 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.320  0.867  0.950  0.902  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.19e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2386"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
##   59   64   63   58   61   87   68   63   73   75  115  144  146  184  208
## 2011 2012
##  198  190
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   51   51   54   43   43   57   54   49   60   67   93  123  125  162  173
## 2011 2012

```

```
## 176 167
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 47 49 52 42 41 52 51 47 58 63 88 120 118 156 167
## 2011 2012
## 168 160
## [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
```

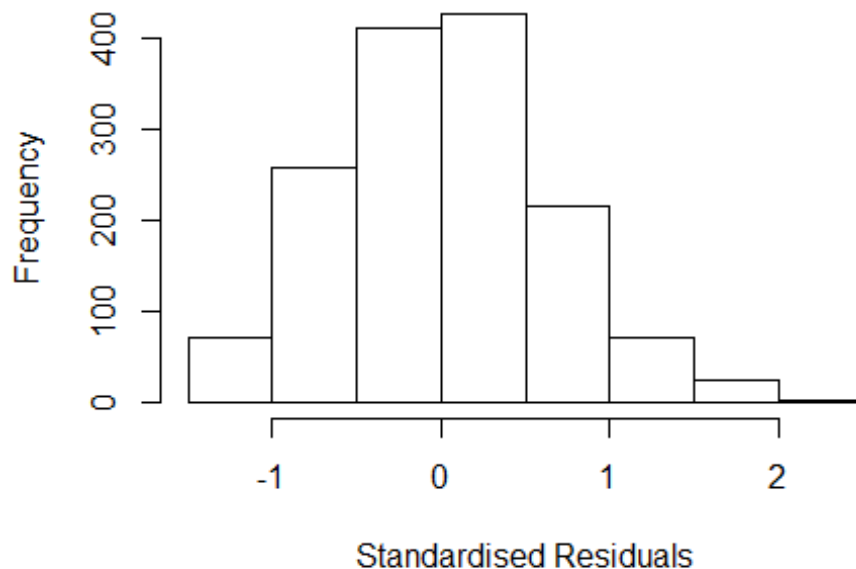


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.84, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 1.84071576079428"
## [1] "Male first author team size 2018 geometric mean: 1.40666678888257"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4100, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.83043503713158"
## [1] "Male last author team size 2018 geometric mean: 1.42377922538571"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4100, 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 2.141 1      1.463
## LastAuthorFemale  2.110 1      1.453
## UniqueAuthors    1.369 4      1.040
## Year              1.400 16     1.011
```

Residuals from first and last author and team size



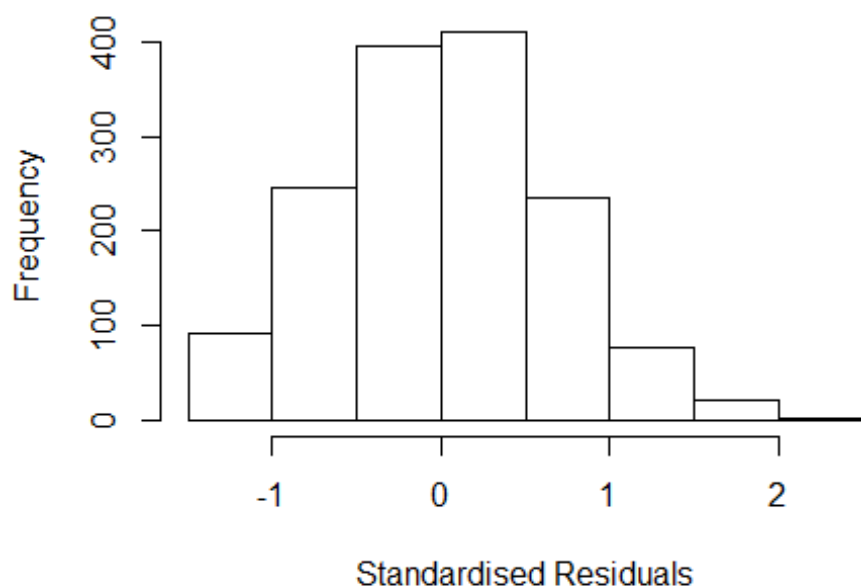
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.39776 -0.44376 -0.00293 0.42384 2.18433
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90692 0.10785 8.41 < 2e-16 ***
## FirstAuthorFemale1 0.11417 0.05135 2.22 0.02634 *
## LastAuthorFemale1 -0.11019 0.05227 -2.11 0.03519 *
## UniqueAuthors2 0.17045 0.04621 3.69 0.00023 ***
## UniqueAuthors3 0.30439 0.05795 5.25 1.7e-07 ***
## UniqueAuthors4 0.20108 0.09475 2.12 0.03400 *
## UniqueAuthors5 0.21051 0.10164 2.07 0.03853 *
## Year1997 0.00493 0.13577 0.04 0.97102
## Year1998 0.02392 0.13789 0.17 0.86232
## Year1999 -0.06171 0.14193 -0.43 0.66381
```

```

## Year2000      -0.09972    0.14542   -0.69  0.49301
## Year2001      0.05382    0.13879    0.39  0.69824
## Year2002     -0.07976    0.13564   -0.59  0.55660
## Year2003      0.17212    0.13167    1.31  0.19136
## Year2004      0.07458    0.13053    0.57  0.56785
## Year2005      0.00721    0.13073    0.06  0.95605
## Year2006      0.15872    0.12267    1.29  0.19593
## Year2007     -0.02086    0.12847   -0.16  0.87102
## Year2008      0.08673    0.12731    0.68  0.49584
## Year2009      0.15987    0.12112    1.32  0.18709
## Year2010      0.02405    0.12165    0.20  0.84329
## Year2011      0.17075    0.11815    1.45  0.14863
## Year2012      0.20623    0.12387    1.66  0.09617 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.646
## Multiple R-squared:  0.0533, Adjusted R-squared:  0.039
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1357 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.230  0.863  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      6.76e-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 2.018 1      1.421
## LastAuthorFemale 2.080 1      1.442
## Year      1.129 16      1.004

```


Residuals from first and last author

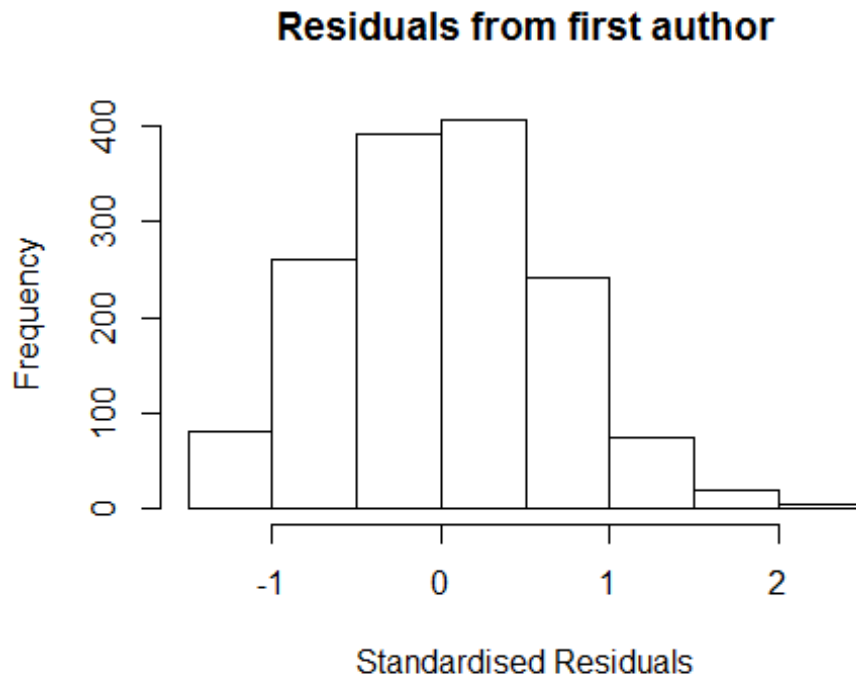


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.36329 -0.44751 0.00625 0.43600 2.10381
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95158 0.10762 8.84 <2e-16 ***
## FirstAuthorFemale1 0.14076 0.05080 2.77 0.0057 **
## LastAuthorFemale1 -0.11285 0.05242 -2.15 0.0315 *
## Year1997 0.01325 0.13570 0.10 0.9222
## Year1998 0.02583 0.13830 0.19 0.8519
## Year1999 -0.06472 0.14381 -0.45 0.6527
## Year2000 -0.10375 0.14826 -0.70 0.4842
## Year2001 0.05855 0.14269 0.41 0.6816
## Year2002 -0.09520 0.13591 -0.70 0.4837
## Year2003 0.16128 0.13395 1.20 0.2288
## Year2004 0.11251 0.12904 0.87 0.3834
## Year2005 -0.00101 0.13136 -0.01 0.9938
```

```

## Year2006          0.15422      0.12466      1.24      0.2162
## Year2007          -0.00994      0.12983     -0.08      0.9390
## Year2008          0.11561      0.12581      0.92      0.3583
## Year2009          0.18717      0.12144      1.54      0.1235
## Year2010          0.05125      0.12211      0.42      0.6747
## Year2011          0.20661      0.11833      1.75      0.0810 .
## Year2012          0.27096      0.12375      2.19      0.0287 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.659
## Multiple R-squared:  0.0304, Adjusted R-squared:  0.0185
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 112 weights are ~= 1. The remaining 1367 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.287  0.871   0.950   0.911   0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.76e-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 1.08 1      1.039
## Year              1.08 16      1.002

```



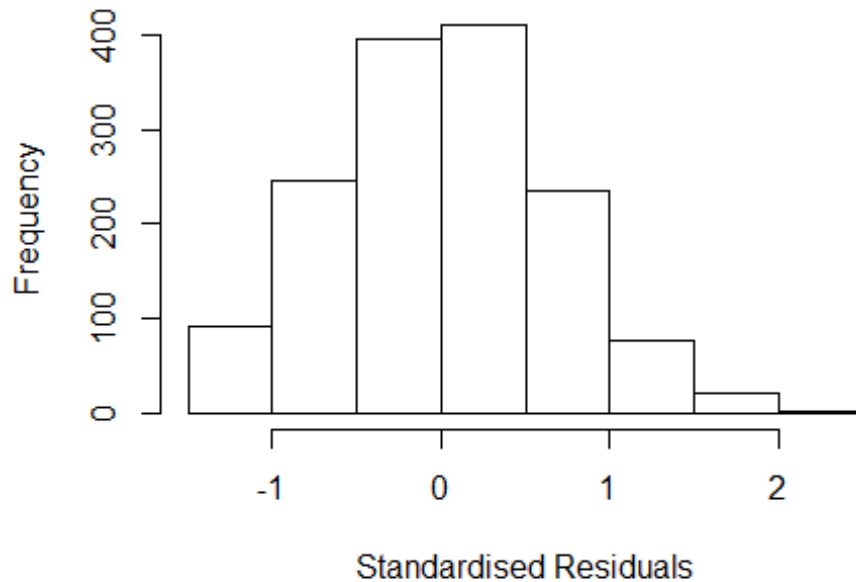
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27045 -0.44607 0.00587 0.44472 2.11787
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94150 0.10701 8.80 <2e-16 ***
## FirstAuthorFemale1 0.06189 0.03734 1.66 0.098 .
## Year1997 0.02384 0.13561 0.18 0.860
## Year1998 0.03289 0.13719 0.24 0.811
## Year1999 -0.06064 0.14304 -0.42 0.672
## Year2000 -0.09893 0.14790 -0.67 0.504
## Year2001 0.05978 0.14326 0.42 0.677
## Year2002 -0.09293 0.13576 -0.68 0.494
## Year2003 0.16259 0.13348 1.22 0.223
## Year2004 0.11388 0.12896 0.88 0.377
## Year2005 -0.00485 0.13116 -0.04 0.971
## Year2006 0.15388 0.12405 1.24 0.215
```

```

## Year2007          -0.00782    0.12952   -0.06    0.952
## Year2008          0.10599    0.12522    0.85    0.397
## Year2009          0.18641    0.12092    1.54    0.123
## Year2010          0.04579    0.12162    0.38    0.707
## Year2011          0.20263    0.11763    1.72    0.085 .
## Year2012          0.26707    0.12325    2.17    0.030 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.661
## Multiple R-squared:  0.0271, Adjusted R-squared:  0.0157
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1360 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.284  0.873  0.949  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      6.76e-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 1.106 1          1.052
## Year            1.106 16          1.003

```

Residuals from last author



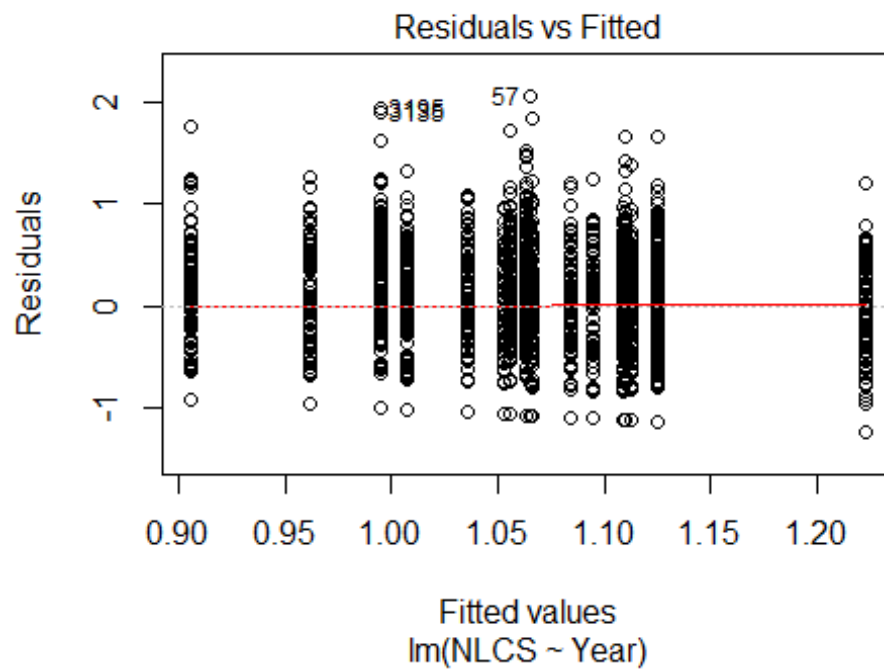
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.242001 -0.452968 0.000414 0.446123 2.088626
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9562 0.1083 8.83 <2e-16 ***
## LastAuthorFemale1 -0.0121 0.0383 -0.32 0.752
## Year1997 0.0267 0.1369 0.20 0.845
## Year1998 0.0343 0.1385 0.25 0.804
## Year1999 -0.0497 0.1440 -0.35 0.730
## Year2000 -0.0949 0.1491 -0.64 0.525
## Year2001 0.0645 0.1443 0.45 0.655
## Year2002 -0.0806 0.1372 -0.59 0.557
## Year2003 0.1686 0.1341 1.26 0.209
## Year2004 0.1242 0.1301 0.95 0.340
## Year2005 0.0148 0.1320 0.11 0.911
## Year2006 0.1641 0.1253 1.31 0.191
```

```

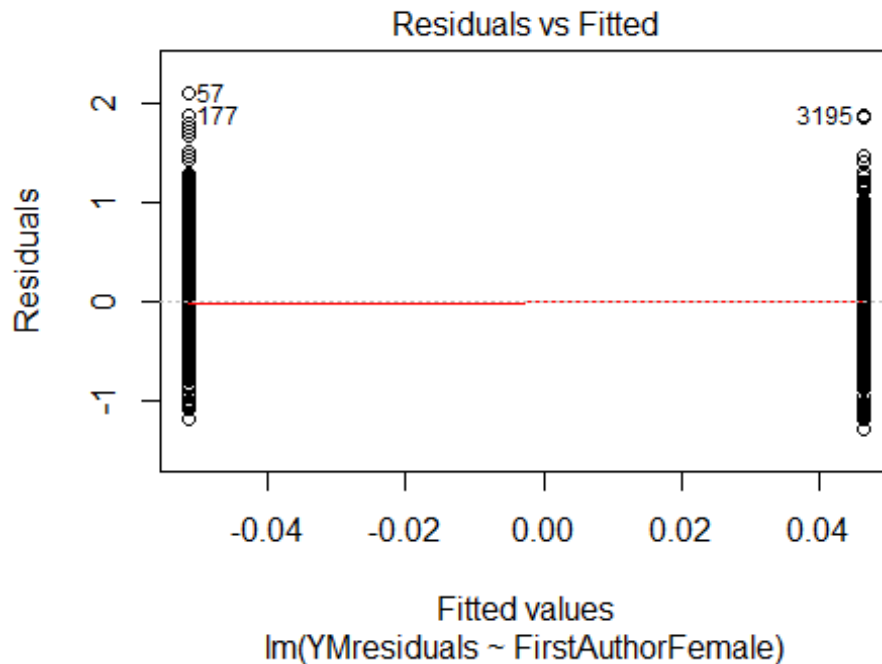
## Year2007          0.0112      0.1300      0.09      0.931
## Year2008          0.1273      0.1267      1.01      0.315
## Year2009          0.2034      0.1221      1.67      0.096 .
## Year2010          0.0672      0.1230      0.55      0.585
## Year2011          0.2172      0.1189      1.83      0.068 .
## Year2012          0.2858      0.1245      2.30      0.022 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.664
## Multiple R-squared:  0.0253, Adjusted R-squared:  0.014
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 115 weights are ~ = 1. The remaining 1364 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.302  0.872  0.949  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      6.76e-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: 1479"
## [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
## 102  95  90  103  119  128  116  90  123  154  160  192  190  250  272
## 2011 2012
## 264  261
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 91  75  71  95  97  102  101  80  105  132  128  167  157  220  238
## 2011 2012

```

```
## 228 225
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 85 70 68 90 92 98 95 73 99 123 119 158 140 211 228
## 2011 2012
## 219 209
## [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
```

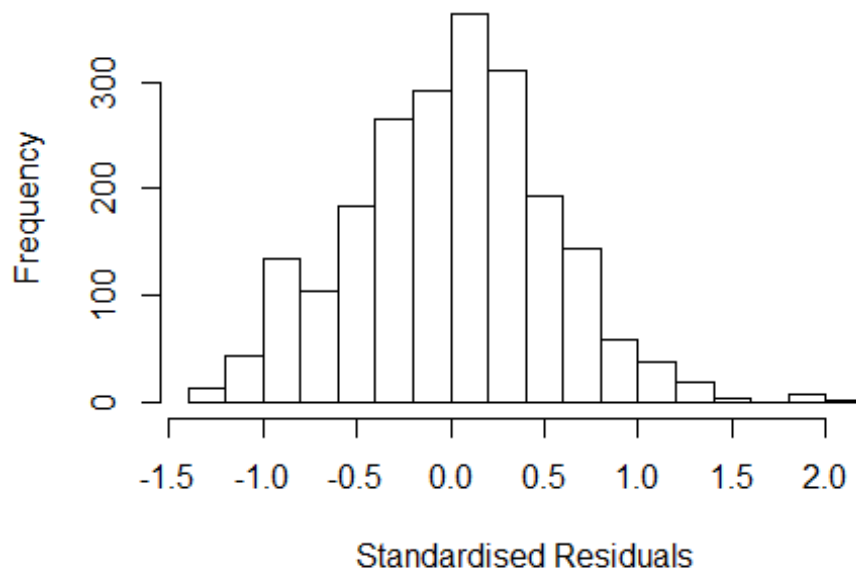


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.6, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 2.41412197354611"
## [1] "Male first author team size 2018 geometric mean: 1.73450794616655"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7700, p-value = 3e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.34055322052228"
## [1] "Male last author team size 2018 geometric mean: 1.87763512973727"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7400, 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 1.603 1          1.266
## LastAuthorFemale  1.588 1          1.260
## UniqueAuthors    1.170 4          1.020
## Year              1.229 16         1.006
```


Residuals from first and last author and team size



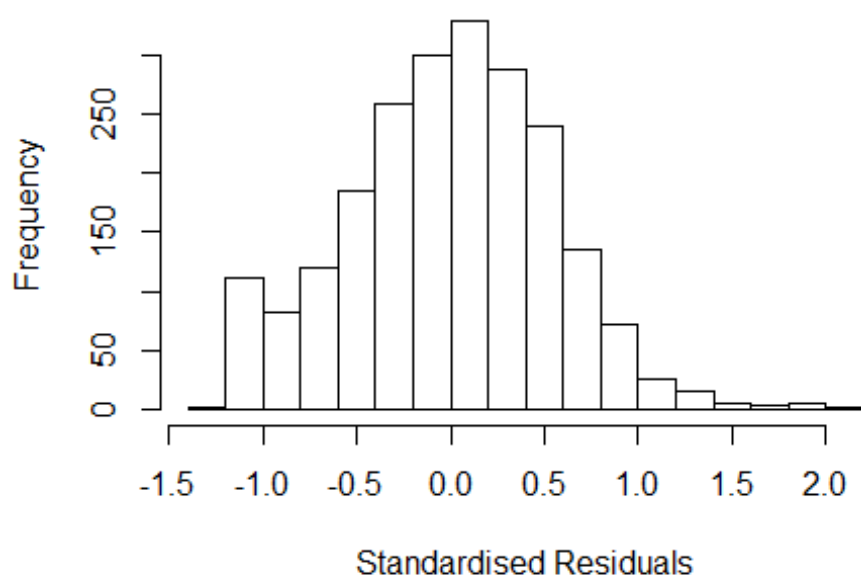
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2857 -0.3451 0.0303 0.3407 2.1563
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9567 0.0683 14.00 < 2e-16 ***
## FirstAuthorFemale1 0.0579 0.0294 1.97 0.049 *
## LastAuthorFemale1 0.0366 0.0291 1.26 0.209
## UniqueAuthors2 0.2106 0.0285 7.38 2.2e-13 ***
## UniqueAuthors3 0.2901 0.0359 8.09 1.0e-15 ***
## UniqueAuthors4 0.3307 0.0447 7.40 1.9e-13 ***
## UniqueAuthors5 0.3600 0.0469 7.68 2.4e-14 ***
## Year1997 -0.0549 0.0876 -0.63 0.531
## Year1998 -0.0322 0.0935 -0.34 0.731
## Year1999 -0.0767 0.0890 -0.86 0.389
```

```

## Year2000          -0.2429      0.0967    -2.51      0.012 *
## Year2001          -0.1369      0.0844    -1.62      0.105
## Year2002           0.1178      0.0839     1.40      0.161
## Year2003          -0.0327      0.0942    -0.35      0.729
## Year2004          -0.0513      0.0818    -0.63      0.531
## Year2005          -0.1254      0.0824    -1.52      0.128
## Year2006          -0.0315      0.0788    -0.40      0.689
## Year2007          -0.0354      0.0764    -0.46      0.643
## Year2008          -0.1104      0.0828    -1.33      0.183
## Year2009          -0.0394      0.0792    -0.50      0.619
## Year2010          -0.0338      0.0758    -0.45      0.656
## Year2011          -0.1146      0.0784    -1.46      0.144
## Year2012          -0.1781      0.0790    -2.25      0.024 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.0882, Adjusted R-squared:  0.0789
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 183 weights are ~= 1. The remaining 1994 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.043  0.866  0.952   0.903   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.59e-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 1.573 1          1.254
## LastAuthorFemale  1.566 1          1.252
## Year              1.086 16          1.003

```

Residuals from first and last author



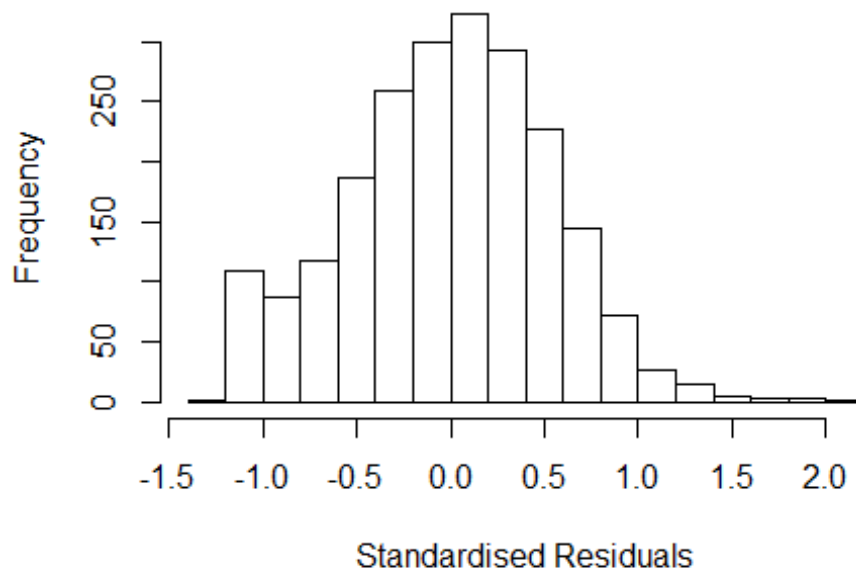
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3108 -0.3738  0.0172  0.3636  2.1048
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.00821    0.06702   15.04  <2e-16 ***
## FirstAuthorFemale1  0.09496    0.03037    3.13   0.0018 **
## LastAuthorFemale1  0.03868    0.03009    1.29   0.1988
## Year1997        -0.03509    0.08622   -0.41   0.6841
## Year1998         0.00407    0.09397    0.04   0.9655
## Year1999        -0.03473    0.08975   -0.39   0.6988
## Year2000        -0.20639    0.09771   -2.11   0.0348 *
## Year2001        -0.10586    0.08483   -1.25   0.2122
## Year2002         0.16899    0.08420    2.01   0.0449 *
## Year2003         0.02552    0.09254    0.28   0.7828
## Year2004        -0.00207    0.08117   -0.03   0.9797
## Year2005        -0.09176    0.08298   -1.11   0.2689
```

```

## Year2006          0.02394    0.07895    0.30    0.7618
## Year2007          0.03430    0.07549    0.45    0.6496
## Year2008         -0.04458    0.08230   -0.54    0.5881
## Year2009          0.03659    0.07716    0.47    0.6354
## Year2010          0.02977    0.07511    0.40    0.6919
## Year2011         -0.04350    0.07872   -0.55    0.5806
## Year2012         -0.09878    0.07920   -1.25    0.2125
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.0294, Adjusted R-squared:  0.0213
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 196 weights are ~= 1. The remaining 1981 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.101  0.869  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      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 1.047 1      1.023
## Year              1.047 16      1.001

```

Residuals from first author



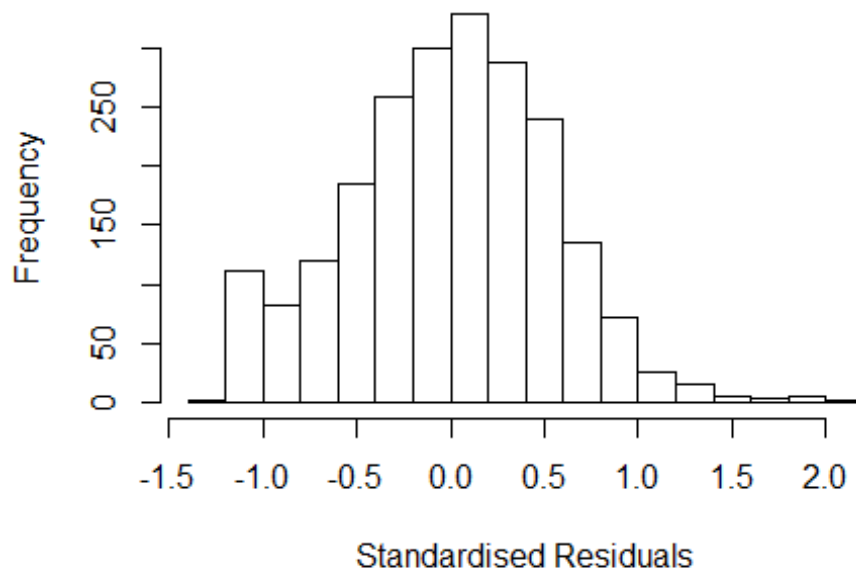
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2979 -0.3762  0.0198  0.3684  2.1002
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.01285    0.06673   15.18 < 2e-16 ***
## FirstAuthorFemale1 0.11843    0.02482    4.77 1.9e-06 ***
## Year1997        -0.03478    0.08616   -0.40  0.686
## Year1998         0.00486    0.09420    0.05  0.959
## Year1999        -0.03111    0.08969   -0.35  0.729
## Year2000        -0.20173    0.09759   -2.07  0.039 *
## Year2001        -0.10104    0.08463   -1.19  0.233
## Year2002         0.16663    0.08402    1.98  0.047 *
## Year2003         0.02866    0.09254    0.31  0.757
## Year2004        -0.00353    0.08093   -0.04  0.965
## Year2005        -0.08866    0.08294   -1.07  0.285
## Year2006         0.02681    0.07892    0.34  0.734
```

```

## Year2007          0.03825    0.07544    0.51    0.612
## Year2008          -0.04078    0.08209   -0.50    0.619
## Year2009          0.03995    0.07698    0.52    0.604
## Year2010          0.03119    0.07515    0.42    0.678
## Year2011          -0.03929    0.07859   -0.50    0.617
## Year2012          -0.09512    0.07907   -1.20    0.229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.545
## Multiple R-squared:  0.0286, Adjusted R-squared:  0.0209
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 193 weights are ~= 1. The remaining 1984 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.104  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      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 1.042 1      1.021
## Year              1.042 16      1.001

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3012 -0.3704 0.0246 0.3716 2.0882
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02476 0.06807 15.05 <2e-16 ***
## LastAuthorFemale1 0.09602 0.02468 3.89 0.0001 ***
## Year1997 -0.03359 0.08828 -0.38 0.7036
## Year1998 0.00128 0.09444 0.01 0.9892
## Year1999 -0.04003 0.09134 -0.44 0.6613
## Year2000 -0.20604 0.09920 -2.08 0.0379 *
## Year2001 -0.11436 0.08599 -1.33 0.1837
## Year2002 0.18037 0.08577 2.10 0.0356 *
## Year2003 0.03347 0.09337 0.36 0.7200
## Year2004 0.00916 0.08252 0.11 0.9116
## Year2005 -0.08828 0.08387 -1.05 0.2926
## Year2006 0.02922 0.07995 0.37 0.7148
```

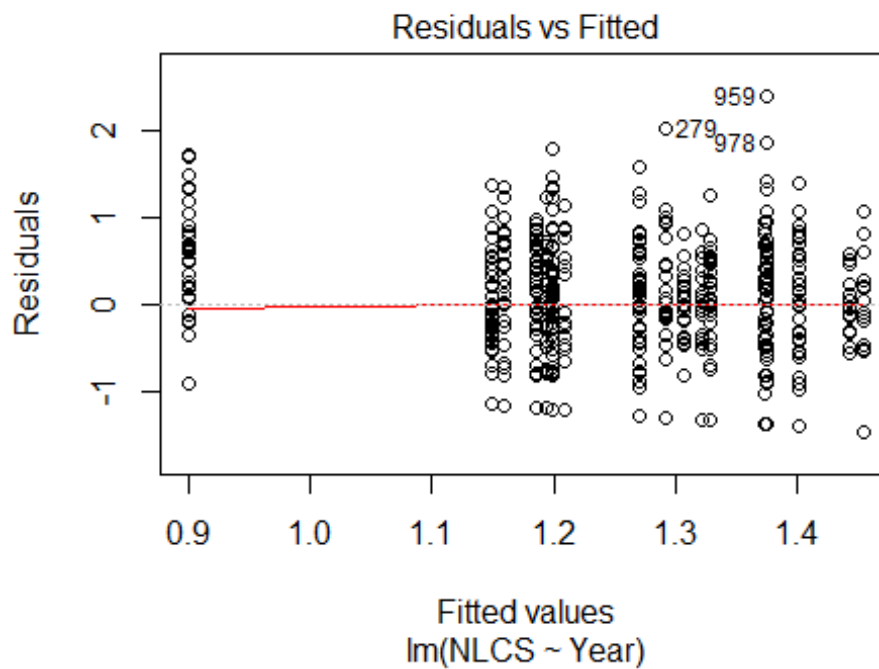
```

## Year2007      0.04104      0.07639      0.54      0.5911
## Year2008     -0.04138      0.08369     -0.49      0.6211
## Year2009      0.04218      0.07841      0.54      0.5907
## Year2010      0.03892      0.07603      0.51      0.6088
## Year2011     -0.03771      0.08003     -0.47      0.6375
## Year2012     -0.08891      0.08045     -1.11      0.2692
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.545
## Multiple R-squared:  0.0249, Adjusted R-squared:  0.0172
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 187 weights are ~= 1. The remaining 1990 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.109  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      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 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
##   17   20   24   40   23   28   36   37   50   36   46   66   62   52   58
## 2011 2012
##   60   58
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   14   22   19   16   19   30   29   38   27   34   55   53   37   47
## 2011 2012

```



```
## 51 53
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 13 18 17 15 18 28 24 36 27 31 54 52 32 47
## 2011 2012
## 50 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 = 40, df = 16, p-value = 7e-04
```



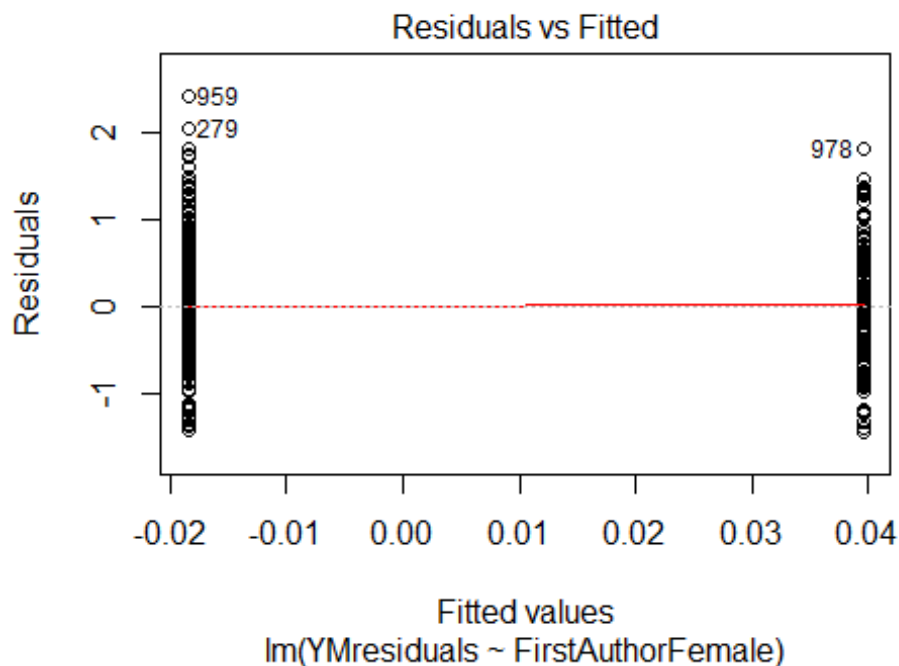
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.39, df = 1, p-value = 0.5

## [1] "Female first author team size 2018 geometric mean: 1.64375182951723"
## [1] "Male first author team size 2018 geometric mean: 1.5397382892659"

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

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 480, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.57517674960554"
## [1] "Male last author team size 2018 geometric mean: 1.58541391028459"

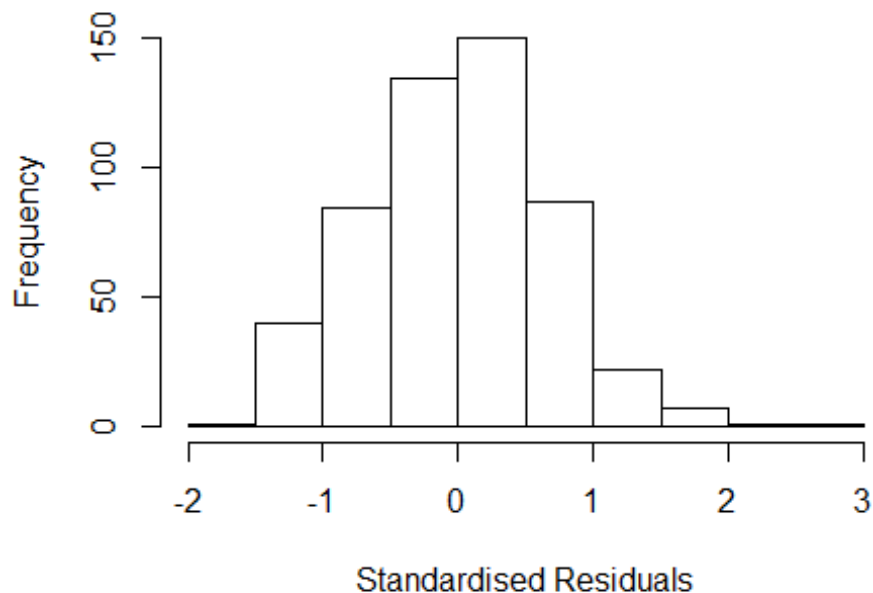
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 440, p-value = 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	1.509	1	1.228
LastAuthorFemale	1.484	1	1.218
UniqueAuthors	2.169	4	1.102
Year	2.286	16	1.026

Residuals from first and last author and team size



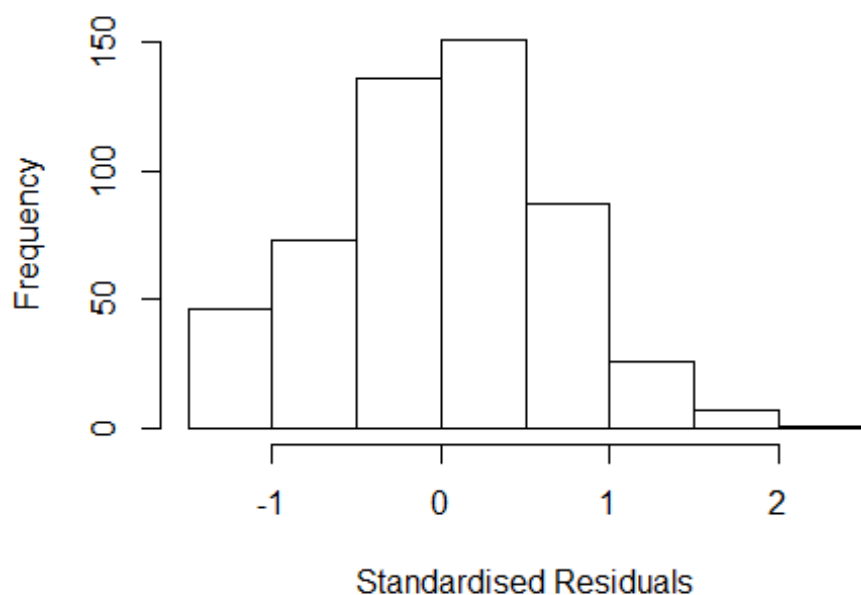
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 959 84859871117 3.764 2012    1202      3      2.608
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.53933 -0.45992  0.00768  0.45460  2.60842
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3722    0.1635   8.39 4.9e-16 ***
## FirstAuthorFemale1  0.1456    0.0739   1.97  0.0494 *
## LastAuthorFemale1 -0.1504    0.0782  -1.92  0.0550 .
## UniqueAuthors2     0.1672    0.0905   1.85  0.0654 .
## UniqueAuthors3     0.4827    0.0923   5.23 2.5e-07 ***
## UniqueAuthors4     0.5393    0.1207   4.47 9.8e-06 ***
## UniqueAuthors5     0.3890    0.1221   3.19  0.0015 **
## Year1997          -0.0789    0.1927  -0.41  0.6822
## Year1998          -0.0506    0.2269  -0.22  0.8237
## Year1999          -0.1189    0.2054  -0.58  0.5629
```

```

## Year2000      -0.2553      0.2779      -0.92      0.3587
## Year2001      -0.1793      0.1926      -0.93      0.3524
## Year2002      -0.3425      0.2104      -1.63      0.1042
## Year2003      -0.2151      0.2548      -0.84      0.3989
## Year2004      -0.1468      0.1962      -0.75      0.4547
## Year2005      -0.2329      0.1937      -1.20      0.2298
## Year2006      -0.0761      0.1944      -0.39      0.6957
## Year2007      -0.6129      0.2055      -2.98      0.0030 **
## Year2008      -0.2570      0.1902      -1.35      0.1772
## Year2009      -0.3630      0.2064      -1.76      0.0793 .
## Year2010      -0.3535      0.1832      -1.93      0.0542 .
## Year2011      -0.4361      0.1973      -2.21      0.0276 *
## Year2012      -0.2166      0.2013      -1.08      0.2824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.685
## Multiple R-squared:  0.119, Adjusted R-squared:  0.0802
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 49 weights are ~= 1. The remaining 478 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.115  0.882  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      1.90e-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 1.529 1      1.237
## LastAuthorFemale  1.539 1      1.241
## Year              1.181 16      1.005

```

Residuals from first and last author

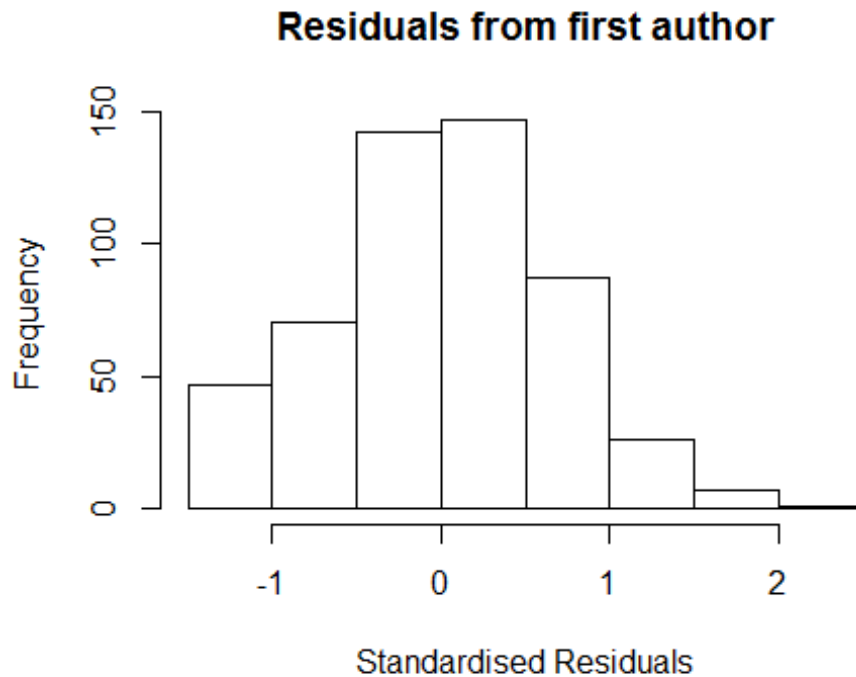


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4345 -0.4642 0.0217 0.4773 2.4797
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4345 0.1569 9.14 <2e-16 ***
## FirstAuthorFemale1 0.2338 0.0776 3.01 0.0027 **
## LastAuthorFemale1 -0.2064 0.0830 -2.49 0.0132 *
## Year1997 -0.0534 0.1776 -0.30 0.7638
## Year1998 -0.0891 0.2269 -0.39 0.6946
## Year1999 -0.1009 0.1983 -0.51 0.6111
## Year2000 -0.2004 0.2736 -0.73 0.4643
## Year2001 -0.1674 0.1763 -0.95 0.3429
## Year2002 -0.2833 0.2133 -1.33 0.1847
## Year2003 -0.2791 0.2519 -1.11 0.2683
## Year2004 -0.0893 0.1964 -0.45 0.6495
## Year2005 -0.2511 0.1883 -1.33 0.1828
```

```

## Year2006          -0.0822      0.1872   -0.44   0.6608
## Year2007          -0.6155      0.2054   -3.00   0.0029 **
## Year2008          -0.1945      0.1826   -1.07   0.2873
## Year2009          -0.2278      0.2028   -1.12   0.2619
## Year2010          -0.3033      0.1772   -1.71   0.0876 .
## Year2011          -0.3072      0.1877   -1.64   0.1023
## Year2012          -0.1502      0.1986   -0.76   0.4499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.689
## Multiple R-squared:  0.0582, Adjusted R-squared:  0.0248
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 475 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.168  0.875   0.950   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      1.90e-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 1.091 1      1.045
## Year              1.091 16      1.003

```

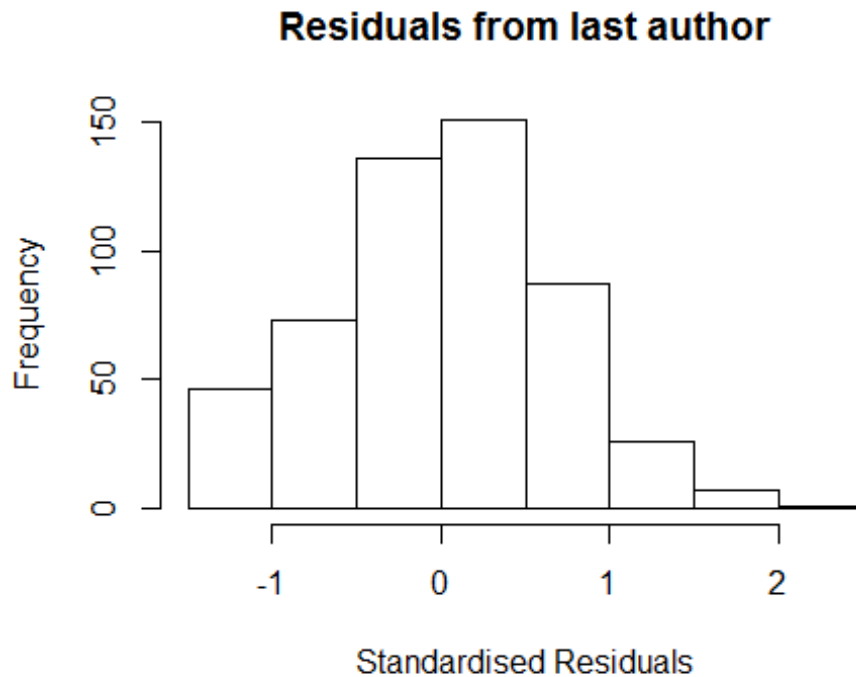


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4288 -0.4479  0.0271  0.4681  2.4825
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4265     0.1594   8.95  <2e-16 ***
## FirstAuthorFemale1  0.1044     0.0673   1.55   0.1212
## Year1997         -0.0416     0.1823  -0.23   0.8198
## Year1998         -0.0850     0.2311  -0.37   0.7133
## Year1999         -0.1176     0.2014  -0.58   0.5596
## Year2000         -0.1928     0.2740  -0.70   0.4820
## Year2001         -0.1820     0.1812  -1.00   0.3157
## Year2002         -0.2959     0.2161  -1.37   0.1714
## Year2003         -0.2941     0.2604  -1.13   0.2594
## Year2004         -0.1021     0.2009  -0.51   0.6115
## Year2005         -0.2562     0.1886  -1.36   0.1751
## Year2006         -0.0944     0.1903  -0.50   0.6200
```

```

## Year2007          -0.6224      0.2057   -3.03   0.0026 **
## Year2008          -0.1950      0.1840   -1.06   0.2898
## Year2009          -0.2498      0.2045   -1.22   0.2226
## Year2010          -0.3339      0.1793   -1.86   0.0632 .
## Year2011          -0.3157      0.1897   -1.66   0.0967 .
## Year2012          -0.1449      0.2022   -0.72   0.4737
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.689
## Multiple R-squared:  0.05,   Adjusted R-squared:  0.0183
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 477 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.167  0.880  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      1.90e-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 1.103 1          1.050
## Year            1.103 16          1.003

```

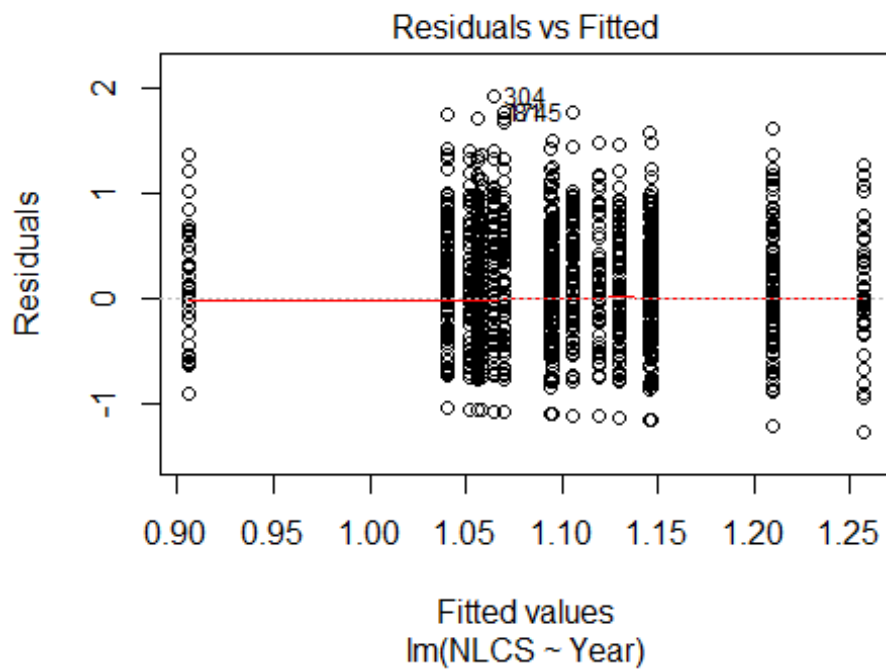
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4860 -0.4826 0.0255 0.4794 2.4405
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4860 0.1638 9.07 <2e-16 ***
## LastAuthorFemale1 -0.0470 0.0706 -0.67 0.5056
## Year1997 -0.0819 0.1867 -0.44 0.6613
## Year1998 -0.1188 0.2362 -0.50 0.6153
## Year1999 -0.1289 0.2029 -0.64 0.5255
## Year2000 -0.2293 0.2681 -0.86 0.3929
## Year2001 -0.1980 0.1835 -1.08 0.2810
## Year2002 -0.3003 0.2214 -1.36 0.1756
## Year2003 -0.2957 0.2562 -1.15 0.2490
## Year2004 -0.1029 0.2045 -0.50 0.6149
## Year2005 -0.2808 0.1956 -1.44 0.1518
## Year2006 -0.1165 0.1946 -0.60 0.5496
```

```

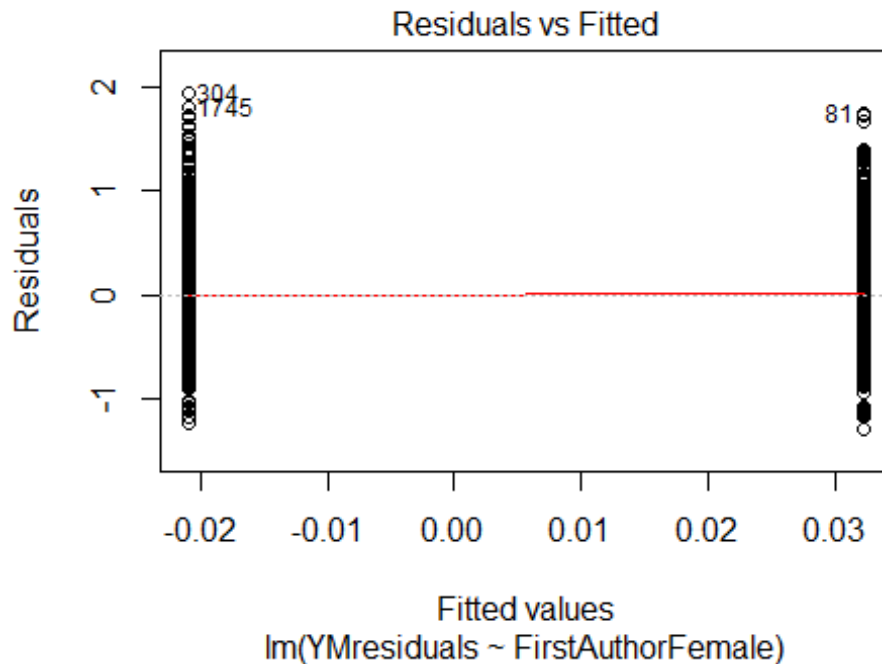
## Year2007          -0.6326      0.2158   -2.93   0.0035 **
## Year2008          -0.2110      0.1887   -1.12   0.2640
## Year2009          -0.2704      0.2091   -1.29   0.1966
## Year2010          -0.3250      0.1841   -1.77   0.0781 .
## Year2011          -0.3234      0.1930   -1.68   0.0944 .
## Year2012          -0.1625      0.2051   -0.79   0.4286
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.687
## Multiple R-squared:  0.0458, Adjusted R-squared:  0.014
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 479 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.181  0.865  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      1.90e-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: 527"
## [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
##   105   99  103   98   89   94  115  124  143  140  127  128  138  148  174
## 2011 2012
##   168  146
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   89   75   81   71   51   44   92   94  117  119  105  104  117  126  142
## 2011 2012

```

```
## 148 120
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 85 74 79 67 51 42 87 88 108 115 102 99 110 119 132
## 2011 2012
## 135 108
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 17, df = 16, p-value = 0.4
```

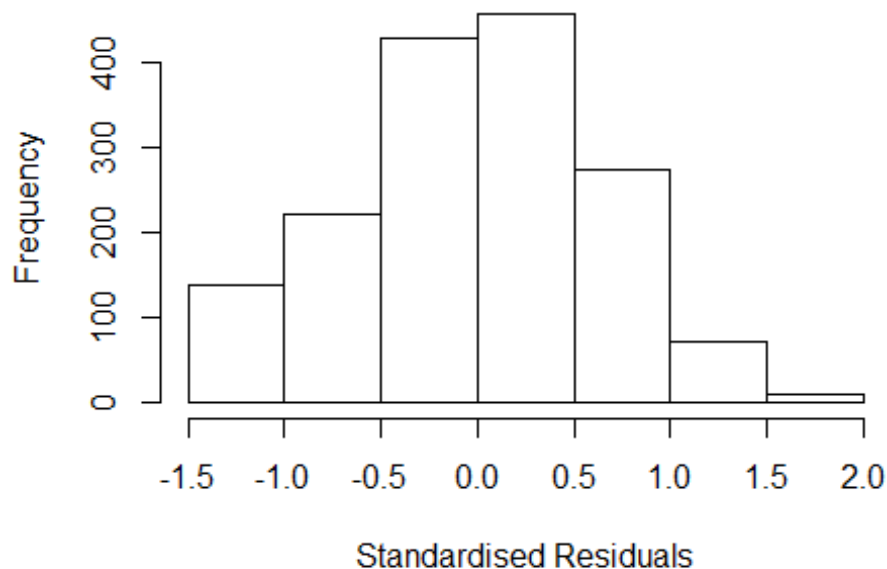


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.9, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 1.75046584735332"
## [1] "Male first author team size 2018 geometric mean: 1.55994183198369"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.55407193274284"
## [1] "Male last author team size 2018 geometric mean: 1.7110114699818"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, 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 1.750 1      1.323
## LastAuthorFemale  1.748 1      1.322
## UniqueAuthors    1.211 4      1.024
## Year              1.274 16     1.008
```

Residuals from first and last author and team size



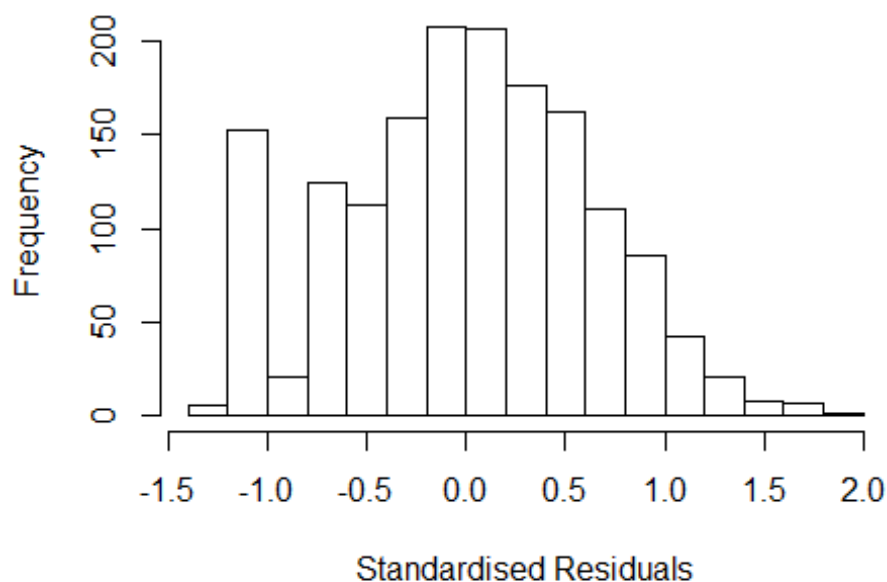
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4316 -0.4186 0.0059 0.4408 1.9923
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00875 0.08534 11.82 < 2e-16 ***
## FirstAuthorFemale1 0.06420 0.04241 1.51 0.13029
## LastAuthorFemale1 -0.02011 0.04320 -0.47 0.64159
## UniqueAuthors2 0.12362 0.03984 3.10 0.00195 **
## UniqueAuthors3 0.19558 0.05533 3.53 0.00042 ***
## UniqueAuthors4 0.01774 0.12440 0.14 0.88664
## UniqueAuthors5 0.14296 0.11107 1.29 0.19824
## Year1997 -0.03717 0.11529 -0.32 0.74715
## Year1998 -0.01703 0.12695 -0.13 0.89330
## Year1999 0.07875 0.12431 0.63 0.52648
```

```

## Year2000      0.22725    0.12366    1.84  0.06629 .
## Year2001     -0.14020    0.13366   -1.05  0.29435
## Year2002      0.01996    0.10948    0.18  0.85535
## Year2003     -0.04318    0.10795   -0.40  0.68921
## Year2004     -0.00555    0.10866   -0.05  0.95929
## Year2005      0.04382    0.10271    0.43  0.66967
## Year2006      0.07837    0.10262    0.76  0.44519
## Year2007     -0.02315    0.10619   -0.22  0.82745
## Year2008      0.10410    0.10469    0.99  0.32019
## Year2009      0.02285    0.10574    0.22  0.82891
## Year2010     -0.02983    0.10280   -0.29  0.77174
## Year2011      0.07800    0.10029    0.78  0.43686
## Year2012      0.05706    0.10372    0.55  0.58229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.645
## Multiple R-squared:  0.0229, Adjusted R-squared:  0.00923
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 146 weights are ~= 1. The remaining 1455 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.320  0.873  0.947  0.913  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.25e-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 1.772 1      1.331
## LastAuthorFemale  1.753 1      1.324
## Year              1.092 16      1.003

```

Residuals from first and last author



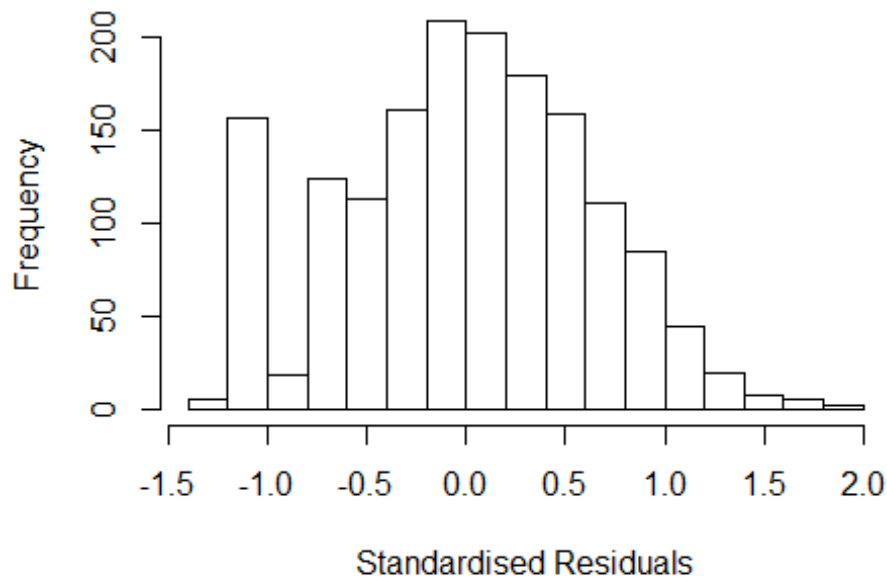
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3092 -0.4420  0.0106  0.4342  1.9482
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03130    0.08509   12.12  <2e-16 ***
## FirstAuthorFemale1  0.07871    0.04304    1.83   0.068 .
## LastAuthorFemale1 -0.03334    0.04360   -0.76   0.445
## Year1997        -0.01531    0.11413   -0.13   0.893
## Year1998         0.00455    0.12820    0.04   0.972
## Year1999         0.08770    0.12430    0.71   0.481
## Year2000         0.23254    0.12445    1.87   0.062 .
## Year2001        -0.11342    0.13394   -0.85   0.397
## Year2002         0.02508    0.10992    0.23   0.820
## Year2003        -0.02202    0.10693   -0.21   0.837
## Year2004         0.02217    0.10764    0.21   0.837
## Year2005         0.04744    0.10323    0.46   0.646
```

```

## Year2006          0.10738    0.10209    1.05    0.293
## Year2007          0.00404    0.10650    0.04    0.970
## Year2008          0.13778    0.10514    1.31    0.190
## Year2009          0.05300    0.10595    0.50    0.617
## Year2010         -0.01430    0.10289   -0.14    0.889
## Year2011          0.09890    0.10036    0.99    0.325
## Year2012          0.08819    0.10390    0.85    0.396
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.646
## Multiple R-squared:  0.0123, Adjusted R-squared:  0.00111
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 143 weights are ~= 1. The remaining 1458 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.342  0.872  0.950  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.064 1         1.032
## Year              1.064 16         1.002

```


Residuals from first author



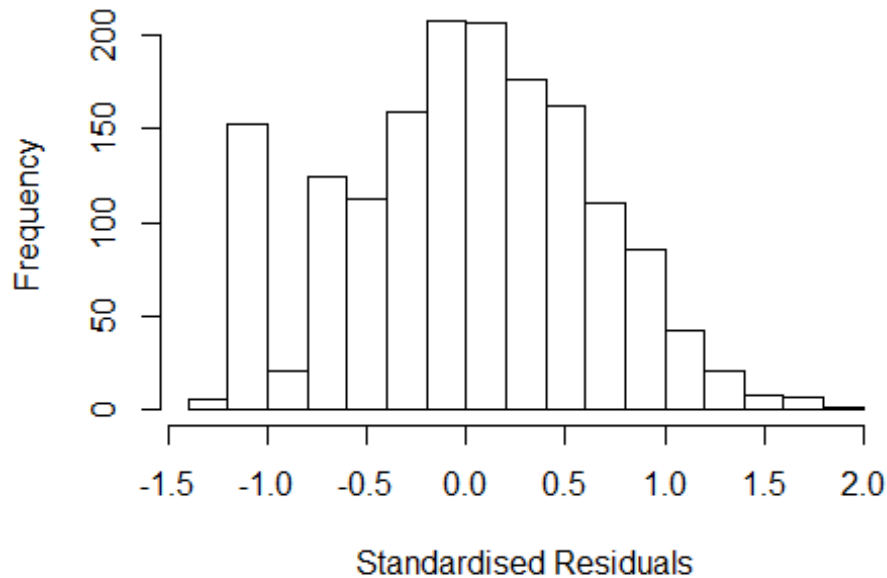
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3151 -0.4261  0.0112  0.4367  1.9505
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.030896   0.085442   12.07  <2e-16 ***
## FirstAuthorFemale1 0.055170   0.033577    1.64   0.101
## Year1997      -0.018791   0.114463   -0.16   0.870
## Year1998       0.002561   0.128375    0.02   0.984
## Year1999       0.084737   0.124503    0.68   0.496
## Year2000       0.228991   0.124365    1.84   0.066 .
## Year2001      -0.113473   0.134396   -0.84   0.399
## Year2002       0.022074   0.110167    0.20   0.841
## Year2003      -0.026188   0.106959   -0.24   0.807
## Year2004       0.017138   0.107434    0.16   0.873
## Year2005       0.044002   0.103309    0.43   0.670
## Year2006       0.104380   0.102224    1.02   0.307
```

```

## Year2007          -0.000676    0.106534   -0.01    0.995
## Year2008          0.134594    0.105203    1.28    0.201
## Year2009          0.048787    0.106039    0.46    0.646
## Year2010         -0.016414    0.103153   -0.16    0.874
## Year2011          0.094039    0.100334    0.94    0.349
## Year2012          0.086901    0.103989    0.84    0.403
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.646
## Multiple R-squared:  0.012, Adjusted R-squared:  0.00144
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1452 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.341  0.871  0.949  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 1.057 1          1.028
## Year            1.057 16          1.002

```

Residuals from last author



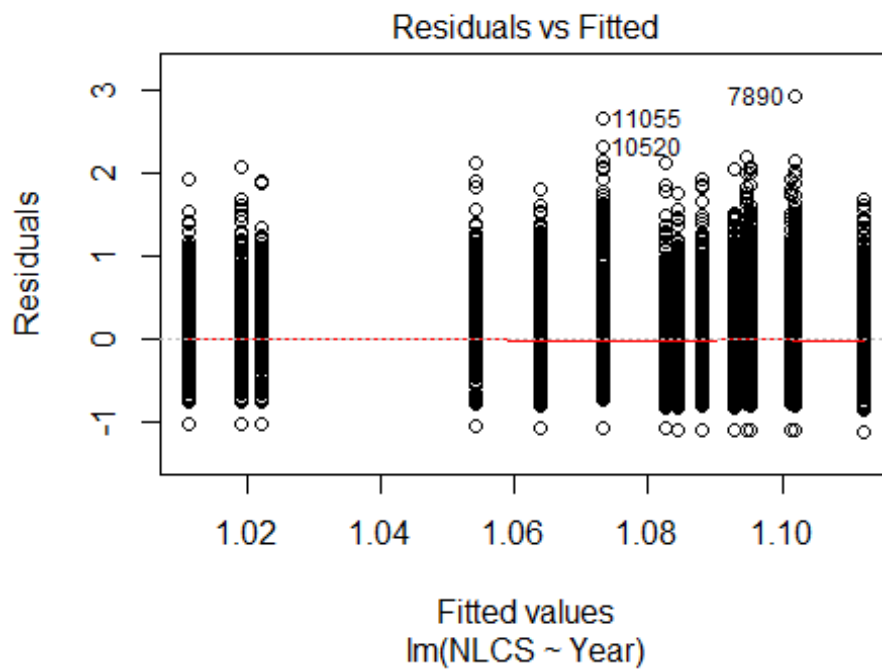
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2926 -0.4350 0.0117 0.4471 1.9422
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04000 0.08571 12.13 <2e-16 ***
## LastAuthorFemale1 0.02300 0.03396 0.68 0.498
## Year1997 -0.01723 0.11530 -0.15 0.881
## Year1998 0.00175 0.12867 0.01 0.989
## Year1999 0.08911 0.12582 0.71 0.479
## Year2000 0.22964 0.12524 1.83 0.067 .
## Year2001 -0.11249 0.13550 -0.83 0.407
## Year2002 0.02698 0.11037 0.24 0.807
## Year2003 -0.02350 0.10774 -0.22 0.827
## Year2004 0.02089 0.10813 0.19 0.847
## Year2005 0.04728 0.10384 0.46 0.649
## Year2006 0.10852 0.10260 1.06 0.290
```

```

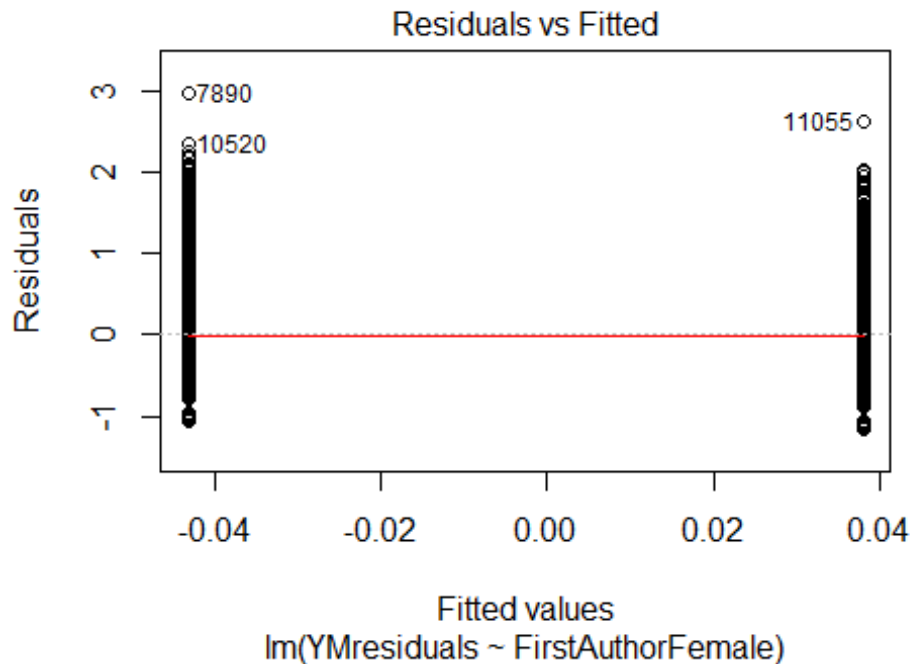
## Year2007      0.00364    0.10730    0.03    0.973
## Year2008      0.14371    0.10570    1.36    0.174
## Year2009      0.05581    0.10646    0.52    0.600
## Year2010     -0.01022    0.10366   -0.10    0.921
## Year2011      0.09738    0.10110    0.96    0.336
## Year2012      0.09481    0.10452    0.91    0.364
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.647
## Multiple R-squared:  0.0106, Adjusted R-squared:  2.46e-05
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 153 weights are ~= 1. The remaining 1448 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.347  0.874  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      6.25e-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: 1601"
## [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
## 312 353 356 351 376 402 420 380 464 479 547 651 728 963 906
## 2011 2012
## 940 987
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 252 275 277 272 301 310 364 317 392 410 454 556 625 822 772
## 2011 2012

```

```
## 806 838
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 232 257 264 250 284 290 348 300 359 389 422 520 582 764 713
## 2011 2012
## 755 777
## [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
```

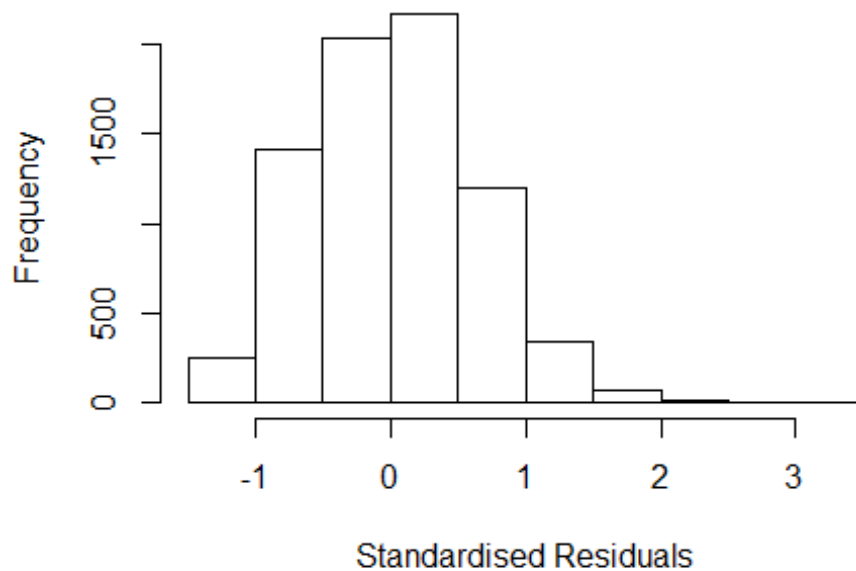


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 14, df = 1, p-value = 2e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 2.20537443988424"
## [1] "Male first author team size 2018 geometric mean: 1.94678056599268"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 74000, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.12103474057015"
## [1] "Male last author team size 2018 geometric mean: 2.09289861663085"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 71000, 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 1.466 1          1.211
## LastAuthorFemale  1.453 1          1.205
## UniqueAuthors    1.051 4          1.006
## Year              1.062 16         1.002
```

Residuals from first and last author and team size



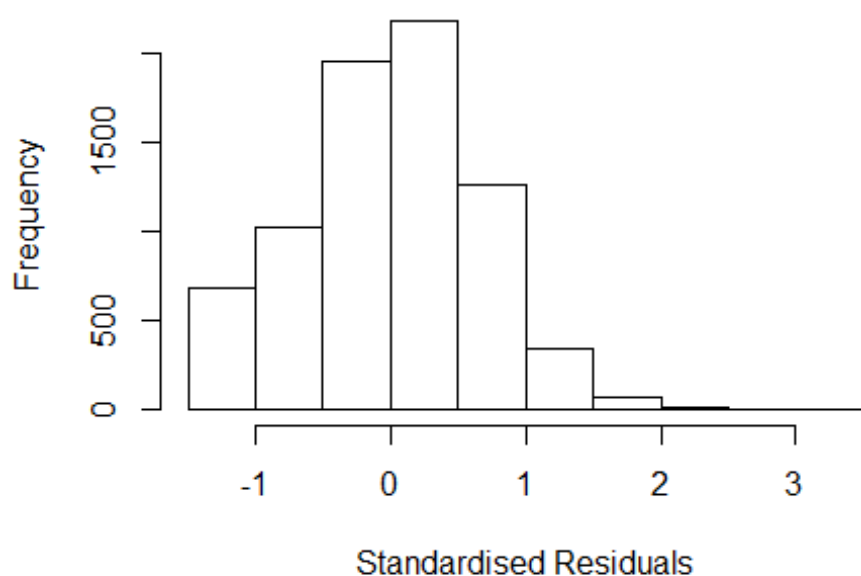
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 7890 78649906763 4.032 2010      2700      2      3.17
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.40221 -0.43662  0.00939  0.43098  3.17044
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.933405   0.049583   18.83  <2e-16 ***
## FirstAuthorFemale1 0.036190   0.018096    2.00  0.0455 *
## LastAuthorFemale1 0.047894   0.017943    2.67  0.0076 **
## UniqueAuthors2    0.225568   0.018580   12.14  <2e-16 ***
## UniqueAuthors3    0.315790   0.022391   14.10  <2e-16 ***
## UniqueAuthors4    0.389577   0.028575   13.63  <2e-16 ***
## UniqueAuthors5    0.432615   0.030637   14.12  <2e-16 ***
## Year1997         -0.028497   0.065868   -0.43  0.6653
## Year1998          0.000156   0.062978    0.00  0.9980
## Year1999         -0.076178   0.063770   -1.19  0.2323
```

```

## Year2000      -0.022759    0.060899   -0.37    0.7086
## Year2001      -0.081599    0.060432   -1.35    0.1770
## Year2002      -0.044299    0.059280   -0.75    0.4549
## Year2003      -0.124474    0.062392   -2.00    0.0461 *
## Year2004      -0.054063    0.058232   -0.93    0.3532
## Year2005      -0.126216    0.057849   -2.18    0.0292 *
## Year2006      -0.071564    0.056707   -1.26    0.2070
## Year2007      -0.040446    0.055495   -0.73    0.4661
## Year2008      -0.064386    0.054435   -1.18    0.2369
## Year2009      -0.043407    0.054106   -0.80    0.4224
## Year2010      -0.071844    0.054293   -1.32    0.1858
## Year2011      -0.081341    0.054448   -1.49    0.1352
## Year2012      -0.115193    0.055246   -2.09    0.0371 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.642
## Multiple R-squared:  0.0655, Adjusted R-squared:  0.0628
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 5316 is an outlier with |weight| = 0 ( < 1.3e-05);
## 621 weights are ~= 1. The remaining 6884 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.126  0.861  0.950  0.913  0.986  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.33e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.469 1 1.212
## LastAuthorFemale 1.468 1 1.212
## Year 1.020 16 1.001

```


Residuals from first and last author



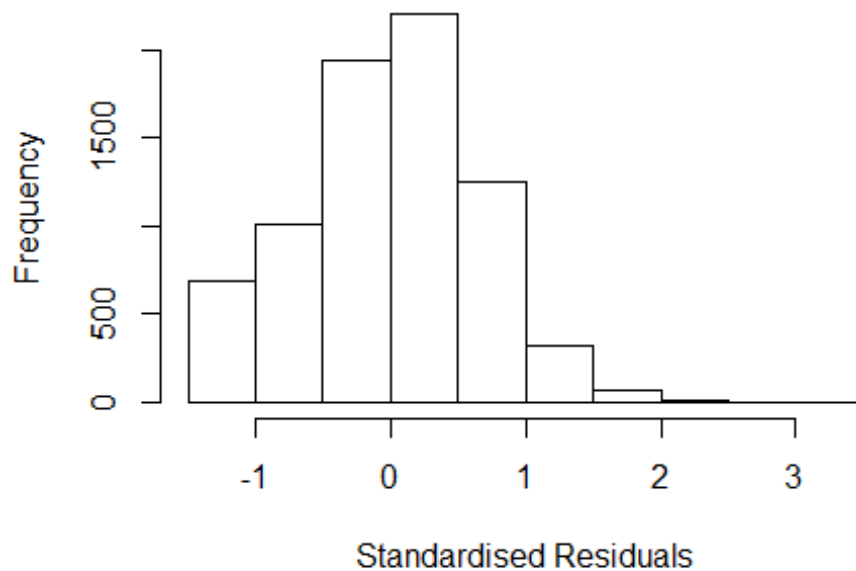
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 7890  78649906763 4.032 2010      2700      2      3.023
## 11055 80052565202 3.737 2012      1700      2      2.645
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1551 -0.4474  0.0217  0.4471  3.0232
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.014028   0.049322  20.56 < 2e-16 ***
## FirstAuthorFemale1 0.065605   0.018739   3.50 0.00047 ***
## LastAuthorFemale1 0.047285   0.018647   2.54 0.01124 *
## Year1997        -0.010110   0.065591  -0.15 0.87751
## Year1998         0.026565   0.063238   0.42 0.67444
## Year1999        -0.059884   0.064520  -0.93 0.35336
## Year2000         0.008633   0.060973   0.14 0.88741
## Year2001        -0.054524   0.061937  -0.88 0.37871
## Year2002        -0.005091   0.060619  -0.08 0.93307
## Year2003        -0.086087   0.063458  -1.36 0.17495
## Year2004         0.008001   0.059000   0.14 0.89214
```

```

## Year2005      -0.075584    0.058337   -1.30  0.19514
## Year2006      -0.017447    0.057447   -0.30  0.76136
## Year2007       0.028197    0.055769    0.51  0.61316
## Year2008       0.005596    0.054752    0.10  0.91860
## Year2009       0.019451    0.054158    0.36  0.71949
## Year2010      -0.005220    0.054416   -0.10  0.92358
## Year2011      -0.000981    0.054474   -0.02  0.98564
## Year2012      -0.034938    0.055611   -0.63  0.52986
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.659
## Multiple R-squared:  0.0084, Adjusted R-squared:  0.00602
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 644 weights are ~= 1. The remaining 6862 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0018 0.8720 0.9500 0.9120 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.012 1      1.006
## Year      1.012 16      1.000

```

Residuals from first author



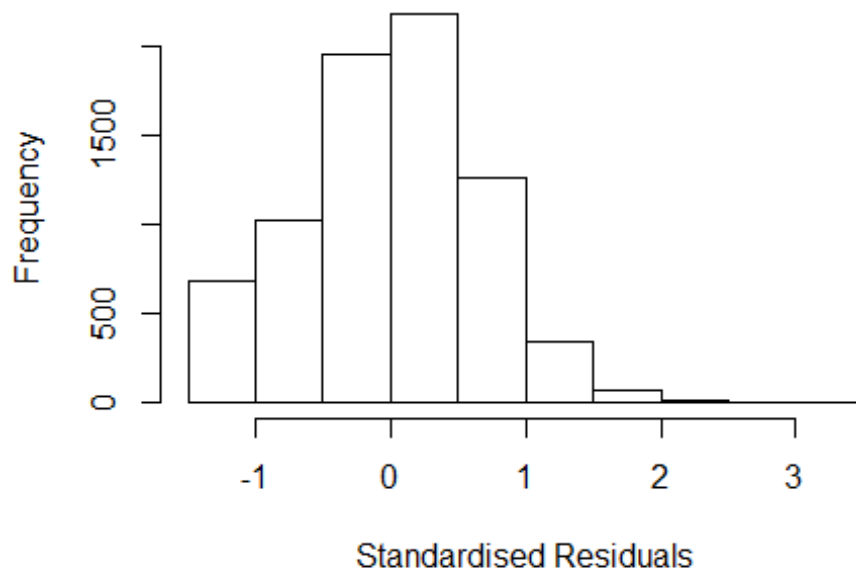
```
## [1] "List of 2 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 7890  78649906763 4.032 2010    2700    2    3.023
## 11055 80052565202 3.737 2012    1700    2    2.645
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.146 -0.441  0.020  0.444  3.014
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.022961   0.049159  20.81  < 2e-16 ***
## FirstAuthorFemale1 0.093131   0.015582   5.98  2.4e-09 ***
## Year1997        -0.008355   0.065673  -0.13    0.90
## Year1998         0.028350   0.063414   0.45    0.65
## Year1999        -0.060447   0.064665  -0.93    0.35
## Year2000         0.007926   0.061097   0.13    0.90
## Year2001        -0.053869   0.061978  -0.87    0.38
## Year2002        -0.006728   0.060612  -0.11    0.91
## Year2003        -0.085835   0.063560  -1.35    0.18
## Year2004         0.007986   0.059015   0.14    0.89
## Year2005        -0.077272   0.058397  -1.32    0.19
```

```

## Year2006      -0.015627    0.057510   -0.27    0.79
## Year2007      0.029948    0.055781    0.54    0.59
## Year2008      0.006404    0.054806    0.12    0.91
## Year2009      0.020839    0.054184    0.38    0.70
## Year2010     -0.004767    0.054456   -0.09    0.93
## Year2011      0.000368    0.054524    0.01    0.99
## Year2012     -0.033189    0.055652   -0.60    0.55
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.659
## Multiple R-squared:  0.00753,    Adjusted R-squared:  0.00527
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 627 weights are ~= 1. The remaining 6879 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0023 0.8730 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      1.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.012 1      1.006
## Year      1.012 16      1.000

```

Residuals from last author



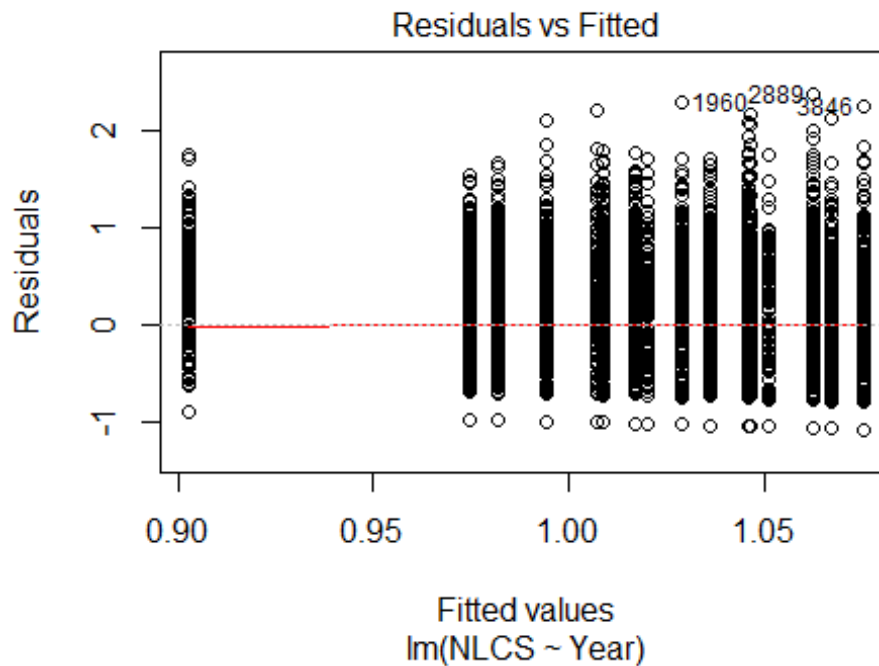
```
## [1] "List of 2 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 7890  78649906763 4.032 2010    2700    2    3.023
## 11055 80052565202 3.737 2012    1700    2    2.645
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1422 -0.4449  0.0196  0.4453  3.0051
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.026534   0.049164   20.88 < 2e-16 ***
## LastAuthorFemale1 0.085212   0.015515    5.49 4.1e-08 ***
## Year1997       -0.010116   0.065541   -0.15  0.88
## Year1998        0.025383   0.063179    0.40  0.69
## Year1999       -0.058874   0.064357   -0.91  0.36
## Year2000        0.011869   0.060929    0.19  0.85
## Year2001       -0.054193   0.062019   -0.87  0.38
## Year2002       -0.000167   0.060727    0.00  1.00
## Year2003       -0.085093   0.063497   -1.34  0.18
## Year2004        0.010225   0.059085    0.17  0.86
## Year2005       -0.070345   0.058359   -1.21  0.23
```

```

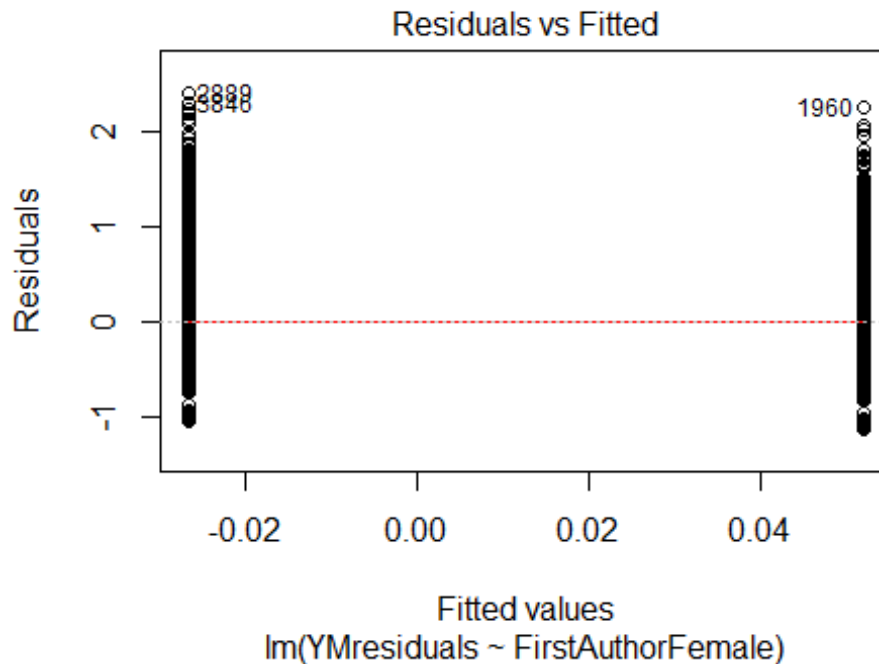
## Year2006      -0.015257    0.057325   -0.27    0.79
## Year2007      0.030434    0.055782    0.55    0.59
## Year2008      0.009640    0.054697    0.18    0.86
## Year2009      0.022610    0.054159    0.42    0.68
## Year2010      0.000377    0.054383    0.01    0.99
## Year2011      0.003014    0.054442    0.06    0.96
## Year2012     -0.029928    0.055557   -0.54    0.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.00673,    Adjusted R-squared:  0.00448
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 631 weights are ~= 1. The remaining 6875 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.003  0.873  0.950  0.912  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##      trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 7506"
## [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
## 243 275 234 231 262 363 373 430 440 359 447 444 476 571 560
## 2011 2012
## 550 548
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 179 193 161 145 138 186 315 362 360 305 383 376 401 468 462

```

```
## 2011 2012
## 464 461
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 174 188 155 137 131 175 306 341 339 292 366 358 382 443 438
## 2011 2012
## 429 432
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 30, df = 16, p-value = 0.02
```

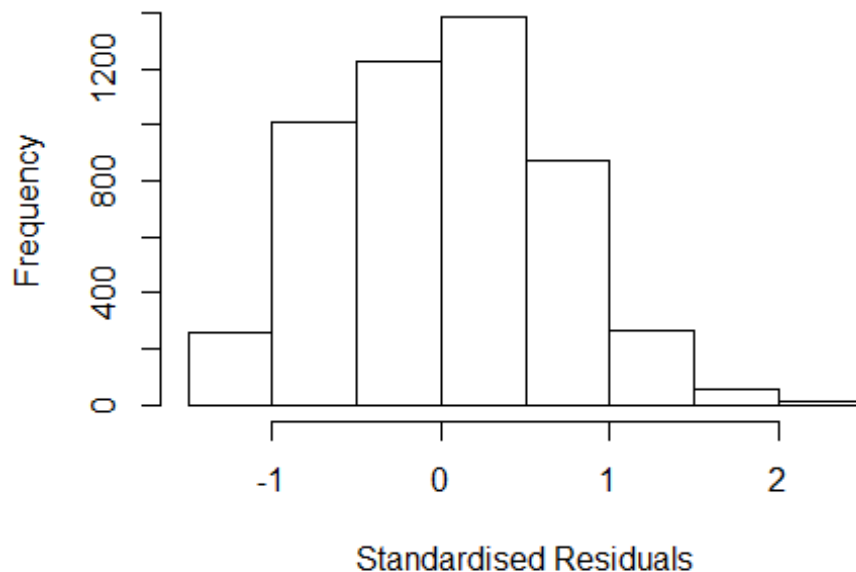


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.2, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 1.7610454632561"
## [1] "Male first author team size 2018 geometric mean: 1.701942515733"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.71019849961161"
## [1] "Male last author team size 2018 geometric mean: 1.73746548415072"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 23000, 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 1.743 1      1.320
## LastAuthorFemale  1.699 1      1.304
## UniqueAuthors    1.148 4      1.017
## Year              1.156 16     1.005
```


Residuals from first and last author and team size



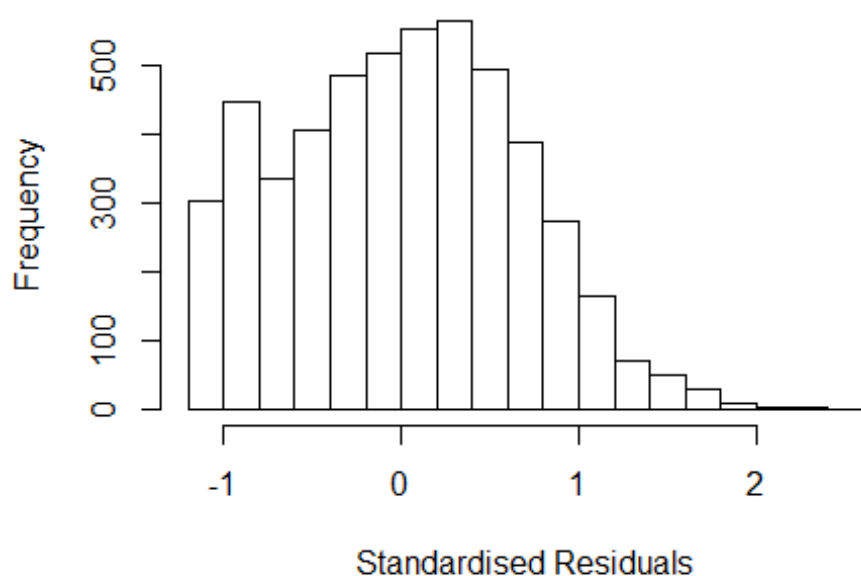
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4371 -0.4989 0.0188 0.4750 2.4894
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9678 0.0575 16.83 < 2e-16 ***
## FirstAuthorFemale1 0.0481 0.0265 1.81 0.070 .
## LastAuthorFemale1 0.0475 0.0269 1.76 0.078 .
## UniqueAuthors2 0.1275 0.0240 5.30 1.2e-07 ***
## UniqueAuthors3 0.1709 0.0344 4.97 6.8e-07 ***
## UniqueAuthors4 0.2286 0.0483 4.73 2.3e-06 ***
## UniqueAuthors5 0.3570 0.0483 7.38 1.8e-13 ***
## Year1997 -0.0597 0.0796 -0.75 0.454
## Year1998 -0.0017 0.0874 -0.02 0.984
## Year1999 -0.0345 0.0839 -0.41 0.681
```

```

## Year2000          0.0242      0.0849      0.29      0.775
## Year2001         -0.1727      0.0796     -2.17      0.030 *
## Year2002         -0.0314      0.0707     -0.44      0.657
## Year2003         -0.0909      0.0690     -1.32      0.187
## Year2004         -0.0302      0.0704     -0.43      0.668
## Year2005         -0.0124      0.0679     -0.18      0.855
## Year2006          0.0166      0.0681      0.24      0.807
## Year2007         -0.0883      0.0685     -1.29      0.197
## Year2008         -0.0856      0.0674     -1.27      0.204
## Year2009         -0.0995      0.0656     -1.52      0.129
## Year2010         -0.0759      0.0653     -1.16      0.245
## Year2011         -0.0482      0.0661     -0.73      0.466
## Year2012         -0.0703      0.0666     -1.06      0.291
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.71
## Multiple R-squared:  0.0227, Adjusted R-squared:  0.0185
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 433 weights are ~= 1. The remaining 4653 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.194  0.866  0.950  0.919  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.734 1      1.317
## LastAuthorFemale  1.710 1      1.308
## Year              1.044 16      1.001

```

Residuals from first and last author



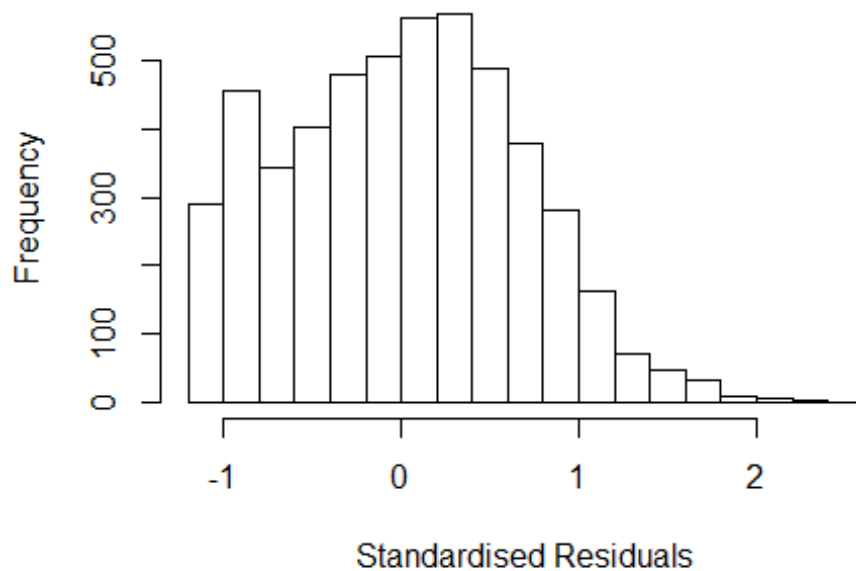
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1465 -0.5146 0.0218 0.4849 2.4420
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99443 0.05788 17.18 <2e-16 ***
## FirstAuthorFemale1 0.07022 0.02676 2.62 0.0087 **
## LastAuthorFemale1 0.03424 0.02730 1.25 0.2098
## Year1997 -0.03966 0.08004 -0.50 0.6203
## Year1998 0.01379 0.08793 0.16 0.8754
## Year1999 -0.01841 0.08429 -0.22 0.8271
## Year2000 0.04761 0.08595 0.55 0.5796
## Year2001 -0.12565 0.08050 -1.56 0.1186
## Year2002 -0.01780 0.07098 -0.25 0.8020
## Year2003 -0.07427 0.06961 -1.07 0.2860
## Year2004 -0.00940 0.07111 -0.13 0.8948
## Year2005 0.02363 0.06812 0.35 0.7287
```

```

## Year2006          0.03464      0.06863      0.50      0.6137
## Year2007          -0.05227      0.06906     -0.76      0.4491
## Year2008          -0.05713      0.06773     -0.84      0.3990
## Year2009          -0.05961      0.06592     -0.90      0.3659
## Year2010          -0.02605      0.06565     -0.40      0.6915
## Year2011          -0.00755      0.06654     -0.11      0.9097
## Year2012          -0.01444      0.06704     -0.22      0.8295
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.72
## Multiple R-squared:  0.0073, Adjusted R-squared:  0.00377
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 407 weights are ~= 1. The remaining 4679 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.226  0.864  0.948  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      1.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.033 1          1.016
## Year              1.033 16          1.001

```

Residuals from first author



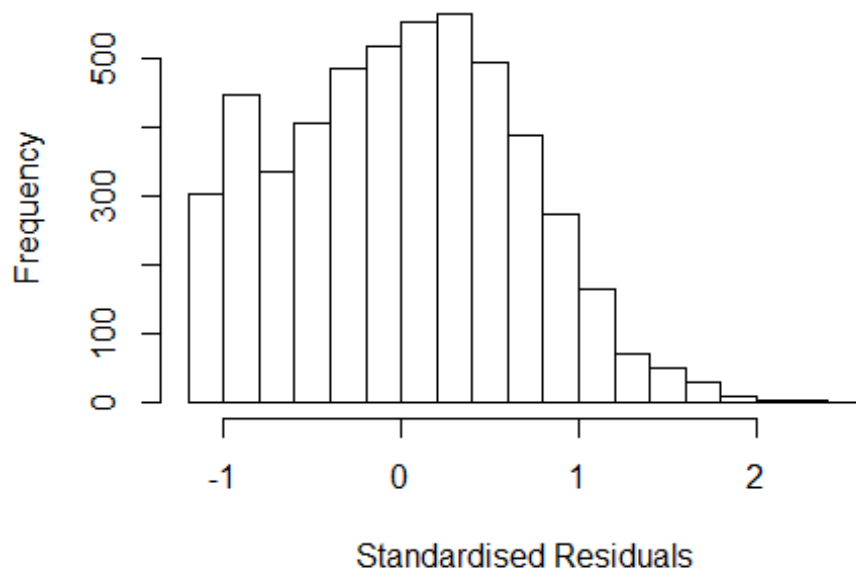
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1361 -0.5140 0.0223 0.4830 2.4392
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99723 0.05785 17.24 < 2e-16 ***
## FirstAuthorFemale1 0.09268 0.02066 4.49 7.4e-06 ***
## Year1997 -0.03816 0.08009 -0.48 0.63
## Year1998 0.01298 0.08782 0.15 0.88
## Year1999 -0.01638 0.08433 -0.19 0.85
## Year2000 0.04615 0.08589 0.54 0.59
## Year2001 -0.12502 0.08049 -1.55 0.12
## Year2002 -0.01800 0.07093 -0.25 0.80
## Year2003 -0.07315 0.06959 -1.05 0.29
## Year2004 -0.00942 0.07110 -0.13 0.89
## Year2005 0.02406 0.06809 0.35 0.72
## Year2006 0.03530 0.06859 0.51 0.61
```

```

## Year2007          -0.05168    0.06910   -0.75    0.45
## Year2008          -0.05718    0.06773   -0.84    0.40
## Year2009          -0.05897    0.06591   -0.89    0.37
## Year2010          -0.02545    0.06565   -0.39    0.70
## Year2011          -0.00695    0.06652   -0.10    0.92
## Year2012          -0.01286    0.06701   -0.19    0.85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.72
## Multiple R-squared:  0.00698,    Adjusted R-squared:  0.00365
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 428 weights are ~= 1. The remaining 4658 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.228  0.864  0.947  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      1.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.02 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
## -1.1292 -0.5096 0.0198 0.4802 2.4329
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99875 0.05774 17.30 < 2e-16 ***
## LastAuthorFemale1 0.08152 0.02106 3.87 0.00011 ***
## Year1997 -0.04031 0.07986 -0.50 0.61378
## Year1998 0.01739 0.08788 0.20 0.84315
## Year1999 -0.01823 0.08413 -0.22 0.82848
## Year2000 0.04895 0.08617 0.57 0.57005
## Year2001 -0.12578 0.08068 -1.56 0.11906
## Year2002 -0.01383 0.07088 -0.20 0.84534
## Year2003 -0.07209 0.06954 -1.04 0.29994
## Year2004 -0.00462 0.07103 -0.07 0.94812
## Year2005 0.02833 0.06800 0.42 0.67695
## Year2006 0.04042 0.06853 0.59 0.55537
```

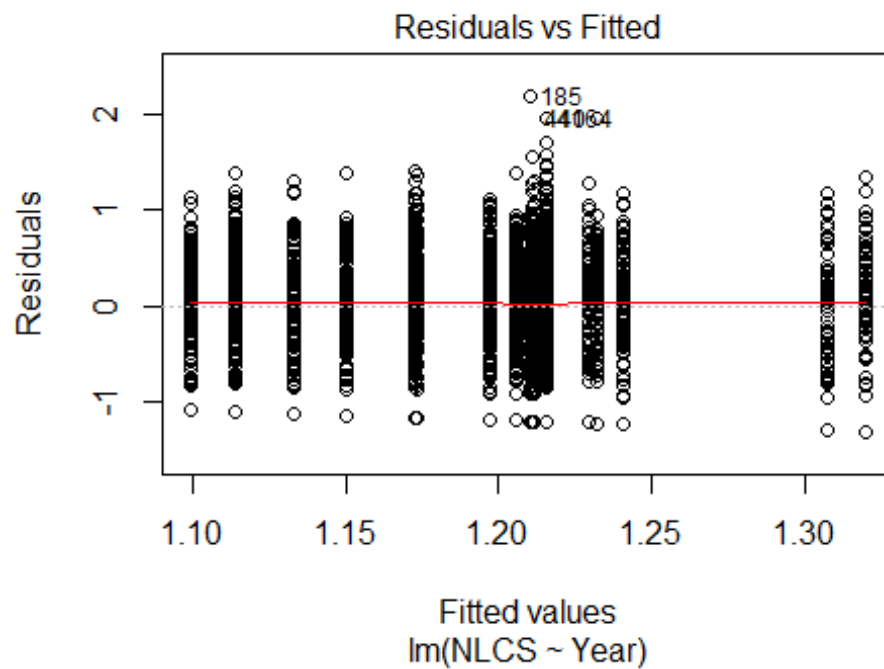
```

## Year2007          -0.04647      0.06886    -0.67  0.49974
## Year2008          -0.05073      0.06750    -0.75  0.45236
## Year2009          -0.05141      0.06570    -0.78  0.43399
## Year2010          -0.02042      0.06551    -0.31  0.75521
## Year2011          -0.00163      0.06634    -0.02  0.98045
## Year2012          -0.00877      0.06688    -0.13  0.89573
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.72
## Multiple R-squared:  0.00601,    Adjusted R-squared:  0.00268
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 405 weights are ~= 1. The remaining 4681 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.230  0.865  0.948  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      1.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5086"
## [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
## 100 117 129 119 111 123 169 146 158 214 234 255 273 343 372
## 2011 2012
## 363 429
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 78 104 82 93 79 74 134 122 138 177 199 213 215 286 331
## 2011 2012

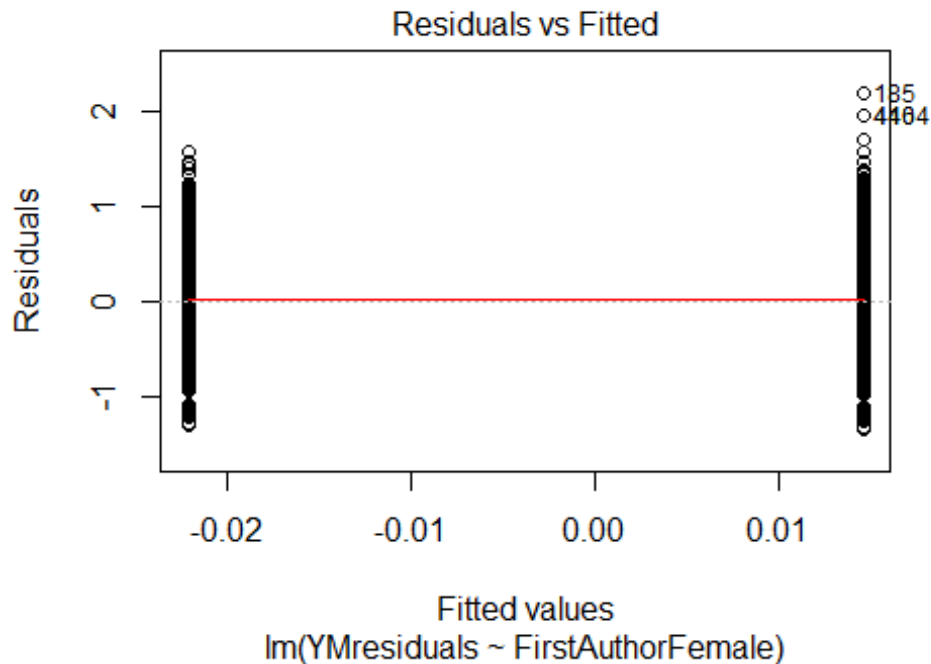
```



```
## 293 370
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 73 99 77 84 69 69 127 106 128 157 182 199 200 267 307
## 2011 2012
## 274 344
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.03
```

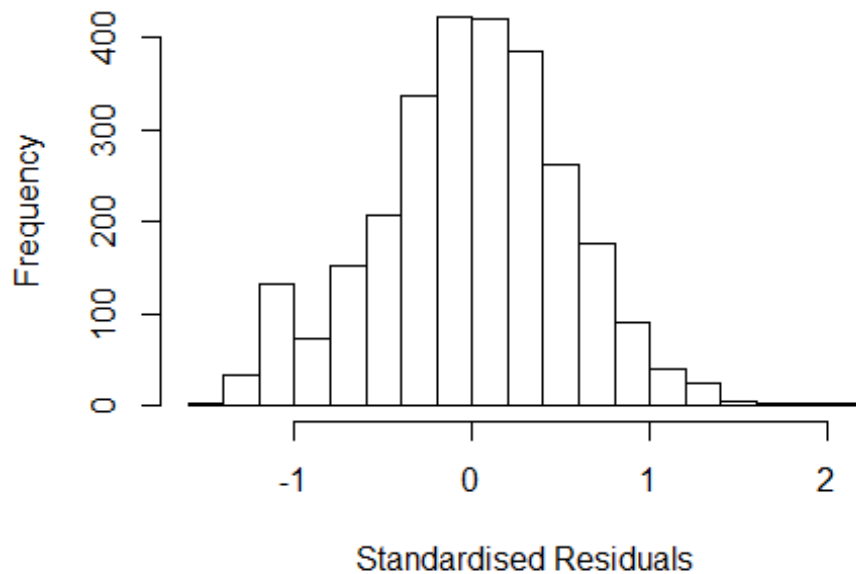


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.6, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 2.9058077157659"
## [1] "Male first author team size 2018 geometric mean: 2.66622444652537"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.80885331411581"
## [1] "Male last author team size 2018 geometric mean: 2.84488285055812"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, 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 1.199  1      1.095
## LastAuthorFemale  1.198  1      1.094
## UniqueAuthors    1.196  4      1.023
## Year             1.238 16      1.007
```

Residuals from first and last author and team size



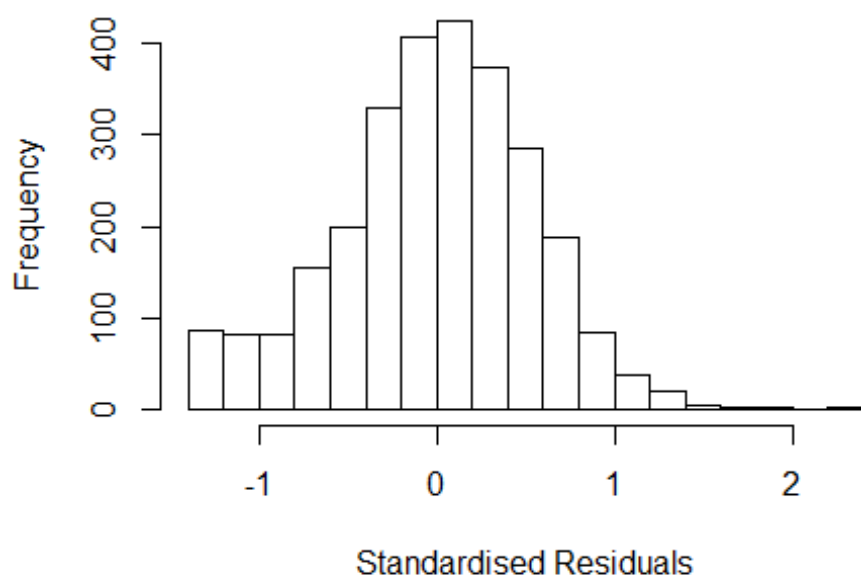
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.511 -0.339 0.013 0.347 2.108
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21995 0.08118 15.03 < 2e-16 ***
## FirstAuthorFemale1 -0.00577 0.02342 -0.25 0.80533
## LastAuthorFemale1 0.04064 0.02273 1.79 0.07387 .
## UniqueAuthors2 0.12328 0.03184 3.87 0.00011 ***
## UniqueAuthors3 0.26497 0.03182 8.33 < 2e-16 ***
## UniqueAuthors4 0.24996 0.03558 7.03 2.7e-12 ***
## UniqueAuthors5 0.27293 0.03075 8.87 < 2e-16 ***
## Year1997 -0.14393 0.09602 -1.50 0.13397
## Year1998 -0.02082 0.10244 -0.20 0.83897
## Year1999 -0.10783 0.09889 -1.09 0.27564
```

```

## Year2000      -0.13816      0.10689      -1.29      0.19625
## Year2001      -0.20567      0.10851      -1.90      0.05815 .
## Year2002      -0.22763      0.09269      -2.46      0.01412 *
## Year2003      -0.13867      0.09387      -1.48      0.13972
## Year2004      -0.18781      0.09368      -2.00      0.04508 *
## Year2005      -0.31518      0.08780      -3.59      0.00034 ***
## Year2006      -0.18670      0.08823      -2.12      0.03442 *
## Year2007      -0.19569      0.08764      -2.23      0.02564 *
## Year2008      -0.18225      0.08602      -2.12      0.03421 *
## Year2009      -0.17678      0.08420      -2.10      0.03585 *
## Year2010      -0.21460      0.08416      -2.55      0.01083 *
## Year2011      -0.28008      0.08609      -3.25      0.00115 **
## Year2012      -0.18373      0.08548      -2.15      0.03169 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.0538, Adjusted R-squared:  0.0462
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 250 weights are ~= 1. The remaining 2512 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0593 0.8620 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          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 1.185 1 1.088
## LastAuthorFemale 1.175 1 1.084
## Year 1.076 16 1.002

```

Residuals from first and last author



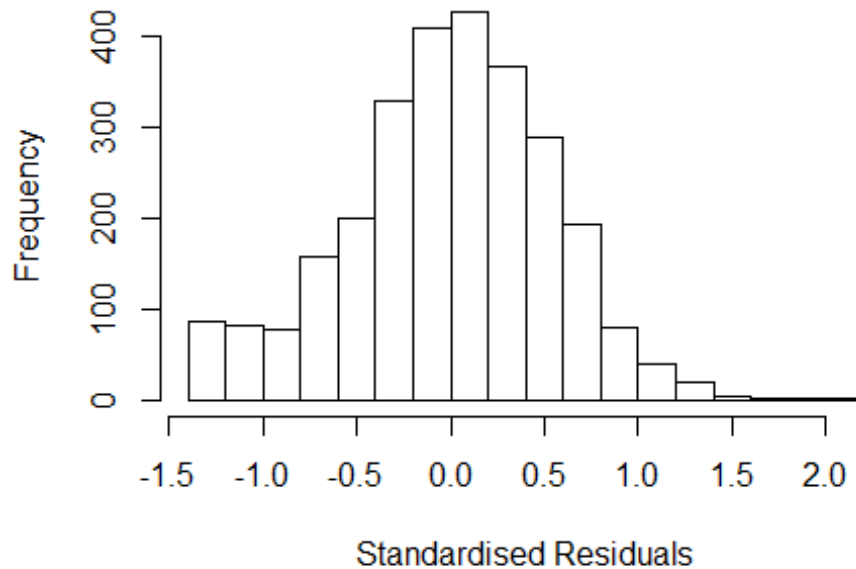
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3699 -0.3447 0.0187 0.3565 2.2028
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3346 0.0728 18.33 <2e-16 ***
## FirstAuthorFemale1 0.0191 0.0239 0.80 0.4242
## LastAuthorFemale1 0.0163 0.0230 0.71 0.4780
## Year1997 -0.1515 0.0909 -1.67 0.0959 .
## Year1998 -0.0162 0.0971 -0.17 0.8675
## Year1999 -0.0830 0.0954 -0.87 0.3841
## Year2000 -0.1106 0.1009 -1.10 0.2731
## Year2001 -0.1708 0.1050 -1.63 0.1038
## Year2002 -0.2104 0.0883 -2.38 0.0173 *
## Year2003 -0.1086 0.0897 -1.21 0.2260
## Year2004 -0.1698 0.0888 -1.91 0.0560 .
## Year2005 -0.2707 0.0832 -3.25 0.0012 **
```

```

## Year2006          -0.1500      0.0832   -1.80   0.0715 .
## Year2007          -0.1436      0.0819   -1.75   0.0796 .
## Year2008          -0.1352      0.0807   -1.68   0.0937 .
## Year2009          -0.1326      0.0787   -1.69   0.0921 .
## Year2010          -0.1730      0.0785   -2.20   0.0277 *
## Year2011          -0.2337      0.0808   -2.89   0.0039 **
## Year2012          -0.1330      0.0800   -1.66   0.0964 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.525
## Multiple R-squared:  0.0112, Adjusted R-squared:  0.00471
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 248 weights are ~= 1. The remaining 2514 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0397 0.8620 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      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 1.042 1      1.021
## Year              1.042 16      1.001

```

Residuals from first author



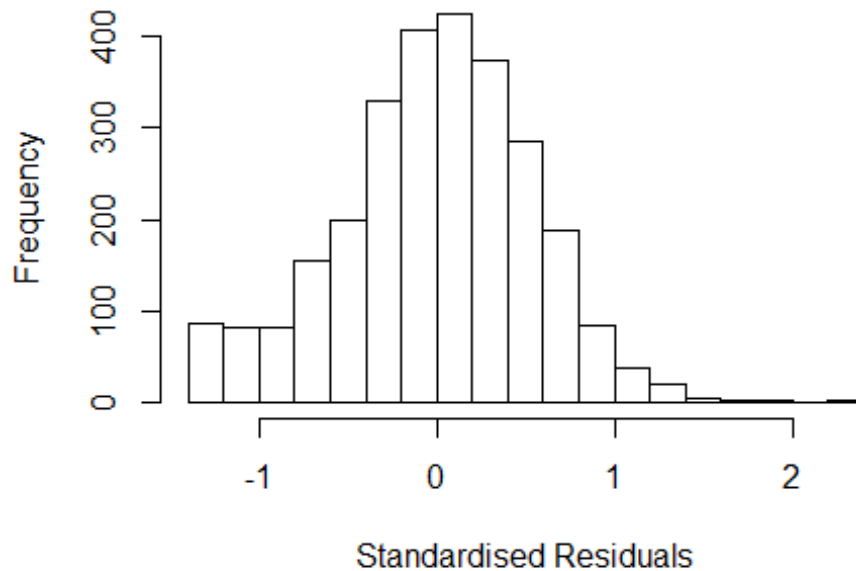
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3629 -0.3437  0.0187  0.3555  2.1910
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3370     0.0727   18.40  <2e-16 ***
## FirstAuthorFemale1  0.0259     0.0224    1.16  0.2481
## Year1997        -0.1490     0.0905   -1.65  0.1000
## Year1998        -0.0126     0.0964   -0.13  0.8959
## Year1999        -0.0815     0.0950   -0.86  0.3909
## Year2000        -0.1094     0.1007   -1.09  0.2771
## Year2001        -0.1694     0.1051   -1.61  0.1070
## Year2002        -0.2093     0.0880   -2.38  0.0174 *
## Year2003        -0.1055     0.0891   -1.18  0.2368
## Year2004        -0.1687     0.0886   -1.90  0.0570 .
## Year2005        -0.2691     0.0829   -3.25  0.0012 **
## Year2006        -0.1477     0.0827   -1.79  0.0744 .
```

```

## Year2007          -0.1420      0.0816   -1.74   0.0819 .
## Year2008          -0.1326      0.0802   -1.65   0.0983 .
## Year2009          -0.1301      0.0781   -1.67   0.0960 .
## Year2010          -0.1709      0.0781   -2.19   0.0287 *
## Year2011          -0.2312      0.0803   -2.88   0.0040 **
## Year2012          -0.1303      0.0795   -1.64   0.1011
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.525
## Multiple R-squared:  0.011, Adjusted R-squared:  0.00489
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 245 weights are ~= 1. The remaining 2517 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0431 0.8630 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      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 1.034 1          1.017
## Year            1.034 16          1.001

```


Residuals from last author



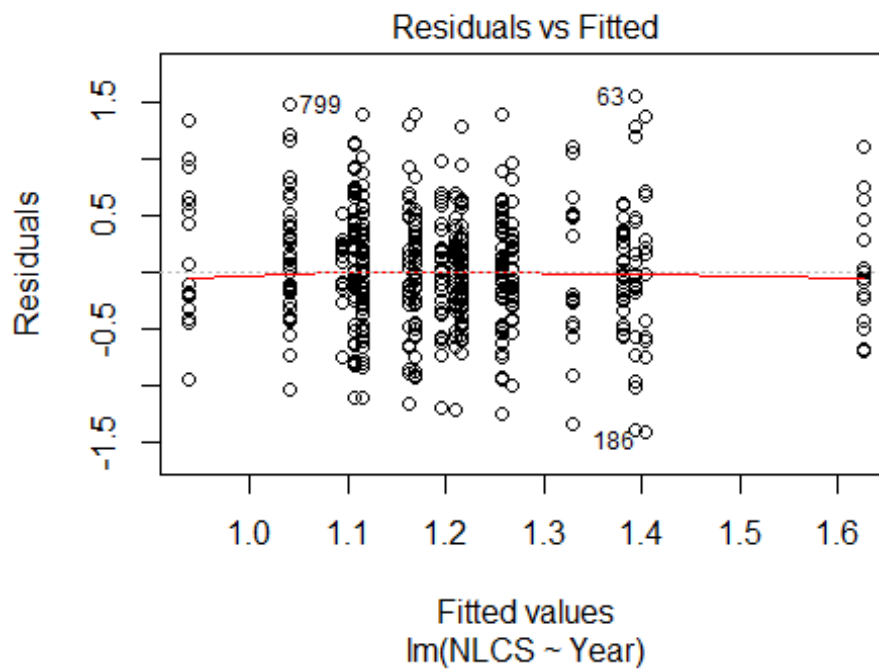
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3655 -0.3444 0.0171 0.3549 2.2147
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3416 0.0722 18.58 <2e-16 ***
## LastAuthorFemale1 0.0239 0.0216 1.11 0.2676
## Year1997 -0.1513 0.0912 -1.66 0.0972 .
## Year1998 -0.0174 0.0973 -0.18 0.8580
## Year1999 -0.0819 0.0957 -0.86 0.3921
## Year2000 -0.1094 0.1008 -1.09 0.2778
## Year2001 -0.1696 0.1052 -1.61 0.1069
## Year2002 -0.2096 0.0885 -2.37 0.0180 *
## Year2003 -0.1107 0.0899 -1.23 0.2180
## Year2004 -0.1688 0.0891 -1.89 0.0582 .
## Year2005 -0.2699 0.0834 -3.24 0.0012 **
## Year2006 -0.1509 0.0834 -1.81 0.0703 .
```

```

## Year2007          -0.1429      0.0821   -1.74   0.0818 .
## Year2008          -0.1361      0.0809   -1.68   0.0924 .
## Year2009          -0.1321      0.0789   -1.67   0.0942 .
## Year2010          -0.1715      0.0788   -2.18   0.0295 *
## Year2011          -0.2324      0.0810   -2.87   0.0042 **
## Year2012          -0.1321      0.0802   -1.65   0.0996 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.525
## Multiple R-squared:  0.0109, Adjusted R-squared:  0.0048
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 248 weights are ~= 1. The remaining 2514 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0361 0.8610 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      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: 2762"
## [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
##   26   29   37   27   40   29   26   23   52   39   54   60   68   73   74
## 2011 2012
##   56   63
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19   18   24   16   29   13   16   16   37   28   36   42   50   57   58
## 2011 2012

```

```
## 45 51
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 18 15 23 15 25 13 15 13 34 25 32 39 44 52 53
## 2011 2012
## 43 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 = 56, df = 16, p-value = 3e-06
```



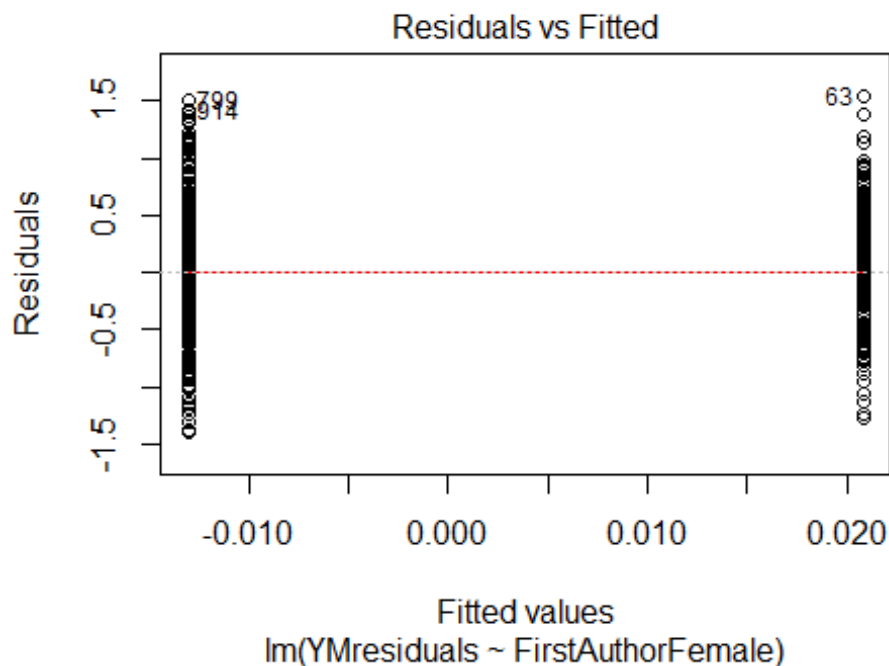
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.8, df = 1, p-value = 0.03

## [1] "Female first author team size 2018 geometric mean: 2.95895197963173"
## [1] "Male first author team size 2018 geometric mean: 3.26801292275476"

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

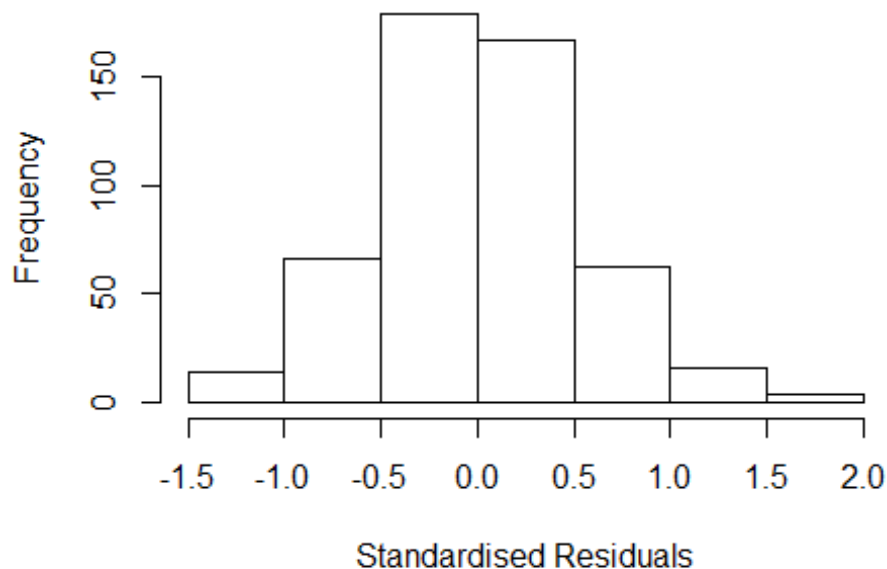
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 280, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.56607497078363"
## [1] "Male last author team size 2018 geometric mean: 3.56752101645312"

## Warning in wilcox.test.default(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] "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  1.261  1      1.123
## UniqueAuthors    2.237  4      1.106
## Year              2.839 16      1.033
```

Residuals from first and last author and team size



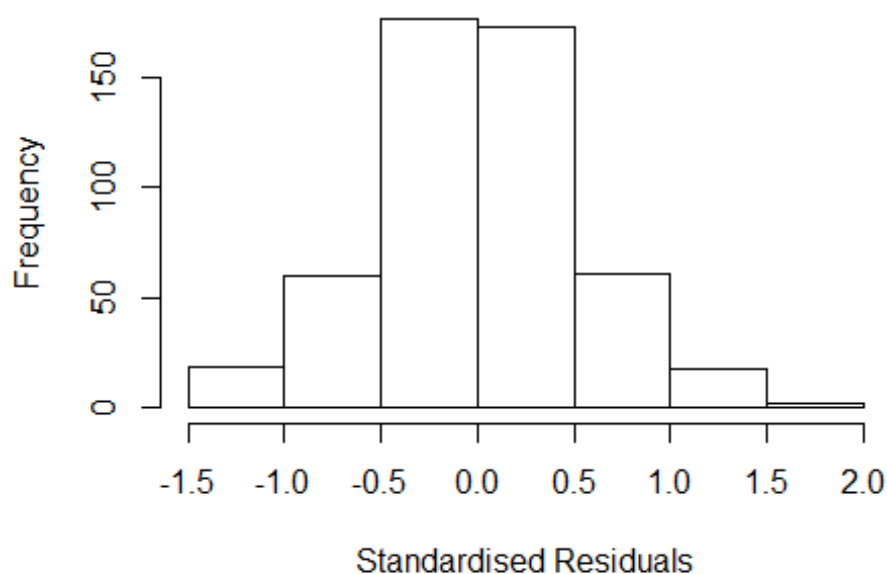
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2708 -0.2889 -0.0133 0.3342 1.7742
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6730 0.1727 3.90 0.00011 ***
## FirstAuthorFemale1 0.0224 0.0473 0.47 0.63679
## LastAuthorFemale1 0.1121 0.0512 2.19 0.02916 *
## UniqueAuthors2 0.1755 0.0753 2.33 0.02008 *
## UniqueAuthors3 0.2295 0.0809 2.84 0.00473 **
## UniqueAuthors4 0.1692 0.0933 1.81 0.07028 .
## UniqueAuthors5 0.2587 0.0805 3.21 0.00140 **
## Year1997 0.5011 0.3345 1.50 0.13478
## Year1998 0.2884 0.2100 1.37 0.17027
## Year1999 0.2766 0.1977 1.40 0.16247
```

```

## Year2000          0.5469      0.1852      2.95  0.00330 **
## Year2001          0.5308      0.2662      1.99  0.04670 *
## Year2002          0.3950      0.2546      1.55  0.12142
## Year2003          0.7156      0.2301      3.11  0.00199 **
## Year2004          0.3666      0.1876      1.95  0.05123 .
## Year2005          0.3275      0.2076      1.58  0.11528
## Year2006          0.3763      0.1878      2.00  0.04568 *
## Year2007          0.2665      0.1902      1.40  0.16170
## Year2008          0.3023      0.1822      1.66  0.09775 .
## Year2009          0.2103      0.1998      1.05  0.29306
## Year2010          0.1855      0.1924      0.96  0.33554
## Year2011          0.0898      0.2051      0.44  0.66166
## Year2012          0.3999      0.1944      2.06  0.04017 *
## ---
## 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:  0.0657
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 462 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.150  0.861  0.953  0.894  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      1.97e-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 1.132 1      1.064
## LastAuthorFemale  1.192 1      1.092
## Year              1.253 16      1.007

```

Residuals from first and last author



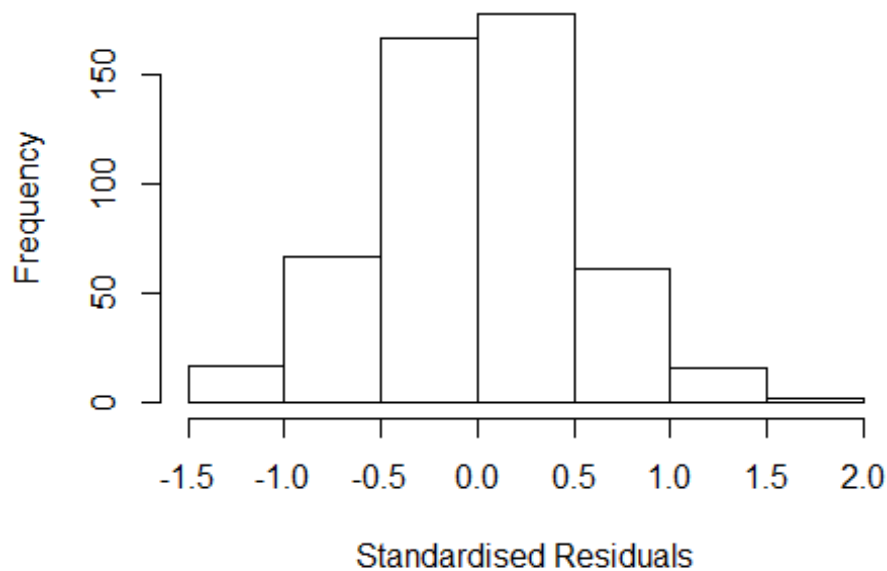
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3651 -0.3164 -0.0024 0.3505 1.5909
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7722 0.2013 3.84 0.00014 ***
## FirstAuthorFemale1 0.0313 0.0478 0.66 0.51258
## LastAuthorFemale1 0.0998 0.0516 1.93 0.05366 .
## Year1997 0.5878 0.3496 1.68 0.09336 .
## Year1998 0.3470 0.2314 1.50 0.13431
## Year1999 0.3404 0.2206 1.54 0.12340
## Year2000 0.5430 0.2088 2.60 0.00960 **
## Year2001 0.5929 0.3020 1.96 0.05015 .
## Year2002 0.4496 0.2762 1.63 0.10427
## Year2003 0.7995 0.2442 3.27 0.00114 **
## Year2004 0.4352 0.2074 2.10 0.03635 *
## Year2005 0.3996 0.2222 1.80 0.07276 .
```

```

## Year2006          0.4128      0.2097      1.97  0.04955 *
## Year2007          0.3163      0.2123      1.49  0.13694
## Year2008          0.3686      0.2039      1.81  0.07128 .
## Year2009          0.2790      0.2196      1.27  0.20446
## Year2010          0.2804      0.2133      1.31  0.18927
## Year2011          0.1739      0.2217      0.78  0.43301
## Year2012          0.4837      0.2132      2.27  0.02374 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.493
## Multiple R-squared:  0.0802, Adjusted R-squared:  0.0464
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 461 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.276  0.862  0.951  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      1.97e-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 1.109 1      1.053
## Year              1.109 16      1.003

```


Residuals from first author



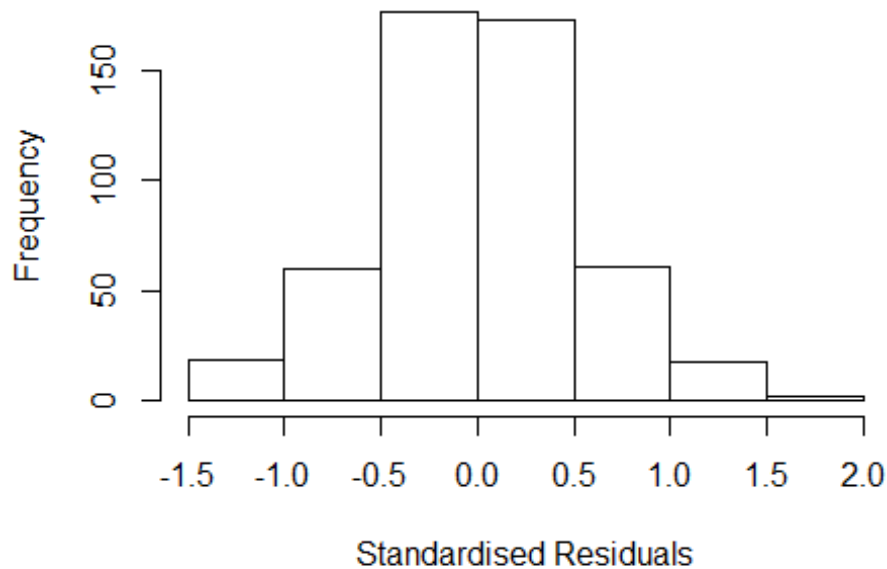
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.40230 -0.30117 0.00305 0.33090 1.56983
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8067 0.2044 3.95 9.1e-05 ***
## FirstAuthorFemale1 0.0537 0.0474 1.13 0.2577
## Year1997 0.5407 0.3547 1.52 0.1280
## Year1998 0.3603 0.2395 1.50 0.1331
## Year1999 0.3012 0.2230 1.35 0.1774
## Year2000 0.5397 0.2152 2.51 0.0125 *
## Year2001 0.5956 0.3069 1.94 0.0528 .
## Year2002 0.4277 0.2747 1.56 0.1200
## Year2003 0.7934 0.2438 3.25 0.0012 **
## Year2004 0.4250 0.2114 2.01 0.0449 *
## Year2005 0.3849 0.2262 1.70 0.0895 .
## Year2006 0.4017 0.2132 1.88 0.0602 .
```

```

## Year2007          0.3038      0.2164      1.40      0.1611
## Year2008          0.3489      0.2079      1.68      0.0939 .
## Year2009          0.2672      0.2257      1.18      0.2370
## Year2010          0.2665      0.2167      1.23      0.2194
## Year2011          0.1605      0.2261      0.71      0.4782
## Year2012          0.4708      0.2166      2.17      0.0302 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.0733, Adjusted R-squared:  0.0411
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 466 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.278  0.862  0.950  0.893  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.97e-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 1.166 1      1.080
## Year      1.166 16      1.005

```

Residuals from last author



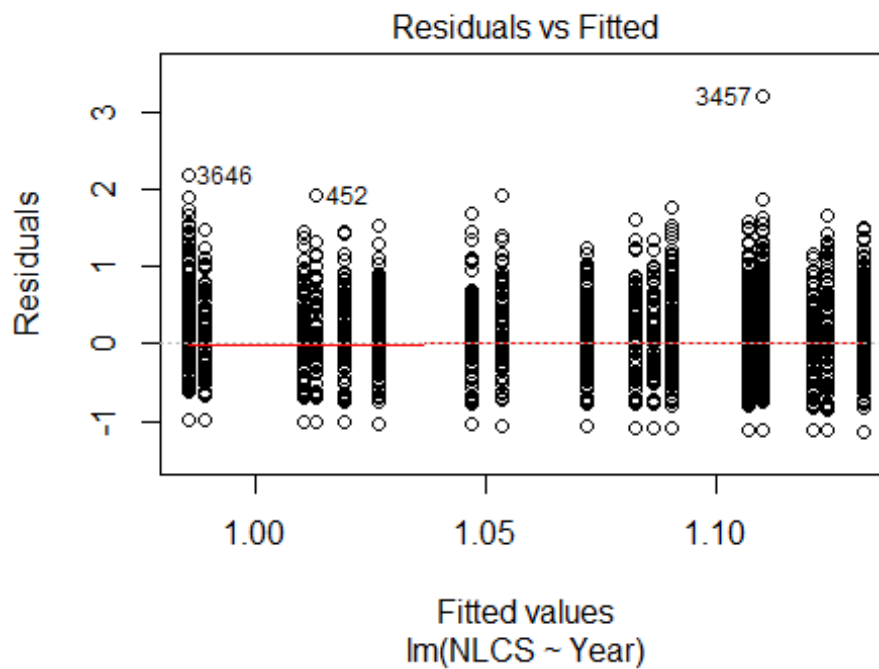
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3712 -0.3185 -0.0126 0.3431 1.5809
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7811 0.2024 3.86 0.00013 ***
## LastAuthorFemale1 0.1077 0.0511 2.11 0.03534 *
## Year1997 0.5902 0.3486 1.69 0.09111 .
## Year1998 0.3486 0.2325 1.50 0.13439
## Year1999 0.3394 0.2219 1.53 0.12687
## Year2000 0.5438 0.2104 2.59 0.01002 *
## Year2001 0.5875 0.3028 1.94 0.05292 .
## Year2002 0.4553 0.2774 1.64 0.10145
## Year2003 0.8021 0.2447 3.28 0.00112 **
## Year2004 0.4336 0.2091 2.07 0.03859 *
## Year2005 0.3973 0.2241 1.77 0.07688 .
## Year2006 0.4156 0.2112 1.97 0.04962 *
```

```

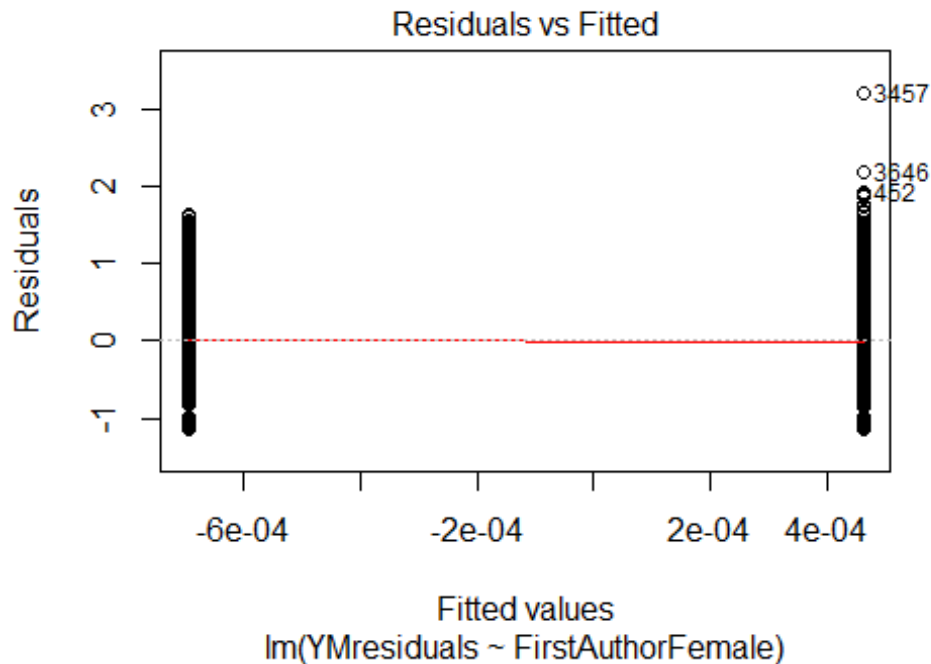
## Year2007          0.3175      0.2142      1.48  0.13891
## Year2008          0.3728      0.2055      1.81  0.07028 .
## Year2009          0.2789      0.2213      1.26  0.20821
## Year2010          0.2845      0.2148      1.32  0.18584
## Year2011          0.1750      0.2237      0.78  0.43434
## Year2012          0.4859      0.2153      2.26  0.02447 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.0801, Adjusted R-squared:  0.0482
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 460 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.270  0.858  0.946  0.892  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      1.97e-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: 508"
## [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
##  108  119  133  114  127  143  130  123  120  157  192  240  264  349  372
## 2011 2012
##  362  370
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   90   95   98   77   83  107   96  104  103  135  163  198  228  301  318
## 2011 2012

```

```
## 313 312
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 83 92 90 75 81 105 88 100 97 133 152 186 217 285 305
## 2011 2012
## 297 303
## [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
```

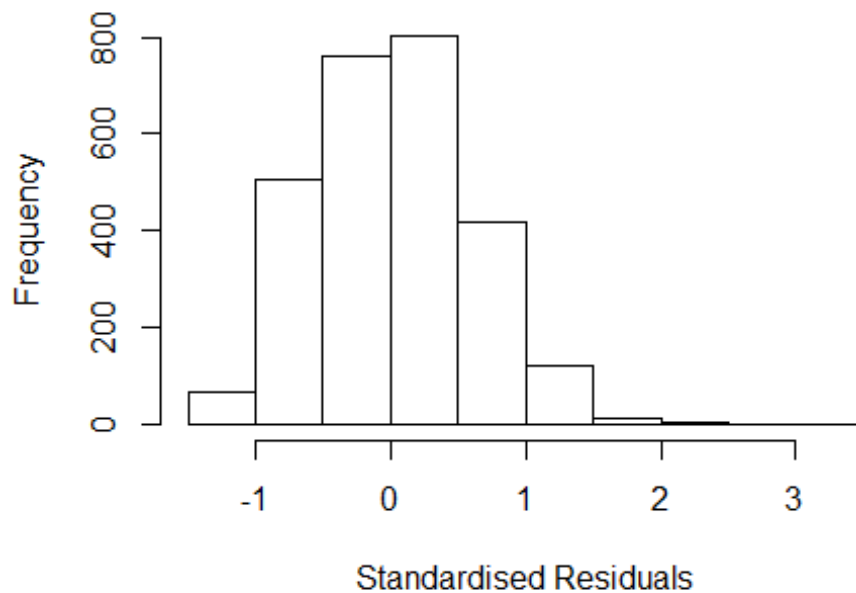


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.97, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 1.82290688805833"
## [1] "Male first author team size 2018 geometric mean: 1.59376976475169"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.89325569491182"
## [1] "Male last author team size 2018 geometric mean: 1.53604801804096"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, p-value = 0.003
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.725 1      1.313
## LastAuthorFemale  1.693 1      1.301
## UniqueAuthors    1.146 4      1.017
## Year              1.161 16     1.005
```

Residuals from first and last author and team size



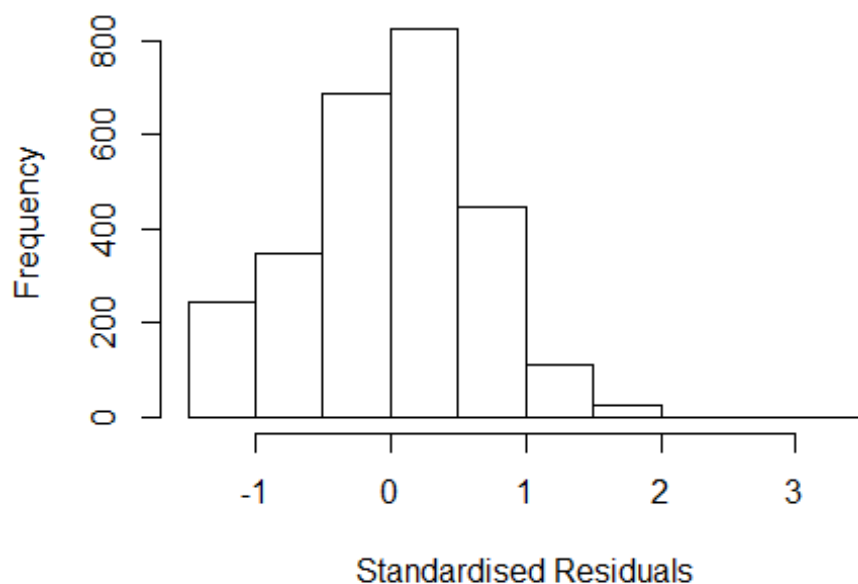
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 3457 79952786437 4.31 2011      1211      3      3.105
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.43731 -0.39792  0.00684  0.42338  3.10538
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.89096    0.07724   11.54 < 2e-16 ***
## FirstAuthorFemale1 -0.02252    0.03229   -0.70    0.49
## LastAuthorFemale1 -0.03186    0.03214   -0.99    0.32
## UniqueAuthors2     0.27824    0.02991    9.30 < 2e-16 ***
## UniqueAuthors3     0.43108    0.04008   10.76 < 2e-16 ***
## UniqueAuthors4     0.43407    0.05608    7.74 1.4e-14 ***
## UniqueAuthors5     0.51391    0.05640    9.11 < 2e-16 ***
## Year1997          0.02066    0.10050    0.21    0.84
## Year1998         -0.02180    0.09555   -0.23    0.82
## Year1999          0.01528    0.10295    0.15    0.88
```

```

## Year2000      0.07739      0.10604      0.73      0.47
## Year2001      0.08435      0.10037      0.84      0.40
## Year2002      0.04350      0.10163      0.43      0.67
## Year2003      0.04635      0.09504      0.49      0.63
## Year2004      0.08687      0.09413      0.92      0.36
## Year2005      0.03587      0.09189      0.39      0.70
## Year2006      0.05212      0.08756      0.60      0.55
## Year2007      0.00628      0.08867      0.07      0.94
## Year2008      0.06430      0.08762      0.73      0.46
## Year2009      0.10708      0.08510      1.26      0.21
## Year2010      0.06985      0.08320      0.84      0.40
## Year2011      0.03542      0.08575      0.41      0.68
## Year2012     -0.04887      0.08407     -0.58      0.56
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.619
## Multiple R-squared:  0.0842, Adjusted R-squared:  0.0766
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 2368 is an outlier with |weight| = 0 ( < 3.7e-05);
## 242 weights are ~= 1. The remaining 2446 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.237  0.858  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           3.72e-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 1.732 1 1.316
## LastAuthorFemale 1.722 1 1.312
## Year 1.061 16 1.002

```


Residuals from first and last author



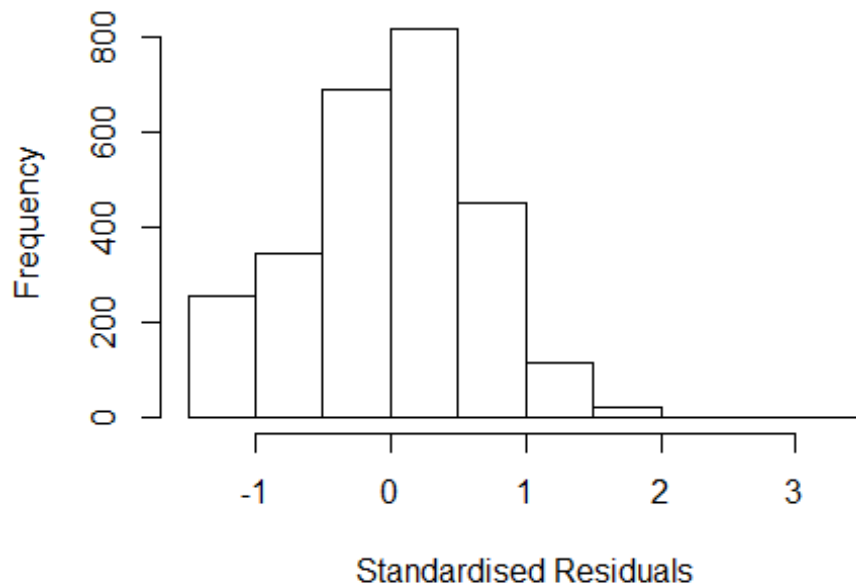
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 3457 79952786437 4.31 2011      1211      3      3.249
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.157 -0.442  0.032  0.430  3.249
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.01251    0.08273   12.24  <2e-16 ***
## FirstAuthorFemale1  0.04152    0.03392    1.22    0.22
## LastAuthorFemale1 -0.03762    0.03406   -1.10    0.27
## Year1997          -0.02329    0.10633   -0.22    0.83
## Year1998          -0.03744    0.10186   -0.37    0.71
## Year1999          -0.02434    0.11169   -0.22    0.83
## Year2000           0.06014    0.11163    0.54    0.59
## Year2001           0.09995    0.10754    0.93    0.35
## Year2002           0.02422    0.10772    0.22    0.82
## Year2003           0.01124    0.10277    0.11    0.91
## Year2004           0.08792    0.10007    0.88    0.38
## Year2005           0.00300    0.09846    0.03    0.98
```

```

## Year2006          0.04426    0.09371    0.47    0.64
## Year2007          0.00843    0.09397    0.09    0.93
## Year2008          0.06428    0.09286    0.69    0.49
## Year2009          0.10299    0.09101    1.13    0.26
## Year2010          0.08730    0.08931    0.98    0.33
## Year2011          0.04857    0.09220    0.53    0.60
## Year2012         -0.04831    0.09091   -0.53    0.60
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.642
## Multiple R-squared:  0.00664,    Adjusted R-squared:  -5.49e-05
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2368 is an outlier with |weight| = 0 ( < 3.7e-05);
## 227 weights are ~= 1. The remaining 2461 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.335  0.868  0.950  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.72e-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 1.037 1          1.018
## Year              1.037 16          1.001

```

Residuals from first author



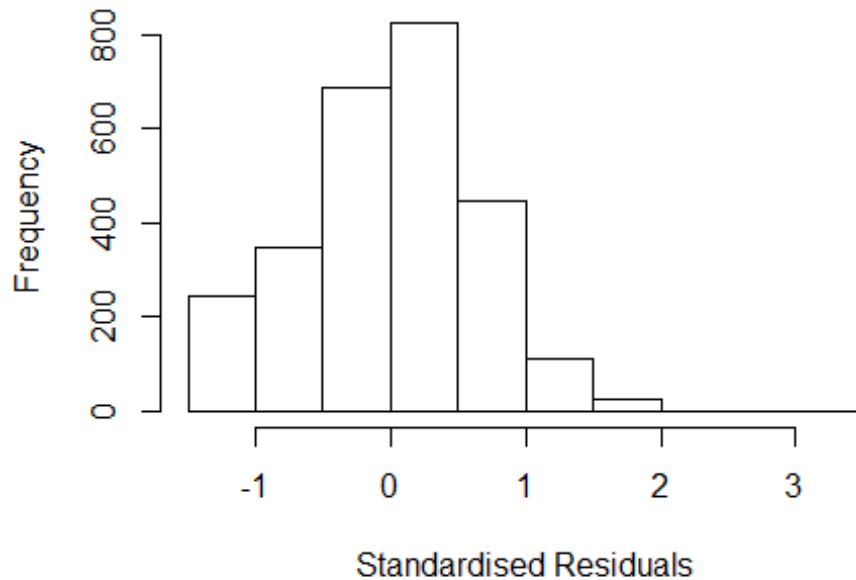
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 3457 79952786437 4.31 2011      1211      3      3.249
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1274 -0.4444  0.0317  0.4297  3.2545
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.00977    0.08283   12.19  <2e-16 ***
## FirstAuthorFemale1 0.01762    0.02629    0.67    0.50
## Year1997      -0.02416    0.10649   -0.23    0.82
## Year1998      -0.03678    0.10191   -0.36    0.72
## Year1999      -0.02550    0.11199   -0.23    0.82
## Year2000       0.05961    0.11168    0.53    0.59
## Year2001       0.09723    0.10770    0.90    0.37
## Year2002       0.02362    0.10822    0.22    0.83
## Year2003       0.01019    0.10288    0.10    0.92
## Year2004       0.08723    0.10025    0.87    0.38
## Year2005       0.00112    0.09868    0.01    0.99
## Year2006       0.04094    0.09383    0.44    0.66
```

```

## Year2007          0.00578    0.09407    0.06    0.95
## Year2008          0.06232    0.09310    0.67    0.50
## Year2009          0.10003    0.09109    1.10    0.27
## Year2010          0.08580    0.08942    0.96    0.34
## Year2011          0.04568    0.09238    0.49    0.62
## Year2012         -0.04945    0.09110   -0.54    0.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.642
## Multiple R-squared:  0.00617,    Adjusted R-squared:  -0.000156
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2368 is an outlier with |weight| = 0 ( < 3.7e-05);
## 216 weights are ~= 1. The remaining 2472 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.334  0.870  0.949  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.72e-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 1.032 1          1.016
## Year            1.032 16          1.001

```

Residuals from last author



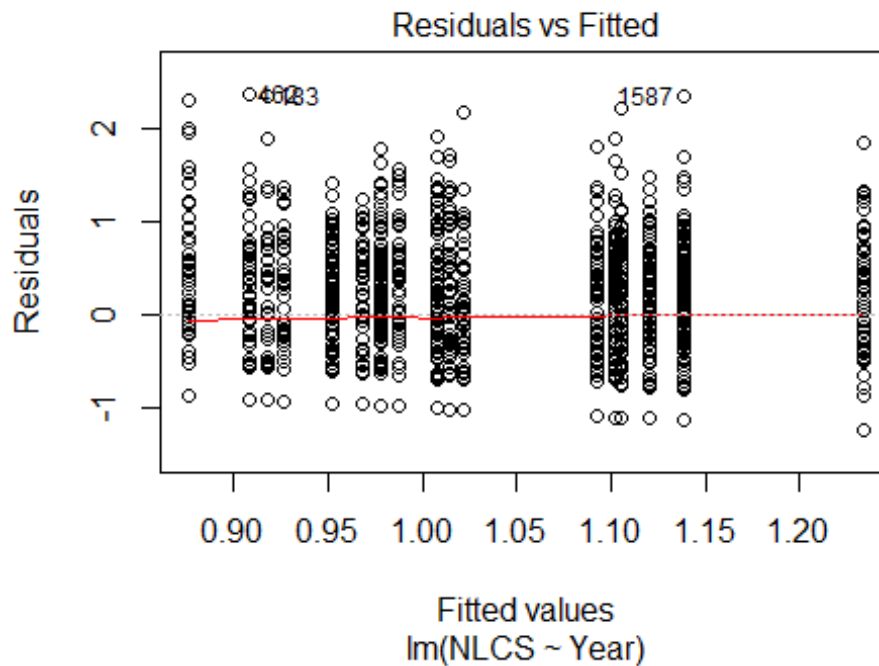
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 3457 79952786437 4.31 2011      1211      3      3.249
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1207 -0.4420  0.0337  0.4309  3.2427
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.01980    0.08270   12.33  <2e-16 ***
## LastAuthorFemale1 -0.01078    0.02633   -0.41    0.68
## Year1997        -0.02412    0.10662   -0.23    0.82
## Year1998        -0.03875    0.10199   -0.38    0.70
## Year1999        -0.02631    0.11208   -0.23    0.81
## Year2000         0.05838    0.11157    0.52    0.60
## Year2001         0.09630    0.10776    0.89    0.37
## Year2002         0.02263    0.10819    0.21    0.83
## Year2003         0.01159    0.10304    0.11    0.91
## Year2004         0.08624    0.10041    0.86    0.39
## Year2005         0.00130    0.09865    0.01    0.99
## Year2006         0.04116    0.09391    0.44    0.66
```

```

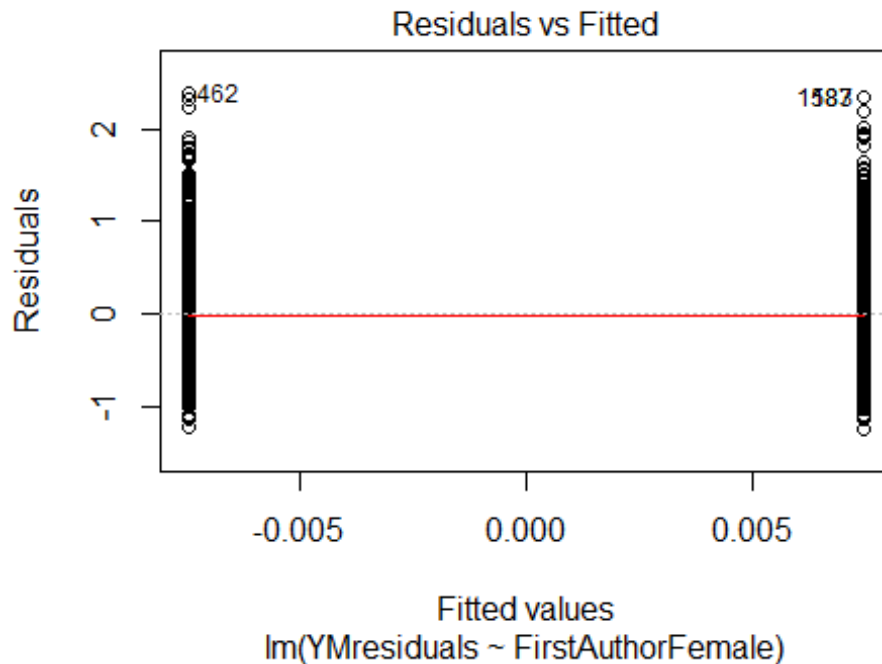
## Year2007      0.00608    0.09415    0.06    0.95
## Year2008      0.06515    0.09313    0.70    0.48
## Year2009      0.10091    0.09120    1.11    0.27
## Year2010      0.08885    0.08954    0.99    0.32
## Year2011      0.04751    0.09250    0.51    0.61
## Year2012     -0.04672    0.09125   -0.51    0.61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.644
## Multiple R-squared:  0.00607,    Adjusted R-squared:  -0.000259
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 2368 is an outlier with |weight| = 0 ( < 3.7e-05);
## 223 weights are ~= 1. The remaining 2465 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.340  0.870  0.950  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.72e-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: 2689"
## [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
##   89   77   81   81   97   98   88   76   80  106  130  136  143  185  163
## 2011 2012
##  162  184
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   77   49   61   61   78   79   70   66   66   86  102  110  117  160  139

```

```
## 2011 2012
## 126 161
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 76 46 57 58 78 76 62 59 63 79 94 98 106 146 133
## 2011 2012
## 120 153
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 19, df = 16, p-value = 0.3
```

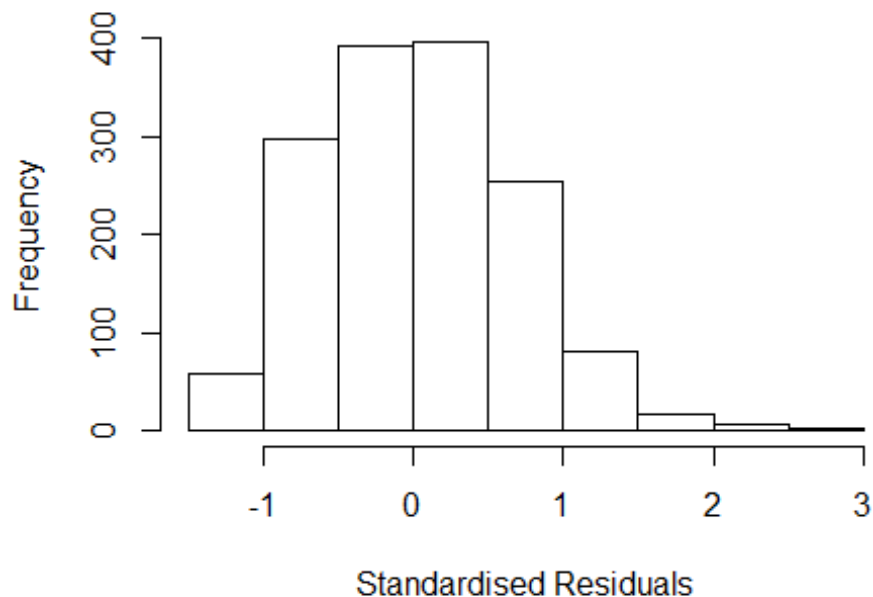


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.093, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 1.64610377208286"
## [1] "Male first author team size 2018 geometric mean: 1.93923159544085"
##
## 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: 1.63322511769283"
## [1] "Male last author team size 2018 geometric mean: 1.94613726327467"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.529 1          1.590
## LastAuthorFemale 2.493 1          1.579
## UniqueAuthors    1.321 4          1.035
## Year             1.367 16          1.010
```


Residuals from first and last author and team size



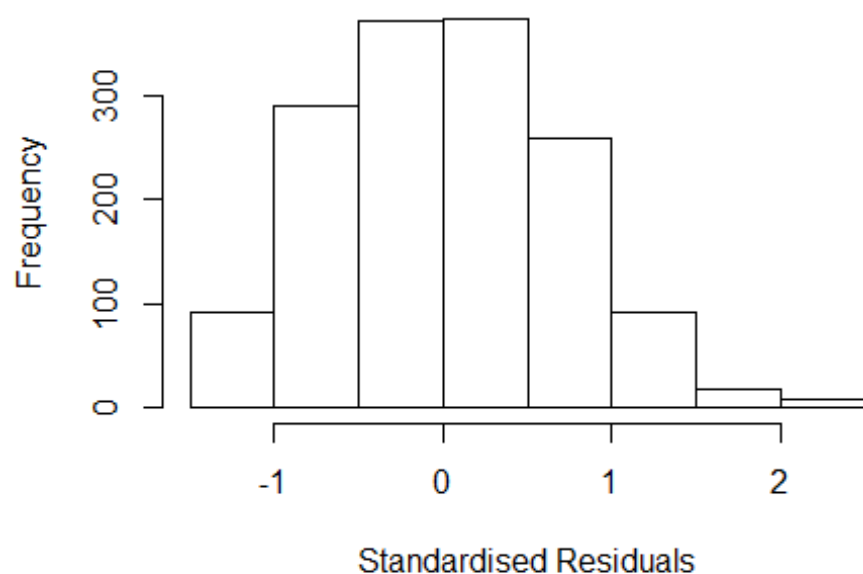
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 462 21844478478 3.281 2000      1710      2      2.532
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.44000 -0.46602  0.00589  0.47080  2.53233
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7002    0.0969   7.23 7.9e-13 ***
## FirstAuthorFemale1  0.0909    0.0560   1.62  0.10476
## LastAuthorFemale1 -0.0547    0.0558  -0.98  0.32721
## UniqueAuthors2    0.3625    0.0436   8.32 < 2e-16 ***
## UniqueAuthors3    0.4600    0.0787   5.84 6.4e-09 ***
## UniqueAuthors4    0.4409    0.0985   4.48 8.2e-06 ***
## UniqueAuthors5    0.2986    0.1257   2.38 0.01766 *
## Year1997          0.0105    0.1379   0.08 0.93930
## Year1998          0.0798    0.1350   0.59 0.55422
## Year1999          0.2069    0.1259   1.64 0.10047
```

```

## Year2000          0.0484      0.1221      0.40  0.69167
## Year2001          0.4289      0.1283      3.34  0.00085 ***
## Year2002          0.1322      0.1528      0.87  0.38708
## Year2003          0.1069      0.1381      0.77  0.43927
## Year2004          0.1184      0.1239      0.96  0.33922
## Year2005          0.1777      0.1192      1.49  0.13627
## Year2006          0.1648      0.1154      1.43  0.15329
## Year2007          0.0384      0.1200      0.32  0.74907
## Year2008          0.1090      0.1116      0.98  0.32919
## Year2009          0.2403      0.1129      2.13  0.03350 *
## Year2010          0.2142      0.1139      1.88  0.06025 .
## Year2011          0.1889      0.1146      1.65  0.09955 .
## Year2012          0.0663      0.1130      0.59  0.55766
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.704
## Multiple R-squared:  0.0918, Adjusted R-squared:  0.0783
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1382 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.168  0.881  0.951  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          6.65e-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 2.138 1          1.462
## LastAuthorFemale 2.129 1          1.459
## Year          1.064 16          1.002

```

Residuals from first and last author



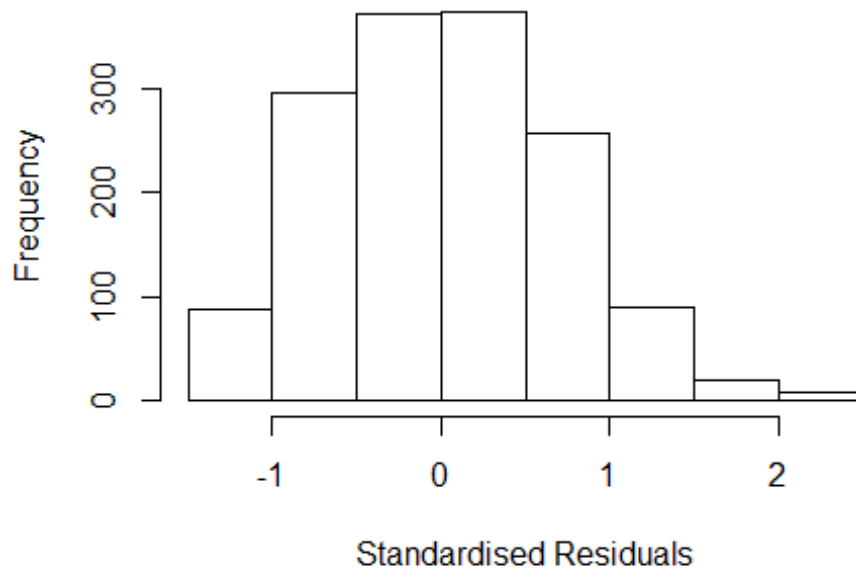
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.25012 -0.50498 -0.00308 0.49981 2.42039
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.79364 0.09782 8.11 1e-15 ***
## FirstAuthorFemale1 0.07821 0.05382 1.45 0.14633
## LastAuthorFemale1 -0.05772 0.05379 -1.07 0.28338
## Year1997 -0.00183 0.14280 -0.01 0.98980
## Year1998 0.07509 0.14275 0.53 0.59895
## Year1999 0.21406 0.12744 1.68 0.09324 .
## Year2000 0.06697 0.12675 0.53 0.59736
## Year2001 0.43599 0.13055 3.34 0.00086 ***
## Year2002 0.15585 0.14844 1.05 0.29390
## Year2003 0.10463 0.14041 0.75 0.45628
## Year2004 0.16159 0.13190 1.23 0.22074
## Year2005 0.23989 0.12465 1.92 0.05448 .
```

```

## Year2006          0.23937    0.11638    2.06  0.03988 *
## Year2007          0.13505    0.12140    1.11  0.26614
## Year2008          0.15202    0.11374    1.34  0.18158
## Year2009          0.29463    0.11428    2.58  0.01003 *
## Year2010          0.31127    0.11484    2.71  0.00679 **
## Year2011          0.29442    0.11819    2.49  0.01284 *
## Year2012          0.16519    0.11327    1.46  0.14496
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.735
## Multiple R-squared:  0.0241, Adjusted R-squared:  0.0123
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1383 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.256  0.871  0.949  0.918  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.65e-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 1.029 1      1.014
## Year              1.029 16      1.001

```

Residuals from first author



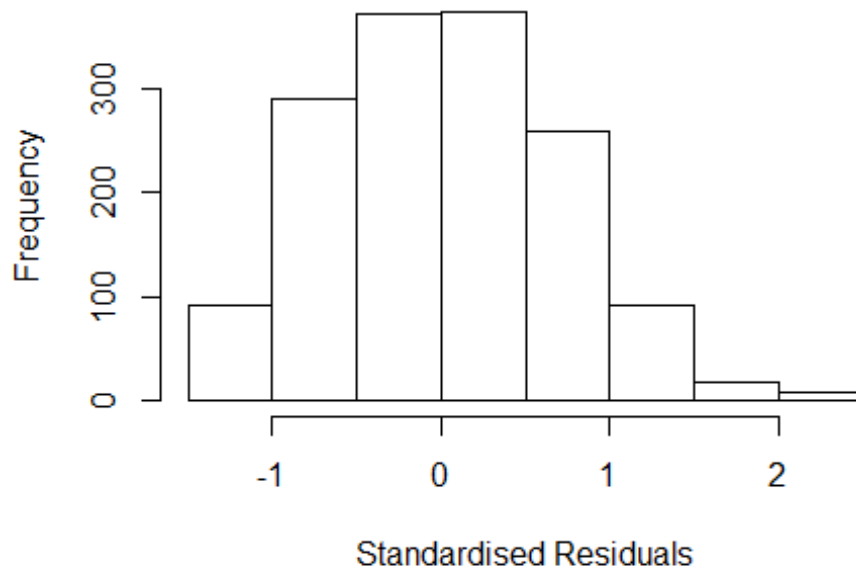
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.25741 -0.50895 -0.00888 0.49630 2.45179
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.78641 0.09720 8.09 1.2e-15 ***
## FirstAuthorFemale1 0.03553 0.03736 0.95 0.34183
## Year1997 -0.00172 0.14204 -0.01 0.99033
## Year1998 0.06912 0.14203 0.49 0.62658
## Year1999 0.21560 0.12667 1.70 0.08894 .
## Year2000 0.06572 0.12624 0.52 0.60275
## Year2001 0.43548 0.13068 3.33 0.00088 ***
## Year2002 0.15220 0.14777 1.03 0.30321
## Year2003 0.10103 0.14055 0.72 0.47238
## Year2004 0.16056 0.13214 1.22 0.22455
## Year2005 0.24257 0.12464 1.95 0.05182 .
## Year2006 0.23785 0.11601 2.05 0.04051 *
```

```

## Year2007          0.13585    0.12121    1.12  0.26258
## Year2008          0.15119    0.11368    1.33  0.18373
## Year2009          0.29502    0.11411    2.59  0.00982 **
## Year2010          0.31056    0.11495    2.70  0.00697 **
## Year2011          0.29159    0.11791    2.47  0.01351 *
## Year2012          0.16655    0.11305    1.47  0.14090
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.734
## Multiple R-squared:  0.0234, Adjusted R-squared:  0.0122
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 115 weights are ~= 1. The remaining 1389 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.242  0.871  0.950  0.918  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.65e-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 1.024 1          1.012
## Year              1.024 16          1.001

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23955 -0.51315 -0.00441 0.49518 2.47038
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.800825 0.098077 8.17 6.8e-16 ***
## LastAuthorFemale1 0.000338 0.037353 0.01 0.99278
## Year1997 0.000799 0.142306 0.01 0.99552
## Year1998 0.070713 0.142582 0.50 0.62000
## Year1999 0.217258 0.127106 1.71 0.08761 .
## Year2000 0.067842 0.126829 0.53 0.59279
## Year2001 0.438389 0.130777 3.35 0.00082 ***
## Year2002 0.158089 0.148008 1.07 0.28564
## Year2003 0.104506 0.141094 0.74 0.45900
## Year2004 0.163986 0.132522 1.24 0.21612
## Year2005 0.245956 0.124593 1.97 0.04856 *
## Year2006 0.244158 0.116340 2.10 0.03601 *
```

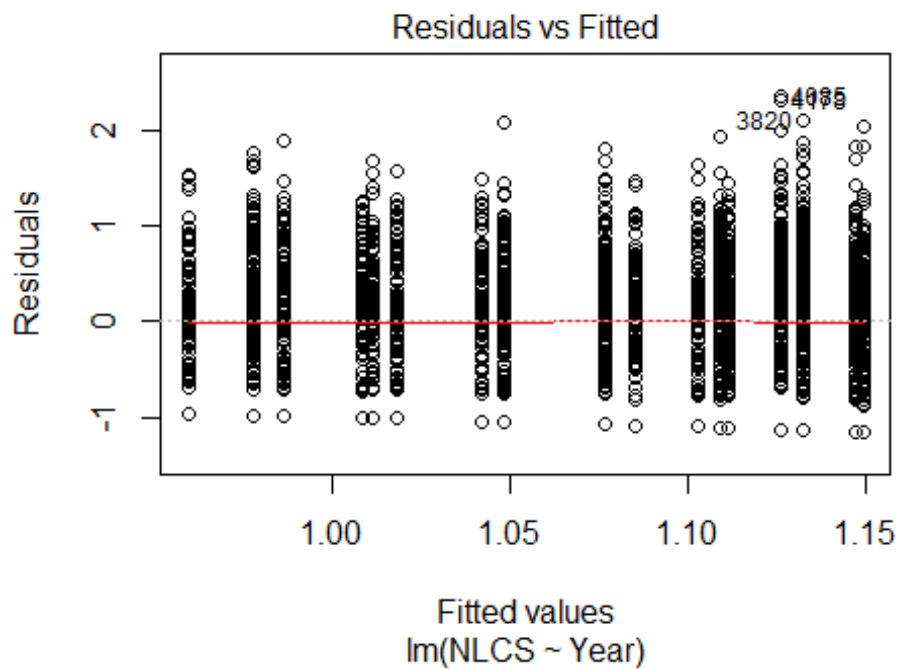
```

## Year2007      0.139328    0.121092    1.15  0.25008
## Year2008      0.154104    0.113922    1.35  0.17635
## Year2009      0.297332    0.114520    2.60  0.00952 **
## Year2010      0.314588    0.115279    2.73  0.00643 **
## Year2011      0.295703    0.118057    2.50  0.01236 *
## Year2012      0.170124    0.113209    1.50  0.13312
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.734
## Multiple R-squared:  0.0228, Adjusted R-squared:  0.0116
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1373 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.234  0.872  0.949  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      6.65e-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: 1504"
## [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
## 138 173 142 147 151 162 193 154 174 218 220 259 282 318 270
## 2011 2012
## 334 313
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 115 139 116 118 116 120 171 130 146 185 191 229 240 273 225
## 2011 2012

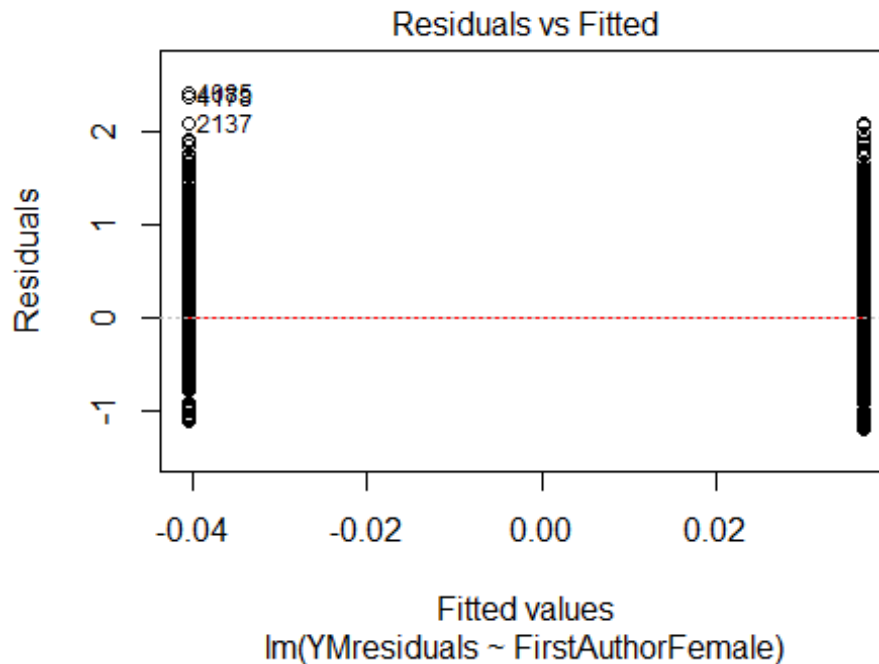
```



```
## 267 255
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 113 134 112 113 108 116 166 124 137 179 183 220 227 255 209
## 2011 2012
## 258 236
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 38, df = 16, p-value = 0.001
```

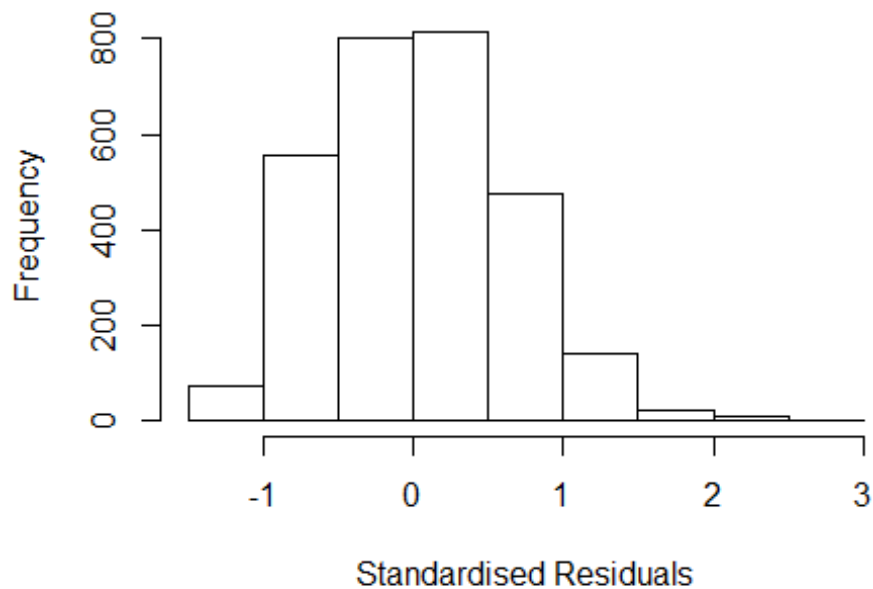


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.5, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 1.645549858944"
## [1] "Male first author team size 2018 geometric mean: 1.67755411821694"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5400, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.69032246795399"
## [1] "Male last author team size 2018 geometric mean: 1.61277405605896"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5600, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.760 1          1.327
## LastAuthorFemale  1.758 1          1.326
## UniqueAuthors    1.173 4          1.020
## Year              1.167 16         1.005
```

Residuals from first and last author and team size



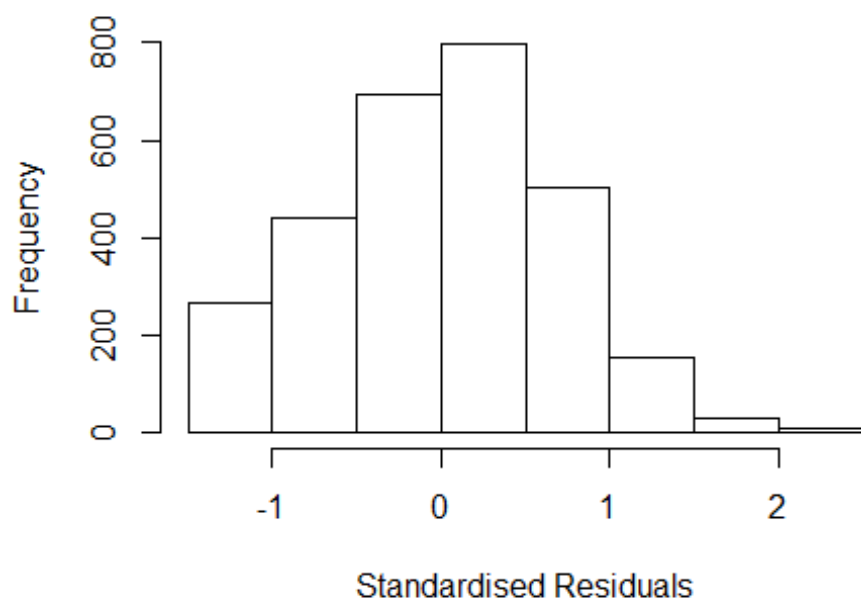
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4178 84856557156 3.439 2012      1203      3      2.583
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.46126 -0.44318  0.00573  0.44220  2.58305
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.87209    0.06228   14.00 < 2e-16 ***
## FirstAuthorFemale1 0.00880    0.03240    0.27  0.79
## LastAuthorFemale1 0.03332    0.03213    1.04  0.30
## UniqueAuthors2    0.43400    0.02905   14.94 < 2e-16 ***
## UniqueAuthors3    0.45401    0.04067   11.16 < 2e-16 ***
## UniqueAuthors4    0.49592    0.05617    8.83 < 2e-16 ***
## UniqueAuthors5    0.39517    0.08092    4.88 1.1e-06 ***
## Year1997          0.04336    0.08087    0.54  0.59
## Year1998         -0.00701    0.08812   -0.08  0.94
## Year1999         -0.09481    0.08384   -1.13  0.26
```

```

## Year2000      -0.00701    0.07987   -0.09    0.93
## Year2001      -0.05809    0.08542   -0.68    0.50
## Year2002      -0.10027    0.08267   -1.21    0.23
## Year2003      -0.04316    0.08064   -0.54    0.59
## Year2004      -0.07692    0.08453   -0.91    0.36
## Year2005       0.06970    0.07732    0.90    0.37
## Year2006       0.05114    0.07629    0.67    0.50
## Year2007      -0.03991    0.07488   -0.53    0.59
## Year2008       0.00581    0.07561    0.08    0.94
## Year2009      -0.05158    0.07206   -0.72    0.47
## Year2010      -0.00508    0.07232   -0.07    0.94
## Year2011       0.00248    0.07590    0.03    0.97
## Year2012      -0.01615    0.08168   -0.20    0.84
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.658
## Multiple R-squared:  0.11, Adjusted R-squared:  0.103
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 261 weights are ~= 1. The remaining 2629 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0891 0.8610 0.9500 0.9130 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.46e-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 1.784 1      1.336
## LastAuthorFemale 1.786 1      1.336
## Year      1.033 16      1.001

```

Residuals from first and last author

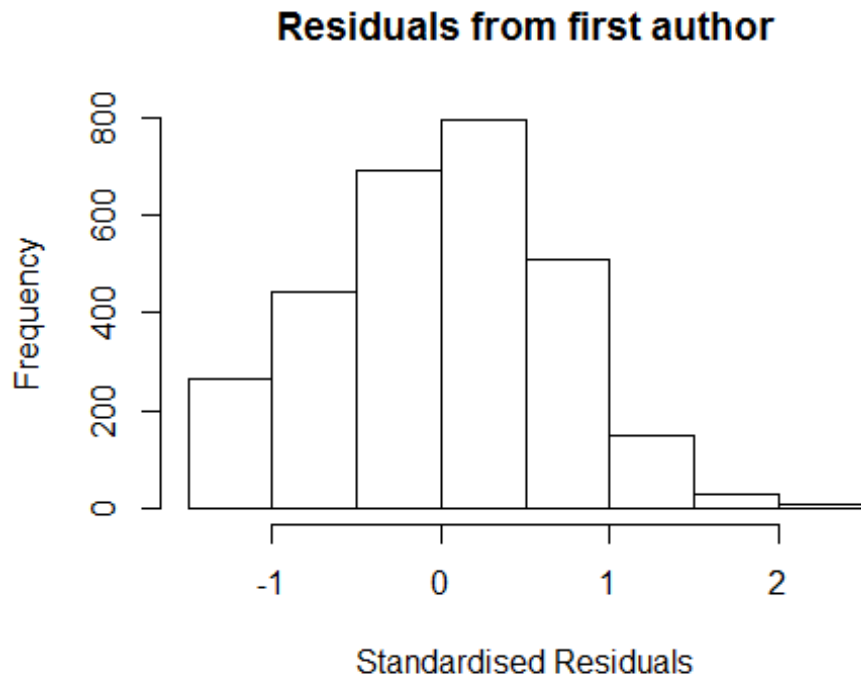


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1849 -0.4799 0.0357 0.4797 2.4573
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96936 0.06512 14.89 <2e-16 ***
## FirstAuthorFemale1 0.08388 0.03485 2.41 0.016 *
## LastAuthorFemale1 0.00632 0.03462 0.18 0.855
## Year1997 0.09567 0.08672 1.10 0.270
## Year1998 -0.00848 0.09235 -0.09 0.927
## Year1999 -0.09227 0.08828 -1.05 0.296
## Year2000 0.04918 0.08641 0.57 0.569
## Year2001 -0.01864 0.08965 -0.21 0.835
## Year2002 -0.07175 0.08786 -0.82 0.414
## Year2003 -0.00708 0.08924 -0.08 0.937
## Year2004 -0.07686 0.08884 -0.87 0.387
## Year2005 0.09745 0.08224 1.18 0.236
```

```

## Year2006          0.12535    0.08061    1.55    0.120
## Year2007          0.02782    0.07817    0.36    0.722
## Year2008          0.08641    0.07895    1.09    0.274
## Year2009          0.01627    0.07500    0.22    0.828
## Year2010          0.09640    0.07574    1.27    0.203
## Year2011          0.08937    0.07997    1.12    0.264
## Year2012          0.05302    0.08635    0.61    0.539
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.703
## Multiple R-squared:  0.0123, Adjusted R-squared:  0.00613
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 219 weights are ~= 1. The remaining 2671 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.859   0.950   0.917   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.46e-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 1.016 1      1.008
## Year              1.016 16      1.000

```

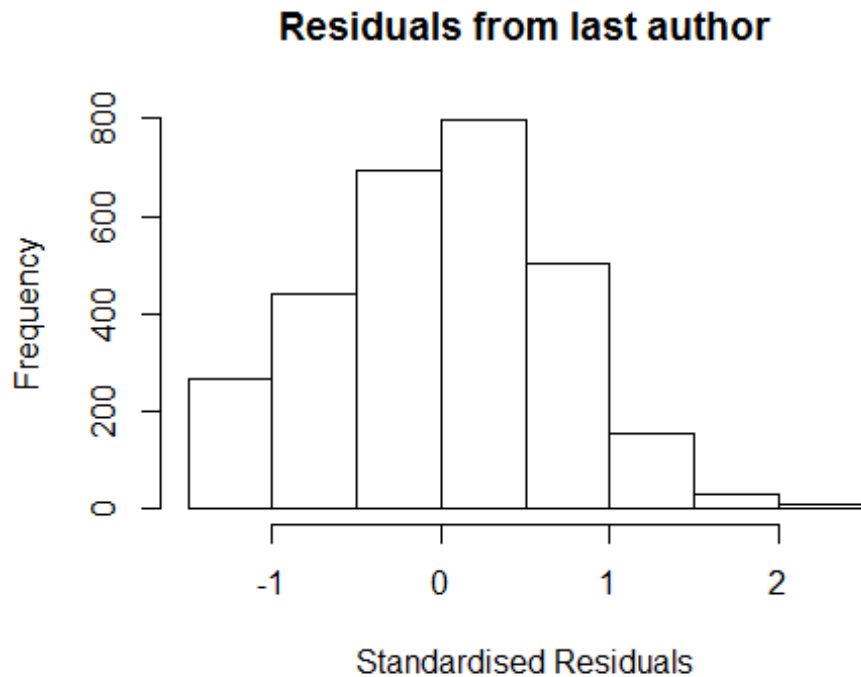


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1838 -0.4797 0.0344 0.4790 2.4628
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97036 0.06467 15.01 < 2e-16 ***
## FirstAuthorFemale1 0.08821 0.02631 3.35 0.00081 ***
## Year1997 0.09572 0.08678 1.10 0.27009
## Year1998 -0.00848 0.09237 -0.09 0.92687
## Year1999 -0.09244 0.08827 -1.05 0.29503
## Year2000 0.04876 0.08632 0.56 0.57216
## Year2001 -0.01860 0.08966 -0.21 0.83568
## Year2002 -0.07217 0.08772 -0.82 0.41075
## Year2003 -0.00706 0.08928 -0.08 0.93702
## Year2004 -0.07719 0.08878 -0.87 0.38466
## Year2005 0.09711 0.08214 1.18 0.23717
## Year2006 0.12523 0.08059 1.55 0.12033
```

```

## Year2007          0.02769    0.07816    0.35  0.72311
## Year2008          0.08639    0.07896    1.09  0.27397
## Year2009          0.01618    0.07502    0.22  0.82925
## Year2010          0.09628    0.07572    1.27  0.20363
## Year2011          0.08908    0.07992    1.11  0.26511
## Year2012          0.05280    0.08634    0.61  0.54087
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.703
## Multiple R-squared:  0.0123, Adjusted R-squared:  0.00647
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 217 weights are ~= 1. The remaining 2673 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.194  0.859  0.950  0.917  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.46e-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 1.016 1          1.008
## Year            1.016 16          1.000

```

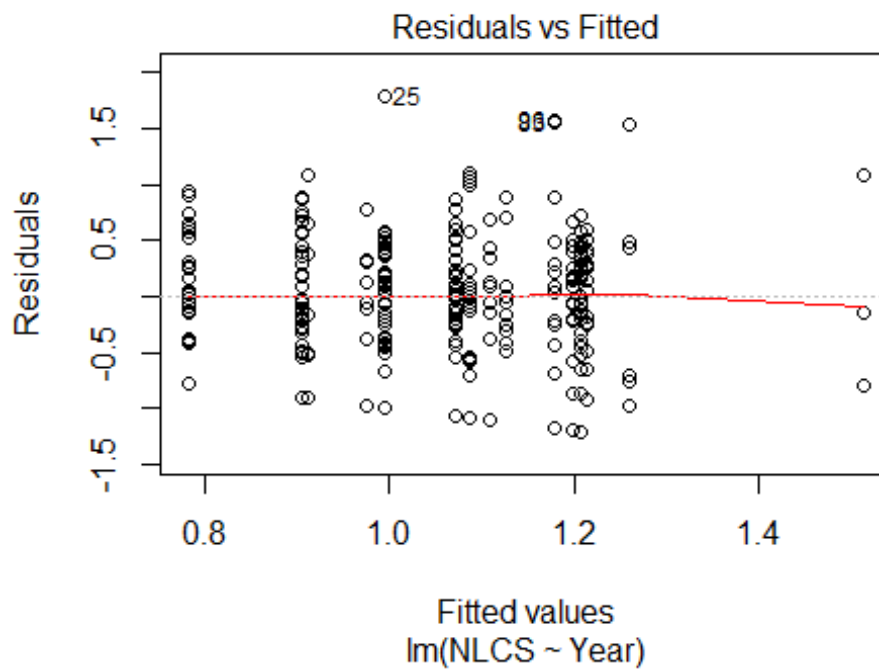
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1724 -0.4840 0.0363 0.4719 2.4034
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97958 0.06496 15.08 <2e-16 ***
## LastAuthorFemale1 0.06326 0.02617 2.42 0.016 *
## Year1997 0.09666 0.08633 1.12 0.263
## Year1998 -0.00255 0.09211 -0.03 0.978
## Year1999 -0.08795 0.08793 -1.00 0.317
## Year2000 0.05350 0.08653 0.62 0.536
## Year2001 -0.01278 0.08954 -0.14 0.887
## Year2002 -0.06186 0.08802 -0.70 0.482
## Year2003 -0.00827 0.08889 -0.09 0.926
## Year2004 -0.06809 0.08857 -0.77 0.442
## Year2005 0.10575 0.08211 1.29 0.198
## Year2006 0.12959 0.08035 1.61 0.107
```

```

## Year2007      0.03576    0.07801    0.46    0.647
## Year2008      0.09128    0.07876    1.16    0.247
## Year2009      0.02361    0.07493    0.32    0.753
## Year2010      0.10423    0.07540    1.38    0.167
## Year2011      0.09709    0.07982    1.22    0.224
## Year2012      0.05605    0.08626    0.65    0.516
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.703
## Multiple R-squared:  0.0104, Adjusted R-squared:  0.00452
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 226 weights are ~= 1. The remaining 2664 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.219  0.859  0.950  0.917  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.46e-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: 2890"
## [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
##   17   5   9  11  12   7  14  19  10  15  24  24  38  49  39
## 2011 2012
##   46  43
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7    4    6    9    8    6   10   15    9   14   19   17   35   39   33
## 2011 2012

```

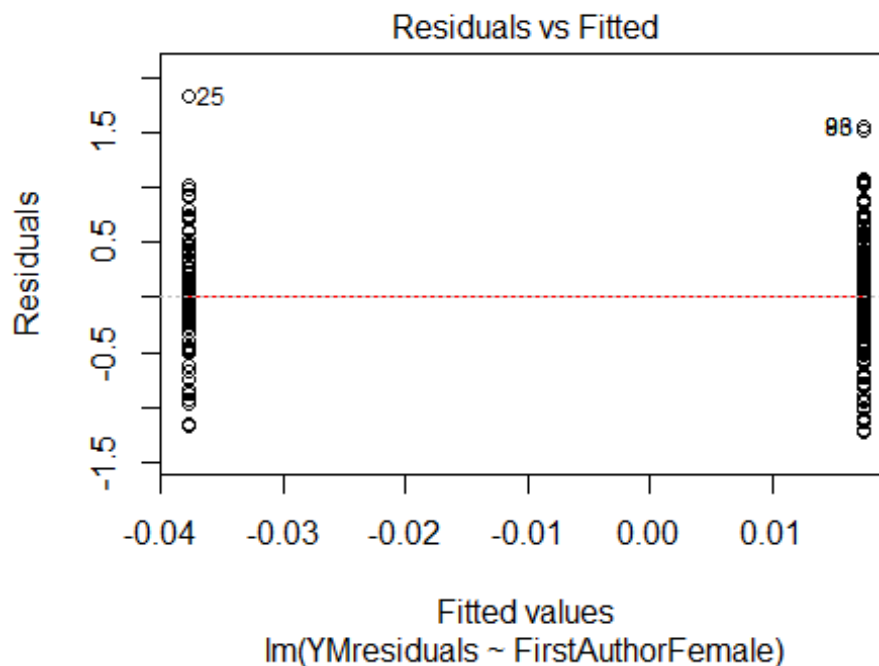
```
## 35 38
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 4 6 9 7 6 8 13 6 14 19 16 27 35 29
## 2011 2012
## 35 35
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 30, df = 16, p-value = 0.02
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0015, df = 1, p-value = 1
## [1] "Female first author team size 2018 geometric mean: 2.76943730371007"
## [1] "Male first author team size 2018 geometric mean: 2.57605388080925"
## Warning in wilcox.test.default(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 = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.30699068678021"
## [1] "Male last author team size 2018 geometric mean: 2.85310752846956"

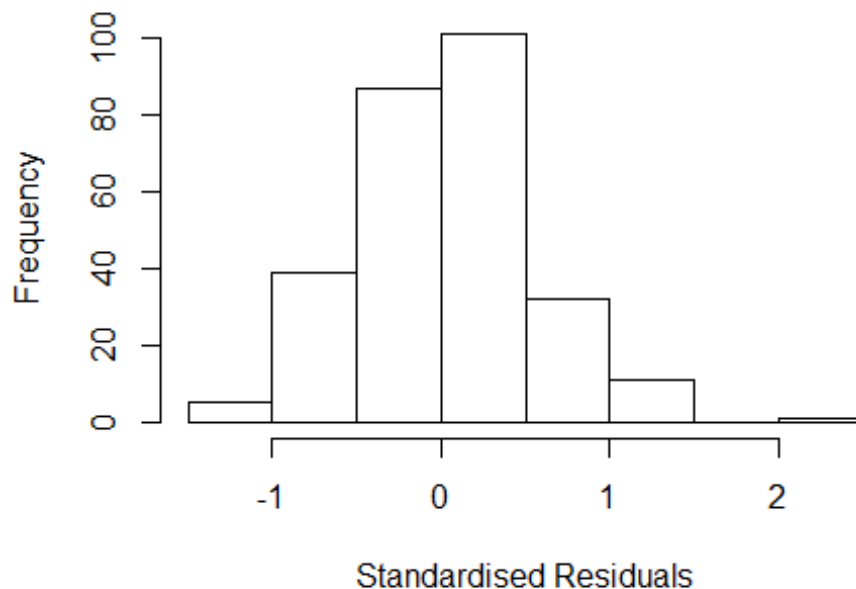
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	2.173	1	1.474
LastAuthorFemale	2.142	1	1.463
UniqueAuthors	3.833	4	1.183
Year	4.395	16	1.047

Residuals from first and last author and team size



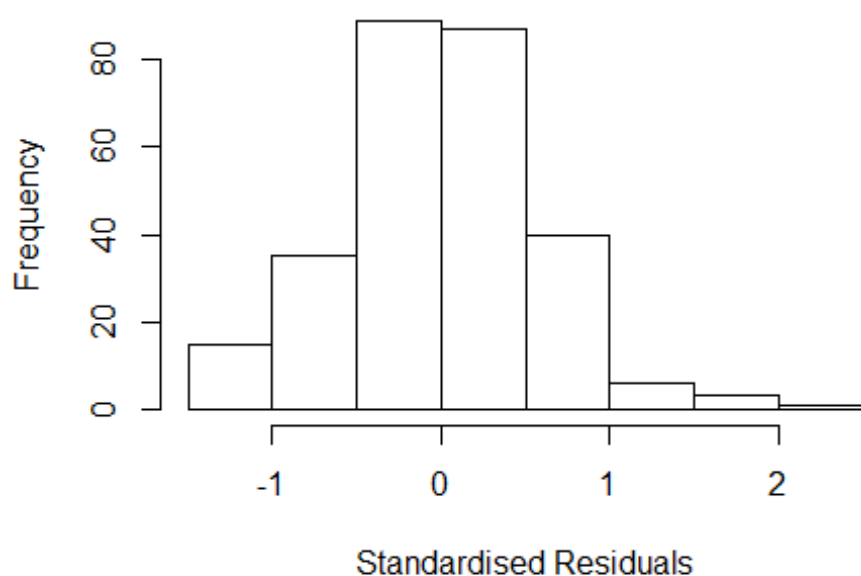
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2537 -0.3418 0.0203 0.3490 2.3628
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.67424 0.27746 2.43 0.0158 *
## FirstAuthorFemale1 -0.14302 0.09407 -1.52 0.1297
## LastAuthorFemale1 0.09240 0.09826 0.94 0.3479
## UniqueAuthors2 0.28036 0.08432 3.32 0.0010 **
## UniqueAuthors3 0.39280 0.12752 3.08 0.0023 **
## UniqueAuthors4 -0.00412 0.11972 -0.03 0.9726
## UniqueAuthors5 0.27057 0.10054 2.69 0.0076 **
## Year1997 0.72389 0.40708 1.78 0.0766 .
## Year1998 -0.20442 0.31011 -0.66 0.5104
## Year1999 0.29192 0.39129 0.75 0.4563
```

```

## Year2000      0.05129      0.35355      0.15      0.8848
## Year2001      0.48900      0.55865      0.88      0.3822
## Year2002      0.44031      0.31052      1.42      0.1574
## Year2003      0.24745      0.43011      0.58      0.5656
## Year2004      0.29254      0.35057      0.83      0.4048
## Year2005      0.29249      0.31366      0.93      0.3520
## Year2006      0.38931      0.29399      1.32      0.1866
## Year2007      0.42418      0.30451      1.39      0.1648
## Year2008      0.30397      0.29163      1.04      0.2983
## Year2009      0.20669      0.28831      0.72      0.4741
## Year2010      0.45719      0.28660      1.60      0.1119
## Year2011      -0.05594      0.28610     -0.20      0.8451
## Year2012      0.15189      0.28833      0.53      0.5988
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.504
## Multiple R-squared:  0.17, Adjusted R-squared:  0.0982
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 12 is an outlier with |weight| <= 4.6e-08 ( < 0.00036);
## 30 weights are ~= 1. The remaining 245 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.397  0.868  0.948  0.901  0.984  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.62e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.128 1 1.459
## LastAuthorFemale 2.142 1 1.464
## Year 1.531 16 1.013

```

Residuals from first and last author



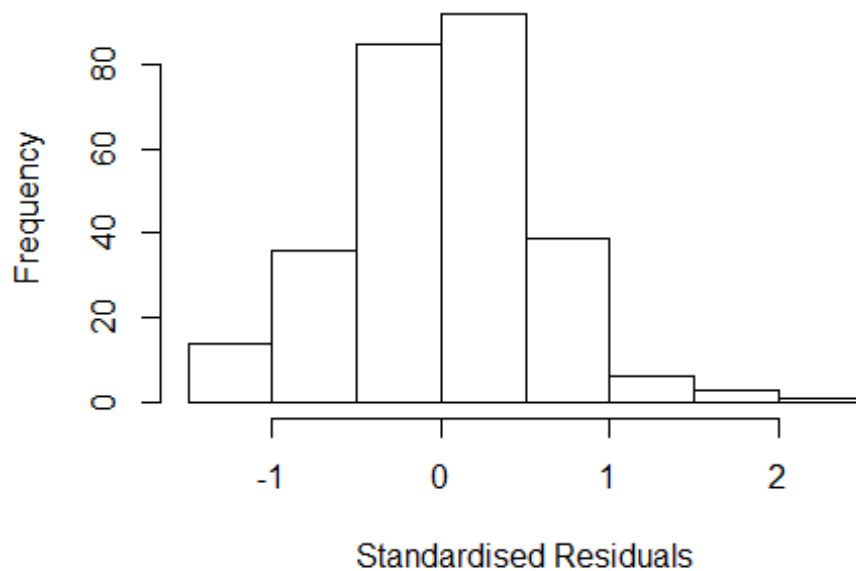
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26455 -0.32985 -0.00227 0.33973 2.14612
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9007 0.2989 3.01 0.0028 **
## FirstAuthorFemale1 -0.1284 0.0976 -1.32 0.1894
## LastAuthorFemale1 0.0840 0.1011 0.83 0.4067
## Year1997 0.5117 0.4816 1.06 0.2889
## Year1998 -0.2204 0.3665 -0.60 0.5482
## Year1999 0.1711 0.4085 0.42 0.6757
## Year2000 0.0147 0.3665 0.04 0.9680
## Year2001 0.1957 0.6049 0.32 0.7466
## Year2002 0.2475 0.3464 0.71 0.4755
## Year2003 0.1744 0.4666 0.37 0.7089
## Year2004 0.1446 0.3543 0.41 0.6835
## Year2005 0.1440 0.3403 0.42 0.6726
```

```

## Year2006          0.3639      0.3142      1.16      0.2479
## Year2007          0.3623      0.3165      1.14      0.2534
## Year2008          0.1933      0.3120      0.62      0.5361
## Year2009          0.1064      0.3062      0.35      0.7285
## Year2010          0.3635      0.3076      1.18      0.2385
## Year2011         -0.0925      0.3077     -0.30      0.7639
## Year2012          0.0541      0.3104      0.17      0.8619
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.51
## Multiple R-squared:  0.0932, Adjusted R-squared:  0.0297
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 253 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0374 0.8600 0.9470 0.8950 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      3.62e-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 1.206 1      1.098
## Year              1.206 16      1.006

```


Residuals from first author



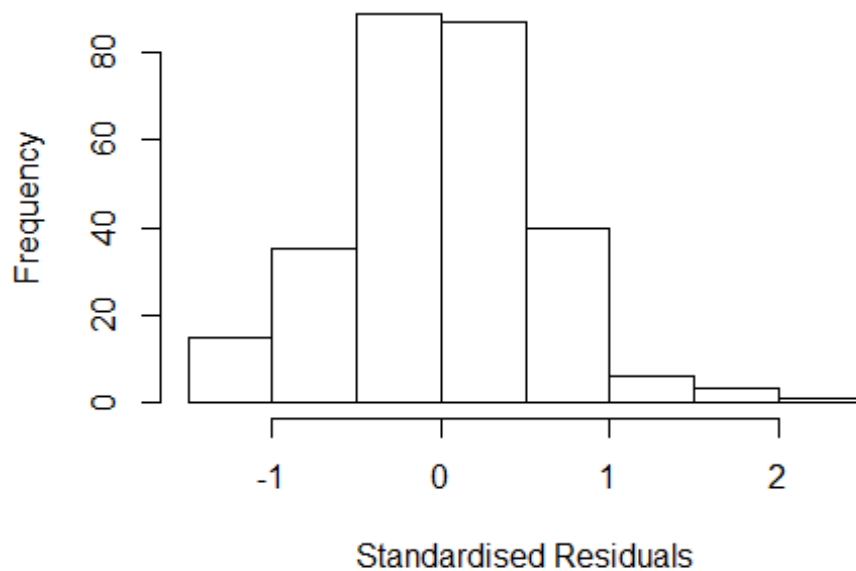
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2731 -0.3133 0.0109 0.3431 2.1717
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9103 0.3117 2.92 0.0038 **
## FirstAuthorFemale1 -0.0786 0.0730 -1.08 0.2828
## Year1997 0.5005 0.4894 1.02 0.3074
## Year1998 -0.2214 0.3802 -0.58 0.5608
## Year1999 0.1611 0.4202 0.38 0.7017
## Year2000 0.0360 0.3679 0.10 0.9221
## Year2001 0.1569 0.6091 0.26 0.7969
## Year2002 0.2385 0.3543 0.67 0.5014
## Year2003 0.1794 0.4803 0.37 0.7091
## Year2004 0.1604 0.3620 0.44 0.6581
## Year2005 0.1551 0.3501 0.44 0.6580
## Year2006 0.3628 0.3261 1.11 0.2669
```

```

## Year2007          0.3620      0.3276      1.11      0.2701
## Year2008          0.1917      0.3237      0.59      0.5543
## Year2009          0.1031      0.3178      0.32      0.7458
## Year2010          0.3583      0.3206      1.12      0.2647
## Year2011         -0.0987      0.3203     -0.31      0.7583
## Year2012          0.0544      0.3222      0.17      0.8661
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.0906, Adjusted R-squared:  0.0307
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 252 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0273 0.8590 0.9450 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      3.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.186 1      1.089
## Year              1.186 16      1.005

```

Residuals from last author



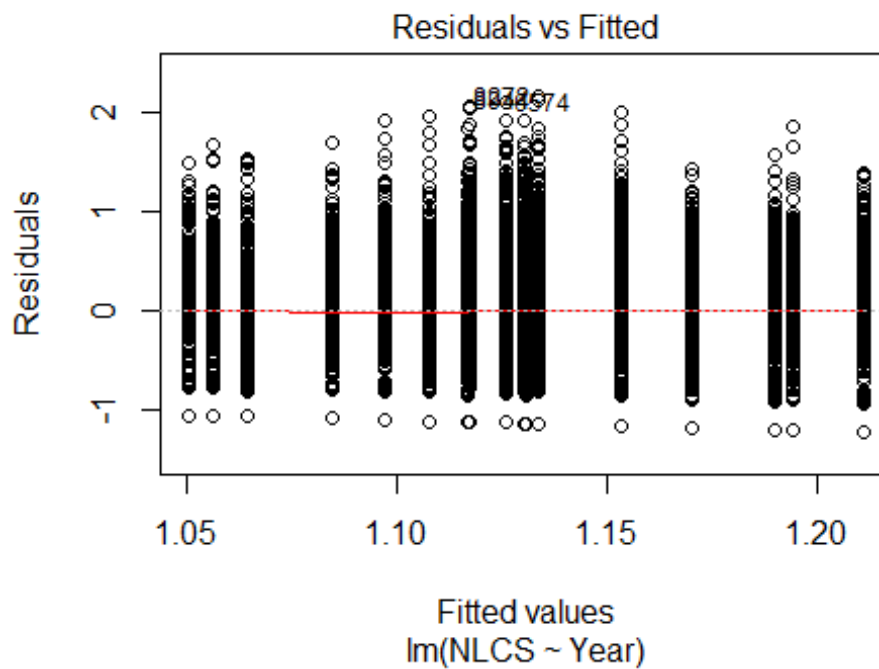
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24007 -0.31602 0.00225 0.33825 2.10755
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.87084 0.32753 2.66 0.0083 **
## LastAuthorFemale1 0.00529 0.07607 0.07 0.9446
## Year1997 0.54435 0.49979 1.09 0.2771
## Year1998 -0.20168 0.38646 -0.52 0.6022
## Year1999 0.20077 0.42565 0.47 0.6376
## Year2000 0.06946 0.38291 0.18 0.8562
## Year2001 0.16417 0.65658 0.25 0.8028
## Year2002 0.26305 0.36829 0.71 0.4757
## Year2003 0.19537 0.49169 0.40 0.6914
## Year2004 0.17988 0.37568 0.48 0.6325
## Year2005 0.18582 0.36429 0.51 0.6104
## Year2006 0.36478 0.34300 1.06 0.2886
```

```

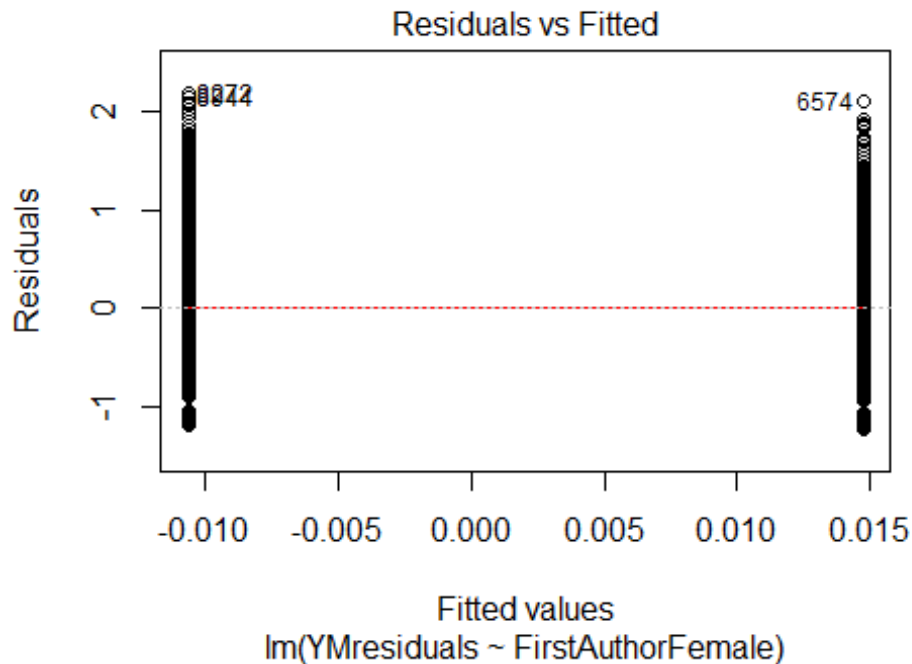
## Year2007      0.37940      0.34513      1.10      0.2727
## Year2008      0.20235      0.34144      0.59      0.5539
## Year2009      0.10740      0.33624      0.32      0.7497
## Year2010      0.36393      0.33900      1.07      0.2840
## Year2011     -0.08950      0.33766     -0.27      0.7912
## Year2012      0.05556      0.33963      0.16      0.8702
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.0854, Adjusted R-squared:  0.0251
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 245 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0563 0.8560 0.9450 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      3.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 276"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
## 325 324 324 358 357 367 411 359 366 410 475 547 689 801 843
## 2011 2012
## 932 832
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 283 284 265 281 289 266 359 314 310 362 415 477 616 701 723
## 2011 2012

```

```
## 809 718
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 275 272 253 270 274 250 335 305 291 345 397 459 582 667 684
## 2011 2012
## 782 685
## [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
```

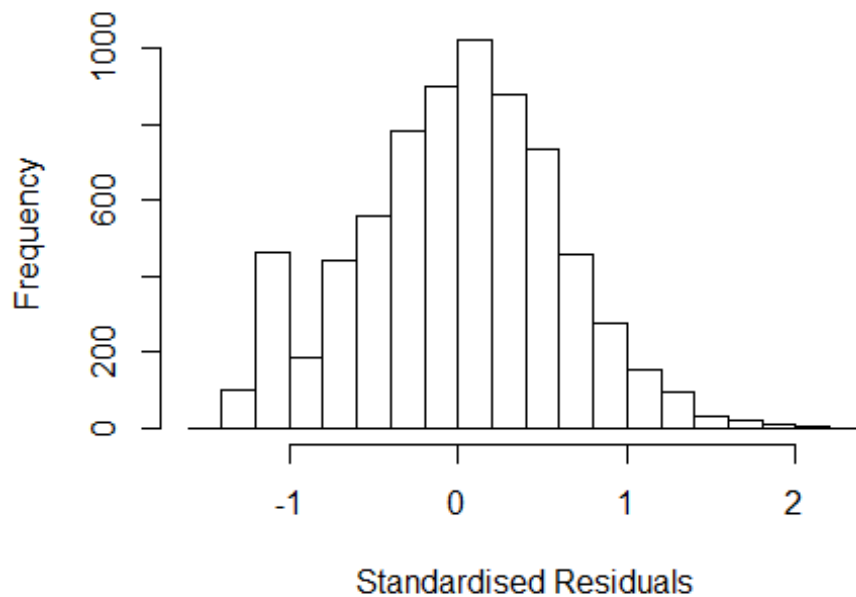


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 12, df = 1, p-value = 6e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 1.668315997371"
## [1] "Male first author team size 2018 geometric mean: 1.44887983167246"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 81000, p-value = 6e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.62910911430023"
## [1] "Male last author team size 2018 geometric mean: 1.4901736384165"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 78000, 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 1.860 1          1.364
## LastAuthorFemale  1.835 1          1.355
## UniqueAuthors    1.112 4          1.013
## Year              1.100 16         1.003
```

Residuals from first and last author and team size



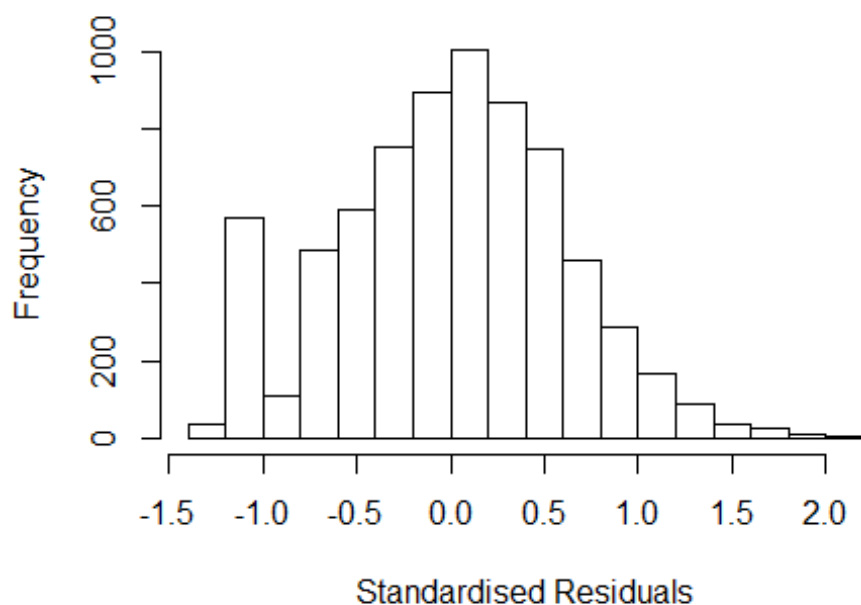
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4052 -0.3925 0.0251 0.4014 2.2233
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01394 0.03951 25.66 < 2e-16 ***
## FirstAuthorFemale1 0.02603 0.01986 1.31 0.190
## LastAuthorFemale1 -0.01544 0.01977 -0.78 0.435
## UniqueAuthors2 0.18063 0.01814 9.96 < 2e-16 ***
## UniqueAuthors3 0.23945 0.02445 9.79 < 2e-16 ***
## UniqueAuthors4 0.24460 0.03796 6.44 1.2e-10 ***
## UniqueAuthors5 0.24050 0.03989 6.03 1.7e-09 ***
## Year1997 -0.01543 0.05451 -0.28 0.777
## Year1998 0.00880 0.05727 0.15 0.878
## Year1999 -0.01514 0.05232 -0.29 0.772
```

```

## Year2000      0.01604      0.05204      0.31      0.758
## Year2001     -0.04569      0.05443     -0.84      0.401
## Year2002      0.12067      0.05069      2.38      0.017 *
## Year2003      0.04454      0.05158      0.86      0.388
## Year2004      0.10851      0.04998      2.17      0.030 *
## Year2005      0.00760      0.05056      0.15      0.881
## Year2006      0.07979      0.04797      1.66      0.096 .
## Year2007      0.08873      0.04672      1.90      0.058 .
## Year2008      0.05276      0.04730      1.12      0.265
## Year2009      0.00542      0.04528      0.12      0.905
## Year2010      0.03048      0.04578      0.67      0.506
## Year2011      0.02472      0.04575      0.54      0.589
## Year2012     -0.00441      0.04720     -0.09      0.926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0336, Adjusted R-squared:  0.0306
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 565 weights are ~= 1. The remaining 6561 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.120  0.870  0.951  0.905  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.40e-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 1.850 1      1.360
## LastAuthorFemale  1.851 1      1.360
## Year              1.039 16      1.001

```


Residuals from first and last author



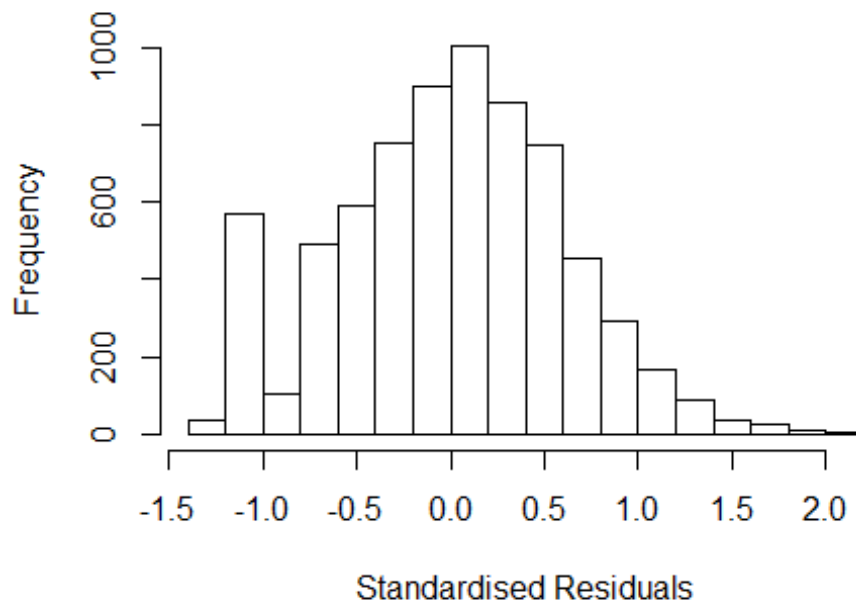
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2599 -0.4026 0.0218 0.4105 2.1494
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06506 0.04004 26.60 <2e-16 ***
## FirstAuthorFemale1 0.05183 0.02019 2.57 0.0103 *
## LastAuthorFemale1 -0.02241 0.02021 -1.11 0.2676
## Year1997 -0.03205 0.05528 -0.58 0.5620
## Year1998 0.00385 0.05746 0.07 0.9466
## Year1999 -0.02168 0.05308 -0.41 0.6830
## Year2000 0.02447 0.05296 0.46 0.6440
## Year2001 -0.03118 0.05544 -0.56 0.5739
## Year2002 0.14306 0.05146 2.78 0.0054 **
## Year2003 0.05592 0.05260 1.06 0.2878
## Year2004 0.12264 0.05092 2.41 0.0161 *
## Year2005 0.02003 0.05111 0.39 0.6951
```

```

## Year2006          0.09819    0.04857    2.02    0.0433 *
## Year2007          0.11912    0.04722    2.52    0.0117 *
## Year2008          0.06944    0.04776    1.45    0.1460
## Year2009          0.03381    0.04608    0.73    0.4631
## Year2010          0.04978    0.04652    1.07    0.2847
## Year2011          0.04751    0.04639    1.02    0.3059
## Year2012          0.02467    0.04787    0.52    0.6063
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.00686,    Adjusted R-squared:  0.00435
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 626 weights are ~= 1. The remaining 6500 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.166  0.865   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      1.40e-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 1.025 1         1.012
## Year              1.025 16         1.001

```

Residuals from first author

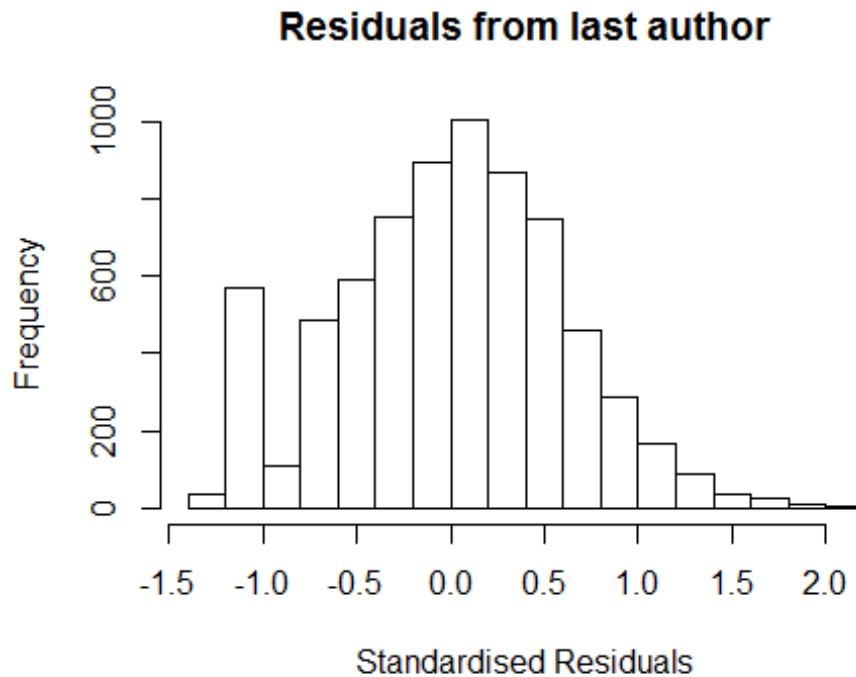


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2419 -0.4025 0.0207 0.4077 2.1527
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06325 0.04001 26.58 <2e-16 ***
## FirstAuthorFemale1 0.03634 0.01504 2.42 0.0157 *
## Year1997 -0.03271 0.05530 -0.59 0.5542
## Year1998 0.00426 0.05747 0.07 0.9409
## Year1999 -0.02195 0.05310 -0.41 0.6793
## Year2000 0.02432 0.05301 0.46 0.6464
## Year2001 -0.03188 0.05548 -0.57 0.5656
## Year2002 0.14231 0.05149 2.76 0.0057 **
## Year2003 0.05536 0.05258 1.05 0.2925
## Year2004 0.12308 0.05094 2.42 0.0157 *
## Year2005 0.01823 0.05106 0.36 0.7211
## Year2006 0.09700 0.04854 2.00 0.0457 *
```

```

## Year2007          0.11780      0.04718      2.50      0.0126 *
## Year2008          0.06868      0.04779      1.44      0.1507
## Year2009          0.03248      0.04605      0.71      0.4806
## Year2010          0.04919      0.04654      1.06      0.2905
## Year2011          0.04608      0.04637      0.99      0.3204
## Year2012          0.02402      0.04785      0.50      0.6157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.00671,    Adjusted R-squared:  0.00434
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 616 weights are ~= 1. The remaining 6510 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.165  0.865  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      1.40e-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 1.026 1          1.013
## Year            1.026 16          1.001

```



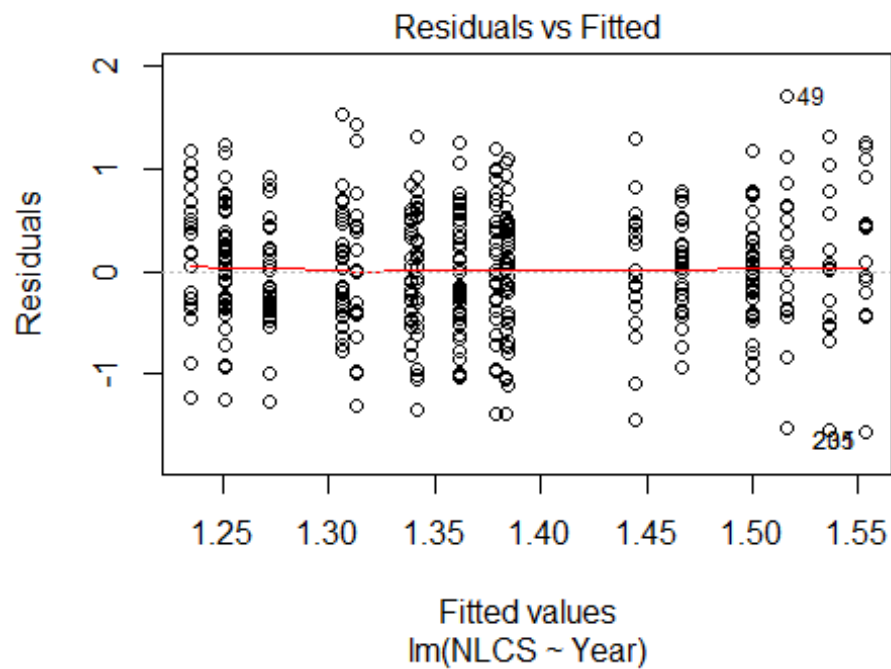
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2296 -0.3973 0.0231 0.4067 2.1428
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07210 0.03992 26.86 <2e-16 ***
## LastAuthorFemale1 0.01381 0.01506 0.92 0.3590
## Year1997 -0.03459 0.05539 -0.62 0.5323
## Year1998 0.00266 0.05753 0.05 0.9632
## Year1999 -0.02217 0.05323 -0.42 0.6771
## Year2000 0.02421 0.05314 0.46 0.6487
## Year2001 -0.03324 0.05551 -0.60 0.5493
## Year2002 0.14369 0.05152 2.79 0.0053 **
## Year2003 0.05584 0.05260 1.06 0.2885
## Year2004 0.12330 0.05102 2.42 0.0157 *
## Year2005 0.01864 0.05106 0.37 0.7151
## Year2006 0.09863 0.04856 2.03 0.0423 *
```

```

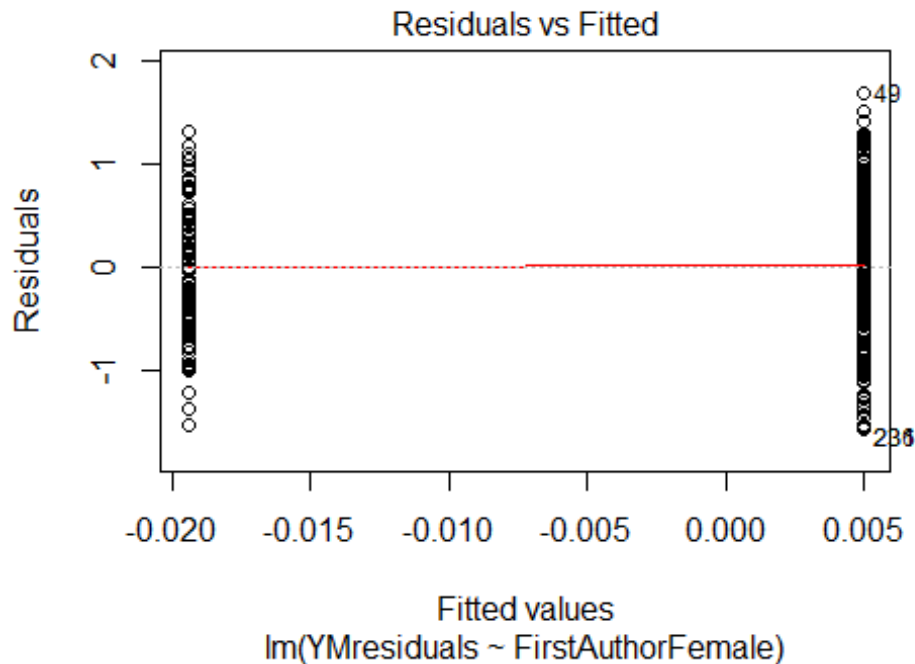
## Year2007      0.11896    0.04722    2.52    0.0118 *
## Year2008      0.06963    0.04780    1.46    0.1452
## Year2009      0.03423    0.04607    0.74    0.4575
## Year2010      0.05124    0.04654    1.10    0.2710
## Year2011      0.04713    0.04643    1.02    0.3100
## Year2012      0.02741    0.04790    0.57    0.5671
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.597
## Multiple R-squared:  0.00599,    Adjusted R-squared:  0.00361
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 640 weights are ~= 1. The remaining 6486 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.170  0.865  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      1.40e-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: 7126"
## [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
##   24   32   46   30   33   30   27   28   36   35   41   54   50   52   55
## 2011 2012
##   53   69
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   14   23   33   18   25   16   18   18   22   26   28   33   36   32   46
## 2011 2012

```

```
## 40 47
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 11 23 30 16 23 14 14 16 18 23 24 27 31 27 37
## 2011 2012
## 36 37
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 36, df = 16, p-value = 0.003
```

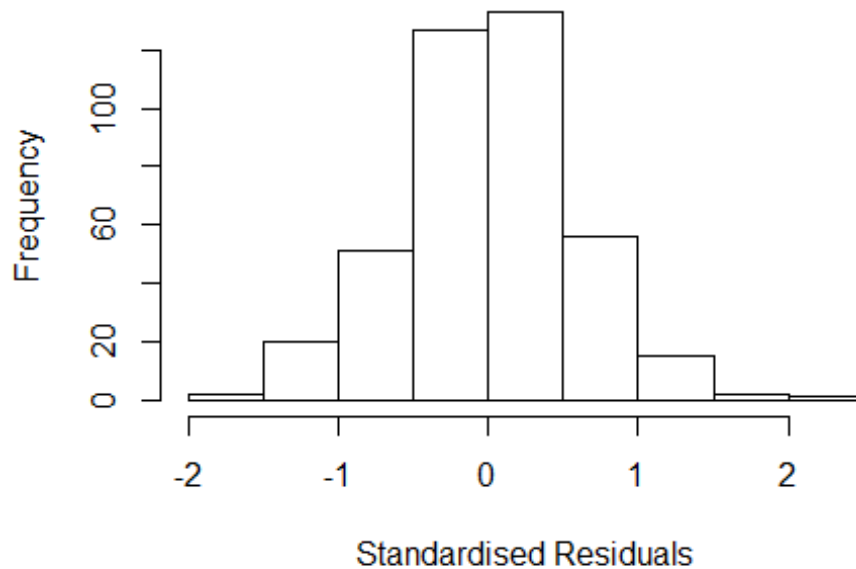


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0026, df = 1, p-value = 1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.66789976384271"
## [1] "Male first author team size 2018 geometric mean: 2.35947958412732"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 750, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.93570896606199"
## [1] "Male last author team size 2018 geometric mean: 2.36093053499168"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 560, 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 1.390 1          1.179
## LastAuthorFemale  1.600 1          1.265
## UniqueAuthors    2.565 4          1.125
## Year              3.173 16         1.037
```


Residuals from first and last author and team size



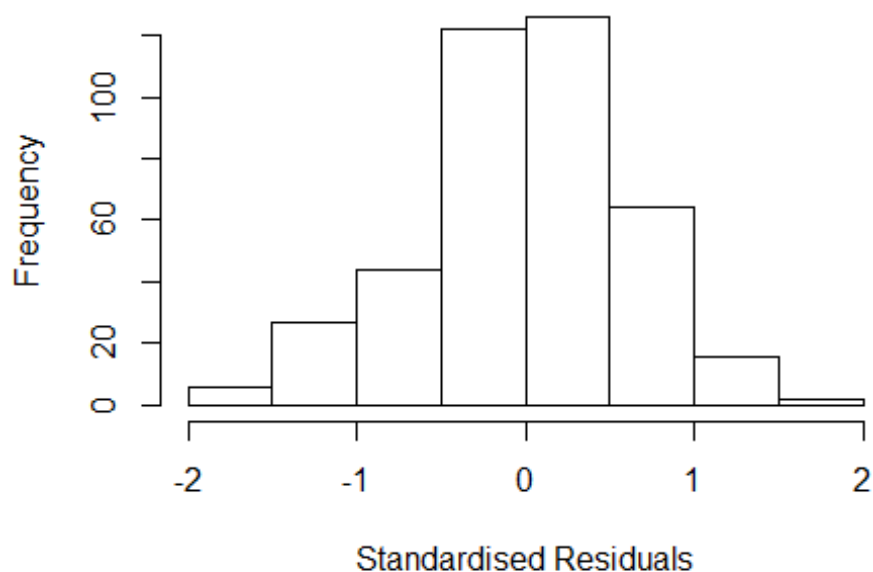
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.78618 -0.36238 0.00622 0.39439 2.00465
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3162 0.2486 5.29 2.0e-07 ***
## FirstAuthorFemale1 -0.0933 0.0836 -1.12 0.27
## LastAuthorFemale1 0.0771 0.0955 0.81 0.42
## UniqueAuthors2 0.3734 0.0799 4.67 4.1e-06 ***
## UniqueAuthors3 0.4760 0.0943 5.05 7.0e-07 ***
## UniqueAuthors4 0.2262 0.1598 1.42 0.16
## UniqueAuthors5 0.4674 0.1157 4.04 6.5e-05 ***
## Year1997 -0.1099 0.2856 -0.38 0.70
## Year1998 -0.1246 0.2790 -0.45 0.66
## Year1999 -0.0803 0.2822 -0.28 0.78
```

```

## Year2000          -0.3147      0.3105    -1.01      0.31
## Year2001          -0.0832      0.4068    -0.20      0.84
## Year2002          -0.2750      0.3022    -0.91      0.36
## Year2003          -0.1606      0.3215    -0.50      0.62
## Year2004          -0.1519      0.2577    -0.59      0.56
## Year2005          -0.2073      0.2650    -0.78      0.43
## Year2006          -0.1753      0.2579    -0.68      0.50
## Year2007          -0.2873      0.2523    -1.14      0.26
## Year2008          -0.2747      0.2560    -1.07      0.28
## Year2009          -0.2238      0.2748    -0.81      0.42
## Year2010          -0.3939      0.2629    -1.50      0.13
## Year2011          -0.1282      0.2502    -0.51      0.61
## Year2012          -0.2133      0.2620    -0.81      0.42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.575
## Multiple R-squared:  0.125, Adjusted R-squared:  0.0748
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 370 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.200  0.875  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      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 1.268 1      1.126
## LastAuthorFemale  1.400 1      1.183
## Year              1.427 16      1.011

```

Residuals from first and last author



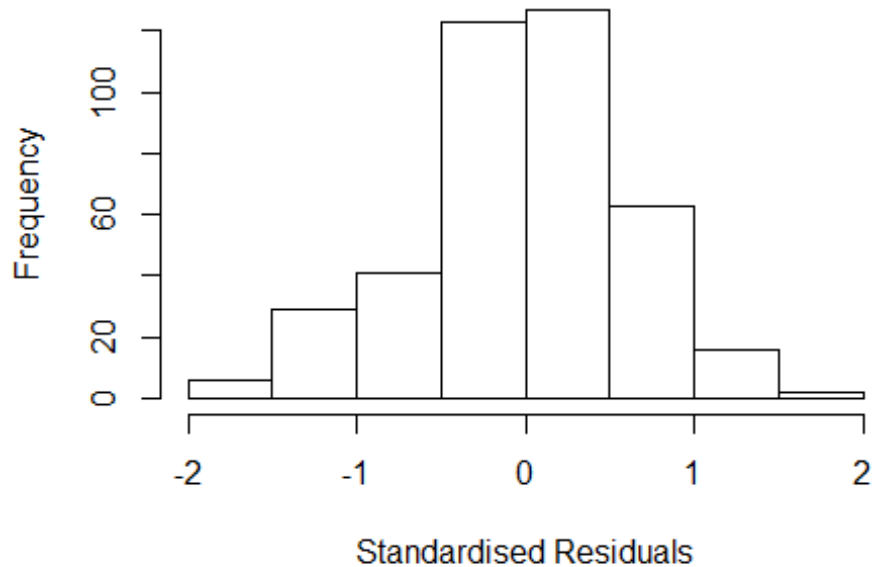
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6311 -0.3628 0.0136 0.3962 1.6842
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5144 0.2280 6.64 1e-10 ***
## FirstAuthorFemale1 -0.0704 0.0851 -0.83 0.41
## LastAuthorFemale1 0.0418 0.0925 0.45 0.65
## Year1997 0.0124 0.2672 0.05 0.96
## Year1998 -0.0880 0.2714 -0.32 0.75
## Year1999 -0.1151 0.2728 -0.42 0.67
## Year2000 -0.2879 0.3086 -0.93 0.35
## Year2001 0.0750 0.4278 0.18 0.86
## Year2002 -0.1623 0.3017 -0.54 0.59
## Year2003 -0.1292 0.3115 -0.41 0.68
## Year2004 -0.0457 0.2495 -0.18 0.85
## Year2005 -0.1140 0.2476 -0.46 0.65
```

```

## Year2006          -0.0868      0.2462   -0.35      0.72
## Year2007          -0.2199      0.2416   -0.91      0.36
## Year2008          -0.2947      0.2456   -1.20      0.23
## Year2009          -0.1535      0.2688   -0.57      0.57
## Year2010          -0.2682      0.2509   -1.07      0.29
## Year2011           0.0250      0.2385    0.10      0.92
## Year2012          -0.0678      0.2510   -0.27      0.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0319, Adjusted R-squared:  -0.013
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 381 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.402  0.876  0.954  0.904  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.168 1      1.081
## Year              1.168 16      1.005

```

Residuals from first author



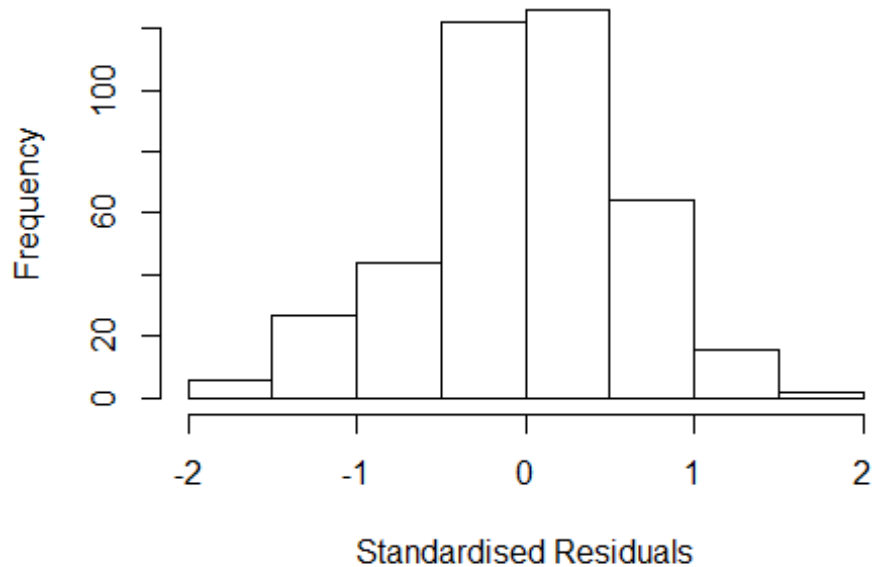
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5921 -0.3716 0.0205 0.3921 1.6832
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5162 0.2302 6.59 1.5e-10 ***
## FirstAuthorFemale1 -0.0577 0.0815 -0.71 0.48
## Year1997 0.0116 0.2689 0.04 0.97
## Year1998 -0.0850 0.2725 -0.31 0.76
## Year1999 -0.1146 0.2736 -0.42 0.68
## Year2000 -0.2839 0.3095 -0.92 0.36
## Year2001 0.0759 0.4293 0.18 0.86
## Year2002 -0.1629 0.3043 -0.54 0.59
## Year2003 -0.1278 0.3136 -0.41 0.68
## Year2004 -0.0445 0.2510 -0.18 0.86
## Year2005 -0.1139 0.2493 -0.46 0.65
## Year2006 -0.0781 0.2445 -0.32 0.75
```

```

## Year2007          -0.2169      0.2431   -0.89      0.37
## Year2008          -0.2865      0.2443   -1.17      0.24
## Year2009          -0.1518      0.2697   -0.56      0.57
## Year2010          -0.2665      0.2527   -1.05      0.29
## Year2011           0.0267      0.2400     0.11      0.91
## Year2012          -0.0607      0.2513   -0.24      0.81
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.593
## Multiple R-squared:  0.031, Adjusted R-squared:  -0.0113
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 381 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.401  0.876  0.956  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      2.46e-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 1.258 1          1.122
## Year              1.258 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
## -1.6056 -0.3708 0.0142 0.3922 1.6913
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4795 0.2228 6.64 1.1e-10 ***
## LastAuthorFemale1 0.0182 0.0887 0.21 0.84
## Year1997 0.0402 0.2656 0.15 0.88
## Year1998 -0.0613 0.2673 -0.23 0.82
## Year1999 -0.1145 0.2691 -0.43 0.67
## Year2000 -0.2639 0.3025 -0.87 0.38
## Year2001 0.1079 0.4190 0.26 0.80
## Year2002 -0.1390 0.3002 -0.46 0.64
## Year2003 -0.0946 0.3079 -0.31 0.76
## Year2004 -0.0267 0.2454 -0.11 0.91
## Year2005 -0.0873 0.2448 -0.36 0.72
## Year2006 -0.0600 0.2424 -0.25 0.80
```

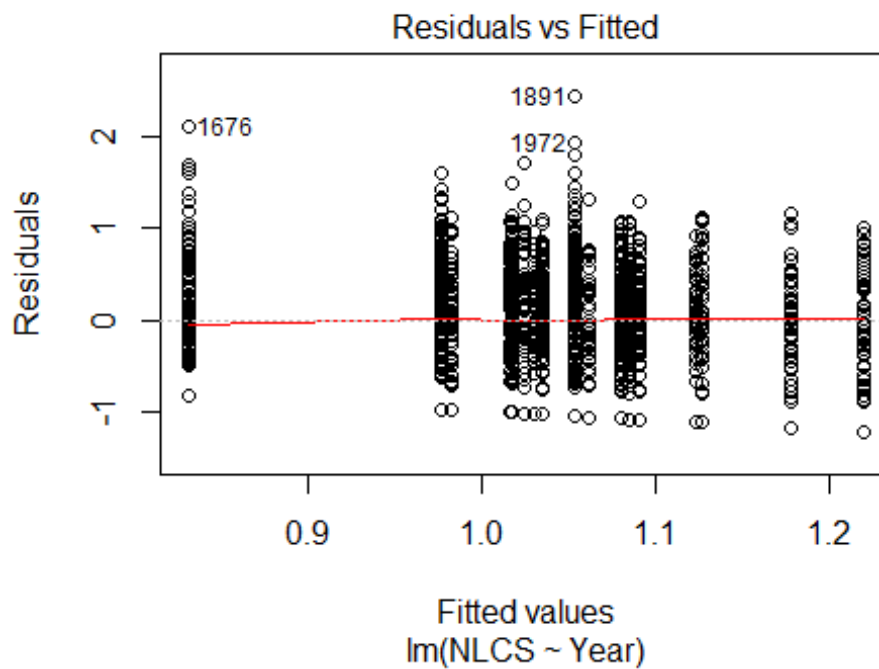
```

## Year2007          -0.1911      0.2381   -0.80      0.42
## Year2008          -0.2683      0.2430   -1.10      0.27
## Year2009          -0.1239      0.2651   -0.47      0.64
## Year2010          -0.2459      0.2489   -0.99      0.32
## Year2011           0.0405      0.2359    0.17      0.86
## Year2012          -0.0422      0.2480   -0.17      0.87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.603
## Multiple R-squared:  0.03,   Adjusted R-squared:  -0.0124
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 380 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.411  0.877  0.954  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      2.46e-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: 407"
## [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
##   53   65   55   68   61   83   66   61   61  100  114  136  133  183  181
## 2011 2012
##  229  219
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   41   60   43   56   44   62   55   52   48   87   95  122  117  152  152
## 2011 2012

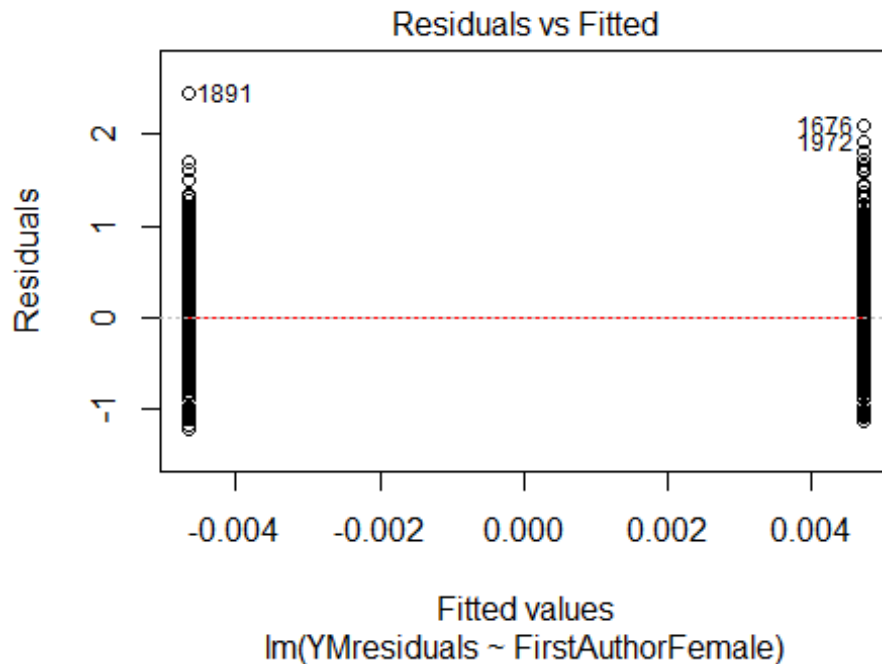
```



```
## 189 175
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 41 56 42 53 42 61 54 50 46 83 88 118 113 139 146
## 2011 2012
## 179 164
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 45, df = 16, p-value = 1e-04
```

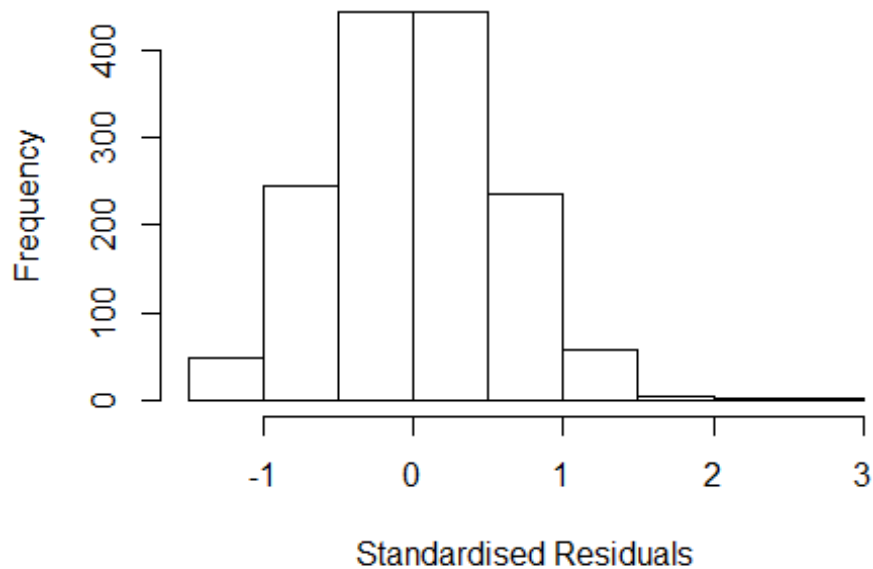


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 1.39018663582352"
## [1] "Male first author team size 2018 geometric mean: 1.21079967661963"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5600, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.30883797540644"
## [1] "Male last author team size 2018 geometric mean: 1.32642585071091"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5100, 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 2.192 1      1.481
## LastAuthorFemale  2.221 1      1.490
## UniqueAuthors    1.261 4      1.029
## Year             1.323 16      1.009
```

Residuals from first and last author and team size



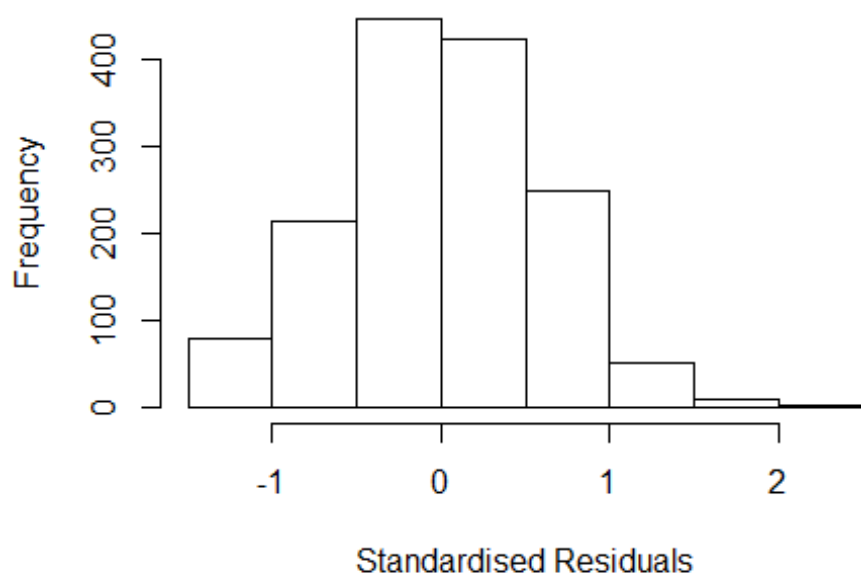
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1891 79953204979 3.499 2011    1201      3      2.582
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q       Max
## -1.331269 -0.379123  0.000993  0.412134  2.582026
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.01615    0.08272   12.28 < 2e-16 ***
## FirstAuthorFemale1  0.01010    0.04593    0.22  0.8259
## LastAuthorFemale1 -0.04056    0.04598   -0.88  0.3779
## UniqueAuthors2     0.25746    0.04164    6.18 8.2e-10 ***
## UniqueAuthors3     0.23181    0.07316    3.17  0.0016 **
## UniqueAuthors4     0.21115    0.08889    2.38  0.0177 *
## UniqueAuthors5     0.22344    0.08060    2.77  0.0056 **
## Year1997         -0.03140    0.13214   -0.24  0.8122
## Year1998          0.06766    0.11026    0.61  0.5396
## Year1999         -0.04209    0.11150   -0.38  0.7059
```

```

## Year2000      -0.03916    0.11060   -0.35    0.7234
## Year2001      -0.09939    0.10752   -0.92    0.3554
## Year2002       0.15618    0.11334    1.38    0.1684
## Year2003       0.10706    0.10656    1.00    0.3152
## Year2004       0.05766    0.12915    0.45    0.6554
## Year2005       0.04632    0.09683    0.48    0.6325
## Year2006      -0.03259    0.09871   -0.33    0.7413
## Year2007       0.02099    0.09215    0.23    0.8199
## Year2008       0.00175    0.09785    0.02    0.9857
## Year2009      -0.07342    0.09838   -0.75    0.4556
## Year2010      -0.27693    0.09623   -2.88    0.0041 **
## Year2011      -0.06872    0.09707   -0.71    0.4791
## Year2012      -0.13369    0.09648   -1.39    0.1661
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0605, Adjusted R-squared:  0.0463
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 126 weights are ~= 1. The remaining 1349 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0193 0.8700 0.9520 0.9110 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.78e-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 2.371 1 1.540
## LastAuthorFemale 2.411 1 1.553
## Year 1.084 16 1.003

```

Residuals from first and last author



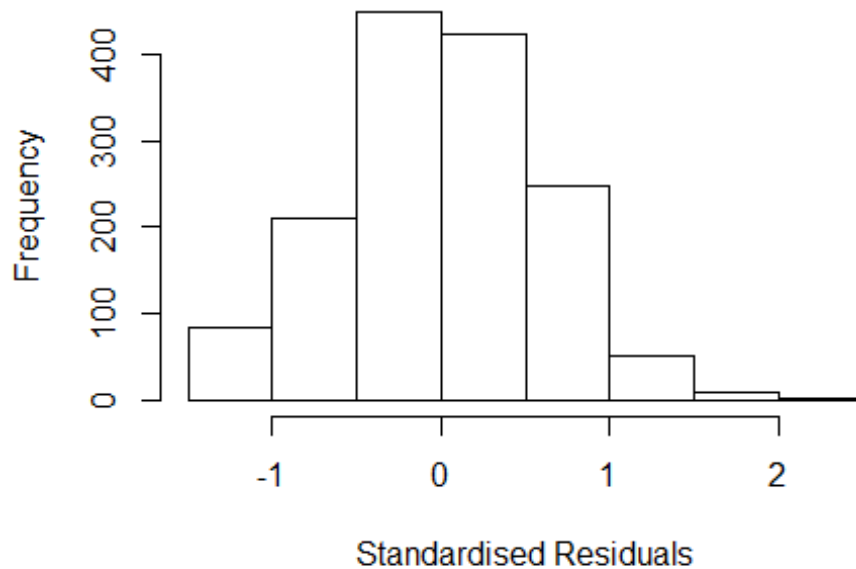
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20712 -0.39631 -0.00119 0.42399 2.49118
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0662 0.0850 12.54 <2e-16 ***
## FirstAuthorFemale1 0.0565 0.0490 1.15 0.2483
## LastAuthorFemale1 -0.0745 0.0492 -1.51 0.1303
## Year1997 -0.0535 0.1346 -0.40 0.6910
## Year1998 0.0641 0.1126 0.57 0.5691
## Year1999 -0.0255 0.1136 -0.22 0.8226
## Year2000 -0.0362 0.1133 -0.32 0.7492
## Year2001 -0.0759 0.1121 -0.68 0.4983
## Year2002 0.1589 0.1180 1.35 0.1784
## Year2003 0.0958 0.1114 0.86 0.3901
## Year2004 0.0572 0.1300 0.44 0.6600
## Year2005 0.0360 0.0995 0.36 0.7179
```

```

## Year2006          -0.0334      0.1005    -0.33    0.7397
## Year2007           0.0351      0.0947     0.37    0.7110
## Year2008           0.0225      0.1003     0.22    0.8227
## Year2009          -0.0703      0.1013    -0.69    0.4882
## Year2010          -0.2802      0.0993    -2.82    0.0048 **
## Year2011          -0.0404      0.1009    -0.40    0.6890
## Year2012          -0.1089      0.0993    -1.10    0.2732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.613
## Multiple R-squared:  0.0291, Adjusted R-squared:  0.0171
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1354 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0618 0.8760 0.9520 0.9140 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.78e-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

```

Residuals from first author



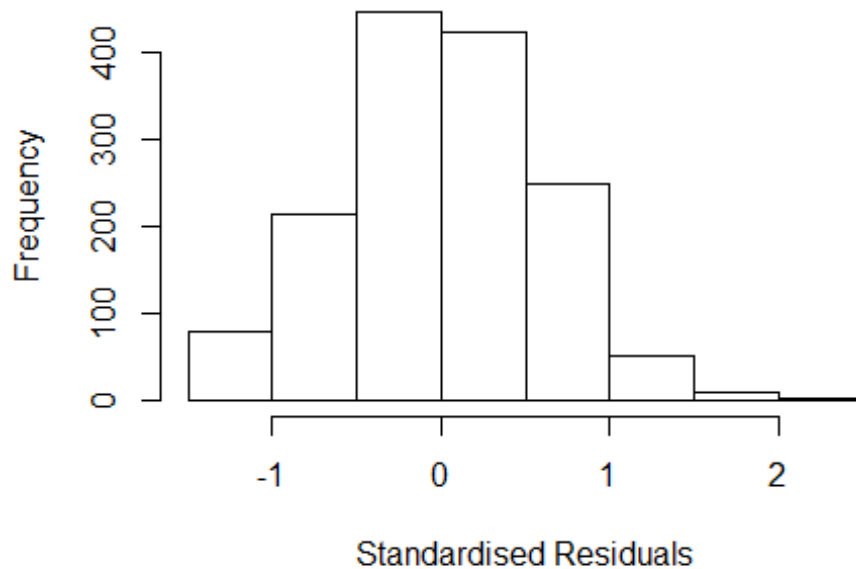
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21764 -0.40254 -0.00492 0.41800 2.48492
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05777 0.08462 12.50 <2e-16 ***
## FirstAuthorFemale1 -0.00145 0.03248 -0.04 0.964
## Year1997 -0.05373 0.13420 -0.40 0.689
## Year1998 0.06689 0.11266 0.59 0.553
## Year1999 -0.01824 0.11383 -0.16 0.873
## Year2000 -0.03388 0.11334 -0.30 0.765
## Year2001 -0.07307 0.11198 -0.65 0.514
## Year2002 0.16131 0.11783 1.37 0.171
## Year2003 0.09819 0.11191 0.88 0.380
## Year2004 0.06667 0.12955 0.51 0.607
## Year2005 0.03710 0.09947 0.37 0.709
## Year2006 -0.02974 0.10050 -0.30 0.767
```

```

## Year2007          0.03546    0.09472    0.37    0.708
## Year2008          0.02587    0.10014    0.26    0.796
## Year2009         -0.06740    0.10123   -0.67    0.506
## Year2010         -0.27853    0.09915   -2.81    0.005 **
## Year2011         -0.04225    0.10087   -0.42    0.675
## Year2012         -0.10761    0.09938   -1.08    0.279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.613
## Multiple R-squared:  0.0277, Adjusted R-squared:  0.0164
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1353 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0633 0.8740 0.9520 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      6.78e-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 1.057 1          1.028
## Year              1.057 16          1.002

```


Residuals from last author



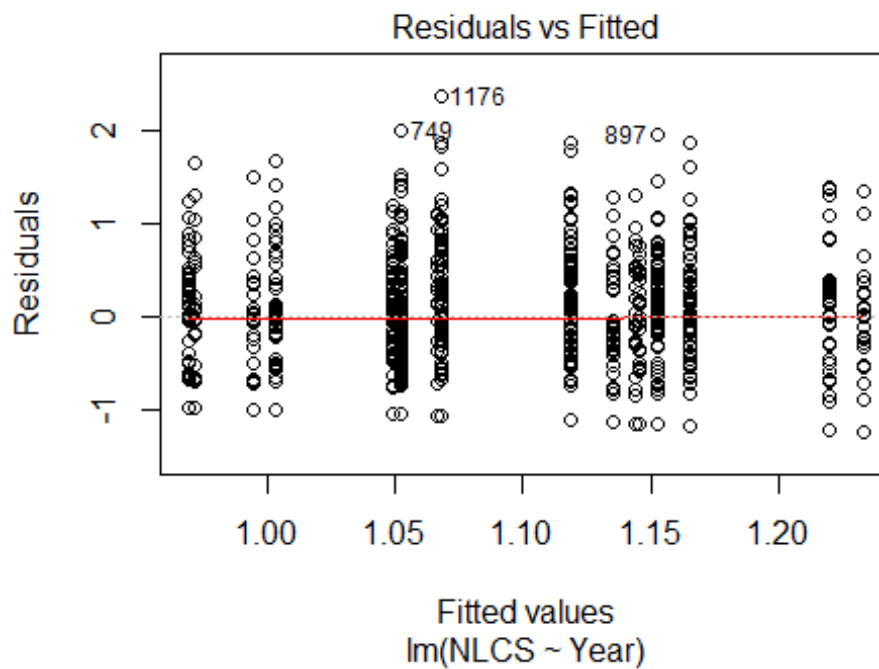
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19993 -0.39261 -0.00297 0.42109 2.49714
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0716 0.0852 12.58 <2e-16 ***
## LastAuthorFemale1 -0.0304 0.0326 -0.93 0.3522
## Year1997 -0.0547 0.1347 -0.41 0.6849
## Year1998 0.0629 0.1127 0.56 0.5769
## Year1999 -0.0224 0.1143 -0.20 0.8444
## Year2000 -0.0336 0.1137 -0.30 0.7677
## Year2001 -0.0740 0.1123 -0.66 0.5099
## Year2002 0.1587 0.1179 1.35 0.1785
## Year2003 0.0985 0.1117 0.88 0.3780
## Year2004 0.0627 0.1294 0.48 0.6283
## Year2005 0.0377 0.0995 0.38 0.7047
## Year2006 -0.0303 0.1007 -0.30 0.7637
```

```

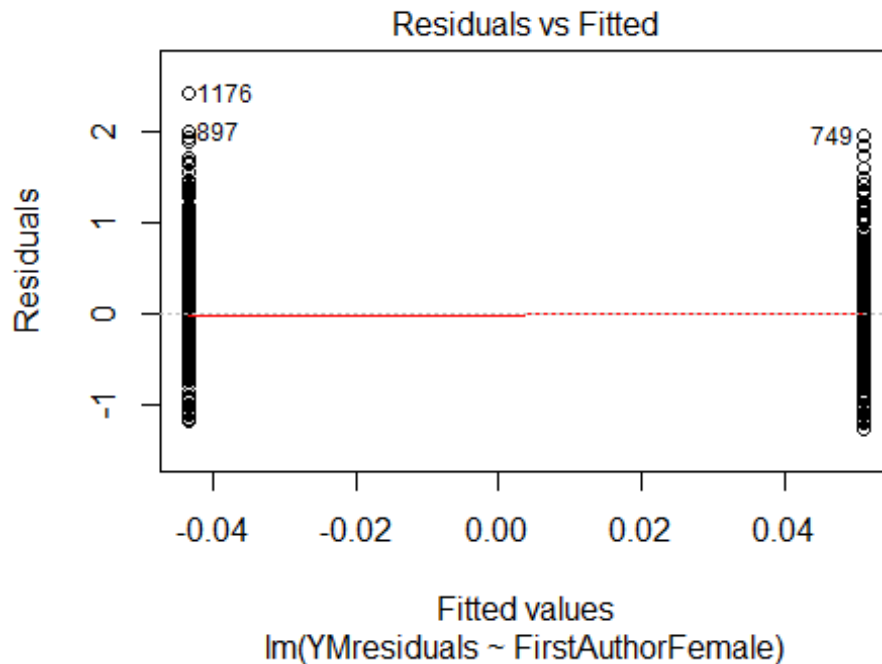
## Year2007          0.0361      0.0949      0.38      0.7035
## Year2008          0.0261      0.1002      0.26      0.7945
## Year2009         -0.0687      0.1015     -0.68      0.4986
## Year2010         -0.2774      0.0992     -2.80      0.0053 **
## Year2011         -0.0394      0.1010     -0.39      0.6966
## Year2012         -0.1052      0.0994     -1.06      0.2902
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.0283, Adjusted R-squared:  0.017
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1355 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0609 0.8750 0.9520 0.9140 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.78e-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: 1475"
## [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
##   26   24   32   32   38   50   43   41   29   58   72   76   87   83  110
## 2011 2012
##  114  145
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   22   23   28   28   24   45   38   37   24   51   55   63   75   73   95
## 2011 2012

```

```
## 99 122
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 22 22 27 27 22 44 34 36 24 45 55 58 70 66 92
## 2011 2012
## 95 118
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 18, df = 16, p-value = 0.3
```

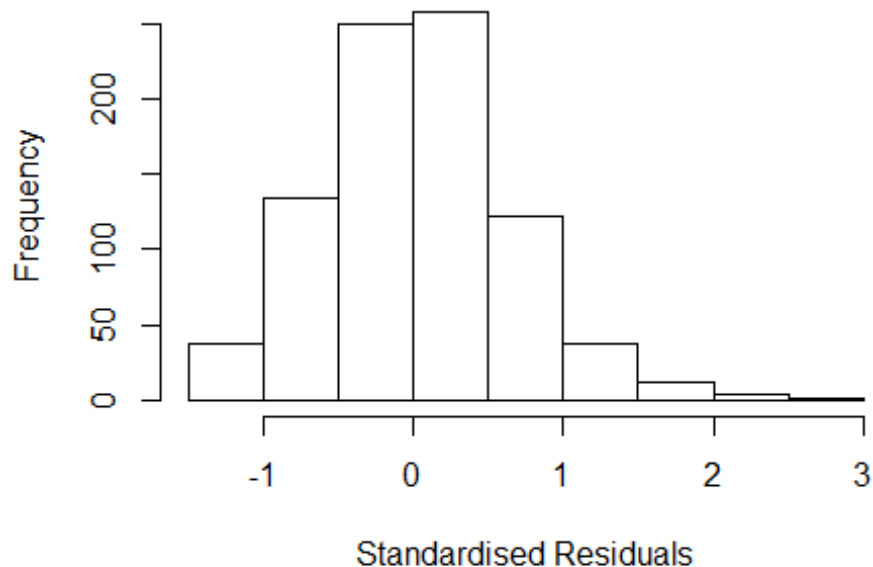


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.95, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 1.78133751186058"
## [1] "Male first author team size 2018 geometric mean: 1.32885270660085"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2900, p-value = 0.001
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.69262129836125"
## [1] "Male last author team size 2018 geometric mean: 1.42976073736751"
##
## 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 1.710 1      1.308
## LastAuthorFemale  1.738 1      1.318
## UniqueAuthors     1.374 4      1.041
## Year              1.501 16     1.013
```

Residuals from first and last author and team size



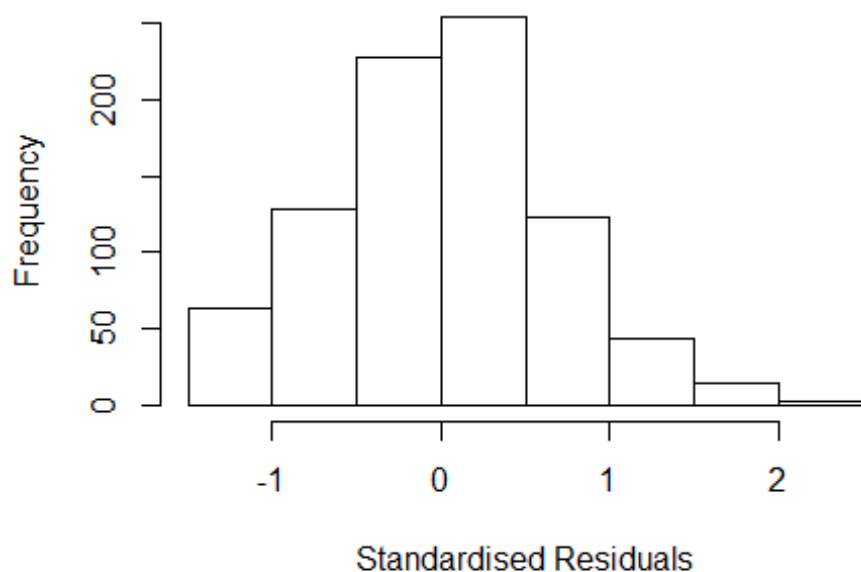
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1176 84856557156 3.439 2012      1203      3      2.575
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4259 -0.3928  0.0149  0.3958  2.5750
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0936    0.1313   8.33 3.4e-16 ***
## FirstAuthorFemale1  0.1234    0.0555   2.22 0.02650 *
## LastAuthorFemale1 -0.0386    0.0556  -0.69 0.48752
## UniqueAuthors2     0.2809    0.0561   5.01 6.8e-07 ***
## UniqueAuthors3     0.3374    0.0695   4.86 1.4e-06 ***
## UniqueAuthors4     0.3548    0.0919   3.86 0.00012 ***
## UniqueAuthors5     0.3096    0.0974   3.18 0.00153 **
## Year1997          -0.1511    0.1837  -0.82 0.41111
## Year1998          -0.2861    0.2033  -1.41 0.15965
## Year1999          -0.2402    0.1574  -1.53 0.12740
```

```

## Year2000          -0.1334      0.1575   -0.85   0.39731
## Year2001          -0.2711      0.1588   -1.71   0.08808 .
## Year2002          -0.1336      0.1810   -0.74   0.46052
## Year2003          -0.1701      0.1597   -1.07   0.28703
## Year2004          -0.0720      0.2071   -0.35   0.72825
## Year2005          -0.3406      0.1653   -2.06   0.03965 *
## Year2006          -0.1884      0.1437   -1.31   0.19019
## Year2007          -0.2447      0.1492   -1.64   0.10122
## Year2008          -0.0818      0.1472   -0.56   0.57840
## Year2009          -0.2893      0.1457   -1.99   0.04741 *
## Year2010          -0.1457      0.1439   -1.01   0.31174
## Year2011          -0.1399      0.1469   -0.95   0.34136
## Year2012          -0.2296      0.1430   -1.61   0.10861
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.0744, Adjusted R-squared:  0.05
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 780 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.039  0.849  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          1.17e-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 1.799 1          1.341
## LastAuthorFemale  1.817 1          1.348
## Year              1.151 16          1.004

```

Residuals from first and last author



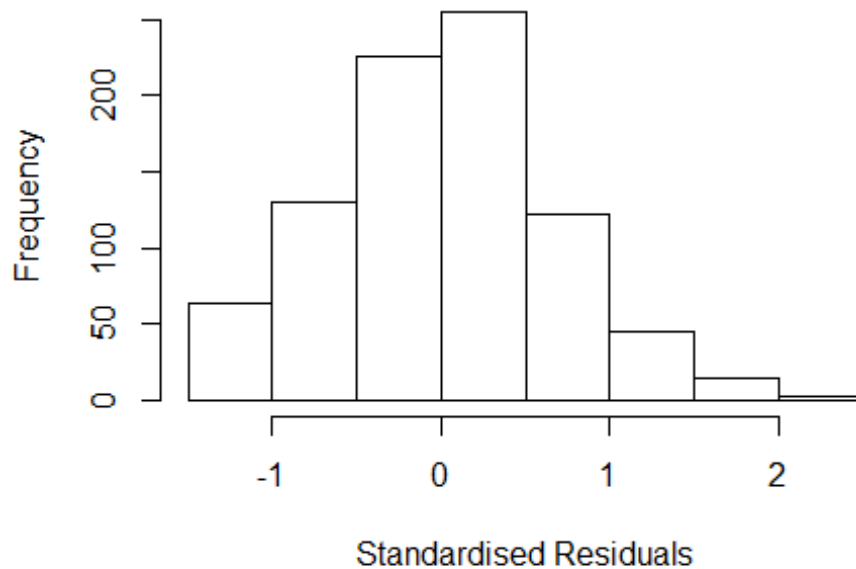
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2398 -0.4408 0.0186 0.4044 2.4695
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1539 0.1319 8.75 <2e-16 ***
## FirstAuthorFemale1 0.1374 0.0596 2.31 0.021 *
## LastAuthorFemale1 -0.0289 0.0595 -0.49 0.627
## Year1997 -0.1552 0.1972 -0.79 0.432
## Year1998 -0.2450 0.1987 -1.23 0.218
## Year1999 -0.2476 0.1642 -1.51 0.132
## Year2000 -0.0848 0.1630 -0.52 0.603
## Year2001 -0.2278 0.1626 -1.40 0.162
## Year2002 -0.0817 0.1805 -0.45 0.651
## Year2003 -0.0901 0.1653 -0.55 0.586
## Year2004 -0.0551 0.2091 -0.26 0.792
## Year2005 -0.2558 0.1681 -1.52 0.128
```

```

## Year2006          -0.1621      0.1459   -1.11    0.267
## Year2007          -0.1992      0.1482   -1.34    0.179
## Year2008          -0.0225      0.1514   -0.15    0.882
## Year2009          -0.2222      0.1486   -1.50    0.135
## Year2010          -0.0762      0.1431   -0.53    0.595
## Year2011          -0.1149      0.1488   -0.77    0.440
## Year2012          -0.1844      0.1451   -1.27    0.204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.635
## Multiple R-squared:  0.0216, Adjusted R-squared:  0.000549
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 780 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0965 0.8570 0.9500 0.9030 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      1.17e-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 1.057 1      1.028
## Year              1.057 16      1.002

```


Residuals from first author



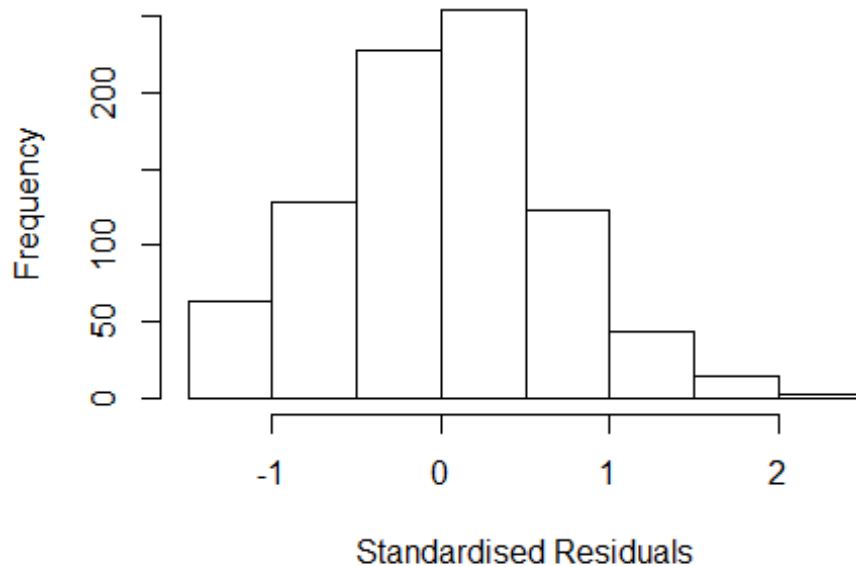
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2437 -0.4385 0.0173 0.4013 2.4735
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1484 0.1311 8.76 <2e-16 ***
## FirstAuthorFemale1 0.1172 0.0458 2.56 0.011 *
## Year1997 -0.1544 0.1967 -0.79 0.433
## Year1998 -0.2385 0.1978 -1.21 0.228
## Year1999 -0.2447 0.1642 -1.49 0.136
## Year2000 -0.0824 0.1629 -0.51 0.613
## Year2001 -0.2254 0.1629 -1.38 0.167
## Year2002 -0.0789 0.1801 -0.44 0.661
## Year2003 -0.0871 0.1649 -0.53 0.598
## Year2004 -0.0535 0.2084 -0.26 0.797
## Year2005 -0.2584 0.1678 -1.54 0.124
## Year2006 -0.1584 0.1456 -1.09 0.277
```

```

## Year2007          -0.1968      0.1483   -1.33    0.185
## Year2008          -0.0219      0.1513   -0.14    0.885
## Year2009          -0.2213      0.1488   -1.49    0.137
## Year2010          -0.0752      0.1433   -0.53    0.600
## Year2011          -0.1150      0.1488   -0.77    0.440
## Year2012          -0.1829      0.1452   -1.26    0.208
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.635
## Multiple R-squared:  0.0214, Adjusted R-squared:  0.00158
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 784 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0953 0.8560 0.9510 0.9040 0.9880 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.17e-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 1.064 1      1.032
## Year      1.064 16      1.002

```

Residuals from last author



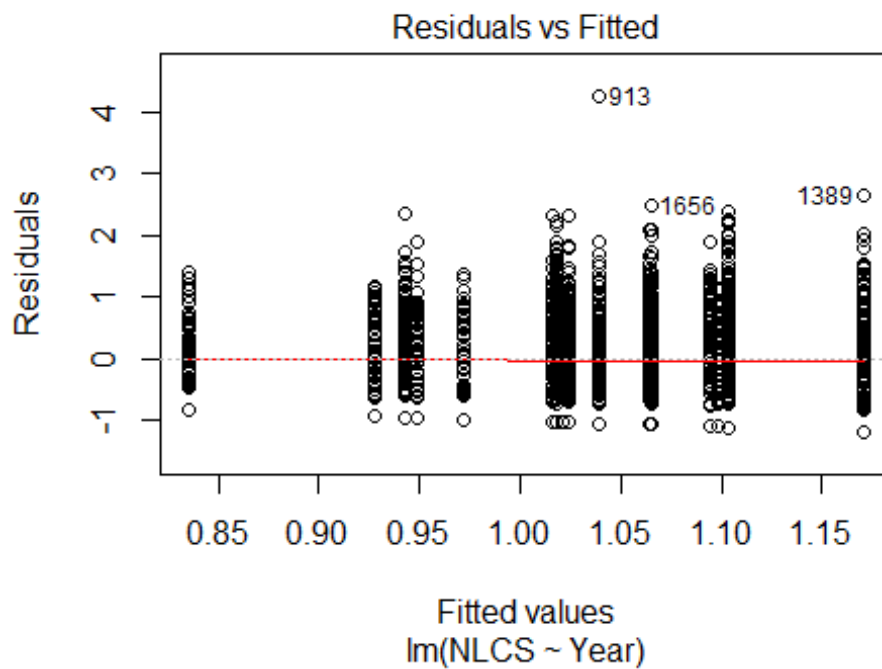
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21215 -0.44053 0.00961 0.41861 2.44844
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1792 0.1312 8.98 <2e-16 ***
## LastAuthorFemale1 0.0677 0.0459 1.48 0.14
## Year1997 -0.1645 0.1965 -0.84 0.40
## Year1998 -0.2345 0.1966 -1.19 0.23
## Year1999 -0.2513 0.1651 -1.52 0.13
## Year2000 -0.0862 0.1657 -0.52 0.60
## Year2001 -0.2438 0.1632 -1.49 0.14
## Year2002 -0.0745 0.1801 -0.41 0.68
## Year2003 -0.0948 0.1645 -0.58 0.56
## Year2004 -0.0604 0.2045 -0.30 0.77
## Year2005 -0.2747 0.1670 -1.65 0.10
## Year2006 -0.1606 0.1462 -1.10 0.27
```

```

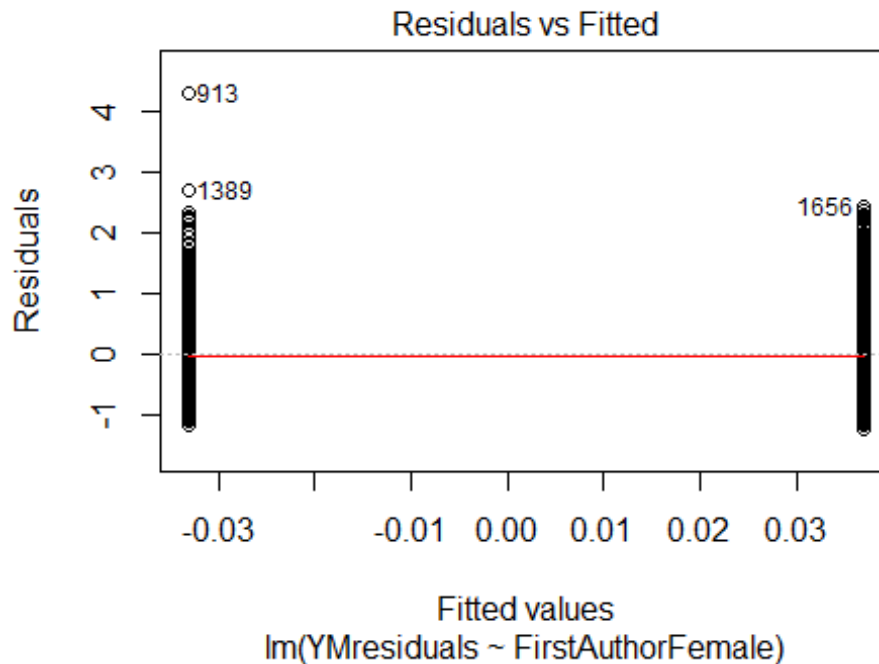
## Year2007          -0.2009      0.1487    -1.35      0.18
## Year2008          -0.0348      0.1510    -0.23      0.82
## Year2009          -0.2296      0.1486    -1.55      0.12
## Year2010          -0.0822      0.1439    -0.57      0.57
## Year2011          -0.1283      0.1483    -0.87      0.39
## Year2012          -0.1887      0.1450    -1.30      0.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.638
## Multiple R-squared:  0.0163, Adjusted R-squared:  -0.00364
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 86 weights are ~= 1. The remaining 771 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.108  0.861  0.948  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      1.17e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 857"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
##   47   75   67   69   86   97  103  119  132  175  195  205  214  246  307
## 2011 2012
##  362  358
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   40   69   58   65   69   77   87  112  112  153  174  179  179  210  255
## 2011 2012

```

```
## 308 302
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 39 69 58 64 65 75 87 112 112 153 172 169 171 198 248
## 2011 2012
## 297 292
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.009
```

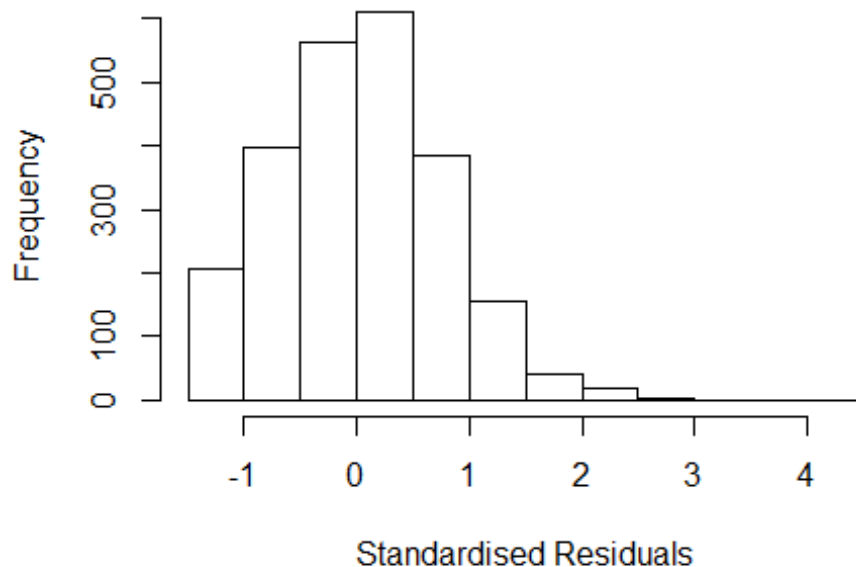


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0037, df = 1, p-value = 1
```



```
## [1] "Female first author team size 2018 geometric mean: 1.25084307025577"
## [1] "Male first author team size 2018 geometric mean: 1.18551696898185"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.25911681345234"
## [1] "Male last author team size 2018 geometric mean: 1.17637298978119"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.385 1          1.840
## LastAuthorFemale 3.311 1          1.820
## UniqueAuthors    1.162 4          1.019
## Year             1.167 16         1.005
```

Residuals from first and last author and team size



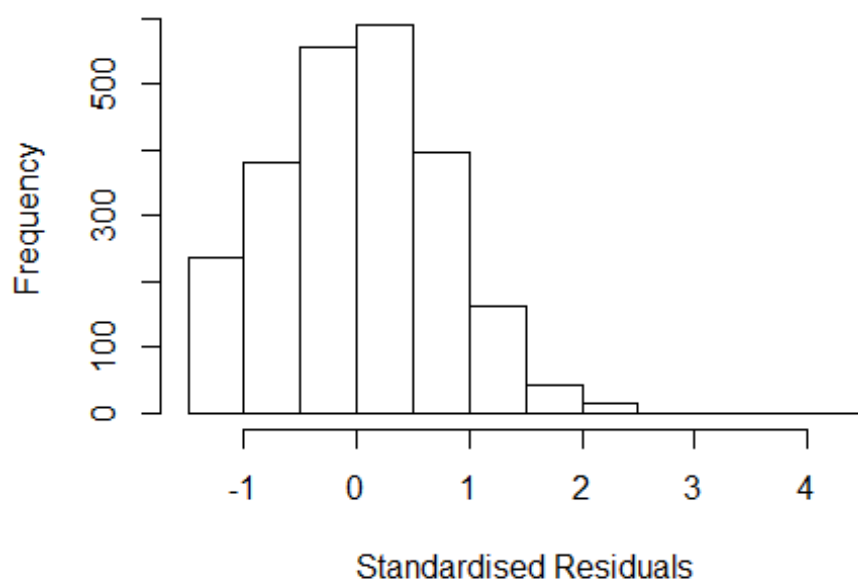
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 913  34247665578 5.284 2005      1202      4      4.333
## 1389 34250205270 3.822 2007      3316      1      2.759
## 1656 77649230762 3.539 2008      1213      2      2.503
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2433 -0.5182  0.0197  0.5113  4.3333
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.90144    0.12484   7.22 6.9e-13 ***
## FirstAuthorFemale1 0.00786    0.05695   0.14 0.89022
## LastAuthorFemale1 0.06415    0.05644   1.14 0.25587
## UniqueAuthors2    0.19304    0.04963   3.89 0.00010 ***
## UniqueAuthors3    0.25595    0.06671   3.84 0.00013 ***
## UniqueAuthors4    0.05500    0.12203   0.45 0.65225
## UniqueAuthors5    0.15154    0.15671   0.97 0.33364
## Year1997        -0.13018    0.15233  -0.85 0.39287
```

```

## Year1998          0.16199    0.15792    1.03  0.30509
## Year1999          -0.03743    0.15844   -0.24  0.81324
## Year2000           0.06751    0.14503    0.47  0.64160
## Year2001          -0.02973    0.14674   -0.20  0.83948
## Year2002           0.03724    0.15356    0.24  0.80841
## Year2003           0.03954    0.14566    0.27  0.78609
## Year2004           0.13303    0.15067    0.88  0.37736
## Year2005           0.04929    0.14071    0.35  0.72614
## Year2006           0.06888    0.13517    0.51  0.61037
## Year2007           0.16107    0.13526    1.19  0.23386
## Year2008           0.06303    0.13711    0.46  0.64577
## Year2009          -0.05978    0.13242   -0.45  0.65172
## Year2010           0.03174    0.13195    0.24  0.80990
## Year2011           0.07678    0.13292    0.58  0.56355
## Year2012          -0.00175    0.13384   -0.01  0.98959
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.78
## Multiple R-squared:  0.0204, Adjusted R-squared:  0.0112
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 726 is an outlier with |weight| = 0 ( < 4.2e-05);
## 222 weights are ~ = 1. The remaining 2158 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.184  0.869  0.950  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           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 3.584 1           1.893
## LastAuthorFemale  3.543 1           1.882
## Year              1.061 16           1.002

```


Residuals from first and last author



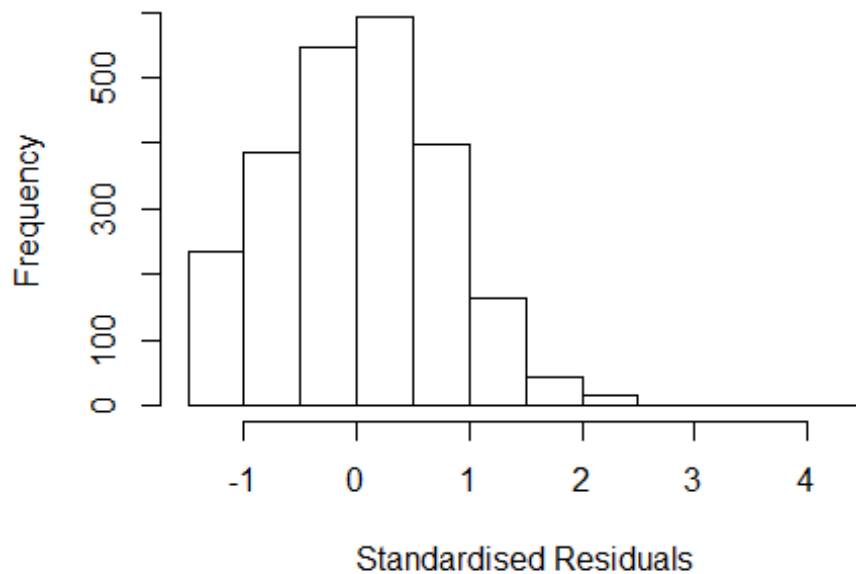
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 913  34247665578 5.284 2005      1202      4      4.321
## 1389 34250205270 3.822 2007      3316      1      2.738
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1708 -0.5257  0.0139  0.5275  4.3212
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9188    0.1250   7.35 2.8e-13 ***
## FirstAuthorFemale1 0.0339    0.0592    0.57  0.57
## LastAuthorFemale1 0.0509    0.0590    0.86  0.39
## Year1997       -0.1399    0.1522   -0.92  0.36
## Year1998        0.1673    0.1589    1.05  0.29
## Year1999       -0.0410    0.1583   -0.26  0.80
## Year2000        0.0858    0.1462    0.59  0.56
## Year2001       -0.0283    0.1482   -0.19  0.85
## Year2002        0.0357    0.1542    0.23  0.82
## Year2003        0.0374    0.1459    0.26  0.80
## Year2004        0.1316    0.1509    0.87  0.38
```

```

## Year2005          0.0440      0.1408      0.31      0.75
## Year2006          0.0773      0.1359      0.57      0.57
## Year2007          0.1655      0.1357      1.22      0.22
## Year2008          0.0760      0.1373      0.55      0.58
## Year2009         -0.0361      0.1329     -0.27      0.79
## Year2010          0.0489      0.1324      0.37      0.71
## Year2011          0.0976      0.1332      0.73      0.46
## Year2012          0.0258      0.1342      0.19      0.85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.785
## Multiple R-squared:  0.0106, Adjusted R-squared:  0.0031
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 726 is an outlier with |weight| = 0 ( < 4.2e-05);
## 208 weights are ~= 1. The remaining 2172 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.199  0.868  0.952   0.918  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 1.038 1          1.019
## Year              1.038 16          1.001

```

Residuals from first author



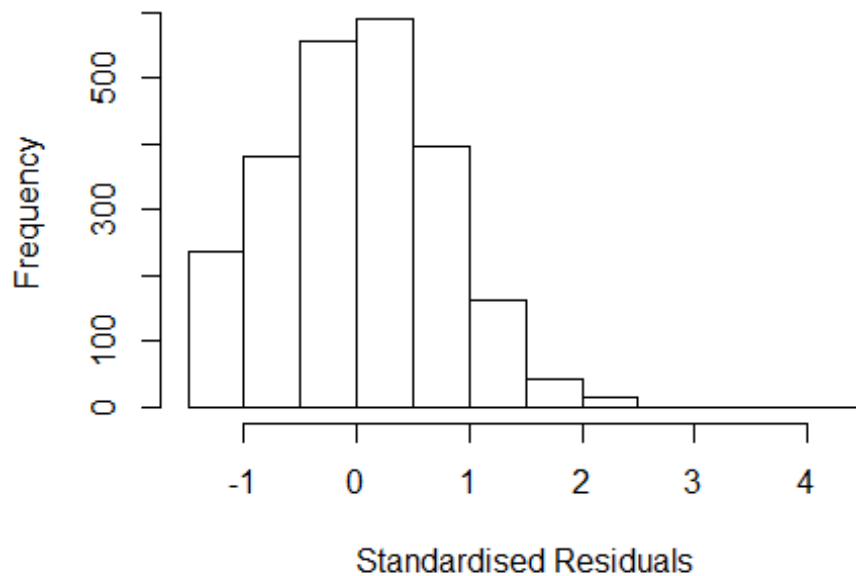
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 913  34247665578 5.284 2005    1202     4     4.321
## 1389 34250205270 3.822 2007    3316     1     2.738
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1658 -0.5296  0.0129  0.5288  4.3178
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9219    0.1252   7.37 2.4e-13 ***
## FirstAuthorFemale1 0.0771    0.0319   2.42  0.016 *
## Year1997      -0.1403    0.1522  -0.92  0.357
## Year1998       0.1668    0.1591   1.05  0.295
## Year1999      -0.0429    0.1583  -0.27  0.786
## Year2000       0.0859    0.1462   0.59  0.557
## Year2001      -0.0274    0.1486  -0.18  0.854
## Year2002       0.0360    0.1543   0.23  0.816
## Year2003       0.0360    0.1461   0.25  0.805
## Year2004       0.1301    0.1512   0.86  0.390
## Year2005       0.0442    0.1411   0.31  0.754
```

```

## Year2006          0.0776      0.1362      0.57      0.569
## Year2007          0.1652      0.1360      1.21      0.225
## Year2008          0.0757      0.1376      0.55      0.582
## Year2009         -0.0364      0.1332     -0.27      0.784
## Year2010          0.0497      0.1327      0.37      0.708
## Year2011          0.0988      0.1335      0.74      0.460
## Year2012          0.0255      0.1345      0.19      0.849
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.785
## Multiple R-squared:  0.0103, Adjusted R-squared:  0.00321
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 726 is an outlier with |weight| = 0 ( < 4.2e-05);
## 205 weights are ~ = 1. The remaining 2175 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.200  0.867   0.952   0.918   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 1.027 1          1.014
## Year            1.027 16          1.001

```

Residuals from last author



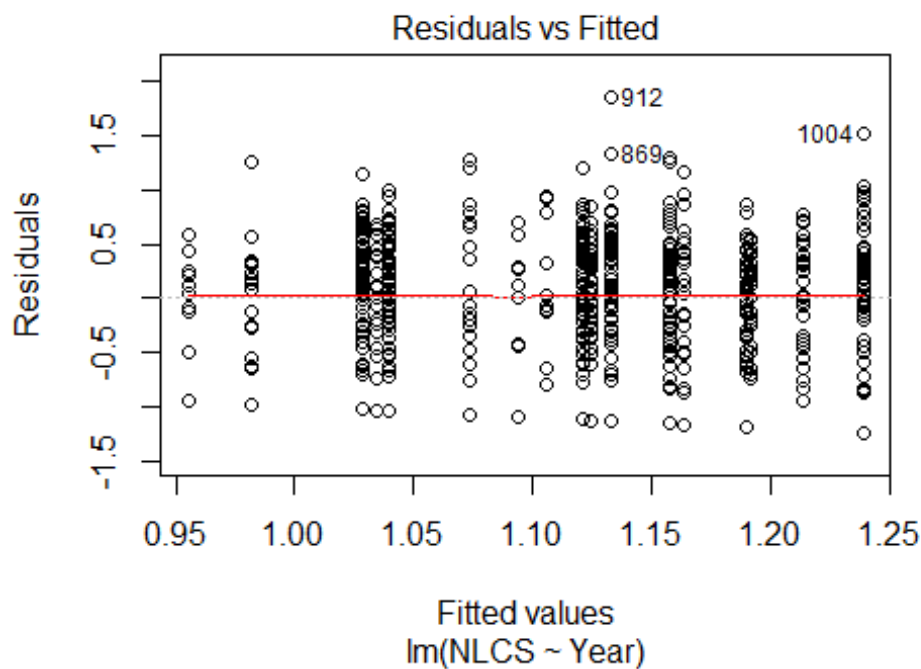
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 913  34247665578 5.284 2005    1202     4    4.321
## 1389 34250205270 3.822 2007    3316     1    2.738
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1676 -0.5266  0.0109  0.5279  4.3189
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9200     0.1249   7.37 2.4e-13 ***
## LastAuthorFemale1  0.0797     0.0318   2.51  0.012 *
## Year1997        -0.1385     0.1520  -0.91  0.363
## Year1998         0.1679     0.1587   1.06  0.290
## Year1999        -0.0383     0.1578  -0.24  0.808
## Year2000         0.0866     0.1462   0.59  0.554
## Year2001        -0.0285     0.1480  -0.19  0.847
## Year2002         0.0367     0.1540   0.24  0.812
## Year2003         0.0390     0.1458   0.27  0.789
## Year2004         0.1339     0.1504   0.89  0.374
## Year2005         0.0451     0.1406   0.32  0.748
```

```

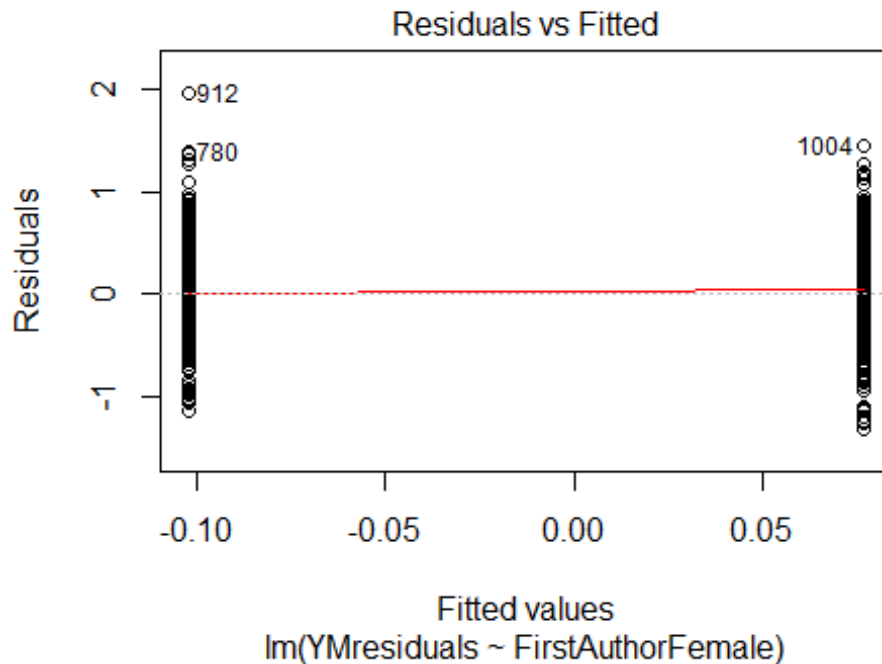
## Year2006          0.0792      0.1355      0.58      0.559
## Year2007          0.1675      0.1354      1.24      0.216
## Year2008          0.0779      0.1370      0.57      0.570
## Year2009         -0.0347      0.1327     -0.26      0.794
## Year2010          0.0501      0.1321      0.38      0.705
## Year2011          0.0986      0.1329      0.74      0.458
## Year2012          0.0276      0.1339      0.21      0.837
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.785
## Multiple R-squared:  0.0105, Adjusted R-squared:  0.00339
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 726 is an outlier with |weight| = 0 ( < 4.2e-05);
## 207 weights are ~ = 1. The remaining 2173 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.200  0.867   0.952   0.918   0.985   0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          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: 2381"
## [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
##   37   33   48   30   35   38   30   46   30   29   50   77   70   92   71
## 2011 2012
##   95   84
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
##      21      18      18      9      13      10      26      39      26      27      42      67      58      82      58
## 2011 2012
##      80      69
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      20      16      16      9      13      9      24      35      24      26      40      61      50      72      53
## 2011 2012
##      75      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
```

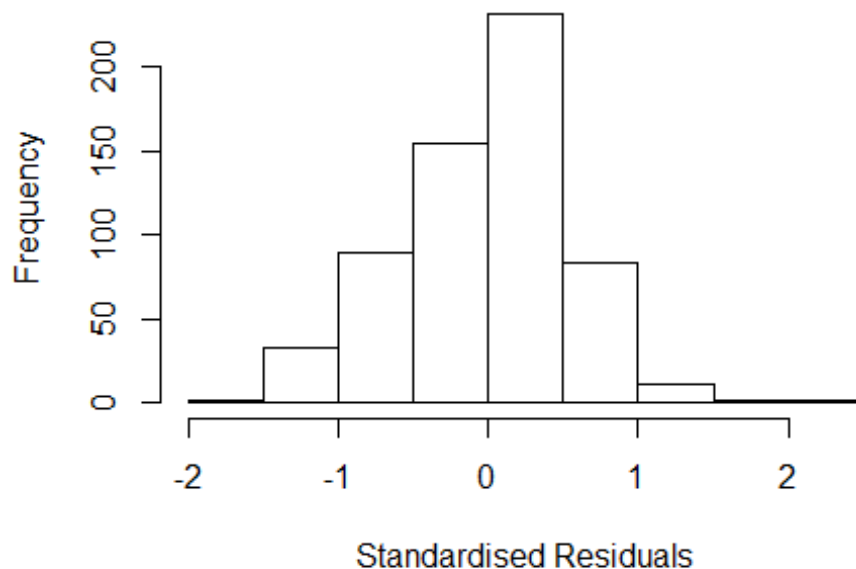


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.55, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 1.74900492766371"
## [1] "Male first author team size 2018 geometric mean: 1.34040805827917"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.55124227854333"
## [1] "Male last author team size 2018 geometric mean: 1.62234836962768"
##
## 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 1.980 1          1.407
## LastAuthorFemale  1.931 1          1.389
## UniqueAuthors    1.946 4          1.087
## Year             2.190 16          1.025
```


Residuals from first and last author and team size



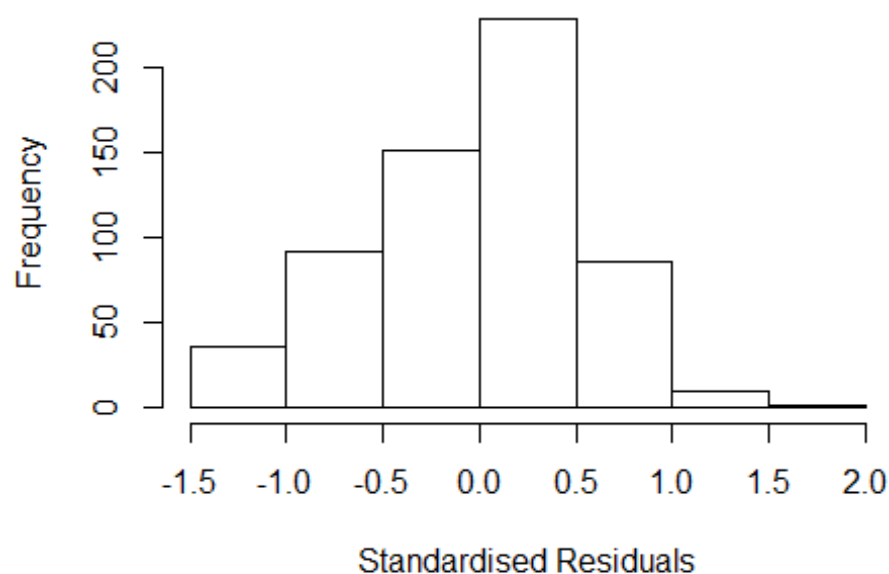
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5590 -0.3810  0.0675  0.3802  2.0752
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.80367    0.16443   4.89 1.3e-06 ***
## FirstAuthorFemale1 0.15979    0.06529   2.45  0.0147 *
## LastAuthorFemale1 0.11068    0.06403   1.73  0.0844 .
## UniqueAuthors2    0.14529    0.05852   2.48  0.0133 *
## UniqueAuthors3    0.11834    0.07849   1.51  0.1322
## UniqueAuthors4    0.27768    0.09429   2.95  0.0034 **
## UniqueAuthors5    0.37674    0.14245   2.64  0.0084 **
## Year1997          0.03953    0.19849   0.20  0.8422
## Year1998         -0.05773    0.20137  -0.29  0.7745
## Year1999          0.12638    0.25728   0.49  0.6235
```

```

## Year2000      0.14126      0.24082      0.59      0.5577
## Year2001     -0.03647      0.22442     -0.16      0.8710
## Year2002      0.17742      0.18368      0.97      0.3345
## Year2003      0.15321      0.18277      0.84      0.4022
## Year2004      0.06654      0.21004      0.32      0.7515
## Year2005      0.21572      0.18907      1.14      0.2544
## Year2006      0.12169      0.17733      0.69      0.4928
## Year2007      0.03315      0.18224      0.18      0.8557
## Year2008      0.14515      0.18614      0.78      0.4358
## Year2009      0.00331      0.17622      0.02      0.9850
## Year2010      0.11517      0.18156      0.63      0.5261
## Year2011      0.10812      0.17807      0.61      0.5440
## Year2012      0.12948      0.18052      0.72      0.4735
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.0939, Adjusted R-squared:  0.0597
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 569 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0993 0.8690 0.9510 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      1.65e-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 1.914 1      1.383
## LastAuthorFemale 1.827 1      1.351
## Year      1.215 16      1.006

```

Residuals from first and last author



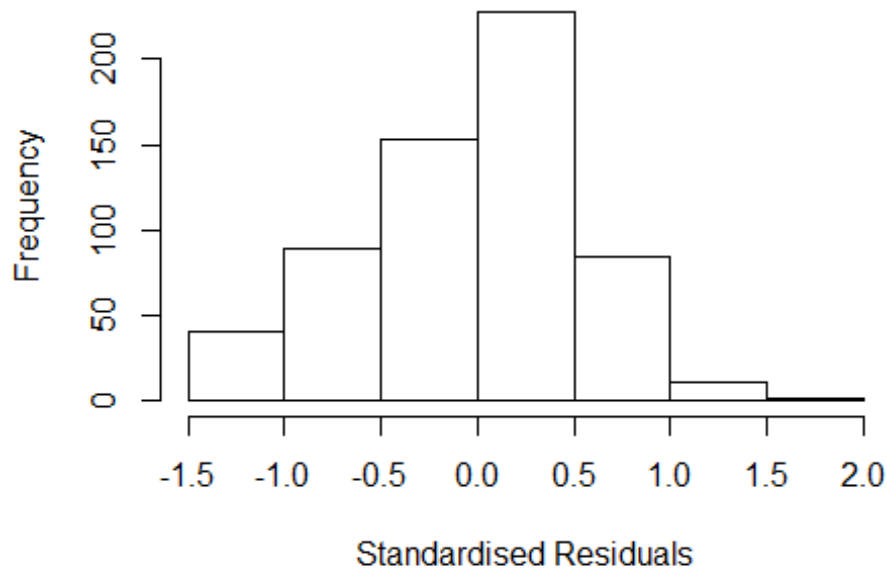
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3202 -0.3570 0.0438 0.3735 1.9906
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.85349 0.17072 5.00 7.6e-07 ***
## FirstAuthorFemale1 0.21414 0.06618 3.24 0.0013 **
## LastAuthorFemale1 0.06596 0.06404 1.03 0.3034
## Year1997 0.04958 0.20452 0.24 0.8085
## Year1998 -0.00649 0.21765 -0.03 0.9762
## Year1999 0.12599 0.24339 0.52 0.6049
## Year2000 0.10208 0.24432 0.42 0.6763
## Year2001 -0.04872 0.22531 -0.22 0.8289
## Year2002 0.18709 0.18512 1.01 0.3126
## Year2003 0.16708 0.18825 0.89 0.3752
## Year2004 0.10369 0.21216 0.49 0.6252
## Year2005 0.23328 0.19598 1.19 0.2344
```

```

## Year2006          0.11887    0.18545    0.64    0.5218
## Year2007          0.04487    0.18927    0.24    0.8127
## Year2008          0.16747    0.19567    0.86    0.3924
## Year2009          0.02962    0.18431    0.16    0.8724
## Year2010          0.11606    0.18834    0.62    0.5380
## Year2011          0.14286    0.18568    0.77    0.4420
## Year2012          0.18661    0.18675    1.00    0.3181
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.0635, Adjusted R-squared:  0.0347
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 549 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.145  0.861  0.947  0.898  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      1.65e-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 1.141 1      1.068
## Year              1.141 16      1.004

```

Residuals from first author



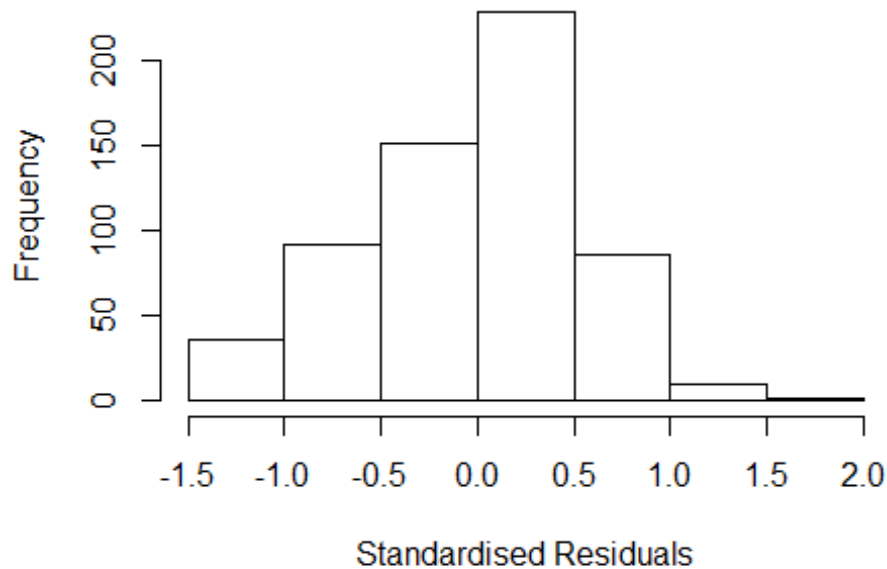
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3085 -0.3615 0.0426 0.3805 1.9866
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.858943 0.169470 5.07 5.4e-07 ***
## FirstAuthorFemale1 0.256483 0.051362 4.99 7.8e-07 ***
## Year1997 0.051287 0.203426 0.25 0.80
## Year1998 -0.000447 0.215732 0.00 1.00
## Year1999 0.133729 0.238270 0.56 0.57
## Year2000 0.108269 0.244342 0.44 0.66
## Year2001 -0.047857 0.223923 -0.21 0.83
## Year2002 0.191581 0.183636 1.04 0.30
## Year2003 0.167547 0.186289 0.90 0.37
## Year2004 0.110418 0.210985 0.52 0.60
## Year2005 0.238228 0.194683 1.22 0.22
## Year2006 0.127149 0.182850 0.70 0.49
```

```

## Year2007      0.051090    0.187584    0.27    0.79
## Year2008      0.173043    0.194323    0.89    0.37
## Year2009      0.035514    0.182686    0.19    0.85
## Year2010      0.121791    0.186521    0.65    0.51
## Year2011      0.141438    0.184172    0.77    0.44
## Year2012      0.193099    0.185332    1.04    0.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.539
## Multiple R-squared:  0.0618, Adjusted R-squared:  0.0347
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 550 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.146  0.856  0.944  0.898  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      1.65e-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 1.101 1      1.049
## Year      1.101 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
## -1.306 -0.388 0.053 0.382 1.928
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.91378 0.18182 5.03 6.7e-07 ***
## LastAuthorFemale1 0.19649 0.05036 3.90 0.00011 ***
## Year1997 0.03572 0.21955 0.16 0.87081
## Year1998 -0.02463 0.22504 -0.11 0.91287
## Year1999 0.11354 0.26699 0.43 0.67079
## Year2000 0.08318 0.25689 0.32 0.74620
## Year2001 -0.03927 0.23693 -0.17 0.86841
## Year2002 0.18368 0.19784 0.93 0.35356
## Year2003 0.17502 0.20138 0.87 0.38514
## Year2004 0.10994 0.22688 0.48 0.62816
## Year2005 0.20895 0.20734 1.01 0.31399
## Year2006 0.11023 0.19759 0.56 0.57716
```

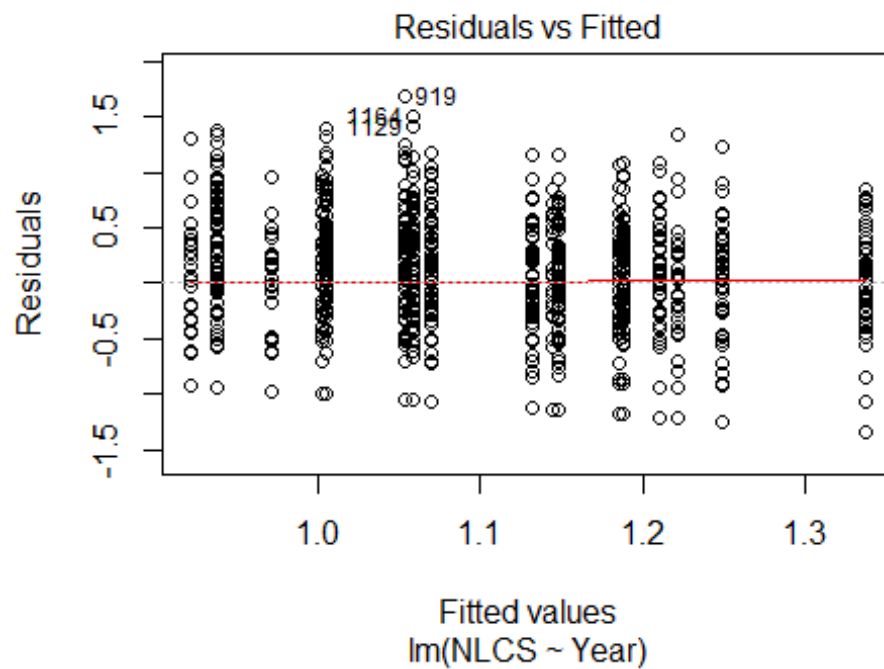
```

## Year2007      0.04863    0.20023    0.24  0.80818
## Year2008      0.13372    0.20605    0.65  0.51662
## Year2009      0.00709    0.19564    0.04  0.97112
## Year2010      0.12377    0.20155    0.61  0.53940
## Year2011      0.14566    0.19763    0.74  0.46140
## Year2012      0.19587    0.19801    0.99  0.32296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared:  0.0434, Adjusted R-squared:  0.0157
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 51 weights are ~= 1. The remaining 554 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.200  0.860  0.948  0.902  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.65e-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: 605"
## [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
##   27   32   28   34   58   48   51   48   48   48   89   96   79  109  124
## 2011 2012
##   94  135
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   26   28   23   30   48   38   44   44   45   43   78   78   70   97  107
## 2011 2012

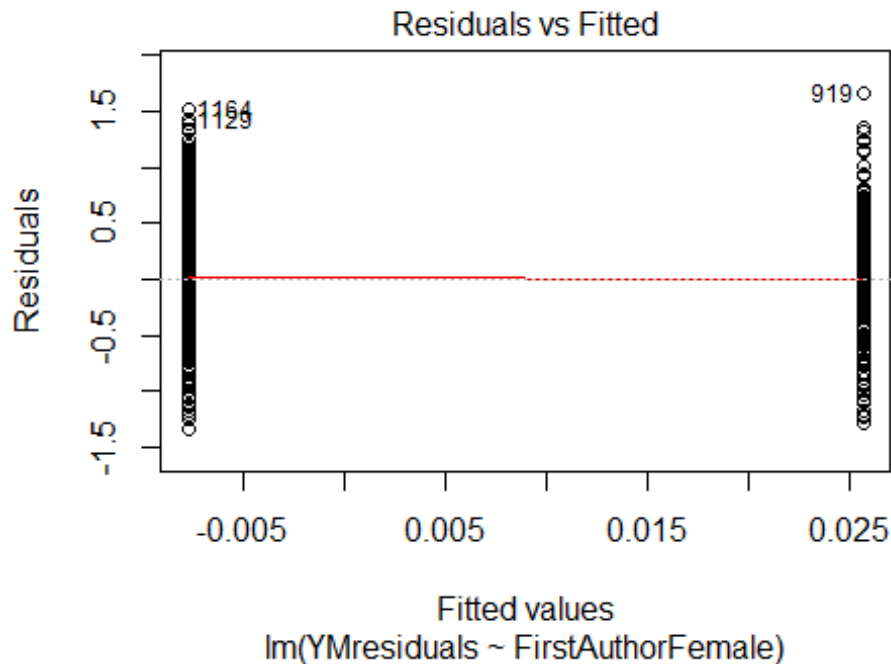
```



```
## 83 122
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 26 27 22 30 47 34 42 43 42 42 72 78 66 93 104
## 2011 2012
## 82 121
## [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
```

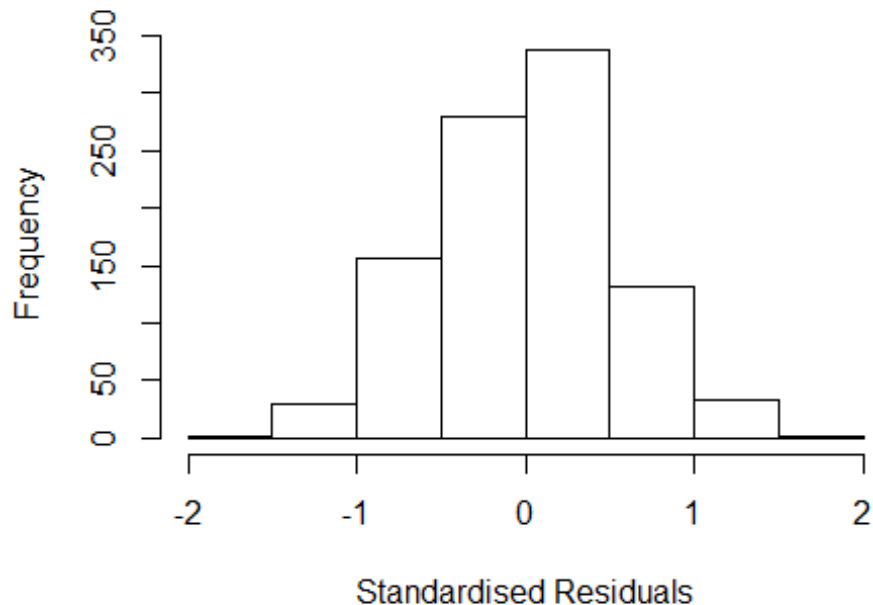


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0087, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 1.7495638687183"
## [1] "Male first author team size 2018 geometric mean: 1.99993957964021"
##
## 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] "Female last author team size 2018 geometric mean: 1.76117226532656"
## [1] "Male last author team size 2018 geometric mean: 1.94987783504709"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, 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 1.703 1          1.305
## LastAuthorFemale  1.794 1          1.339
## UniqueAuthors    1.243 4          1.028
## Year             1.471 16          1.012
```

Residuals from first and last author and team size



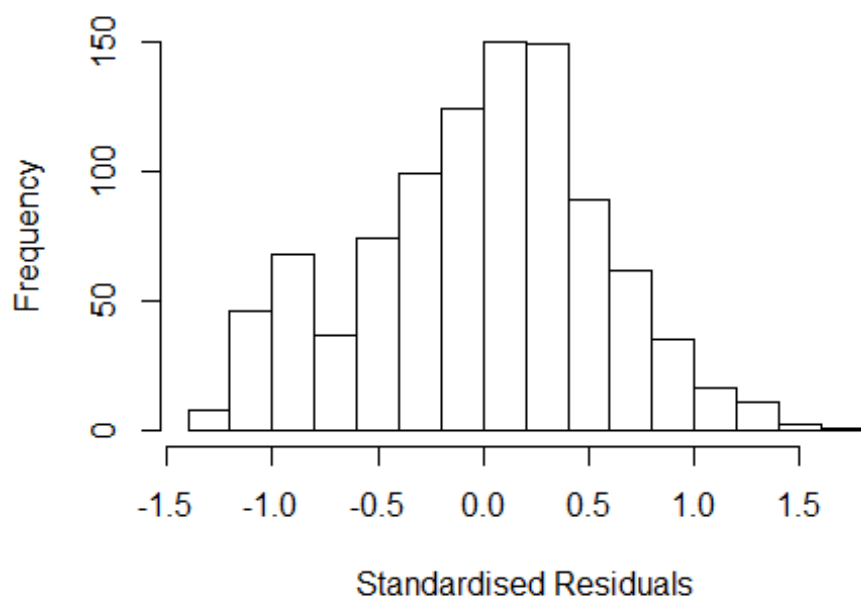
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5236 -0.3872  0.0349  0.3687  1.7632
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9125    0.1052   8.67 < 2e-16 ***
## FirstAuthorFemale1 -0.0179    0.0559  -0.32  0.7497
## LastAuthorFemale1 -0.0165    0.0537  -0.31  0.7589
## UniqueAuthors2    0.2393    0.0456   5.25 1.9e-07 ***
## UniqueAuthors3    0.2337    0.0561   4.17 3.4e-05 ***
## UniqueAuthors4    0.3055    0.1260   2.42  0.0155 *
## UniqueAuthors5    0.4749    0.0949   5.00 6.7e-07 ***
## Year1997         -0.0239    0.1479  -0.16  0.8715
## Year1998          0.2048    0.1642   1.25  0.2124
## Year1999          0.2264    0.1336   1.69  0.0905 .
```

```

## Year2000          0.1598      0.1215      1.32      0.1888
## Year2001          0.2208      0.1411      1.56      0.1180
## Year2002          0.2519      0.1298      1.94      0.0526 .
## Year2003          0.3237      0.1284      2.52      0.0118 *
## Year2004          0.4061      0.1258      3.23      0.0013 **
## Year2005          0.0530      0.1374      0.39      0.7000
## Year2006          0.1690      0.1149      1.47      0.1419
## Year2007          0.2093      0.1111      1.88      0.0599 .
## Year2008          0.0710      0.1228      0.58      0.5635
## Year2009          0.0643      0.1185      0.54      0.5874
## Year2010          0.0210      0.1156      0.18      0.8560
## Year2011          0.0565      0.1173      0.48      0.6300
## Year2012         -0.0326      0.1212     -0.27      0.7882
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.0878, Adjusted R-squared:  0.0667
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 896 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.271  0.854  0.949   0.904   0.987   0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          1.03e-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 1.650 1          1.284
## LastAuthorFemale  1.706 1          1.306
## Year              1.210 16          1.006

```

Residuals from first and last author



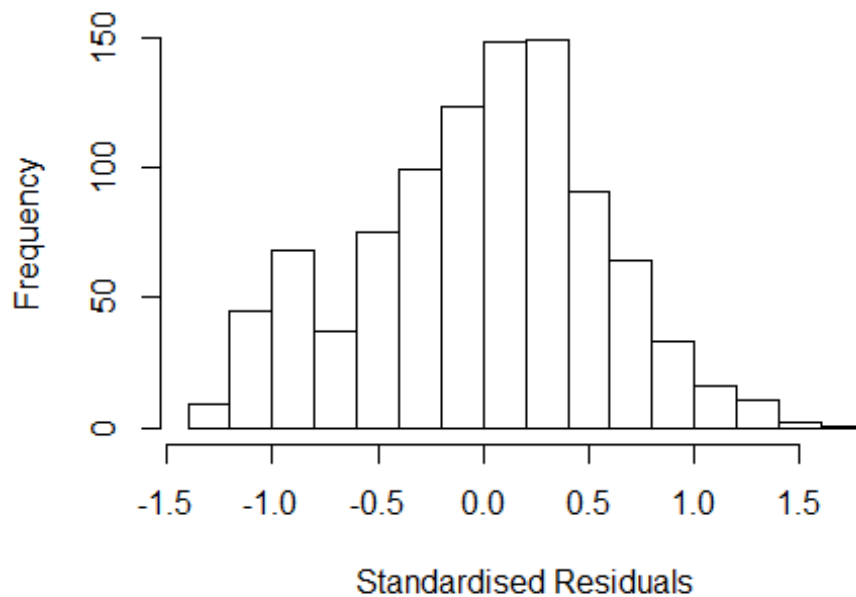
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.369 -0.362 0.034 0.366 1.651
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0059 0.1034 9.73 <2e-16 ***
## FirstAuthorFemale1 0.0136 0.0559 0.24 0.8084
## LastAuthorFemale1 -0.0575 0.0533 -1.08 0.2817
## Year1997 -0.0646 0.1521 -0.42 0.6712
## Year1998 0.2145 0.1729 1.24 0.2150
## Year1999 0.2295 0.1415 1.62 0.1053
## Year2000 0.1446 0.1233 1.17 0.2411
## Year2001 0.1895 0.1350 1.40 0.1607
## Year2002 0.2724 0.1271 2.14 0.0324 *
## Year2003 0.3238 0.1263 2.56 0.0105 *
## Year2004 0.4071 0.1252 3.25 0.0012 **
## Year2005 0.0286 0.1352 0.21 0.8325
```

```

## Year2006          0.1580      0.1146      1.38      0.1681
## Year2007          0.2231      0.1107      2.02      0.0440 *
## Year2008          0.1097      0.1244      0.88      0.3780
## Year2009          0.0832      0.1185      0.70      0.4828
## Year2010          0.0210      0.1156      0.18      0.8561
## Year2011          0.0709      0.1212      0.59      0.5583
## Year2012         -0.0334      0.1238     -0.27      0.7873
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.549
## Multiple R-squared:  0.0431, Adjusted R-squared:  0.025
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 100 weights are ~= 1. The remaining 871 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.345  0.848  0.949  0.898  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      1.03e-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 1.084 1          1.041
## Year              1.084 16          1.003

```

Residuals from first author



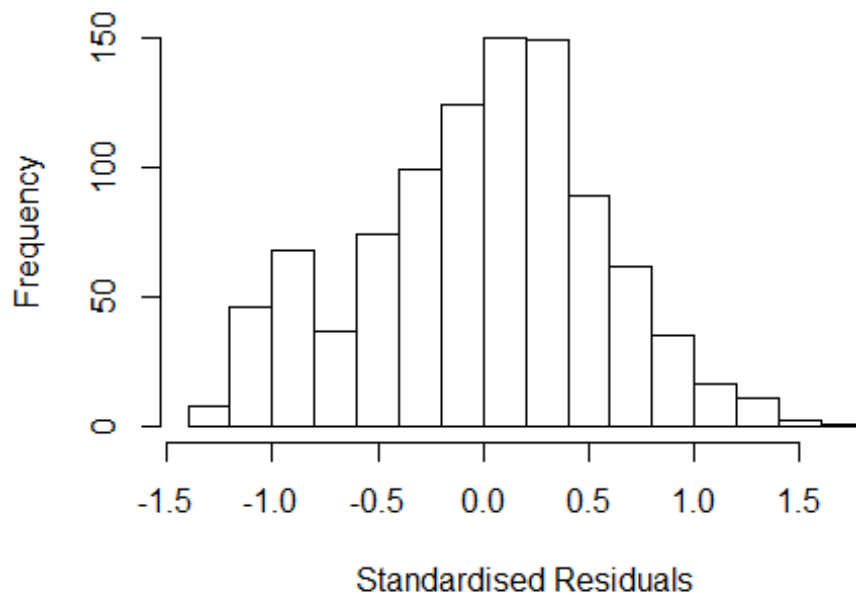
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3789 -0.3761 0.0352 0.3698 1.6663
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0060 0.1042 9.66 <2e-16 ***
## FirstAuthorFemale1 -0.0269 0.0456 -0.59 0.5547
## Year1997 -0.0807 0.1512 -0.53 0.5937
## Year1998 0.2025 0.1736 1.17 0.2436
## Year1999 0.2238 0.1430 1.57 0.1179
## Year2000 0.1355 0.1228 1.10 0.2703
## Year2001 0.1794 0.1343 1.34 0.1820
## Year2002 0.2583 0.1264 2.04 0.0413 *
## Year2003 0.3065 0.1254 2.44 0.0147 *
## Year2004 0.3999 0.1253 3.19 0.0015 **
## Year2005 0.0162 0.1346 0.12 0.9041
## Year2006 0.1428 0.1133 1.26 0.2082
```

```

## Year2007          0.2120      0.1104      1.92      0.0551 .
## Year2008          0.0982      0.1238      0.79      0.4278
## Year2009          0.0677      0.1175      0.58      0.5645
## Year2010          0.0129      0.1161      0.11      0.9114
## Year2011          0.0591      0.1215      0.49      0.6267
## Year2012         -0.0440      0.1240     -0.35      0.7227
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.548
## Multiple R-squared:  0.0421, Adjusted R-squared:  0.025
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 872 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.335  0.849  0.950  0.898  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      1.03e-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 1.121 1      1.059
## Year             1.121 16      1.004

```


Residuals from last author



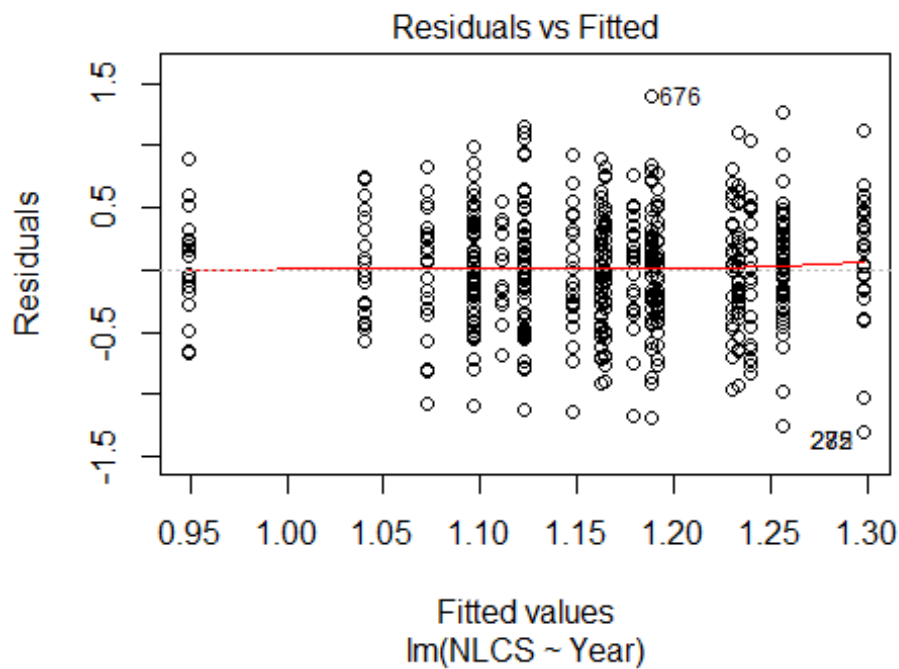
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3688 -0.3617  0.0338  0.3674  1.6477
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0114     0.0994  10.18  <2e-16 ***
## LastAuthorFemale1 -0.0486     0.0436  -1.11  0.2657
## Year1997         -0.0666     0.1517  -0.44  0.6607
## Year1998          0.2135     0.1730   1.23  0.2174
## Year1999          0.2276     0.1418   1.61  0.1086
## Year2000          0.1421     0.1226   1.16  0.2466
## Year2001          0.1877     0.1346   1.39  0.1636
## Year2002          0.2693     0.1256   2.14  0.0323 *
## Year2003          0.3215     0.1261   2.55  0.0110 *
## Year2004          0.4060     0.1253   3.24  0.0012 **
## Year2005          0.0268     0.1351   0.20  0.8427
## Year2006          0.1563     0.1147   1.36  0.1732
```

```

## Year2007          0.2215      0.1105      2.01      0.0452 *
## Year2008          0.1084      0.1240      0.87      0.3824
## Year2009          0.0809      0.1179      0.69      0.4928
## Year2010          0.0203      0.1159      0.18      0.8607
## Year2011          0.0697      0.1212      0.57      0.5655
## Year2012         -0.0346      0.1239     -0.28      0.7802
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.548
## Multiple R-squared:  0.0431, Adjusted R-squared:  0.026
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 98 weights are ~= 1. The remaining 873 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.346  0.848  0.949  0.898  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      1.03e-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: 971"
## [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
##   29   29   33   34   34   38   35   28   28   25   40   51   45   62   58
## 2011 2012
##   65   63
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   23   19   16   21   22   24   34   28   26   22   37   43   37   57   51
## 2011 2012

```

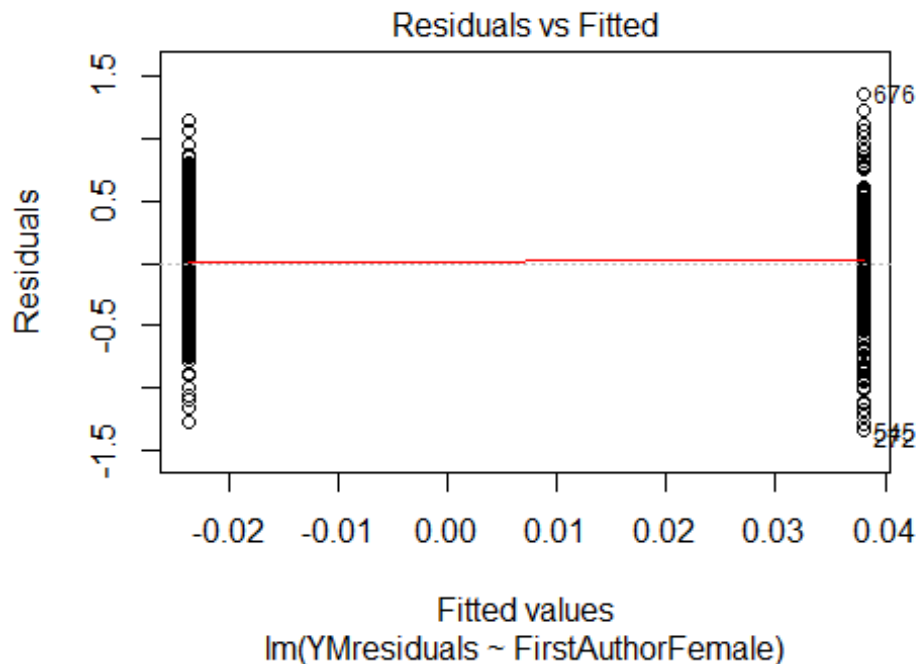
```
## 59 58
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 21 16 15 18 20 20 32 27 24 20 35 39 30 57 48
## 2011 2012
## 56 52
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6
```



```
##
## 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: 2.82431516268199"
## [1] "Male first author team size 2018 geometric mean: 2.84036118951649"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

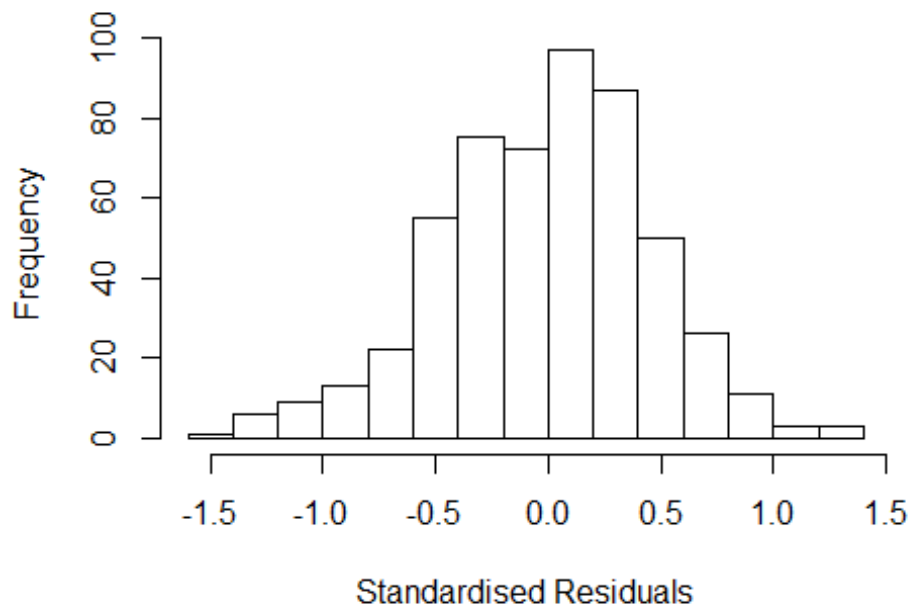
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 480, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.72785187401466"
## [1] "Male last author team size 2018 geometric mean: 2.95855888500566"

## Warning in wilcox.test.default(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 = 0.5
## 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 1.493  1      1.222
## LastAuthorFemale  1.297  1      1.139
## UniqueAuthors    1.846  4      1.080
## Year              2.707 16      1.032
```

Residuals from first and last author and team size



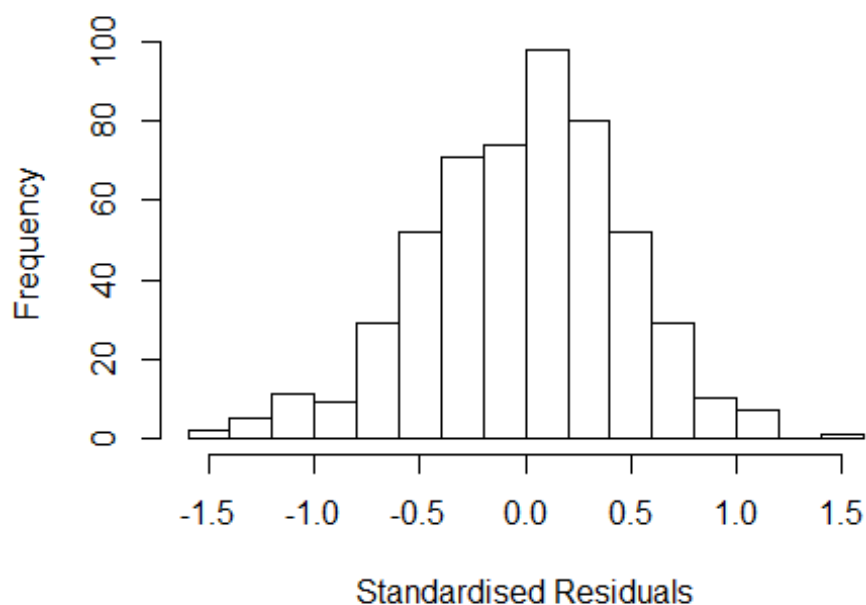
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5096 -0.3169 0.0281 0.2997 1.3645
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1975 0.1177 10.17 < 2e-16 ***
## FirstAuthorFemale1 -0.0424 0.0502 -0.84 0.39923
## LastAuthorFemale1 -0.0596 0.0471 -1.27 0.20626
## UniqueAuthors2 0.1901 0.0536 3.55 0.00042 ***
## UniqueAuthors3 0.2158 0.0658 3.28 0.00111 **
## UniqueAuthors4 0.2229 0.0805 2.77 0.00585 **
## UniqueAuthors5 0.2155 0.0714 3.02 0.00269 **
## Year1997 -0.2449 0.1612 -1.52 0.12940
## Year1998 -0.1437 0.1438 -1.00 0.31785
## Year1999 -0.2349 0.1547 -1.52 0.12947
```

```

## Year2000          -0.2572      0.1621    -1.59   0.11313
## Year2001          -0.0664      0.1565    -0.42   0.67166
## Year2002          -0.0185      0.1300    -0.14   0.88718
## Year2003           0.1389      0.1508     0.92   0.35734
## Year2004          -0.0551      0.1494    -0.37   0.71240
## Year2005          -0.1131      0.1409    -0.80   0.42240
## Year2006          -0.0631      0.1286    -0.49   0.62362
## Year2007          -0.1275      0.1277    -1.00   0.31861
## Year2008          -0.0983      0.1450    -0.68   0.49820
## Year2009          -0.0470      0.1251    -0.38   0.70717
## Year2010          -0.1013      0.1275    -0.79   0.42739
## Year2011          -0.1771      0.1341    -1.32   0.18716
## Year2012          -0.2072      0.1325    -1.56   0.11842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.463
## Multiple R-squared:  0.0823, Adjusted R-squared:  0.0425
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 484 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.265  0.882  0.953  0.906  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 1.462 1      1.209
## LastAuthorFemale  1.241 1      1.114
## Year              1.489 16      1.013

```

Residuals from first and last author



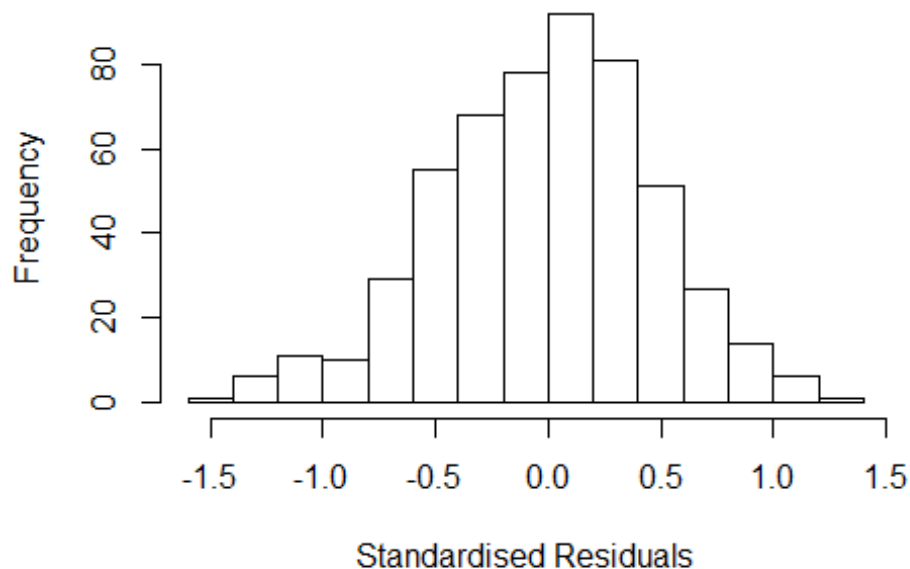
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4705 -0.3175  0.0173  0.3122  1.4088
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.31721    0.12412   10.61  <2e-16 ***
## FirstAuthorFemale1 -0.03540    0.05244   -0.68   0.500
## LastAuthorFemale1 -0.06681    0.04800   -1.39   0.165
## Year1997          -0.30273    0.17141   -1.77   0.078 .
## Year1998          -0.15084    0.14883   -1.01   0.311
## Year1999          -0.23723    0.15878   -1.49   0.136
## Year2000          -0.21406    0.17016   -1.26   0.209
## Year2001          -0.06921    0.16885   -0.41   0.682
## Year2002           0.00219    0.14072    0.02   0.988
## Year2003           0.15331    0.16632    0.92   0.357
## Year2004          -0.05628    0.16193   -0.35   0.728
## Year2005          -0.08530    0.15347   -0.56   0.579
```

```

## Year2006      -0.07003    0.14143   -0.50    0.621
## Year2007      -0.09320    0.13768   -0.68    0.499
## Year2008      -0.08527    0.15221   -0.56    0.576
## Year2009       0.02519    0.13200    0.19    0.849
## Year2010      -0.07718    0.13858   -0.56    0.578
## Year2011      -0.14829    0.14104   -1.05    0.294
## Year2012      -0.17151    0.14153   -1.21    0.226
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.469
## Multiple R-squared:  0.0465, Adjusted R-squared:  0.0129
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 485 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.305  0.877   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      1.89e-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 1.349  1      1.161
## Year              1.349 16      1.009

```


Residuals from first author



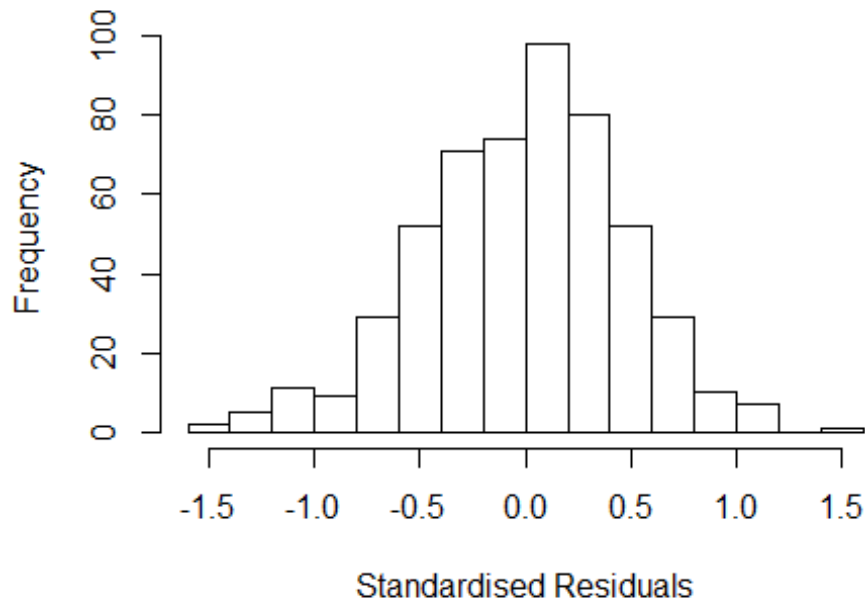
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4370 -0.3124 0.0247 0.3109 1.3619
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30610 0.12185 10.72 <2e-16 ***
## FirstAuthorFemale1 -0.05911 0.05029 -1.18 0.240
## Year1997 -0.31378 0.16947 -1.85 0.065 .
## Year1998 -0.17719 0.14318 -1.24 0.216
## Year1999 -0.25407 0.15695 -1.62 0.106
## Year2000 -0.22484 0.16914 -1.33 0.184
## Year2001 -0.07087 0.16999 -0.42 0.677
## Year2002 -0.00883 0.13873 -0.06 0.949
## Year2003 0.13093 0.16246 0.81 0.421
## Year2004 -0.07802 0.15856 -0.49 0.623
## Year2005 -0.09525 0.15247 -0.62 0.532
## Year2006 -0.08010 0.13973 -0.57 0.567
```

```

## Year2007          -0.11381    0.13539   -0.84    0.401
## Year2008          -0.10793    0.15050   -0.72    0.474
## Year2009           0.01238    0.12989    0.10    0.924
## Year2010          -0.08597    0.13798   -0.62    0.534
## Year2011          -0.15470    0.13912   -1.11    0.267
## Year2012          -0.19079    0.13935   -1.37    0.172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.472
## Multiple R-squared:  0.0418, Adjusted R-squared:  0.00996
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 494 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.334  0.881  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      1.89e-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 1.138 1          1.067
## Year              1.138 16          1.004

```

Residuals from last author



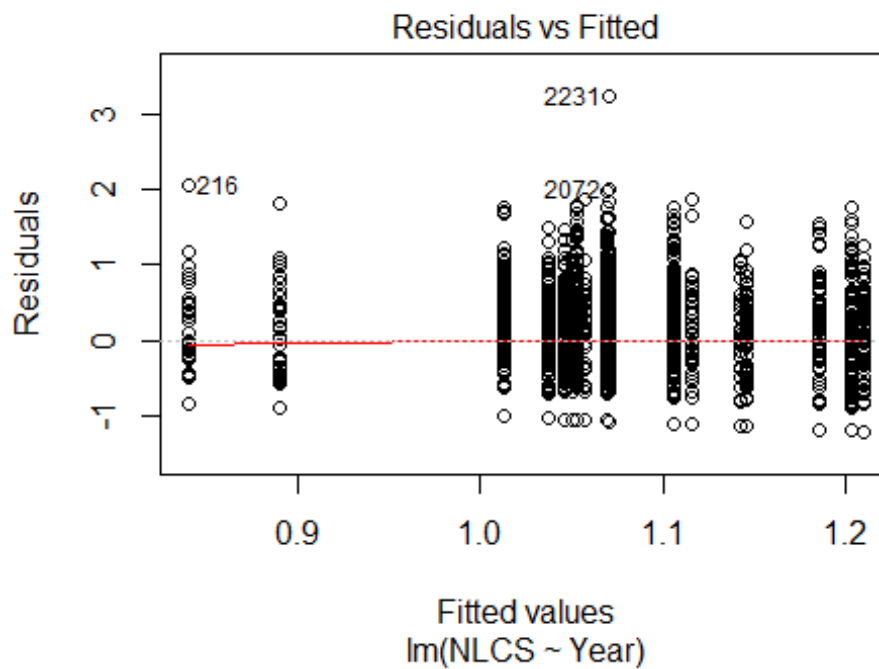
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4583 -0.3277 0.0276 0.3106 1.4387
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30078 0.11914 10.92 <2e-16 ***
## LastAuthorFemale1 -0.07804 0.04585 -1.70 0.089 .
## Year1997 -0.29911 0.17003 -1.76 0.079 .
## Year1998 -0.15088 0.14822 -1.02 0.309
## Year1999 -0.23412 0.15768 -1.48 0.138
## Year2000 -0.20501 0.16693 -1.23 0.220
## Year2001 -0.06152 0.16606 -0.37 0.711
## Year2002 0.00723 0.13835 0.05 0.958
## Year2003 0.15755 0.16403 0.96 0.337
## Year2004 -0.05590 0.16052 -0.35 0.728
## Year2005 -0.08086 0.15245 -0.53 0.596
## Year2006 -0.06997 0.14020 -0.50 0.618
```

```

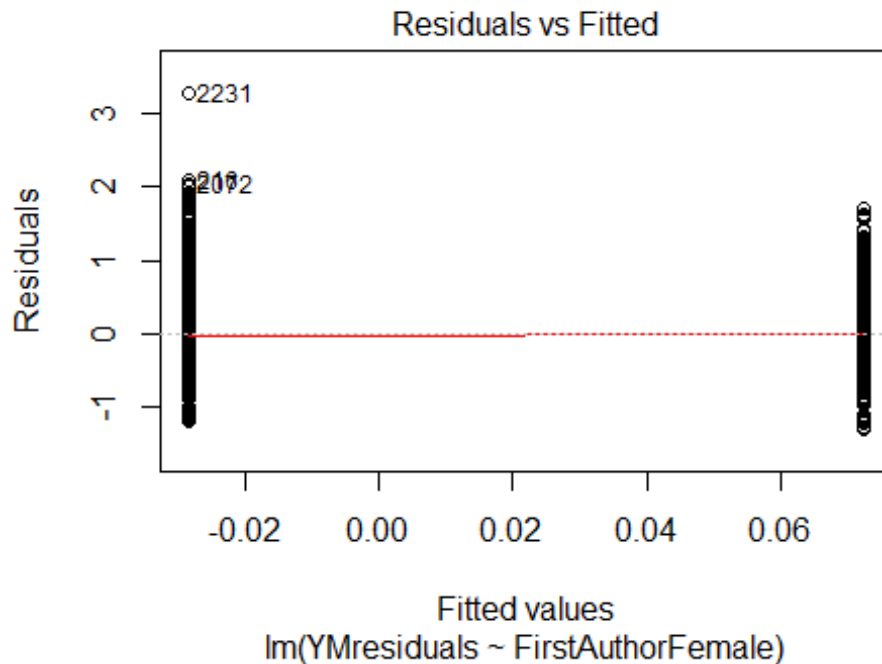
## Year2007          -0.09147      0.13682    -0.67      0.504
## Year2008          -0.08537      0.15086    -0.57      0.572
## Year2009           0.01970      0.13009      0.15      0.880
## Year2010          -0.07946      0.13694    -0.58      0.562
## Year2011          -0.15511      0.13893    -1.12      0.265
## Year2012          -0.17255      0.14038    -1.23      0.220
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0454, Adjusted R-squared:  0.0137
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 488 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.321  0.881  0.951  0.907  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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: 530"
## [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
##   59   59   74   53   65   66   95  104  137   85  143  128  206  220  276
## 2011 2012
##  289  247
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   46   49   57   42   36   45   79   85  117   74  129  114  186  188  239
## 2011 2012

```

```
## 249 220
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 45 49 57 40 36 44 77 84 116 74 125 111 181 187 233
## 2011 2012
## 244 216
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 21, df = 16, p-value = 0.2
```

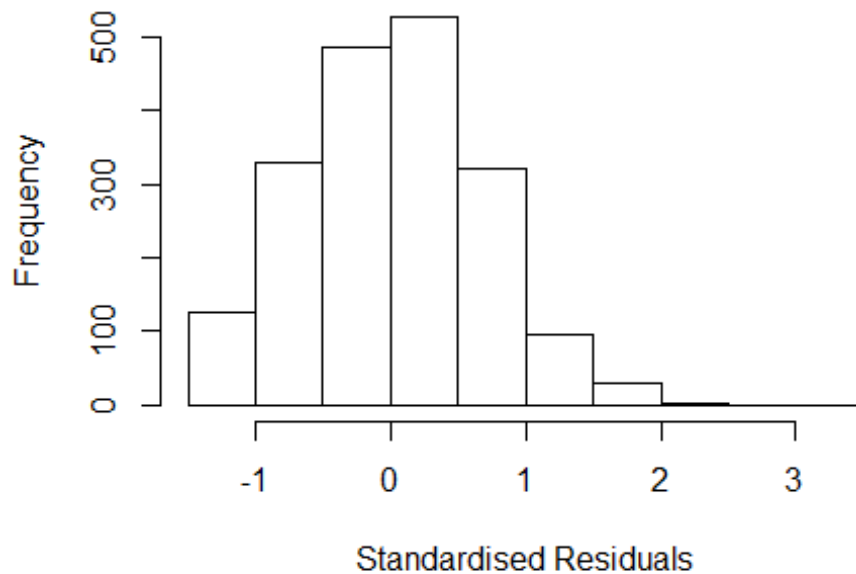


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.2, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 1.16211120097076"
## [1] "Male first author team size 2018 geometric mean: 1.15728012380276"
##
## 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] "Female last author team size 2018 geometric mean: 1.14983404237011"
## [1] "Male last author team size 2018 geometric mean: 1.16392590847024"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5400, 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 3.690 1          1.921
## LastAuthorFemale  3.726 1          1.930
## UniqueAuthors    1.247 4          1.028
## Year             1.187 16          1.005
```

Residuals from first and last author and team size



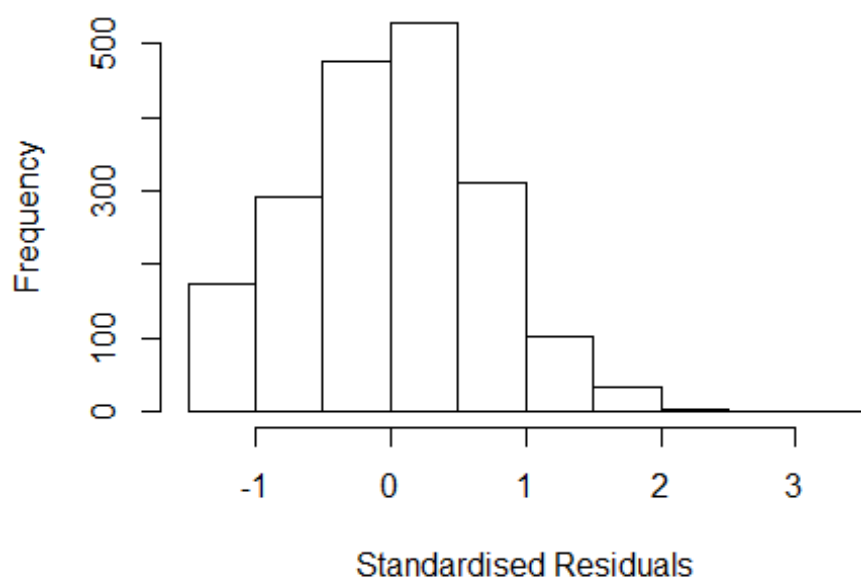
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 2231 79952786437 4.31 2011      1211      3      3.232
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.486 -0.468  0.012  0.463  3.232
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0231     0.1031   9.92 < 2e-16 ***
## FirstAuthorFemale1  0.0722     0.0650   1.11  0.26677
## LastAuthorFemale1  0.0428     0.0641   0.67  0.50418
## UniqueAuthors2     0.0935     0.0485   1.93  0.05395 .
## UniqueAuthors3     0.3033     0.0921   3.29  0.00101 **
## UniqueAuthors4     0.3236     0.2540   1.27  0.20282
## UniqueAuthors5     0.5591     0.1532   3.65  0.00027 ***
## Year1997          -0.0026     0.1673  -0.02  0.98761
## Year1998           0.0332     0.1360   0.24  0.80736
## Year1999          -0.2754     0.1481  -1.86  0.06305 .
```

```

## Year2000          0.0831      0.1427      0.58  0.56033
## Year2001         -0.1840      0.1507     -1.22  0.22223
## Year2002          0.1182      0.1339      0.88  0.37764
## Year2003          0.1343      0.1239      1.08  0.27855
## Year2004          0.0968      0.1218      0.79  0.42704
## Year2005          0.0587      0.1263      0.46  0.64225
## Year2006         -0.0357      0.1171     -0.30  0.76054
## Year2007         -0.0400      0.1201     -0.33  0.73890
## Year2008          0.0108      0.1176      0.09  0.92689
## Year2009         -0.0741      0.1145     -0.65  0.51758
## Year2010         -0.0289      0.1124     -0.26  0.79751
## Year2011         -0.0381      0.1128     -0.34  0.73539
## Year2012         -0.0733      0.1168     -0.63  0.53043
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.687
## Multiple R-squared:  0.0295, Adjusted R-squared:  0.0182
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 1693 is an outlier with |weight| = 0 ( < 5.2e-05);
## 184 weights are ~= 1. The remaining 1734 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.327  0.856  0.949   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      5.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.615 1 1.901
## LastAuthorFemale 3.619 1 1.902
## Year 1.048 16 1.001

```


Residuals from first and last author



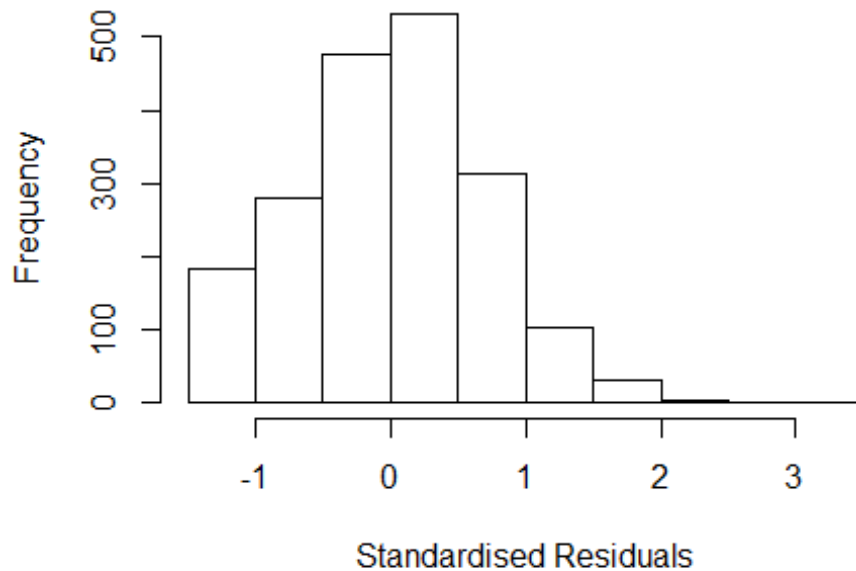
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 2231 79952786437 4.31 2011      1211      3      3.305
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2940 -0.4720  0.0152  0.4704  3.3046
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.02792    0.10341    9.94  <2e-16 ***
## FirstAuthorFemale1  0.05451    0.06469    0.84   0.400
## LastAuthorFemale1  0.07273    0.06358    1.14   0.253
## Year1997          0.00536    0.16709    0.03   0.974
## Year1998          0.04577    0.13593    0.34   0.736
## Year1999         -0.28155    0.14757   -1.91   0.057 .
## Year2000          0.10202    0.14566    0.70   0.484
## Year2001         -0.18365    0.15120   -1.21   0.225
## Year2002          0.12403    0.13424    0.92   0.356
## Year2003          0.13888    0.12414    1.12   0.263
## Year2004          0.11827    0.12222    0.97   0.333
## Year2005          0.07255    0.12639    0.57   0.566
```

```

## Year2006      -0.02895    0.11712   -0.25    0.805
## Year2007      -0.01776    0.11979   -0.15    0.882
## Year2008       0.02406    0.11771    0.20    0.838
## Year2009      -0.04922    0.11454   -0.43    0.667
## Year2010      -0.01169    0.11243   -0.10    0.917
## Year2011      -0.02255    0.11273   -0.20    0.841
## Year2012      -0.05549    0.11704   -0.47    0.635
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.686
## Multiple R-squared:  0.0198, Adjusted R-squared:  0.0105
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1693 is an outlier with |weight| = 0 ( < 5.2e-05);
## 159 weights are ~= 1. The remaining 1759 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.302  0.858  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      5.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.03 1          1.015
## Year              1.03 16          1.001

```

Residuals from first author



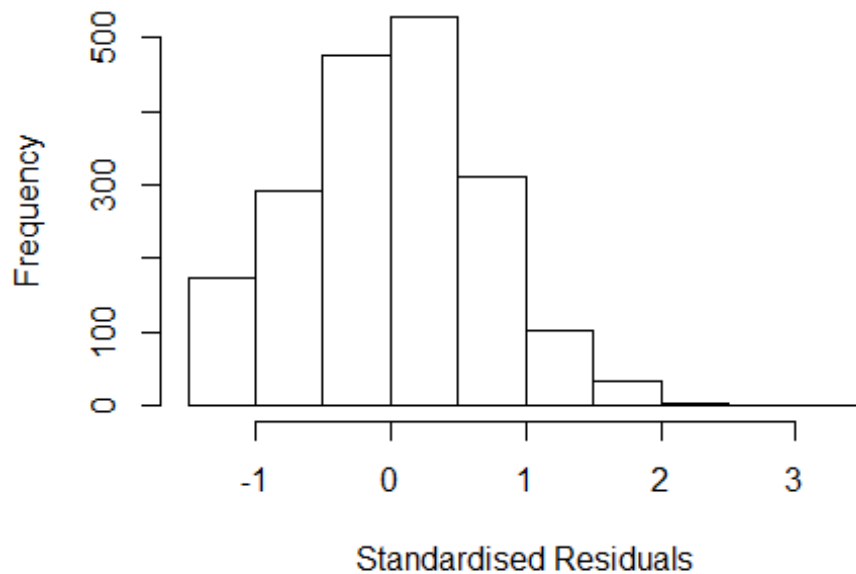
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 2231 79952786437 4.31 2011      1211      3      3.305
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2883 -0.4668  0.0175  0.4708  3.3012
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03213    0.10432   9.89 < 2e-16 ***
## FirstAuthorFemale1 0.11727    0.03448   3.40 0.00069 ***
## Year1997        0.00659    0.16803   0.04 0.96873
## Year1998        0.04518    0.13657   0.33 0.74079
## Year1999       -0.28268    0.14797  -1.91 0.05624 .
## Year2000        0.09488    0.14574   0.65 0.51513
## Year2001       -0.18466    0.15200  -1.21 0.22459
## Year2002        0.12754    0.13466   0.95 0.34368
## Year2003        0.13885    0.12481   1.11 0.26605
## Year2004        0.11614    0.12283   0.95 0.34450
## Year2005        0.07506    0.12702   0.59 0.55463
## Year2006       -0.02841    0.11786  -0.24 0.80955
```

```

## Year2007      -0.01932    0.12051   -0.16  0.87263
## Year2008      0.02590    0.11841    0.22  0.82689
## Year2009     -0.04863    0.11531   -0.42  0.67329
## Year2010     -0.01272    0.11325   -0.11  0.91059
## Year2011     -0.02330    0.11362   -0.21  0.83758
## Year2012     -0.05793    0.11789   -0.49  0.62321
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.687
## Multiple R-squared:  0.0191, Adjusted R-squared:  0.0103
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1693 is an outlier with |weight| = 0 ( < 5.2e-05);
## 157 weights are ~= 1. The remaining 1761 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.305  0.859  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      5.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.032 1          1.016
## Year            1.032 16          1.001

```

Residuals from last author



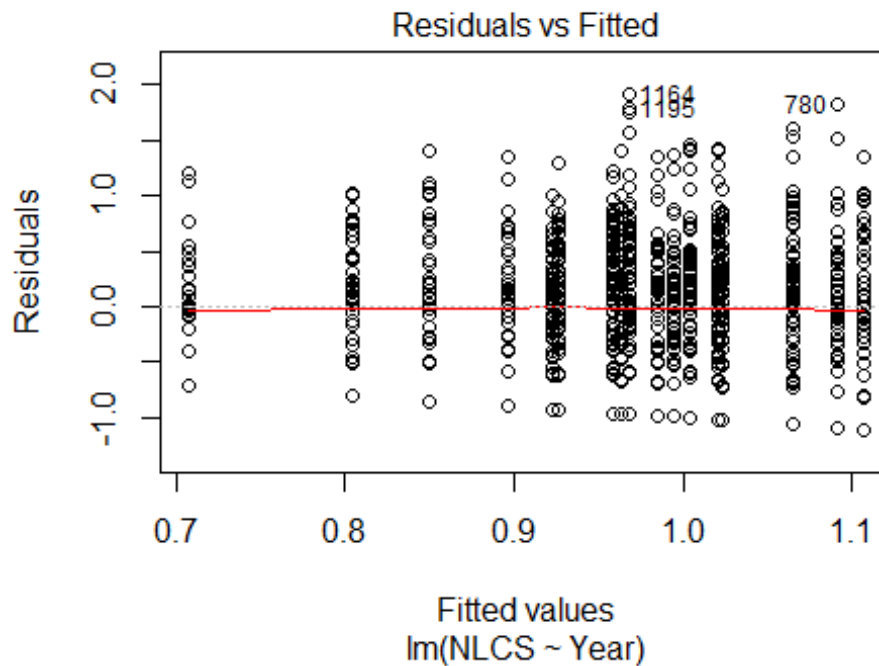
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 2231 79952786437 4.31 2011      1211      3      3.305
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2874 -0.4682  0.0147  0.4675  3.3026
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.02771    0.10294   9.98 < 2e-16 ***
## LastAuthorFemale1 0.11860    0.03391   3.50 0.00048 ***
## Year1997        0.00570    0.16637   0.03 0.97268
## Year1998        0.04670    0.13564   0.34 0.73068
## Year1999       -0.27864    0.14706  -1.89 0.05828 .
## Year2000        0.10634    0.14563   0.73 0.46535
## Year2001       -0.18162    0.15068  -1.21 0.22820
## Year2002        0.12317    0.13396   0.92 0.35801
## Year2003        0.14110    0.12374   1.14 0.25431
## Year2004        0.12259    0.12194   1.01 0.31486
## Year2005        0.07354    0.12600   0.58 0.55952
## Year2006       -0.02718    0.11673  -0.23 0.81587
```

```

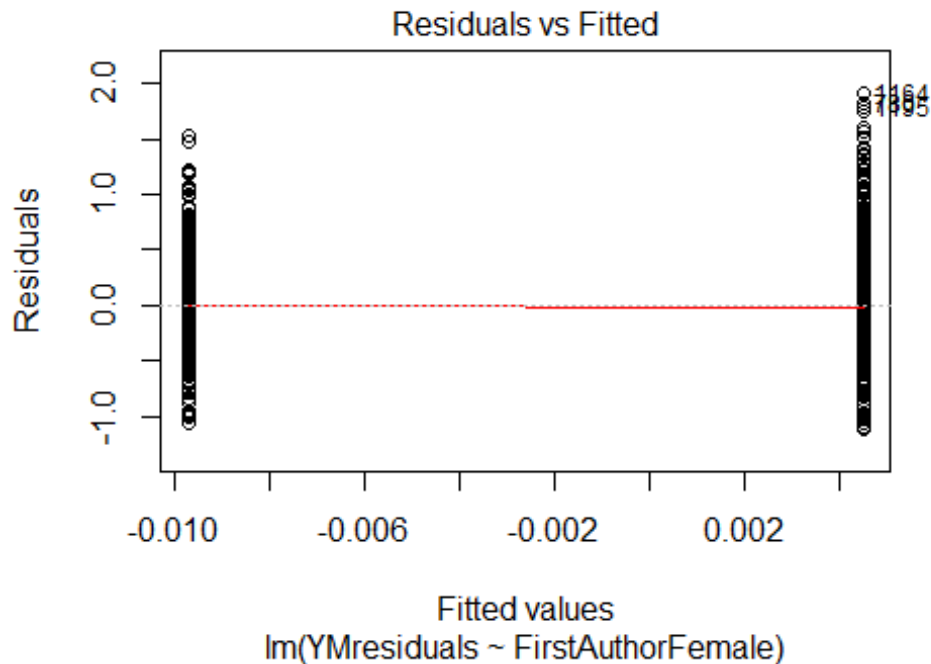
## Year2007          -0.01527      0.11939    -0.13   0.89822
## Year2008           0.02532      0.11726      0.22   0.82907
## Year2009          -0.04810      0.11410    -0.42   0.67338
## Year2010          -0.00901      0.11206    -0.08   0.93593
## Year2011          -0.02035      0.11221    -0.18   0.85609
## Year2012          -0.05181      0.11663    -0.44   0.65693
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.687
## Multiple R-squared:  0.0194, Adjusted R-squared:  0.0106
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1693 is an outlier with |weight| = 0 ( < 5.2e-05);
## 162 weights are ~= 1. The remaining 1756 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.305  0.859   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      5.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1919"
## [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
##   51   68   68   65   72   77   89   68   73   78   69   89  114  125  128
## 2011 2012
##  142   99
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   47   55   54   36   52   48   77   58   63   62   61   72  101  111  111

```

```
## 2011 2012
## 117 88
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 46 50 53 35 50 47 73 56 58 57 59 69 93 105 105
## 2011 2012
## 111 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 = 17, df = 16, p-value = 0.4
```

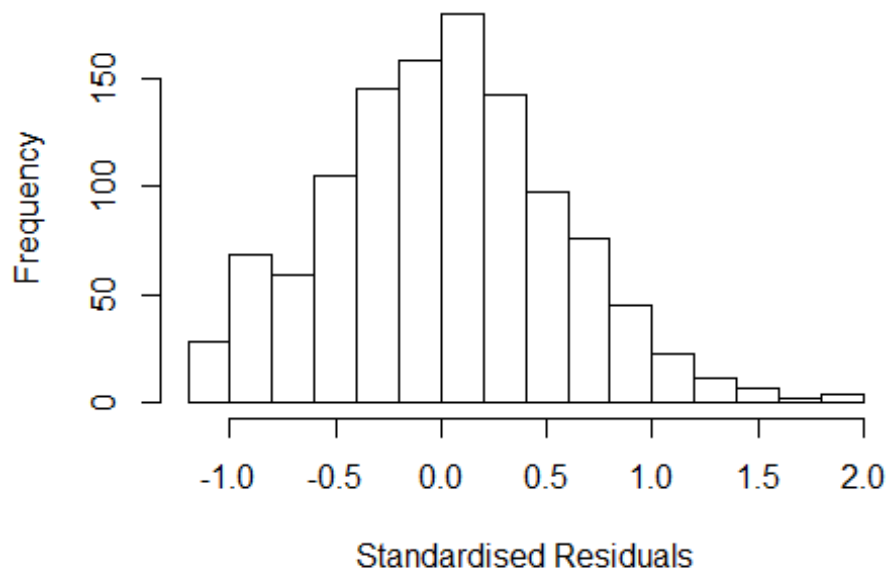


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.9, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 1.83312265890034"
## [1] "Male first author team size 2018 geometric mean: 1.59226195987018"
##
## 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: 1.80595922109521"
## [1] "Male last author team size 2018 geometric mean: 1.61757072228526"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1800, 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 1.749  1      1.323
## LastAuthorFemale  1.695  1      1.302
## UniqueAuthors    1.278  4      1.031
## Year             1.400 16      1.011
```


Residuals from first and last author and team size



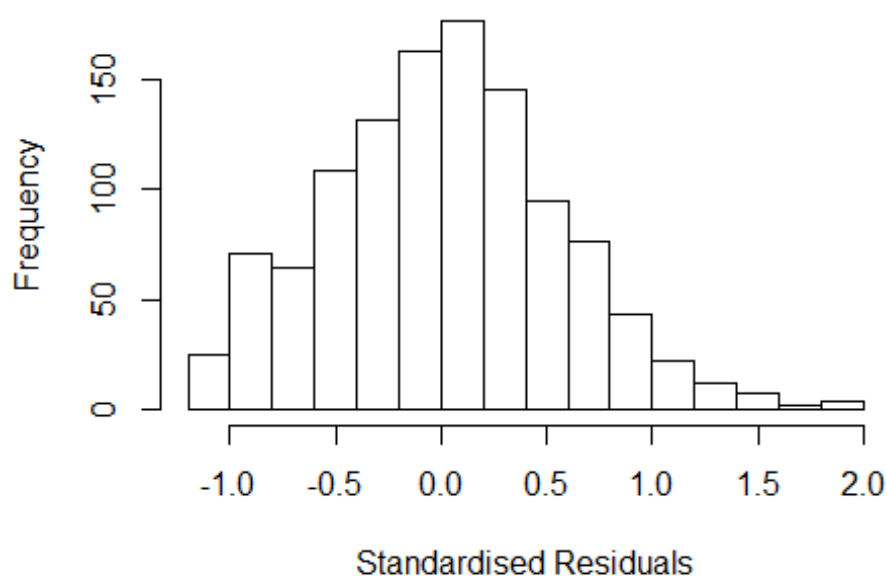
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1539 -0.3758 0.0188 0.3550 1.9827
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.801134 0.099940 8.02 2.7e-15 ***
## FirstAuthorFemale1 0.000947 0.044353 0.02 0.983
## LastAuthorFemale1 0.004165 0.043811 0.10 0.924
## UniqueAuthors2 0.065718 0.039627 1.66 0.098 .
## UniqueAuthors3 0.007993 0.051410 0.16 0.876
## UniqueAuthors4 0.052276 0.103677 0.50 0.614
## UniqueAuthors5 0.224119 0.189254 1.18 0.237
## Year1997 -0.053049 0.123505 -0.43 0.668
## Year1998 0.137507 0.129917 1.06 0.290
## Year1999 -0.111892 0.132622 -0.84 0.399
```

```

## Year2000          0.287062    0.135528    2.12    0.034 *
## Year2001          0.086531    0.131864    0.66    0.512
## Year2002          0.082070    0.116289    0.71    0.480
## Year2003          0.153059    0.123326    1.24    0.215
## Year2004          0.090306    0.116453    0.78    0.438
## Year2005          0.231024    0.122288    1.89    0.059 .
## Year2006          0.164672    0.115543    1.43    0.154
## Year2007          0.181444    0.116477    1.56    0.120
## Year2008          0.194847    0.113068    1.72    0.085 .
## Year2009          0.099203    0.115635    0.86    0.391
## Year2010          0.198939    0.111105    1.79    0.074 .
## Year2011          0.135020    0.114349    1.18    0.238
## Year2012          0.156080    0.121715    1.28    0.200
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.0277, Adjusted R-squared:  0.00869
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 100 weights are ~= 1. The remaining 1048 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.170  0.869  0.953  0.907  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.71e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.733 1          1.316
## LastAuthorFemale  1.685 1          1.298
## Year              1.123 16          1.004

```

Residuals from first and last author



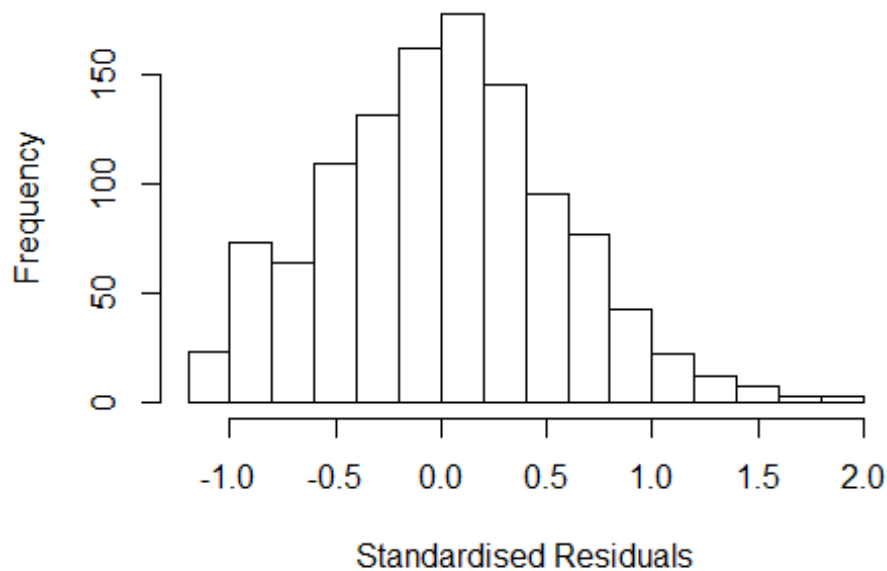
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1034 -0.3621 0.0176 0.3544 1.9532
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.82086 0.09960 8.24 4.7e-16 ***
## FirstAuthorFemale1 0.00180 0.04438 0.04 0.968
## LastAuthorFemale1 0.00583 0.04387 0.13 0.894
## Year1997 -0.05692 0.12370 -0.46 0.646
## Year1998 0.14340 0.12904 1.11 0.267
## Year1999 -0.12063 0.13220 -0.91 0.362
## Year2000 0.28254 0.13576 2.08 0.038 *
## Year2001 0.07946 0.13113 0.61 0.545
## Year2002 0.08374 0.11668 0.72 0.473
## Year2003 0.15422 0.12309 1.25 0.210
## Year2004 0.09441 0.11714 0.81 0.420
## Year2005 0.23074 0.12187 1.89 0.059 .
```

```

## Year2006          0.17169    0.11610    1.48    0.139
## Year2007          0.19212    0.11624    1.65    0.099 .
## Year2008          0.19969    0.11368    1.76    0.079 .
## Year2009          0.10895    0.11620    0.94    0.349
## Year2010          0.20768    0.11106    1.87    0.062 .
## Year2011          0.14143    0.11446    1.24    0.217
## Year2012          0.15757    0.12170    1.29    0.196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared:  0.0235, Adjusted R-squared:  0.00796
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 1060 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.187  0.871  0.953  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      8.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.079 1      1.039
## Year              1.079 16      1.002

```

Residuals from first author



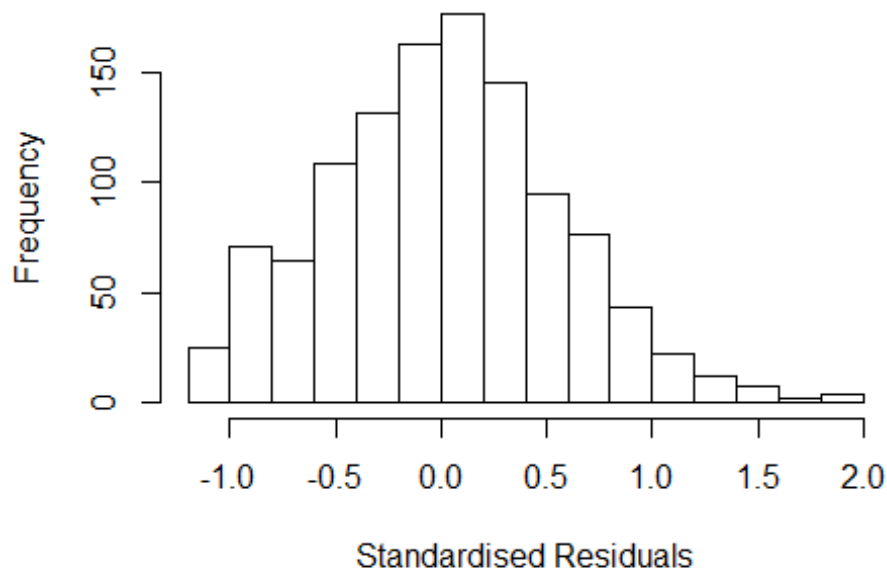
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1039 -0.3625 0.0163 0.3533 1.9526
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.82128 0.09956 8.25 4.4e-16 ***
## FirstAuthorFemale1 0.00541 0.03501 0.15 0.877
## Year1997 -0.05615 0.12350 -0.45 0.649
## Year1998 0.14348 0.12905 1.11 0.266
## Year1999 -0.12016 0.13202 -0.91 0.363
## Year2000 0.28263 0.13577 2.08 0.038 *
## Year2001 0.08012 0.13080 0.61 0.540
## Year2002 0.08417 0.11658 0.72 0.470
## Year2003 0.15378 0.12315 1.25 0.212
## Year2004 0.09463 0.11707 0.81 0.419
## Year2005 0.23080 0.12183 1.89 0.058 .
## Year2006 0.17181 0.11605 1.48 0.139
```

```

## Year2007          0.19318    0.11594    1.67    0.096 .
## Year2008          0.20011    0.11363    1.76    0.078 .
## Year2009          0.10908    0.11618    0.94    0.348
## Year2010          0.20801    0.11105    1.87    0.061 .
## Year2011          0.14185    0.11441    1.24    0.215
## Year2012          0.15812    0.12157    1.30    0.194
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared:  0.0235, Adjusted R-squared:  0.00883
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 86 weights are ~= 1. The remaining 1062 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.186  0.871  0.953  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      8.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.049 1          1.024
## Year            1.049 16          1.001

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1037 -0.3623  0.0177  0.3541  1.9528
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.82121    0.09915   8.28 3.4e-16 ***
## LastAuthorFemale1 0.00691    0.03460   0.20  0.842
## Year1997      -0.05719    0.12323  -0.46  0.643
## Year1998       0.14340    0.12895   1.11  0.266
## Year1999      -0.12078    0.13204  -0.91  0.361
## Year2000       0.28248    0.13570   2.08  0.038 *
## Year2001       0.07930    0.13104   0.61  0.545
## Year2002       0.08370    0.11669   0.72  0.473
## Year2003       0.15427    0.12327   1.25  0.211
## Year2004       0.09420    0.11702   0.80  0.421
## Year2005       0.23077    0.12188   1.89  0.059 .
## Year2006       0.17170    0.11619   1.48  0.140
```

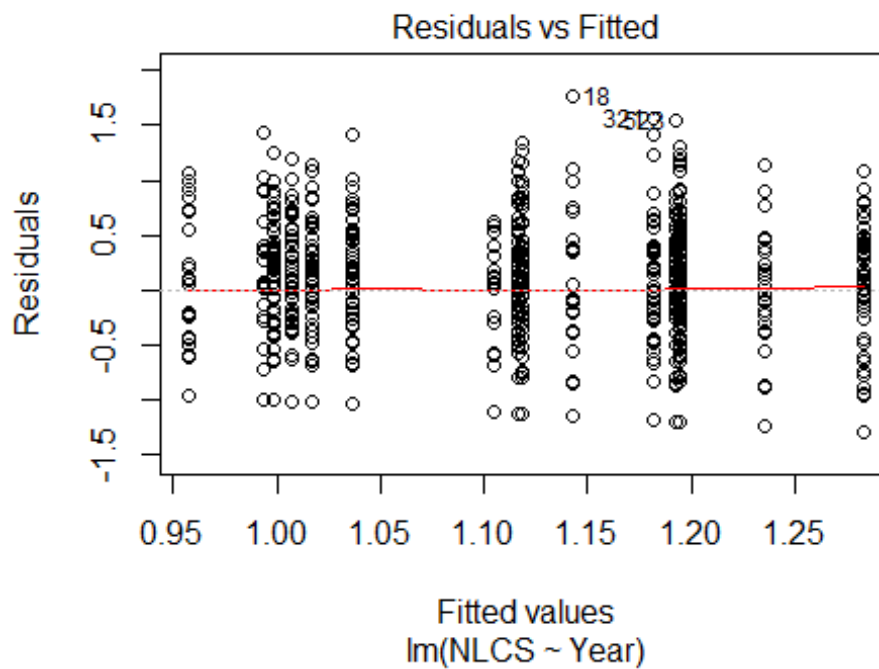
```

## Year2007      0.19187      0.11602      1.65      0.098 .
## Year2008      0.19950      0.11359      1.76      0.079 .
## Year2009      0.10904      0.11616      0.94      0.348
## Year2010      0.20767      0.11108      1.87      0.062 .
## Year2011      0.14129      0.11442      1.23      0.217
## Year2012      0.15761      0.12184      1.29      0.196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.554
## Multiple R-squared:  0.0235, Adjusted R-squared:  0.00882
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 86 weights are ~= 1. The remaining 1062 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.189  0.872  0.953  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      8.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1148"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
##   28   30   30   30   33   38   49   61   64   61   68   71   66   82   83
## 2011 2012
##   83   97
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   26   26   26   24   24   19   45   57   54   56   60   67   58   71   69
## 2011 2012

```



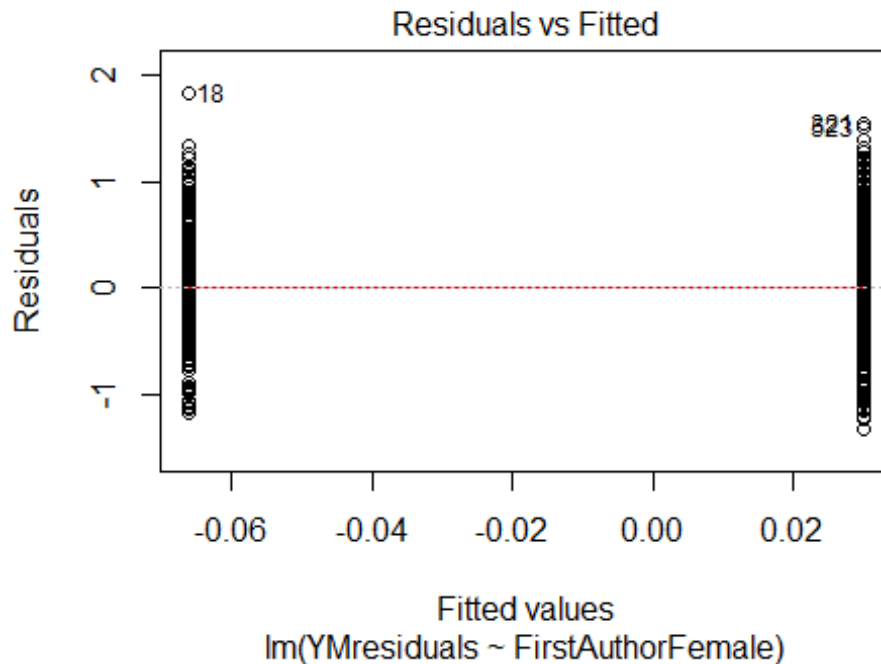
```
## 68 81
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 24 25 25 24 23 17 43 55 52 54 58 61 56 69 66
## 2011 2012
## 62 80
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 8.2, df = 16, p-value = 0.9
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.2, df = 1, p-value = 0.7

## [1] "Female first author team size 2018 geometric mean: 1.72149163042904"
## [1] "Male first author team size 2018 geometric mean: 1.58141906478952"

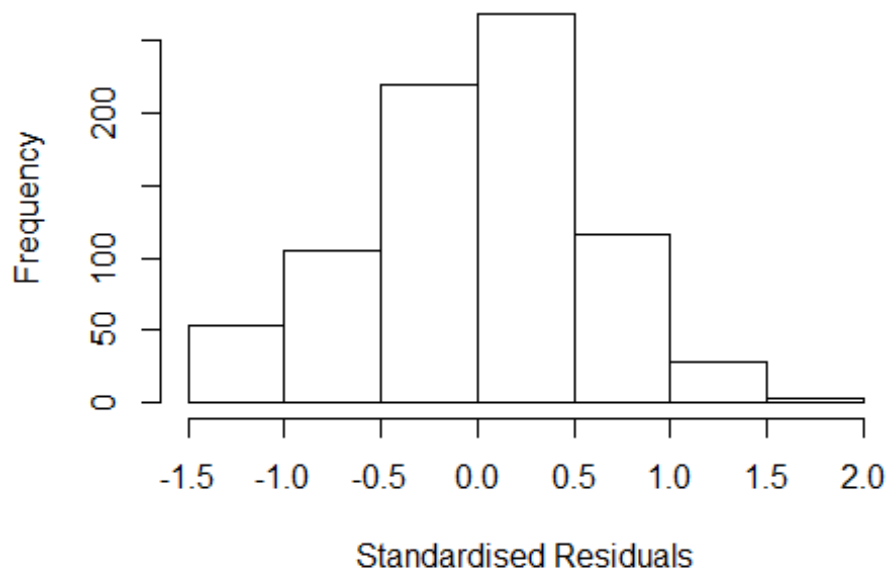
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 760, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.49460848087189"
## [1] "Male last author team size 2018 geometric mean: 1.69858731067491"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.735	1	1.317
LastAuthorFemale	1.717	1	1.310
UniqueAuthors	1.545	4	1.056
Year	1.517	16	1.013

Residuals from first and last author and team size



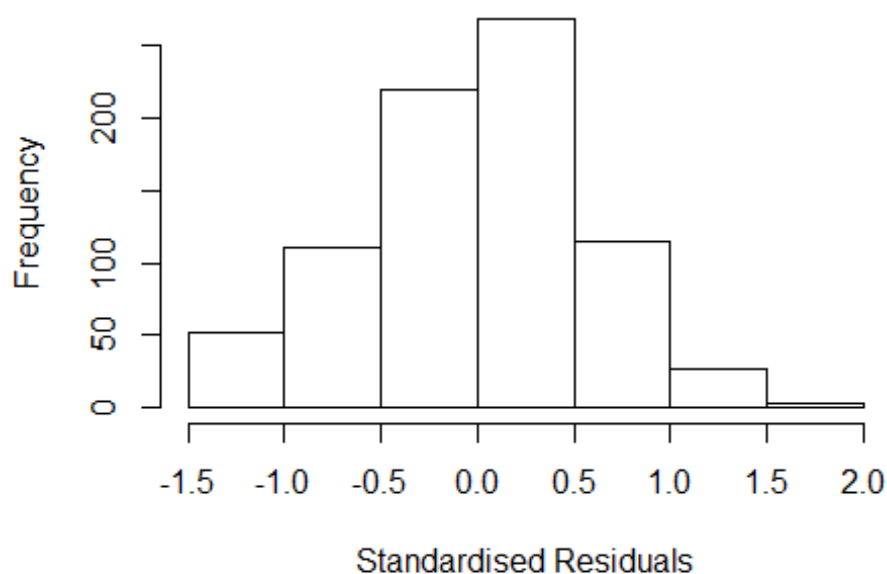
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2860 -0.4222 0.0289 0.3935 1.8463
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1464 0.1508 7.60 8.4e-14 ***
## FirstAuthorFemale1 -0.1143 0.0594 -1.92 0.055 .
## LastAuthorFemale1 0.0256 0.0586 0.44 0.663
## UniqueAuthors2 0.0532 0.0496 1.07 0.283
## UniqueAuthors3 -0.0833 0.0850 -0.98 0.327
## UniqueAuthors4 0.1579 0.1253 1.26 0.208
## UniqueAuthors5 0.0185 0.1407 0.13 0.896
## Year1997 0.1286 0.1885 0.68 0.495
## Year1998 -0.1908 0.2050 -0.93 0.352
## Year1999 -0.0213 0.1768 -0.12 0.904
```

```

## Year2000      -0.1035      0.2201      -0.47      0.638
## Year2001      0.1136      0.2025      0.56      0.575
## Year2002     -0.0133      0.1689     -0.08      0.937
## Year2003      0.0307      0.1726      0.18      0.859
## Year2004      0.1396      0.1714      0.81      0.416
## Year2005     -0.1267      0.1689     -0.75      0.454
## Year2006      0.0429      0.1680      0.26      0.798
## Year2007      0.0699      0.1672      0.42      0.676
## Year2008     -0.0243      0.1747     -0.14      0.890
## Year2009     -0.0752      0.1649     -0.46      0.648
## Year2010     -0.1105      0.1686     -0.66      0.512
## Year2011     -0.1272      0.1716     -0.74      0.459
## Year2012      0.0542      0.1653      0.33      0.743
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.591
## Multiple R-squared:  0.0358, Adjusted R-squared:  0.00829
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 727 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.309  0.871  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      1.26e-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 1.668 1      1.292
## LastAuthorFemale  1.647 1      1.283
## Year              1.083 16      1.002

```

Residuals from first and last author



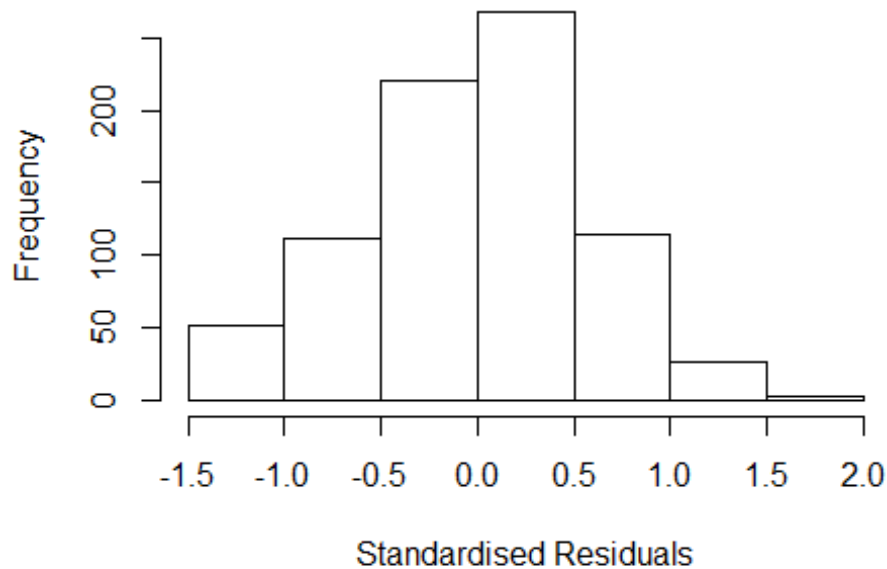
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2988 -0.3947 0.0242 0.3883 1.8534
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14466 0.15047 7.61 8.1e-14 ***
## FirstAuthorFemale1 -0.11569 0.05864 -1.97 0.049 *
## LastAuthorFemale1 0.02167 0.05774 0.38 0.708
## Year1997 0.13640 0.18808 0.73 0.469
## Year1998 -0.17793 0.20382 -0.87 0.383
## Year1999 -0.00446 0.17630 -0.03 0.980
## Year2000 -0.09796 0.21915 -0.45 0.655
## Year2001 0.11990 0.20351 0.59 0.556
## Year2002 -0.01155 0.16981 -0.07 0.946
## Year2003 0.05291 0.17133 0.31 0.758
## Year2004 0.15413 0.17147 0.90 0.369
## Year2005 -0.11410 0.16916 -0.67 0.500
```

```

## Year2006          0.06654    0.16693    0.40    0.690
## Year2007          0.09408    0.16582    0.57    0.571
## Year2008         -0.00930    0.17460   -0.05    0.958
## Year2009         -0.05683    0.16351   -0.35    0.728
## Year2010         -0.09388    0.16988   -0.55    0.581
## Year2011         -0.10506    0.16927   -0.62    0.535
## Year2012          0.07846    0.16419    0.48    0.633
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.593
## Multiple R-squared:  0.0311, Adjusted R-squared:  0.0086
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 729 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.309  0.874  0.952  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      1.26e-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 1.041 1         1.020
## Year              1.041 16         1.001

```

Residuals from first author



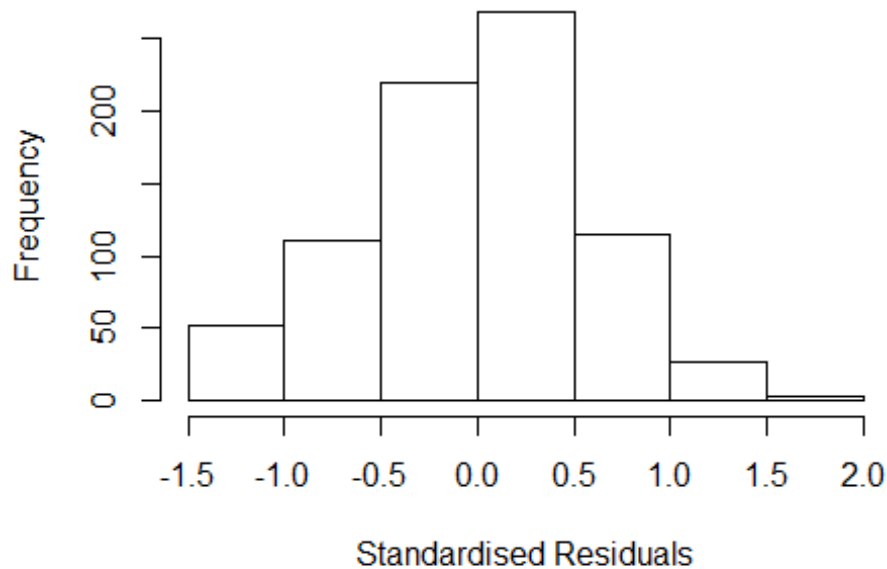
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3004 -0.3981 0.0223 0.3861 1.8567
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14840 0.14958 7.68 4.9e-14 ***
## FirstAuthorFemale1 -0.10115 0.04645 -2.18 0.03 *
## Year1997 0.13533 0.18771 0.72 0.47
## Year1998 -0.18058 0.20366 -0.89 0.38
## Year1999 -0.00542 0.17601 -0.03 0.98
## Year2000 -0.09749 0.21845 -0.45 0.66
## Year2001 0.11832 0.20302 0.58 0.56
## Year2002 -0.01152 0.16948 -0.07 0.95
## Year2003 0.05307 0.17099 0.31 0.76
## Year2004 0.15195 0.17106 0.89 0.37
## Year2005 -0.11555 0.16883 -0.68 0.49
## Year2006 0.06712 0.16675 0.40 0.69
```

```

## Year2007          0.09261    0.16535    0.56    0.58
## Year2008          -0.01175    0.17420   -0.07    0.95
## Year2009          -0.05858    0.16313   -0.36    0.72
## Year2010          -0.09599    0.16938   -0.57    0.57
## Year2011          -0.10526    0.16909   -0.62    0.53
## Year2012           0.07679    0.16355    0.47    0.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.593
## Multiple R-squared:  0.031, Adjusted R-squared:  0.00975
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 732 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.306  0.875   0.953   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      1.26e-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 1.031 1          1.015
## Year            1.031 16          1.001

```


Residuals from last author



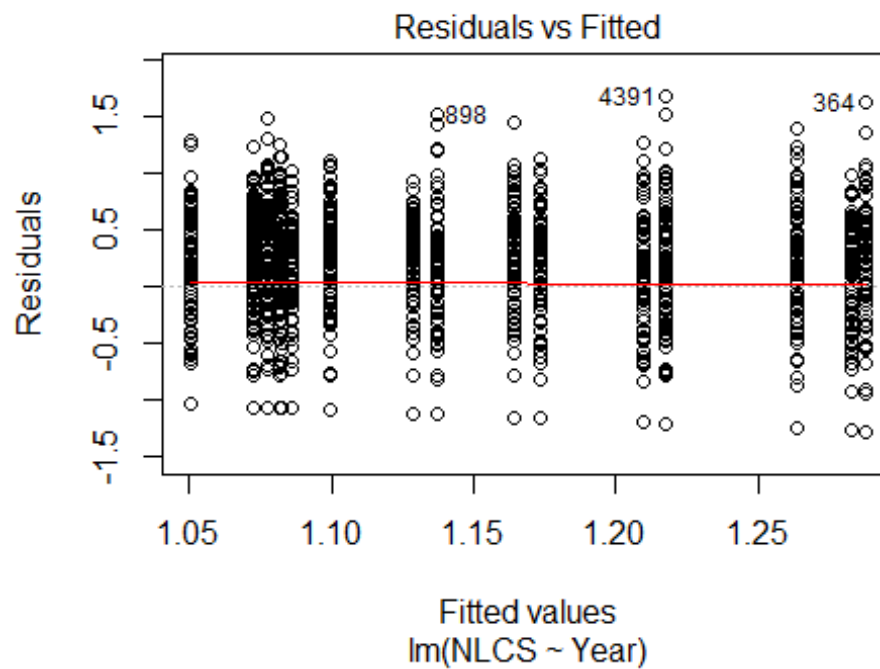
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2911 -0.3818 0.0339 0.3916 1.8244
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13371 0.15005 7.56 1.2e-13 ***
## LastAuthorFemale1 -0.05415 0.04593 -1.18 0.24
## Year1997 0.14111 0.18672 0.76 0.45
## Year1998 -0.18621 0.20335 -0.92 0.36
## Year1999 0.00128 0.17693 0.01 0.99
## Year2000 -0.08622 0.21926 -0.39 0.69
## Year2001 0.11934 0.20411 0.58 0.56
## Year2002 -0.00741 0.16997 -0.04 0.97
## Year2003 0.05556 0.17130 0.32 0.75
## Year2004 0.15735 0.17047 0.92 0.36
## Year2005 -0.11406 0.16888 -0.68 0.50
## Year2006 0.07666 0.16700 0.46 0.65
```

```

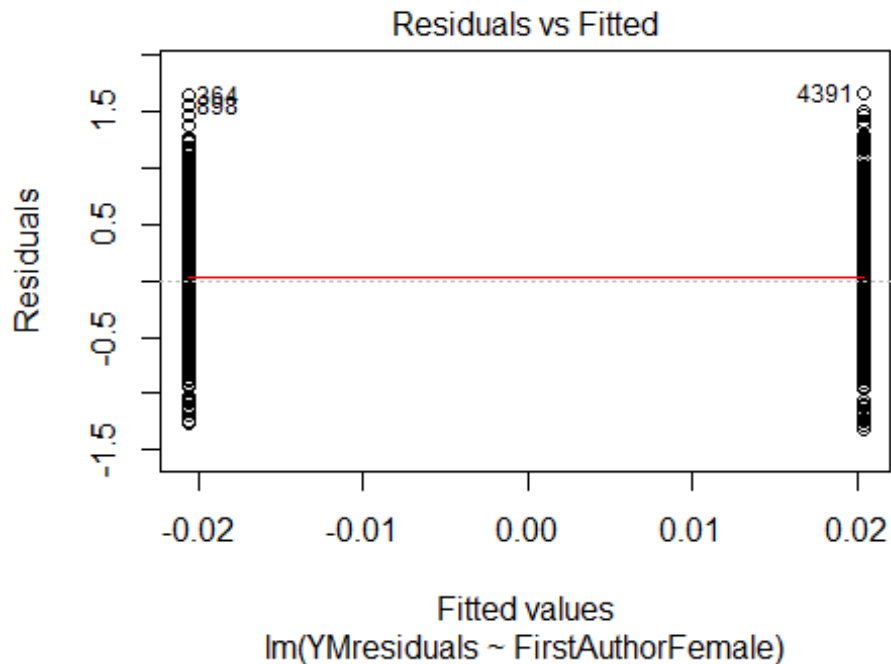
## Year2007      0.08898      0.16557      0.54      0.59
## Year2008     -0.01424      0.17421     -0.08      0.93
## Year2009     -0.06686      0.16335     -0.41      0.68
## Year2010     -0.10074      0.16974     -0.59      0.55
## Year2011     -0.10712      0.16896     -0.63      0.53
## Year2012      0.07736      0.16366      0.47      0.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0266, Adjusted R-squared:  0.00529
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 730 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.325  0.871  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      1.26e-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: 794"
## [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
##  214  200  185  218  189  198  199  179  203  234  231  262  273  221  270
## 2011 2012
##  253  244
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  140  124  110  116   91   98  152  124  140  161  180  177  177  154  183
## 2011 2012

```

```
## 185 174
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 126 111 101 102 83 86 131 116 129 147 159 158 153 140 159
## 2011 2012
## 165 153
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.08
```

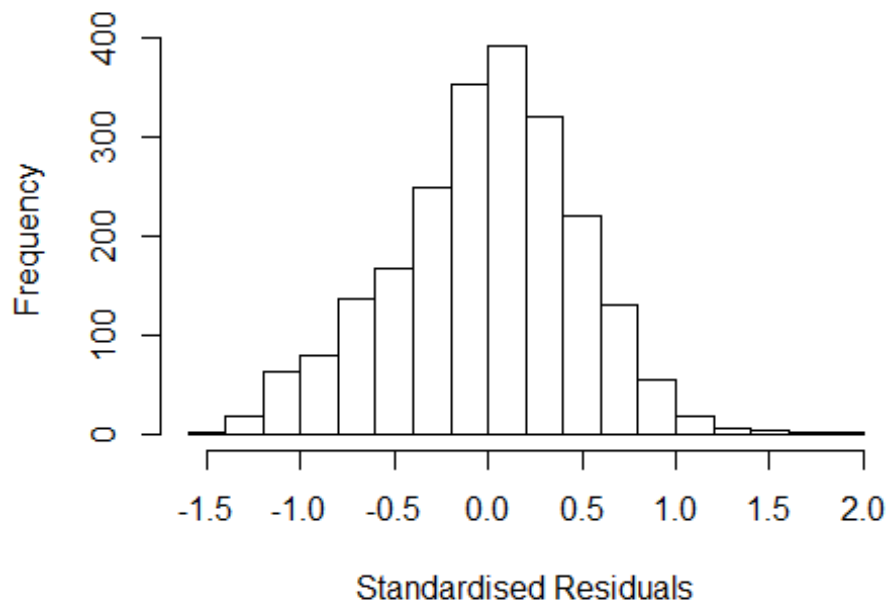


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 3.78389542961577"
## [1] "Male first author team size 2018 geometric mean: 3.37009719366645"
##
## 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: 3.24660447475313"
## [1] "Male last author team size 2018 geometric mean: 3.97607045360794"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, 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 1.096 1      1.047
## LastAuthorFemale  1.153 1      1.074
## UniqueAuthors    1.325 4      1.036
## Year              1.300 16     1.008
```

Residuals from first and last author and team size



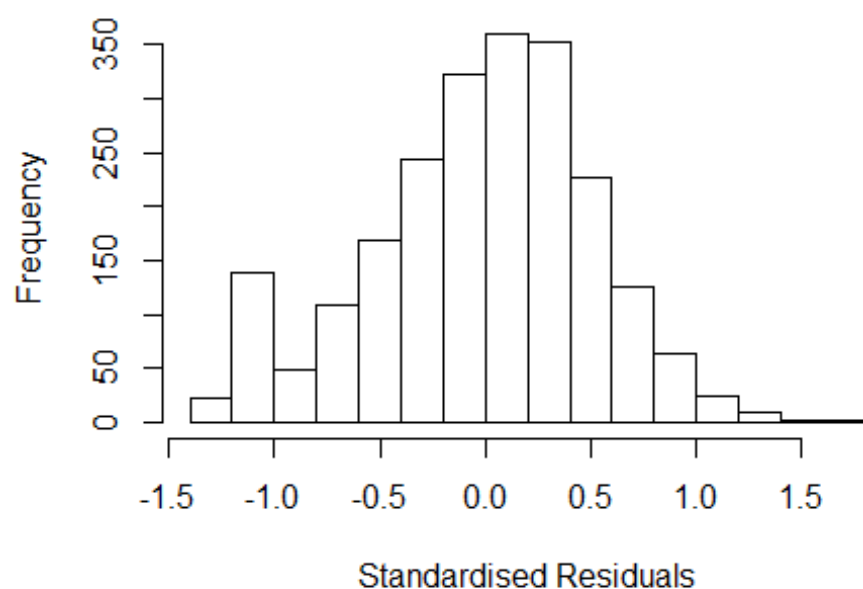
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4577 -0.3198 0.0227 0.3251 1.9651
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9161 0.0623 14.70 < 2e-16 ***
## FirstAuthorFemale1 0.0450 0.0224 2.01 0.04442 *
## LastAuthorFemale1 0.0350 0.0243 1.44 0.14996
## UniqueAuthors2 0.2676 0.0419 6.39 2.0e-10 ***
## UniqueAuthors3 0.3493 0.0376 9.28 < 2e-16 ***
## UniqueAuthors4 0.4617 0.0367 12.56 < 2e-16 ***
## UniqueAuthors5 0.5379 0.0344 15.66 < 2e-16 ***
## Year1997 0.0318 0.0770 0.41 0.68007
## Year1998 -0.0251 0.0760 -0.33 0.74155
## Year1999 -0.1258 0.0763 -1.65 0.09961 .
```

```

## Year2000          0.0625      0.0776      0.81  0.42039
## Year2001         -0.1971      0.0784     -2.51  0.01204 *
## Year2002         -0.0446      0.0681     -0.65  0.51324
## Year2003         -0.0807      0.0708     -1.14  0.25419
## Year2004         -0.1217      0.0677     -1.80  0.07236 .
## Year2005         -0.1791      0.0693     -2.58  0.00985 **
## Year2006         -0.1786      0.0692     -2.58  0.00985 **
## Year2007         -0.2500      0.0660     -3.79  0.00016 ***
## Year2008         -0.1742      0.0713     -2.44  0.01467 *
## Year2009         -0.2491      0.0738     -3.38  0.00075 ***
## Year2010         -0.3127      0.0680     -4.60  4.5e-06 ***
## Year2011         -0.2305      0.0704     -3.27  0.00109 **
## Year2012         -0.1797      0.0704     -2.55  0.01080 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.471
## Multiple R-squared:  0.147, Adjusted R-squared:  0.138
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 176 weights are ~= 1. The remaining 2043 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0428 0.8570 0.9490 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      4.51e-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 1.072 1      1.035
## LastAuthorFemale  1.068 1      1.033
## Year              1.063 16      1.002

```

Residuals from first and last author



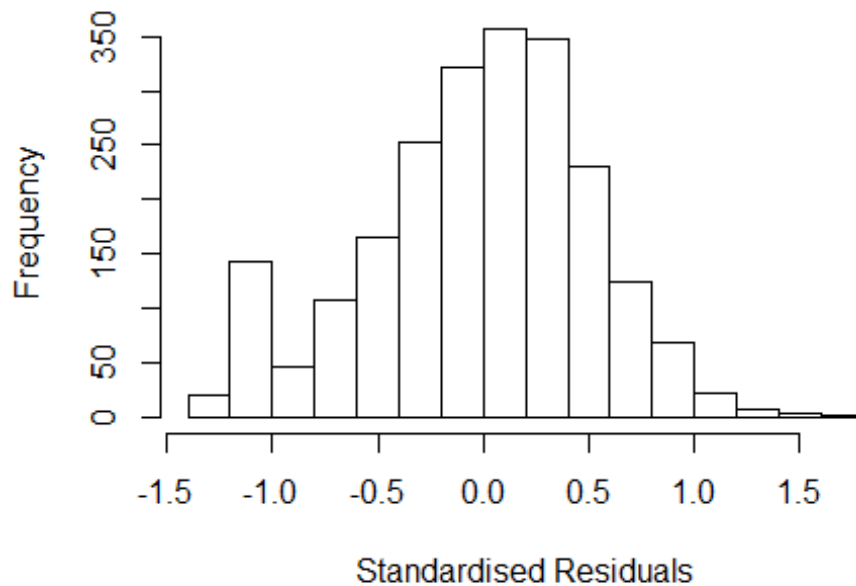
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3292 -0.3363 0.0402 0.3442 1.6401
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2781 0.0543 23.54 < 2e-16 ***
## FirstAuthorFemale1 0.0459 0.0238 1.93 0.05418 .
## LastAuthorFemale1 -0.0521 0.0250 -2.08 0.03741 *
## Year1997 0.0293 0.0787 0.37 0.70962
## Year1998 -0.0293 0.0748 -0.39 0.69517
## Year1999 -0.1391 0.0779 -1.79 0.07415 .
## Year2000 0.0511 0.0779 0.66 0.51206
## Year2001 -0.1903 0.0789 -2.41 0.01590 *
## Year2002 -0.0912 0.0683 -1.33 0.18206
## Year2003 -0.1712 0.0694 -2.47 0.01370 *
## Year2004 -0.1399 0.0668 -2.09 0.03636 *
## Year2005 -0.1952 0.0722 -2.70 0.00690 **
```

```

## Year2006          -0.1845      0.0703    -2.62  0.00875 **
## Year2007          -0.2268      0.0662    -3.43  0.00062 ***
## Year2008          -0.1371      0.0724    -1.89  0.05831 .
## Year2009          -0.2207      0.0748    -2.95  0.00322 **
## Year2010          -0.2482      0.0683    -3.63  0.00029 ***
## Year2011          -0.1747      0.0691    -2.53  0.01155 *
## Year2012          -0.0700      0.0708    -0.99  0.32272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.499
## Multiple R-squared:  0.0282, Adjusted R-squared:  0.0202
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 192 weights are ~= 1. The remaining 2027 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.257  0.859   0.949   0.896   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.51e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.036 1          1.018
## Year              1.036 16          1.001

```


Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3308 -0.3374 0.0359 0.3486 1.6656
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2642 0.0540 23.42 < 2e-16 ***
## FirstAuthorFemale1 0.0363 0.0234 1.55 0.12109
## Year1997 0.0303 0.0783 0.39 0.69872
## Year1998 -0.0299 0.0748 -0.40 0.68972
## Year1999 -0.1335 0.0779 -1.71 0.08675 .
## Year2000 0.0514 0.0775 0.66 0.50692
## Year2001 -0.1889 0.0789 -2.39 0.01680 *
## Year2002 -0.0901 0.0683 -1.32 0.18753
## Year2003 -0.1776 0.0691 -2.57 0.01019 *
## Year2004 -0.1413 0.0668 -2.12 0.03449 *
## Year2005 -0.1937 0.0721 -2.69 0.00727 **
## Year2006 -0.1839 0.0704 -2.61 0.00912 **
```

```

## Year2007          -0.2273      0.0661   -3.44  0.00060 ***
## Year2008          -0.1336      0.0722   -1.85  0.06442 .
## Year2009          -0.2205      0.0749   -2.94  0.00328 **
## Year2010          -0.2463      0.0681   -3.62  0.00031 ***
## Year2011          -0.1727      0.0689   -2.51  0.01231 *
## Year2012          -0.0720      0.0706   -1.02  0.30774
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.0262, Adjusted R-squared:  0.0187
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 180 weights are ~= 1. The remaining 2039 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.241  0.860  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      4.51e-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 1.029 1      1.014
## Year      1.029 16      1.001
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.3435 -0.3404  0.0306  0.3422  1.6627

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2938    0.0541   23.92 < 2e-16 ***
## LastAuthorFemale1 -0.0429    0.0245   -1.75  0.07998 .
## Year1997          0.0322    0.0790    0.41  0.68337
## Year1998         -0.0276    0.0752   -0.37  0.71339
## Year1999         -0.1359    0.0781   -1.74  0.08207 .
## Year2000          0.0497    0.0777    0.64  0.52244
## Year2001         -0.1861    0.0789   -2.36  0.01843 *
## Year2002         -0.0893    0.0684   -1.31  0.19168
## Year2003         -0.1677    0.0694   -2.42  0.01578 *
## Year2004         -0.1316    0.0665   -1.98  0.04785 *
## Year2005         -0.1924    0.0725   -2.65  0.00806 **
## Year2006         -0.1770    0.0703   -2.52  0.01186 *
## Year2007         -0.2232    0.0664   -3.36  0.00078 ***
## Year2008         -0.1299    0.0718   -1.81  0.07039 .
## Year2009         -0.2123    0.0746   -2.85  0.00448 **
## Year2010         -0.2421    0.0684   -3.54  0.00041 ***
## Year2011         -0.1676    0.0690   -2.43  0.01516 *
## Year2012         -0.0625    0.0708   -0.88  0.37711
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.0264, Adjusted R-squared:  0.0188
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 2034 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.243  0.864  0.950  0.896  0.985  0.999
## Algorithmic parameters:
##           tuning.chi             bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           4.51e-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: 2219"
## [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
##    5   12    8    9    4    8    5    2    9    8    6    3    4    7    9
## 2011 2012
##    3    6
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    0    2    0    2    0    0    0    0    0    2    2    0    1    3    2
## 2011 2012
##    1    1
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    0    2    0    2    0    0    0    0    0    1    1    0    1    1    2
## 2011 2012
##    1    1
## [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: 7"
##
## 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] "Female last author team size 2018 geometric mean: 5.29150262212918"
## [1] "Male last author team size 2018 geometric mean: 5"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1, 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: 12"
## [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
##	24	16	17	12	15	10	10	14	14	7	25	14	14	24	16

```
## 2011 2012
```

	2011	2012
##	9	14

```
##
```

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
##	4	2	1	1	0	0	4	8	3	2	7	1	0	0	2

```
## 2011 2012
```

	2011	2012
##	2	2

```
##
```

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
##	4	1	0	1	0	0	4	7	3	2	7	1	0	0	1

```
## 2011 2012
```

	2011	2012
##	1	2

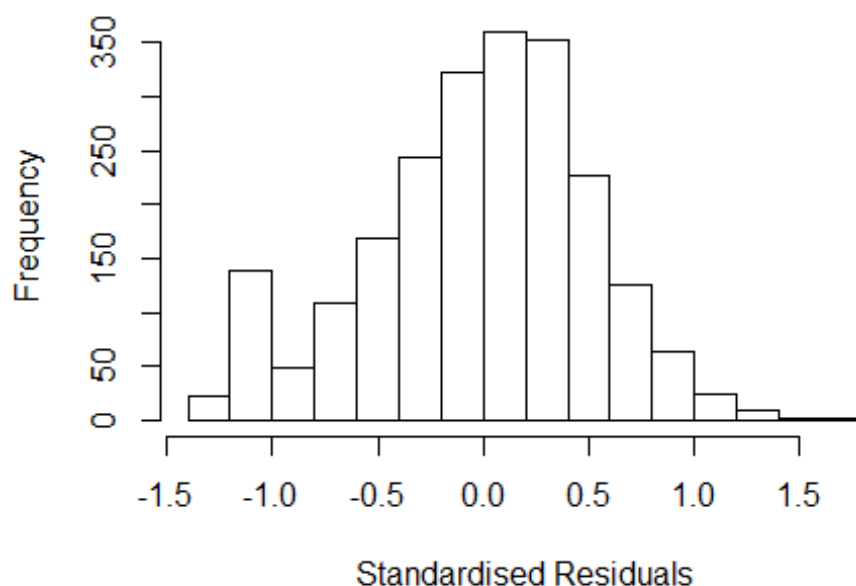
```
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
```

```
## [1] "Female first author team size 2018 geometric mean: 5.94392195276313"
```

```
## [1] "Male first author team size 2018 geometric mean: 6"
```

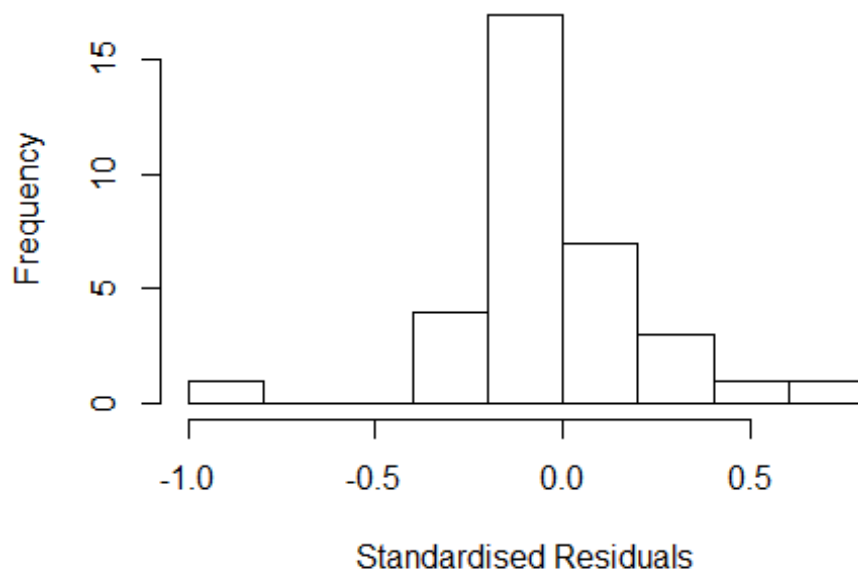
```
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1.5, 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: 5.95789213552895"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale -3.499e+15  1      NaN
## LastAuthorFemale  1.740e+15  1    41718367
## Year              -3.420e+30 11      NaN
```

Residuals from first and last author



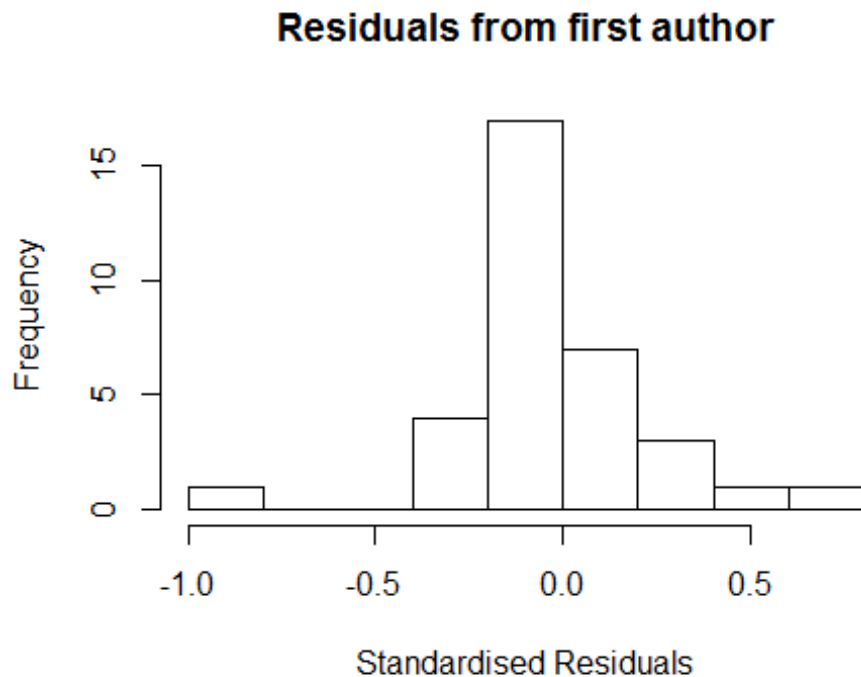
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n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

```

## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97375 -0.09544 -0.00251  0.10552  0.72194
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0096     0.1538   6.56 2.2e-06 ***
## FirstAuthorFemale1 -0.2697     0.1317  -2.05  0.0540 .
## LastAuthorFemale1 -0.0459     0.1147  -0.40  0.6934
## Year1997          -0.0256     0.1538  -0.17  0.8697
## Year1999          -0.1626     0.1538  -1.06  0.3033
## Year2002           0.0565     0.2132   0.27  0.7937
## Year2003           0.6512     0.1152   5.65 1.6e-05 ***
## Year2004           0.2209     0.2889   0.76  0.4535
## Year2005           0.2159     0.1762   1.23  0.2346
## Year2006           0.4727     0.1292   3.66  0.0016 **
## Year2007           0.8000     0.0899   8.90 2.2e-08 ***
## Year2010          -0.3186     0.1538  -2.07  0.0515 .
## Year2011           0.5013     0.1761   2.85  0.0100 **
## Year2012           0.6102     0.2202   2.77  0.0118 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.279
## Multiple R-squared:  0.672, Adjusted R-squared:  0.459
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~ = 1. The remaining 25 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.196  0.929  0.960  0.902  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          2.94e-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 -2.955e+14 1 NaN
## Year -2.955e+14 11 NaN
```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.99840 -0.11158 -0.00308 0.08406 0.74734
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0072 0.1486 6.78 1.1e-06 ***
## FirstAuthorFemale1 -0.3006 0.1235 -2.43 0.0240 *
## Year1997 -0.0232 0.1486 -0.16 0.8774
## Year1999 -0.1602 0.1486 -1.08 0.2932
## Year2002 0.0335 0.2039 0.16 0.8712
## Year2003 0.6782 0.1143 5.93 6.9e-06 ***
## Year2004 0.2076 0.2903 0.71 0.4825
## Year2005 0.2183 0.1718 1.27 0.2177
## Year2006 0.4720 0.1235 3.82 0.0010 ***
```

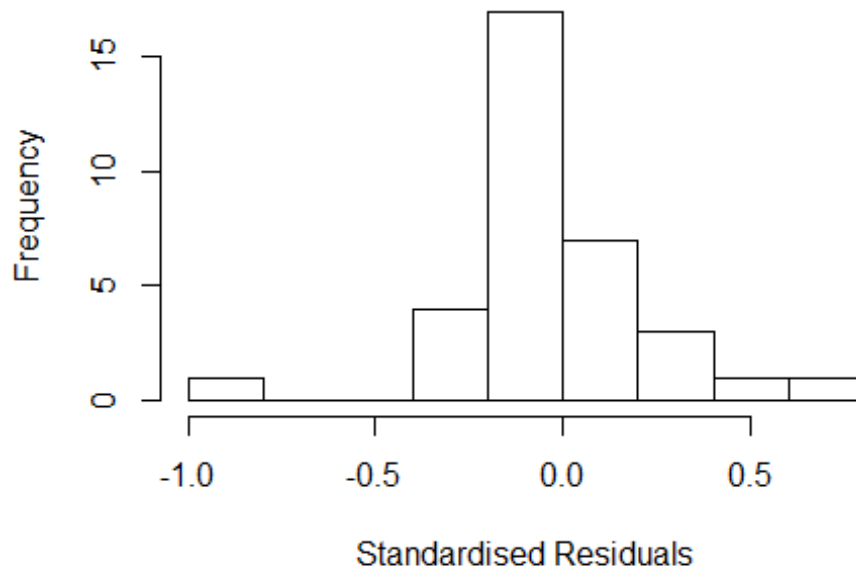


```

## Year2007          0.7874      0.0792      9.94  2.2e-09 ***
## Year2010         -0.3162      0.1486     -2.13   0.0453 *
## Year2011          0.4578      0.1486      3.08   0.0057 **
## Year2012          0.6051      0.2293      2.64   0.0153 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.264
## Multiple R-squared:  0.697, Adjusted R-squared:  0.524
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 24 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.122  0.923  0.954  0.888  0.989  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      2.94e-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 -1.599e+16 1      NaN
## Year             -1.599e+16 11      NaN

```

Residuals from last author



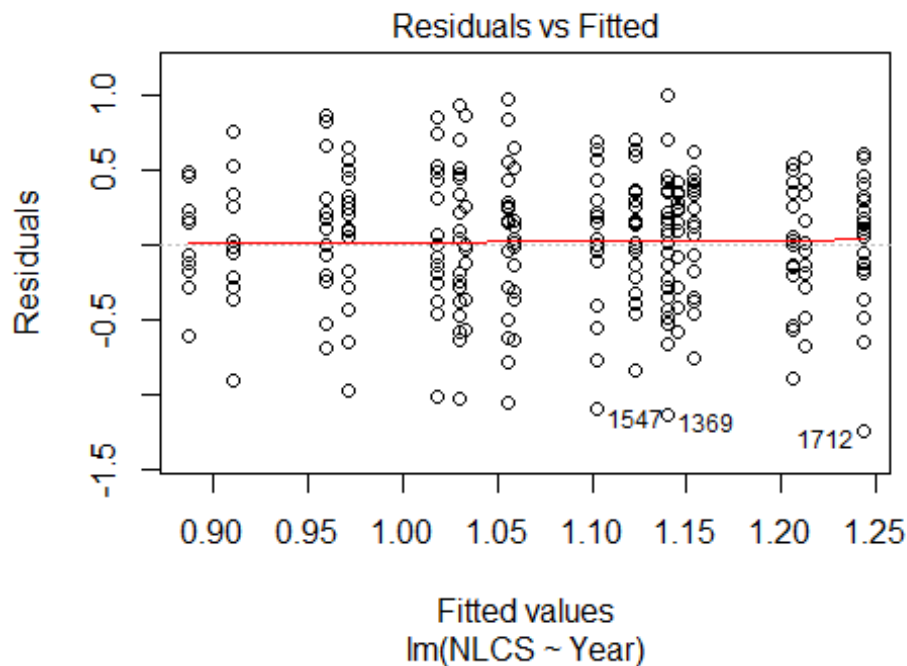
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -7.95e-01 -1.23e-01 1.13e-16 8.44e-02 8.19e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.84131 0.11861 7.09 5.4e-07 ***
## LastAuthorFemale1 -0.18471 0.13013 -1.42 0.17044
## Year1997 0.14269 0.11861 1.20 0.24236
## Year1999 0.00569 0.11861 0.05 0.96219
## Year2002 0.12759 0.23193 0.55 0.58802
## Year2003 0.64053 0.14166 4.52 0.00019 ***
## Year2004 0.30708 0.23480 1.31 0.20506
## Year2005 0.38419 0.14642 2.62 0.01586 *
## Year2006 0.59536 0.14018 4.25 0.00036 ***
## Year2007 0.83740 0.09048 9.26 7.4e-09 ***
## Year2010 -0.15031 0.11861 -1.27 0.21894
## Year2011 0.80840 0.09048 8.93 1.3e-08 ***
```

```

## Year2012          0.71305    0.28414    2.51  0.02036 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.281
## Multiple R-squared:  0.607, Adjusted R-squared:  0.382
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 26 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.375  0.900  0.966  0.901  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          2.94e-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: 34"
## [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
##   102  114  115  102  102   77   83   90   84   51   76   70   84   76   76
## 2011 2012
##    57   67
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   14   19   18   15   11   12   18   12   12   26   19   26   15   20
## 2011 2012
##   22   26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3   11   16   17   14   10   10   16   10   10   26   19   22   15   17
## 2011 2012

```

```
## 20 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 = 11, df = 16, p-value = 0.8
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.5, df = 1, p-value = 0.1

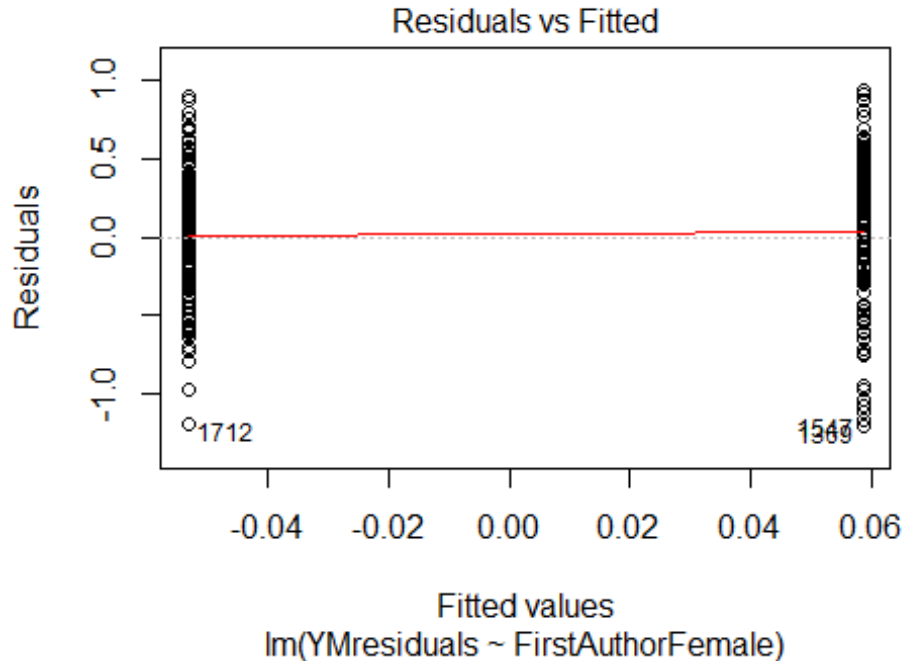
## [1] "Female first author team size 2018 geometric mean: 4.48138952543894"
## [1] "Male first author team size 2018 geometric mean: 5.55536860538186"

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

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 53, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
```

```
##
## [1] "Female last author team size 2018 geometric mean: 4.17502458598268"
## [1] "Male last author team size 2018 geometric mean: 5.23006988083367"

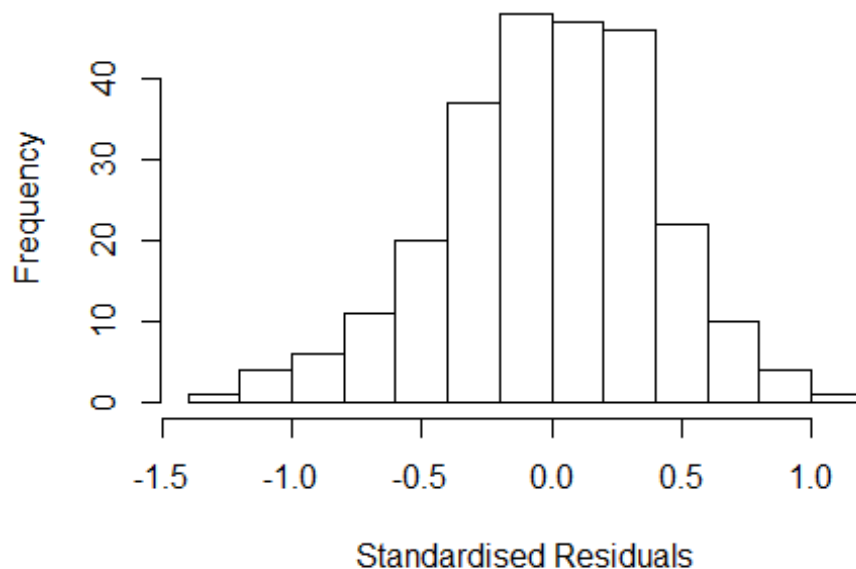
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 90, 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	1.701	1	1.304
LastAuthorFemale	1.545	1	1.243
UniqueAuthors	3.663	4	1.176
Year	4.452	16	1.048

Residuals from first and last author and team size



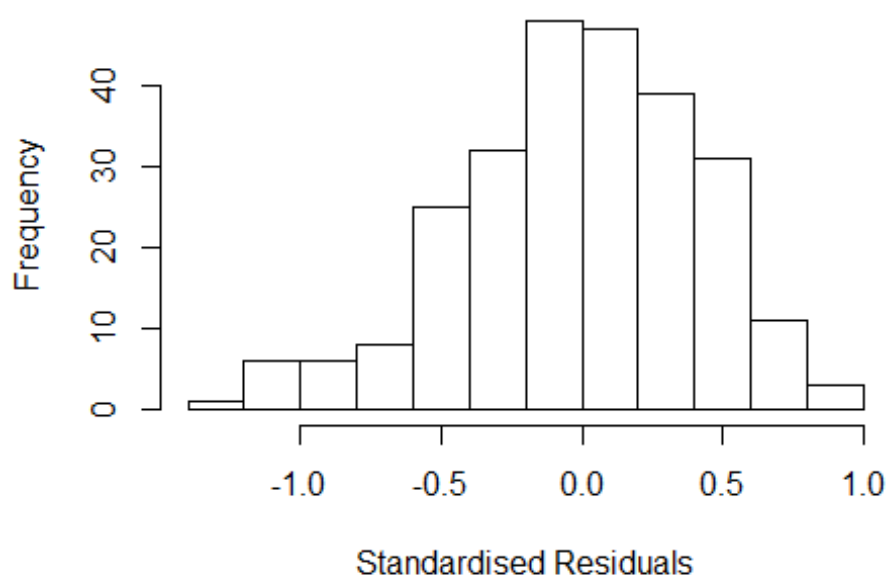
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2292 -0.2890 0.0106 0.2763 1.0467
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95595 0.37165 2.57 0.011 *
## FirstAuthorFemale1 -0.10534 0.06423 -1.64 0.102
## LastAuthorFemale1 -0.02459 0.06879 -0.36 0.721
## UniqueAuthors2 0.18258 0.11811 1.55 0.123
## UniqueAuthors3 0.12034 0.12921 0.93 0.353
## UniqueAuthors4 0.26325 0.13263 1.98 0.048 *
## UniqueAuthors5 0.29390 0.12635 2.33 0.021 *
## Year1997 0.12015 0.36827 0.33 0.745
## Year1998 0.03360 0.37603 0.09 0.929
## Year1999 0.05277 0.39783 0.13 0.895
```

```

## Year2000      0.00382    0.39676    0.01    0.992
## Year2001      0.03033    0.37598    0.08    0.936
## Year2002      0.09886    0.37276    0.27    0.791
## Year2003      0.10399    0.37894    0.27    0.784
## Year2004     -0.17161    0.39062   -0.44    0.661
## Year2005     -0.22578    0.37680   -0.60    0.550
## Year2006      0.08878    0.36657    0.24    0.809
## Year2007     -0.04422    0.37930   -0.12    0.907
## Year2008      0.03460    0.37801    0.09    0.927
## Year2009      0.10259    0.37358    0.27    0.784
## Year2010      0.09210    0.37602    0.24    0.807
## Year2011     -0.14812    0.37700   -0.39    0.695
## Year2012      0.23058    0.37046    0.62    0.534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.414
## Multiple R-squared:  0.119, Adjusted R-squared:  0.0361
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 239 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.359  0.889  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      3.89e-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 1.337 1      1.156
## LastAuthorFemale  1.331 1      1.154
## Year              1.710 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
## -1.24273 -0.26437 0.00173 0.27742 0.97795
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.154788 0.392157 2.94 0.0036 **
## FirstAuthorFemale1 -0.108819 0.060486 -1.80 0.0733 .
## LastAuthorFemale1 -0.041358 0.067346 -0.61 0.5397
## Year1997 0.074993 0.408721 0.18 0.8546
## Year1998 0.023871 0.415052 0.06 0.9542
## Year1999 0.039237 0.422161 0.09 0.9260
## Year2000 -0.060738 0.430768 -0.14 0.8880
## Year2001 -0.000739 0.407395 0.00 0.9986
## Year2002 0.126518 0.404110 0.31 0.7545
## Year2003 0.107968 0.411508 0.26 0.7933
## Year2004 -0.136129 0.423718 -0.32 0.7483
## Year2005 -0.217877 0.409019 -0.53 0.5948
```

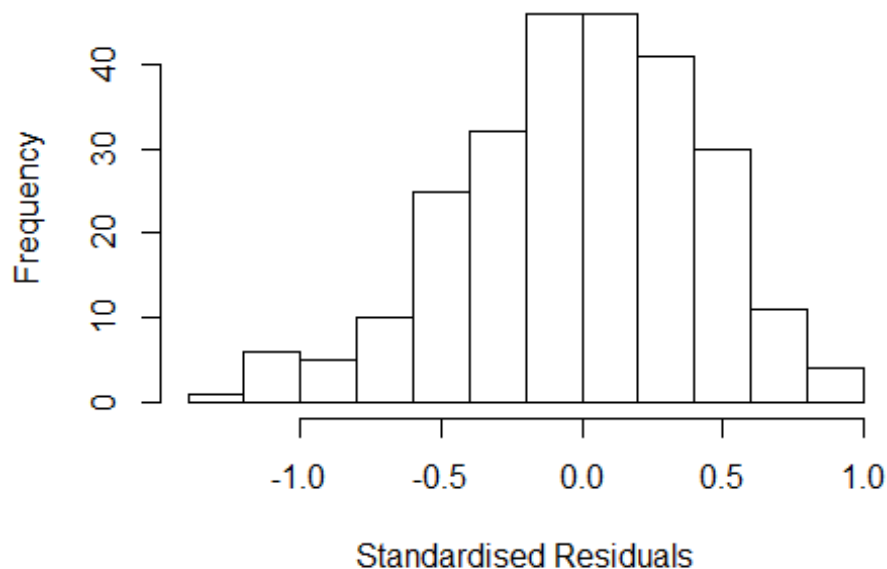


```

## Year2006          0.062190    0.404696    0.15    0.8780
## Year2007         -0.028059    0.418530   -0.07    0.9466
## Year2008          0.041622    0.410904    0.10    0.9194
## Year2009          0.127016    0.409411    0.31    0.7566
## Year2010          0.087942    0.410349    0.21    0.8305
## Year2011         -0.112059    0.415361   -0.27    0.7876
## Year2012          0.255482    0.402445    0.63    0.5262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.0785, Adjusted R-squared:  0.00877
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 236 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.372  0.872  0.955   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      3.89e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.315 1      1.147
## Year              1.315 16      1.009

```

Residuals from first author



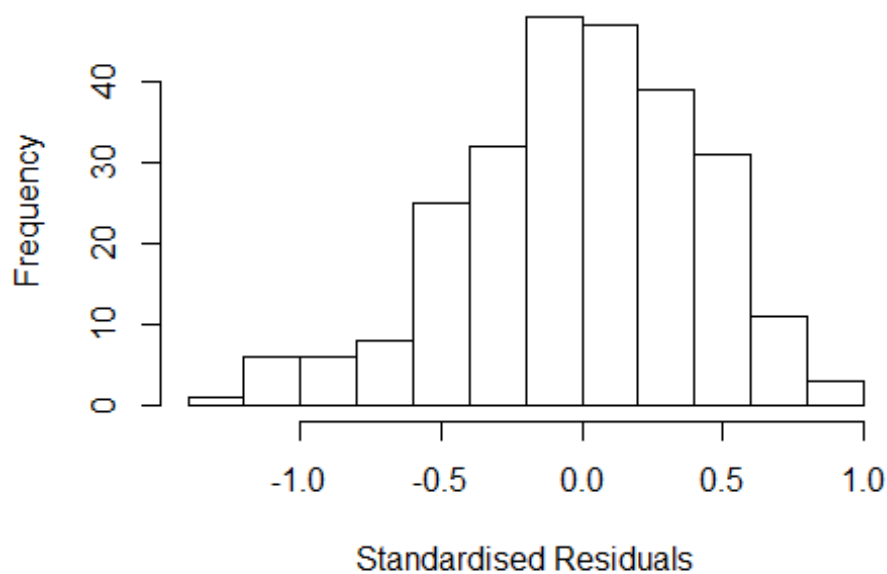
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23199 -0.27662 0.00852 0.27890 0.94888
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15364 0.36772 3.14 0.0019 **
## FirstAuthorFemale1 -0.11502 0.06021 -1.91 0.0573 .
## Year1997 0.06115 0.38323 0.16 0.8734
## Year1998 0.02265 0.39127 0.06 0.9539
## Year1999 0.03237 0.39833 0.08 0.9353
## Year2000 -0.07252 0.40465 -0.18 0.8579
## Year2001 -0.00553 0.38258 -0.01 0.9885
## Year2002 0.12689 0.38073 0.33 0.7392
## Year2003 0.09851 0.38478 0.26 0.7982
## Year2004 -0.14387 0.39878 -0.36 0.7186
## Year2005 -0.23038 0.38277 -0.60 0.5478
## Year2006 0.05263 0.37870 0.14 0.8896
```

```

## Year2007      -0.03740    0.39277   -0.10    0.9242
## Year2008      0.03047    0.38540    0.08    0.9370
## Year2009      0.12097    0.38485    0.31    0.7535
## Year2010      0.07834    0.38466    0.20    0.8388
## Year2011     -0.13059    0.38605   -0.34    0.7355
## Year2012      0.24786    0.37663    0.66    0.5111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0757, Adjusted R-squared:  0.00997
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 234 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.410  0.873  0.956  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      3.89e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.289 1      1.136
## Year            1.289 16      1.008

```

Residuals from last author



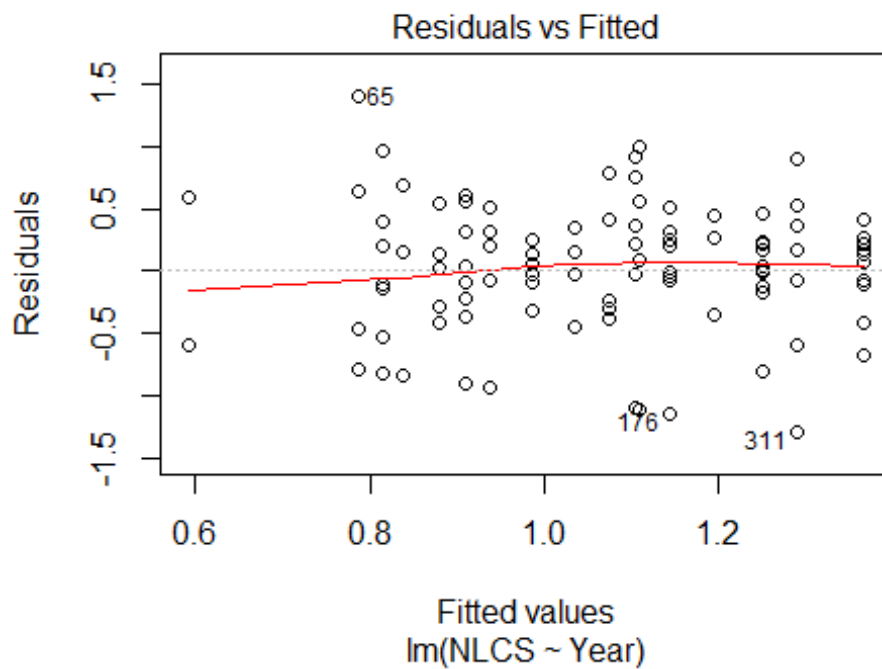
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.18380 -0.27008  0.00274  0.27532  1.06003
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.16302    0.39646    2.93  0.0037 **
## LastAuthorFemale1 -0.06483    0.06777   -0.96  0.3398
## Year1997          0.02134    0.41409    0.05  0.9589
## Year1998         -0.00596    0.41698   -0.01  0.9886
## Year1999         -0.02421    0.42644   -0.06  0.9548
## Year2000         -0.14494    0.42948   -0.34  0.7361
## Year2001         -0.05130    0.41406   -0.12  0.9015
## Year2002          0.07780    0.40901    0.19  0.8493
## Year2003          0.03763    0.41346    0.09  0.9276
## Year2004         -0.17808    0.42401   -0.42  0.6749
## Year2005         -0.26083    0.41316   -0.63  0.5284
## Year2006         -0.00829    0.40753   -0.02  0.9838
```

```

## Year2007          -0.10130      0.41945    -0.24    0.8094
## Year2008          -0.02522      0.41240    -0.06    0.9513
## Year2009           0.07723      0.41439     0.19    0.8523
## Year2010           0.02078      0.41410     0.05    0.9600
## Year2011          -0.18434      0.41589    -0.44    0.6580
## Year2012           0.18316      0.40601     0.45    0.6523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.0654, Adjusted R-squared:  -0.00107
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 239 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.449  0.882  0.956  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      3.89e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 257"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 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
##   23   17   19   17   20   14   11   16   12   15   15   14   16   22   18
## 2011 2012
##   15   20
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    5    5    6    4    3    7   10    5    9   10    6    2    6    9
## 2011 2012

```

```
##      7      7
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      3      2      5      6      4      3      7      9      5      9     10      6      2      6      7
## 2011 2012
##      5      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 = 25, df = 16, p-value = 0.06
```



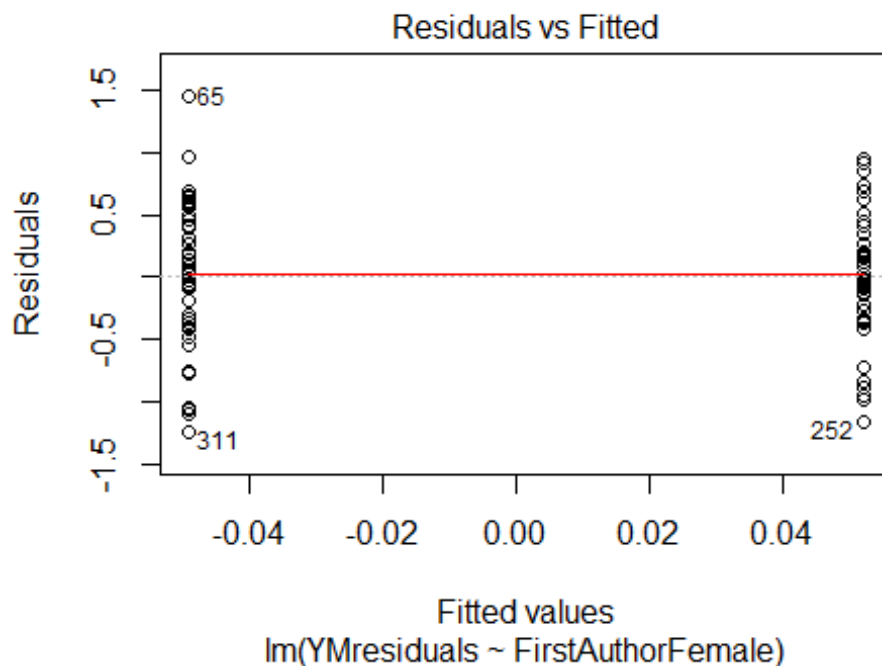
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.8, df = 1, p-value = 0.4

## [1] "Female first author team size 2018 geometric mean: 4.7015092249024"
## [1] "Male first author team size 2018 geometric mean: 3.30192724889463"

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

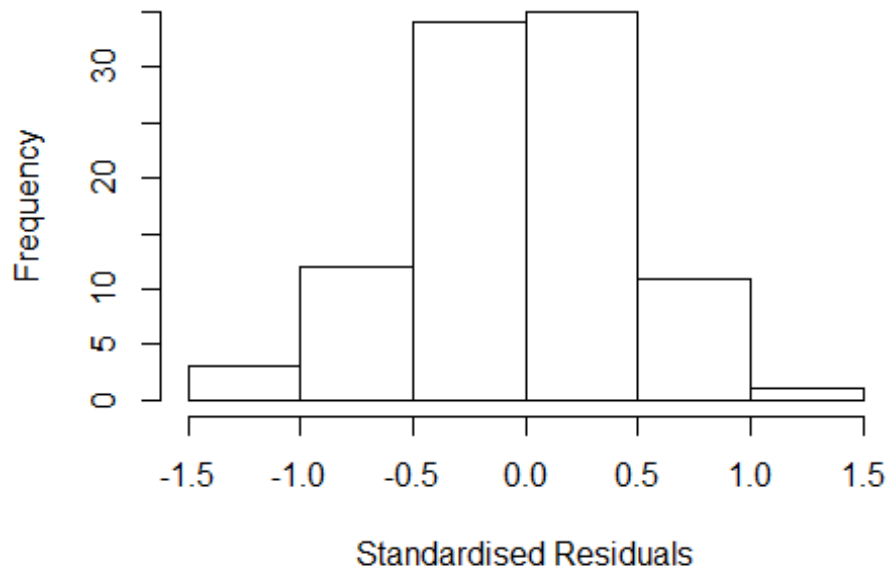
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.77976314968462"
## [1] "Male last author team size 2018 geometric mean: 4.39429035136649"

## Warning in wilcox.test.default(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 = 0.6
## 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  10.83  1      3.291
## LastAuthorFemale  12.77  1      3.574
## UniqueAuthors    224.12  4      1.967
## Year              3355.25 16      1.289
```

Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35067 -0.32550 -0.00492 0.31904 1.48930
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4208 0.4296 3.31 0.0015 **
## FirstAuthorFemale1 -0.1355 0.2028 -0.67 0.5061
## LastAuthorFemale1 -0.1850 0.1947 -0.95 0.3452
## UniqueAuthors2 0.2780 0.2543 1.09 0.2779
## UniqueAuthors3 0.3621 0.2565 1.41 0.1622
## UniqueAuthors4 0.3363 0.2372 1.42 0.1606
## UniqueAuthors5 0.1636 0.2772 0.59 0.5568
## Year1997 -0.4034 0.3246 -1.24 0.2179
## Year1998 -0.9487 0.8101 -1.17 0.2454
## Year1999 -0.5951 0.3676 -1.62 0.1098
```

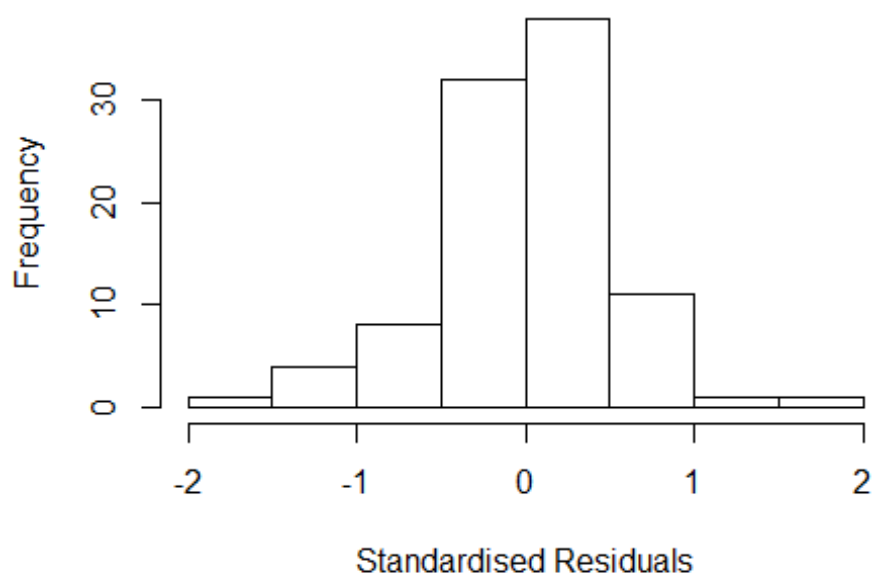


```

## Year2000          -0.3402      0.3242   -1.05    0.2975
## Year2001          -0.7207      0.5330   -1.35    0.1805
## Year2002          -0.7270      0.3795   -1.92    0.0593 .
## Year2003          -0.1490      0.3469   -0.43    0.6688
## Year2004          -0.5384      0.3917   -1.37    0.1734
## Year2005          -0.3254      0.3882   -0.84    0.4046
## Year2006          -0.3062      0.3575   -0.86    0.3945
## Year2007          -0.4763      0.4593   -1.04    0.3031
## Year2008          -0.6482      0.5227   -1.24    0.2189
## Year2009          -0.0701      0.6441   -0.11    0.9137
## Year2010          -0.5623      0.3654   -1.54    0.1282
## Year2011          -0.1442      0.5203   -0.28    0.7824
## Year2012          -0.1576      0.4823   -0.33    0.7448
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.451
## Multiple R-squared:  0.286, Adjusted R-squared:  0.0706
## Convergence in 33 IRWLS iterations
##
## Robustness weights:
## 5 weights are ~= 1. The remaining 91 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.254  0.876  0.951  0.897  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      1.04e-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 7.842 1          2.800
## LastAuthorFemale 11.716 1          3.423
## Year              57.584 16          1.135

```

Residuals from first and last author



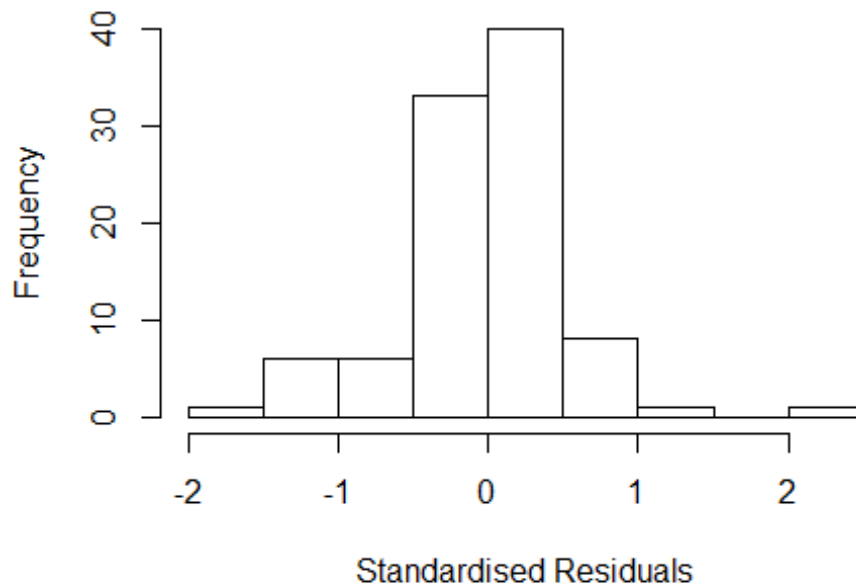
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7454 -0.3036 0.0467 0.2713 1.9050
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.6958 0.2722 6.23 2.3e-08 ***
## FirstAuthorFemale1 -0.1598 0.1779 -0.90 0.372
## LastAuthorFemale1 -0.2105 0.1867 -1.13 0.263
## Year1997 -0.2908 0.2888 -1.01 0.317
## Year1998 -1.2530 0.6855 -1.83 0.071 .
## Year1999 -0.6490 0.2815 -2.31 0.024 *
## Year2000 -0.3928 0.2590 -1.52 0.133
## Year2001 -0.6755 0.5118 -1.32 0.191
## Year2002 -0.7458 0.2984 -2.50 0.015 *
## Year2003 -0.2130 0.2729 -0.78 0.438
## Year2004 -0.6171 0.3203 -1.93 0.058 .
## Year2005 -0.4117 0.2892 -1.42 0.159
```

```

## Year2006          -0.3336      0.2745    -1.22      0.228
## Year2007          -0.5532      0.3575    -1.55      0.126
## Year2008          -0.7345      0.6920    -1.06      0.292
## Year2009           0.0496      0.4332      0.11      0.909
## Year2010          -0.5531      0.2900    -1.91      0.060 .
## Year2011          -0.0513      0.5425    -0.09      0.925
## Year2012          -0.1555      0.4482    -0.35      0.730
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.305, Adjusted R-squared:  0.143
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 89 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0171 0.8610 0.9580 0.8770 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      1.04e-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 4.221 1      2.055
## Year              4.221 16      1.046

```

Residuals from first author



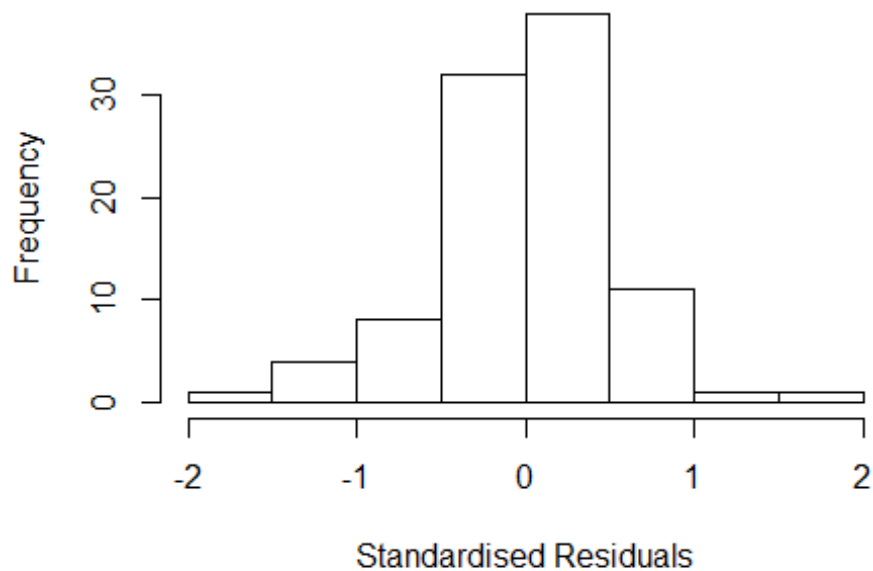
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7554 -0.3017 0.0145 0.2598 2.1104
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5820 0.2630 6.02 5.5e-08 ***
## FirstAuthorFemale1 -0.2558 0.1481 -1.73 0.088 .
## Year1997 -0.3875 0.2658 -1.46 0.149
## Year1998 -1.2485 0.4739 -2.63 0.010 *
## Year1999 -0.5532 0.2729 -2.03 0.046 *
## Year2000 -0.2848 0.2657 -1.07 0.287
## Year2001 -0.4845 0.5495 -0.88 0.381
## Year2002 -0.6329 0.2849 -2.22 0.029 *
## Year2003 -0.0682 0.2478 -0.28 0.784
## Year2004 -0.5685 0.3066 -1.85 0.067 .
## Year2005 -0.3377 0.2852 -1.18 0.240
## Year2006 -0.2434 0.2602 -0.94 0.352
```

```

## Year2007          -0.4102      0.3097   -1.32    0.189
## Year2008          -0.7351      0.7101   -1.04    0.304
## Year2009           0.1734      0.3911    0.44    0.659
## Year2010          -0.5791      0.2824   -2.05    0.044 *
## Year2011          -0.0613      0.6045   -0.10    0.920
## Year2012          -0.1466      0.4732   -0.31    0.758
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.428
## Multiple R-squared:  0.314, Adjusted R-squared:  0.164
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## observation 10 is an outlier with |weight| = 0 ( < 0.001);
## 7 weights are ~= 1. The remaining 88 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0541 0.8780 0.9550 0.8790 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      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 8.29 1          2.879
## Year            8.29 16          1.068

```

Residuals from last author



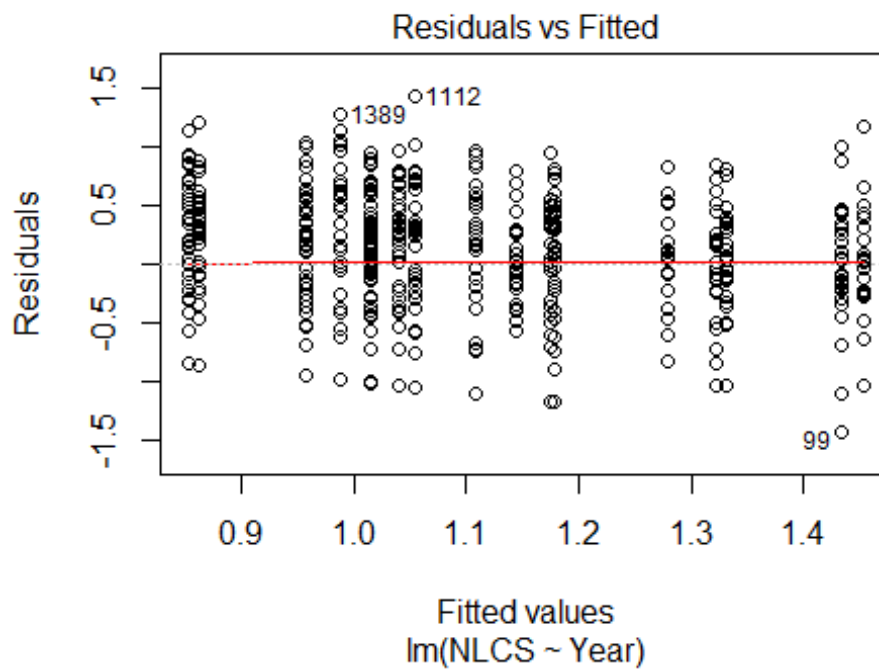
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6235 -0.2918 0.0566 0.2237 1.7059
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.6127 0.2605 6.19 2.6e-08 ***
## LastAuthorFemale1 -0.2872 0.1598 -1.80 0.076 .
## Year1997 -0.1310 0.2197 -0.60 0.553
## Year1998 -1.1305 0.6749 -1.68 0.098 .
## Year1999 -0.5835 0.2785 -2.10 0.039 *
## Year2000 -0.4322 0.2490 -1.74 0.087 .
## Year2001 -0.6632 0.4552 -1.46 0.149
## Year2002 -0.7379 0.3144 -2.35 0.021 *
## Year2003 -0.2150 0.2749 -0.78 0.437
## Year2004 -0.5807 0.3039 -1.91 0.060 .
## Year2005 -0.3285 0.2739 -1.20 0.234
## Year2006 -0.2778 0.2764 -1.01 0.318
```

```

## Year2007          -0.5469      0.3639    -1.50      0.137
## Year2008          -0.7345      0.6926    -1.06      0.292
## Year2009           0.0109      0.4851      0.02      0.982
## Year2010          -0.4653      0.2845    -1.64      0.106
## Year2011           0.0471      0.5321      0.09      0.930
## Year2012          -0.0779      0.4229    -0.18      0.854
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.282, Adjusted R-squared:  0.125
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 90 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0932 0.8420 0.9580 0.8780 0.9910 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      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 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
##   87   86   86   72   49   69   52   38   71   52   59   72   64   69   72
## 2011 2012
##   63   49
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   30   24   28   38   24   27   37   23   51   38   50   49   42   44   48
## 2011 2012

```

```
## 44 36
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 25 20 24 33 19 24 30 18 46 34 47 42 40 40 37
## 2011 2012
## 40 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 = 28, df = 16, p-value = 0.03
```

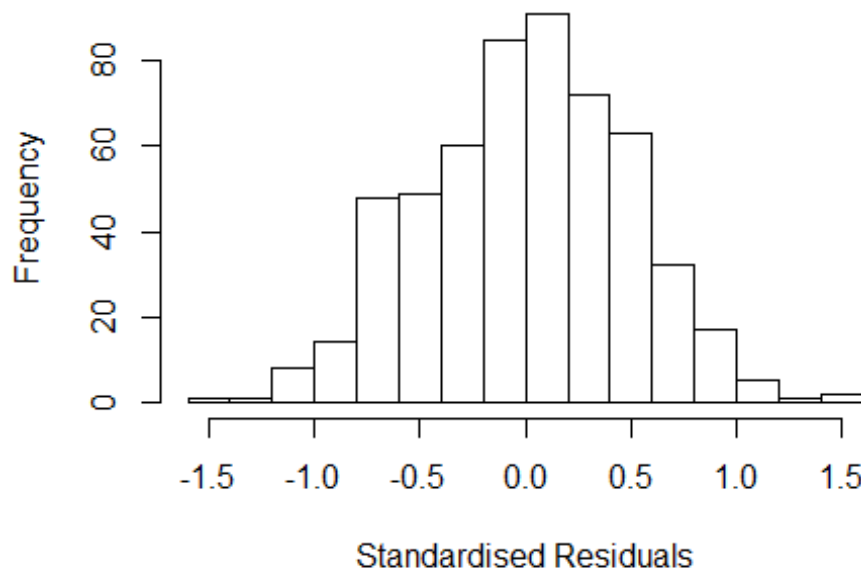


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.7, df = 1, p-value = 0.4

## [1] "Female first author team size 2018 geometric mean: 3.25836787338959"
## [1] "Male first author team size 2018 geometric mean: 3.10369114783072"

## 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



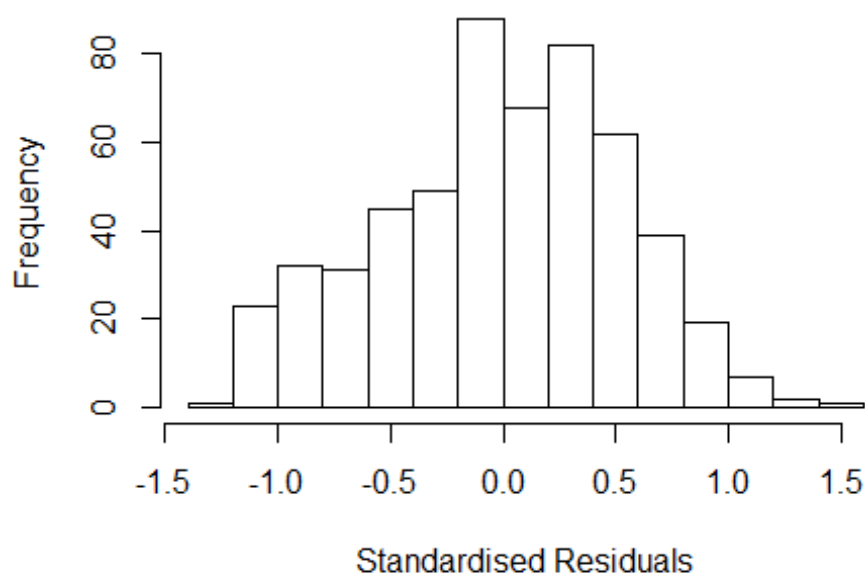
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4055 -0.3504  0.0331  0.3137  1.5730
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.95195    0.10381   9.17 < 2e-16 ***
## FirstAuthorFemale1 0.00480    0.04842   0.10  0.9210
## LastAuthorFemale1 -0.03503    0.05478  -0.64  0.5228
## UniqueAuthors2    0.43217    0.07126   6.06 2.5e-09 ***
## UniqueAuthors3    0.51958    0.06338   8.20 1.9e-15 ***
## UniqueAuthors4    0.53972    0.07776   6.94 1.2e-11 ***
## UniqueAuthors5    0.63411    0.07330   8.65 < 2e-16 ***
## Year1997         -0.02441    0.14109  -0.17  0.8627
## Year1998         -0.00222    0.12672  -0.02  0.9860
## Year1999         -0.08471    0.12334  -0.69  0.4925
```

```

## Year2000      -0.05263    0.14084   -0.37    0.7088
## Year2001      -0.06599    0.14622   -0.45    0.6519
## Year2002      -0.14691    0.11316   -1.30    0.1948
## Year2003       0.10801    0.14785    0.73    0.4654
## Year2004      -0.36336    0.11226   -3.24    0.0013 **
## Year2005      -0.32420    0.11446   -2.83    0.0048 **
## Year2006      -0.45970    0.12293   -3.74    0.0002 ***
## Year2007      -0.25137    0.11930   -2.11    0.0356 *
## Year2008      -0.24921    0.12222   -2.04    0.0420 *
## Year2009      -0.22402    0.11825   -1.89    0.0587 .
## Year2010      -0.32765    0.13333   -2.46    0.0143 *
## Year2011      -0.22177    0.13948   -1.59    0.1124
## Year2012      -0.27240    0.13121   -2.08    0.0384 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.506
## Multiple R-squared:  0.266, Adjusted R-squared:  0.236
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 510 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.314  0.872  0.952   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      1.82e-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 1.154 1      1.074
## LastAuthorFemale  1.138 1      1.067
## Year              1.310 16      1.008

```

Residuals from first and last author



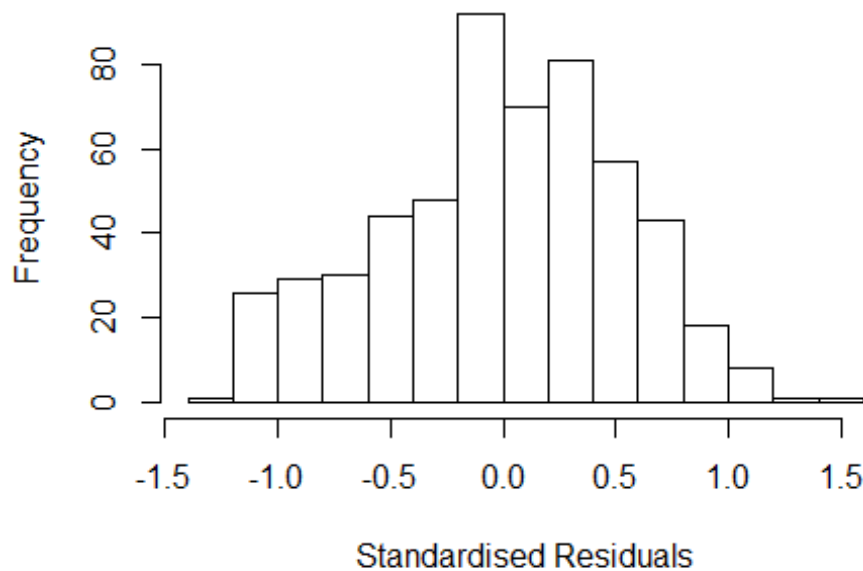
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2161 -0.3736 0.0201 0.3699 1.4067
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4116 0.0801 17.62 < 2e-16 ***
## FirstAuthorFemale1 0.0711 0.0516 1.38 0.16947
## LastAuthorFemale1 -0.0346 0.0624 -0.55 0.57928
## Year1997 -0.1668 0.1426 -1.17 0.24258
## Year1998 -0.1236 0.1138 -1.09 0.27810
## Year1999 -0.2507 0.1040 -2.41 0.01626 *
## Year2000 -0.1331 0.1304 -1.02 0.30780
## Year2001 -0.2287 0.1364 -1.68 0.09420 .
## Year2002 -0.1955 0.1239 -1.58 0.11532
## Year2003 0.1360 0.1292 1.05 0.29298
## Year2004 -0.4681 0.1209 -3.87 0.00012 ***
## Year2005 -0.4345 0.1173 -3.70 0.00024 ***
```

```

## Year2006          -0.5949      0.1207    -4.93  1.1e-06 ***
## Year2007          -0.3885      0.1191    -3.26  0.00117 **
## Year2008          -0.3294      0.1329    -2.48  0.01350 *
## Year2009          -0.3117      0.1302    -2.39  0.01702 *
## Year2010          -0.4922      0.1337    -3.68  0.00025 ***
## Year2011          -0.4140      0.1644    -2.52  0.01207 *
## Year2012          -0.3012      0.1370    -2.20  0.02828 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.0906, Adjusted R-squared:  0.0597
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 516 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.480  0.863  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      1.82e-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 1.155 1      1.075
## Year              1.155 16      1.005

```

Residuals from first author



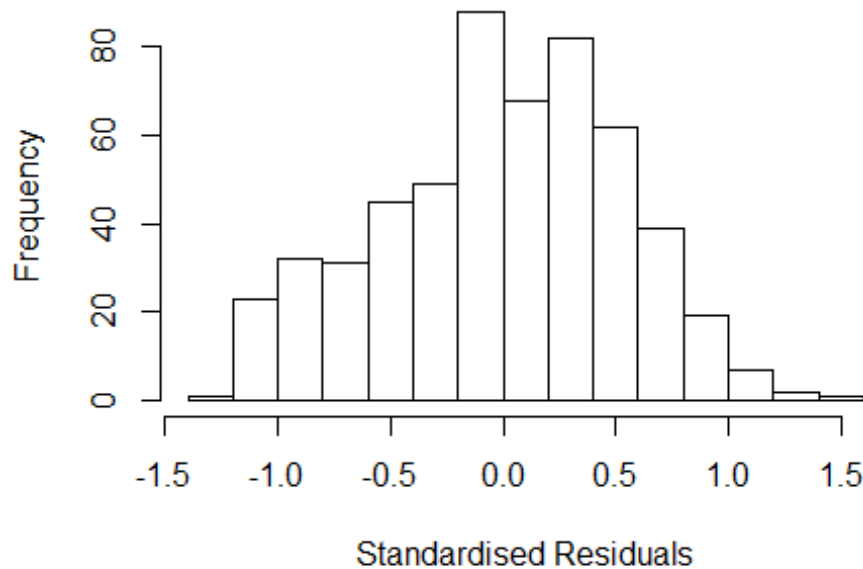
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2128 -0.3773 0.0248 0.3749 1.4175
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4086 0.0797 17.67 < 2e-16 ***
## FirstAuthorFemale1 0.0637 0.0534 1.19 0.23407
## Year1997 -0.1635 0.1425 -1.15 0.25163
## Year1998 -0.1219 0.1134 -1.07 0.28295
## Year1999 -0.2493 0.1037 -2.40 0.01657 *
## Year2000 -0.1333 0.1300 -1.03 0.30553
## Year2001 -0.2350 0.1369 -1.72 0.08667 .
## Year2002 -0.1959 0.1241 -1.58 0.11499
## Year2003 0.1281 0.1274 1.01 0.31519
## Year2004 -0.4695 0.1216 -3.86 0.00013 ***
## Year2005 -0.4415 0.1165 -3.79 0.00017 ***
## Year2006 -0.5950 0.1210 -4.92 1.2e-06 ***
```

```

## Year2007          -0.3897      0.1193   -3.27  0.00116 **
## Year2008          -0.3372      0.1318   -2.56  0.01080 *
## Year2009          -0.3160      0.1300   -2.43  0.01541 *
## Year2010          -0.4965      0.1334   -3.72  0.00022 ***
## Year2011          -0.4147      0.1642   -2.53  0.01182 *
## Year2012          -0.3059      0.1370   -2.23  0.02596 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.541
## Multiple R-squared:  0.0903, Adjusted R-squared:  0.0611
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 35 weights are ~= 1. The remaining 514 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.472  0.862  0.950  0.909  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.82e-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 1.158 1      1.076
## Year              1.158 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
## -1.2354 -0.3815 0.0279 0.3631 1.3928
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4239 0.0819 17.39 < 2e-16 ***
## LastAuthorFemale1 -0.0153 0.0631 -0.24 0.80852
## Year1997 -0.1532 0.1424 -1.08 0.28252
## Year1998 -0.1140 0.1172 -0.97 0.33099
## Year1999 -0.2510 0.1059 -2.37 0.01815 *
## Year2000 -0.1281 0.1314 -0.97 0.33003
## Year2001 -0.2186 0.1353 -1.62 0.10682
## Year2002 -0.1885 0.1262 -1.49 0.13584
## Year2003 0.1306 0.1295 1.01 0.31389
## Year2004 -0.4648 0.1224 -3.80 0.00016 ***
## Year2005 -0.4281 0.1174 -3.65 0.00029 ***
## Year2006 -0.5937 0.1227 -4.84 1.7e-06 ***
```



```

## Year2007          -0.3898      0.1217    -3.20   0.00144 **
## Year2008          -0.3278      0.1346    -2.44   0.01518 *
## Year2009          -0.3018      0.1310    -2.30   0.02161 *
## Year2010          -0.4890      0.1347    -3.63   0.00031 ***
## Year2011          -0.4046      0.1679    -2.41   0.01630 *
## Year2012          -0.2846      0.1374    -2.07   0.03873 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.0882, Adjusted R-squared:  0.059
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 518 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.486  0.862  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      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: 549"
## [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]"
##
## 1997 1998 2003 2005 2006 2009 2010 2011 2012
##    1    2    1    1    2    5    3    4    5
##
## 1997 1998 2003 2005 2006 2009 2010 2011 2012
##    0    0    0    0    0    1    2    1    4
##
## 1997 1998 2003 2005 2006 2009 2010 2011 2012
##    0    0    0    0    0    1    2    0    3

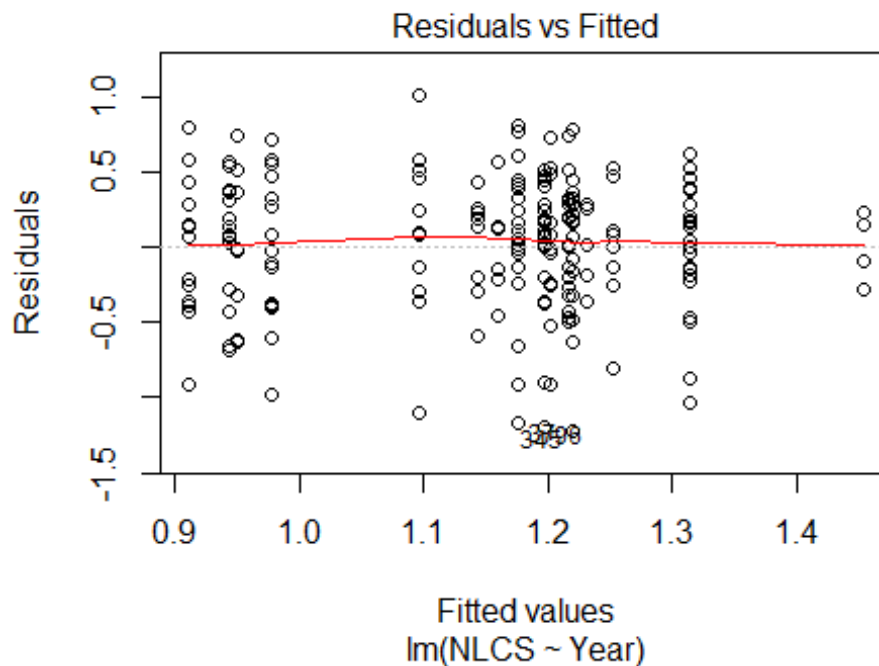
```

```

## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: NaN"
## [1] "Male first author team size 2018 geometric mean: 2"
## [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 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
## 28 29 43 26 22 19 21 22 7 8 7 10 9 20 17
## 2011 2012
## 13 17
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 20 13 22 11 12 12 15 17 4 6 5 8 8 16 14
## 2011 2012
## 10 15
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 16 11 20 11 10 12 14 13 3 6 5 8 5 14 12

```

```
## 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 = 13, df = 16, p-value = 0.7
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.2, df = 1, p-value = 0.7

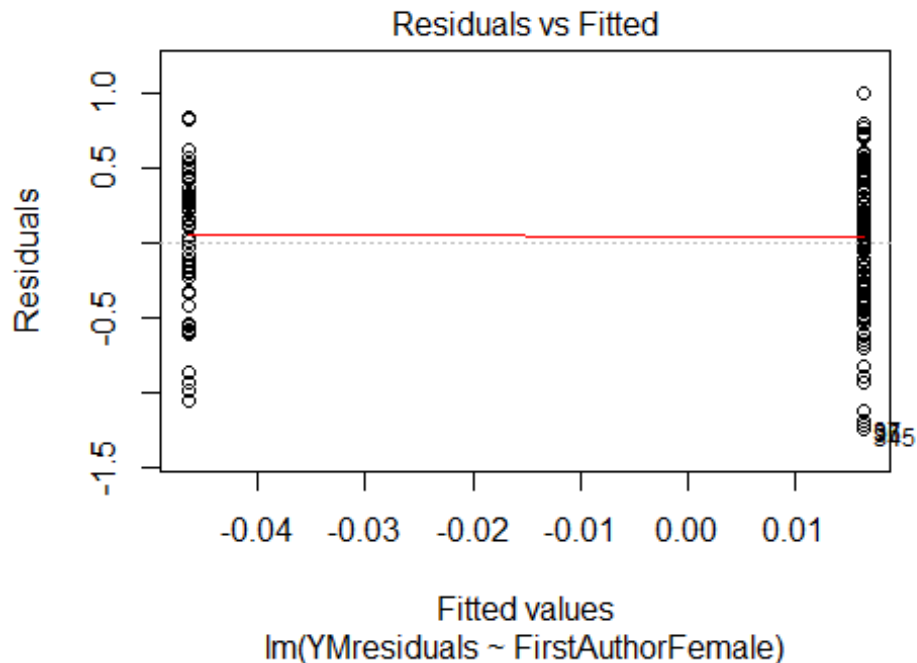
## [1] "Female first author team size 2018 geometric mean: 3.30192724889463"
## [1] "Male first author team size 2018 geometric mean: 3.15377654392853"

## Warning in wilcox.test.default(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 = 0.8
```

```
## 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: 3.6808429993847"

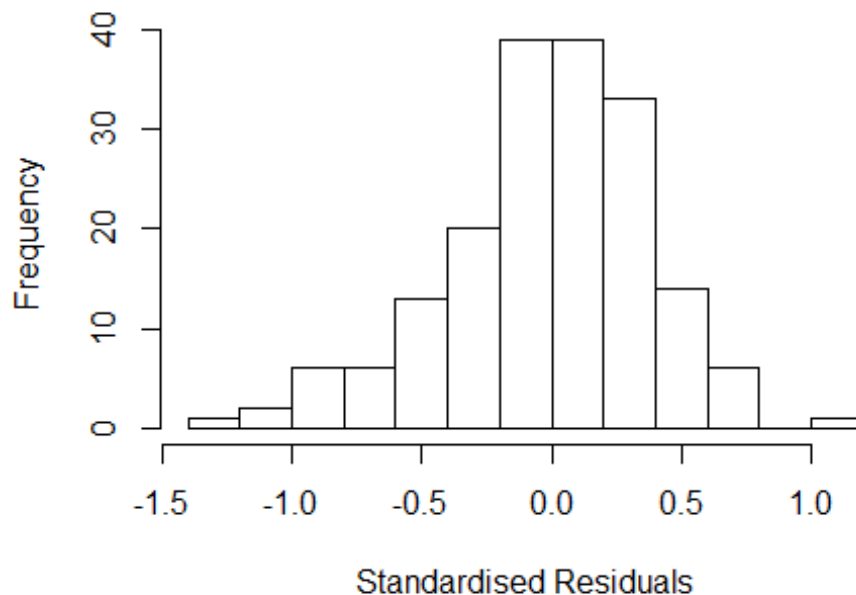
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	1.832	1	1.354
LastAuthorFemale	1.721	1	1.312
UniqueAuthors	7.186	4	1.280
Year	8.569	16	1.069

Residuals from first and last author and team size



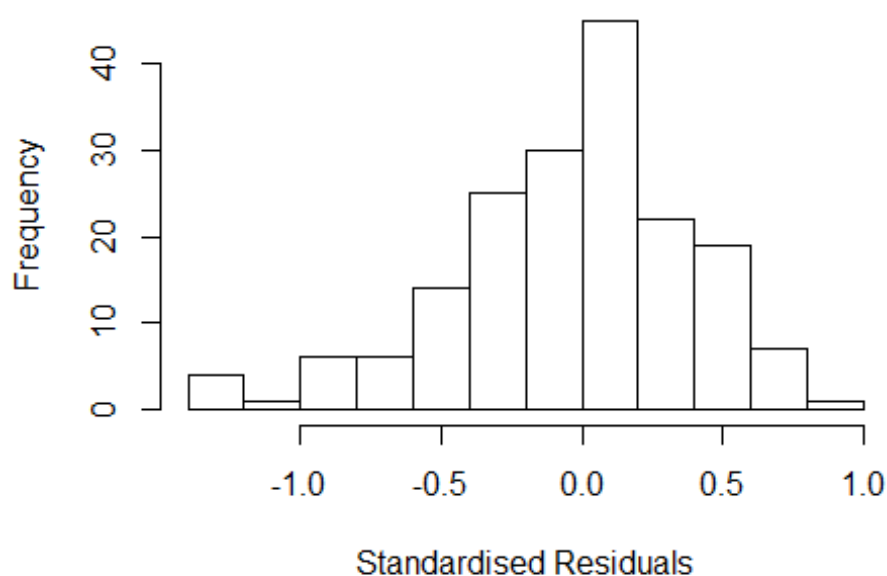
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3713 -0.2085 0.0137 0.2380 1.0790
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.857290 0.199253 4.30 3e-05 ***
## FirstAuthorFemale1 -0.011741 0.079489 -0.15 0.88277
## LastAuthorFemale1 0.026456 0.102328 0.26 0.79633
## UniqueAuthors2 0.239596 0.142420 1.68 0.09449 .
## UniqueAuthors3 0.332161 0.134761 2.46 0.01479 *
## UniqueAuthors4 0.400955 0.145019 2.76 0.00638 **
## UniqueAuthors5 0.536100 0.137119 3.91 0.00014 ***
## Year1997 0.129139 0.177014 0.73 0.46676
## Year1998 0.106270 0.171317 0.62 0.53595
## Year1999 0.084315 0.187356 0.45 0.65331
```

```

## Year2000      0.079157    0.213355    0.37  0.71113
## Year2001      0.138146    0.150776    0.92  0.36095
## Year2002     -0.125412    0.187339   -0.67  0.50420
## Year2003      0.138367    0.150716    0.92  0.35999
## Year2004      0.189798    0.176201    1.08  0.28306
## Year2005     -0.000125    0.182747    0.00  0.99946
## Year2006      0.002714    0.160981    0.02  0.98657
## Year2007      0.061238    0.227562    0.27  0.78820
## Year2008     -0.243910    0.248785   -0.98  0.32840
## Year2009      0.181843    0.184356    0.99  0.32547
## Year2010     -0.275960    0.206649   -1.34  0.18368
## Year2011      0.025953    0.158701    0.16  0.87031
## Year2012     -0.221604    0.179647   -1.23  0.21921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.385
## Multiple R-squared:  0.243, Adjusted R-squared:  0.137
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 156 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.178  0.875  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      5.56e-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 1.311 1      1.145
## LastAuthorFemale  1.379 1      1.174
## Year              1.671 16      1.016

```

Residuals from first and last author



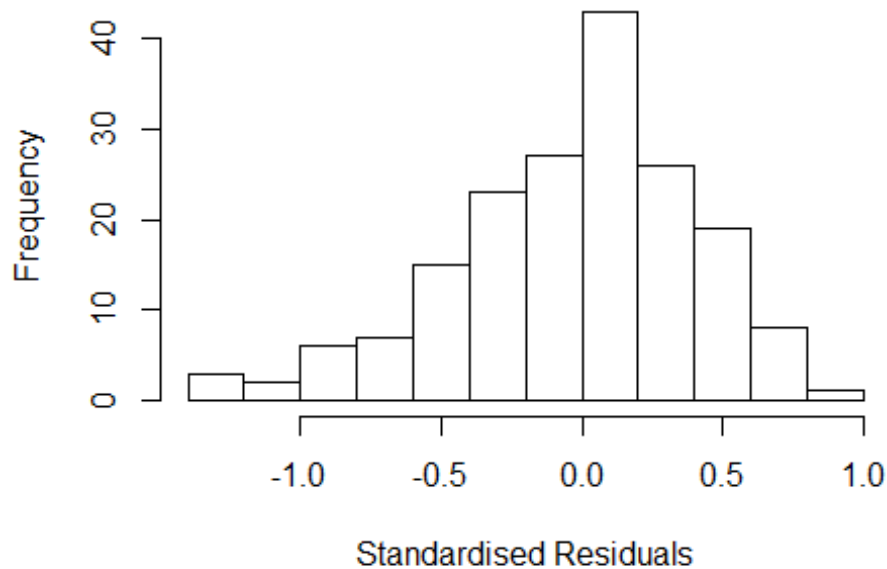
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3206 -0.2579 0.0247 0.2354 0.9323
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20389 0.12322 9.77 <2e-16 ***
## FirstAuthorFemale1 -0.01261 0.07527 -0.17 0.87
## LastAuthorFemale1 0.08811 0.09246 0.95 0.34
## Year1997 0.11092 0.18014 0.62 0.54
## Year1998 0.00829 0.14663 0.06 0.95
## Year1999 0.01817 0.19329 0.09 0.93
## Year2000 -0.03018 0.20202 -0.15 0.88
## Year2001 0.13534 0.15542 0.87 0.39
## Year2002 -0.25099 0.18331 -1.37 0.17
## Year2003 0.07263 0.15400 0.47 0.64
## Year2004 0.31397 0.14372 2.18 0.03 *
## Year2005 -0.04732 0.18628 -0.25 0.80
```

```

## Year2006          0.01284    0.17490    0.07    0.94
## Year2007          0.06135    0.18961    0.32    0.75
## Year2008         -0.21178    0.33747   -0.63    0.53
## Year2009          0.11667    0.16156    0.72    0.47
## Year2010         -0.26097    0.18972   -1.38    0.17
## Year2011          0.04460    0.16025    0.28    0.78
## Year2012         -0.22057    0.16887   -1.31    0.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.108, Adjusted R-squared:  0.00882
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 162 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.280  0.874  0.956  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      5.56e-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 1.25 1         1.118
## Year              1.25 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
## -1.3337 -0.2704 0.0455 0.2504 0.9204
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21882 0.12403 9.83 <2e-16 ***
## FirstAuthorFemale1 -0.00914 0.07606 -0.12 0.904
## Year1997 0.09981 0.17995 0.55 0.580
## Year1998 0.00537 0.14870 0.04 0.971
## Year1999 0.00167 0.19397 0.01 0.993
## Year2000 -0.03325 0.21354 -0.16 0.876
## Year2001 0.12570 0.15911 0.79 0.431
## Year2002 -0.25329 0.18213 -1.39 0.166
## Year2003 0.07509 0.16198 0.46 0.644
## Year2004 0.32921 0.14749 2.23 0.027 *
## Year2005 -0.06311 0.18752 -0.34 0.737
## Year2006 0.01529 0.17147 0.09 0.929
```

```

## Year2007          0.05829    0.19192    0.30    0.762
## Year2008          -0.21818    0.31425   -0.69    0.488
## Year2009          0.11489    0.16473    0.70    0.487
## Year2010          -0.26910    0.18891   -1.42    0.156
## Year2011          0.03807    0.15679    0.24    0.808
## Year2012          -0.21345    0.17842   -1.20    0.233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.418
## Multiple R-squared:  0.101, Adjusted R-squared:  0.00716
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 163 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.289  0.884  0.960  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      5.56e-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 1.341 1          1.158
## Year              1.341 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
## -1.3174 -0.2589  0.0275  0.2301  0.9343

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20052    0.12319   9.75  <2e-16 ***
## LastAuthorFemale1 0.08736    0.09338   0.94    0.35
## Year1997        0.10920    0.17812   0.61    0.54
## Year1998        0.00799    0.14546   0.05    0.96
## Year1999        0.01849    0.19349   0.10    0.92
## Year2000       -0.02881    0.20044  -0.14    0.89
## Year2001        0.13530    0.15448   0.88    0.38
## Year2002       -0.25044    0.18251  -1.37    0.17
## Year2003        0.07616    0.15448   0.49    0.62
## Year2004        0.31760    0.14470   2.19    0.03 *
## Year2005       -0.04812    0.18590  -0.26    0.80
## Year2006        0.01387    0.17288   0.08    0.94
## Year2007        0.05947    0.18611   0.32    0.75
## Year2008       -0.20925    0.33865  -0.62    0.54
## Year2009        0.11687    0.16112   0.73    0.47
## Year2010       -0.26265    0.18998  -1.38    0.17
## Year2011        0.04416    0.15814   0.28    0.78
## Year2012       -0.21979    0.16945  -1.30    0.20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0148
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 161 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.281  0.876   0.956   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      5.56e-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: 180"
## [1] ""

```

```

## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3505"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2007 2008 2009 2010 2011
##    4   14    7   11    3    9    6    6    1    2    1    4    9   10   12
## 2012
##    6
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2007 2008 2009 2010 2011
##    0    1    0    1    1    7    5    5    1    2    0    3    7    9   11
## 2012
##    3
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2007 2008 2009 2010 2011
##    0    1    0    1    1    7    4    3    1    2    0    3    6    8    9
## 2012
##    3
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.44948974278318"
## [1] "Male first author team size 2018 geometric mean: 3.20108587294368"

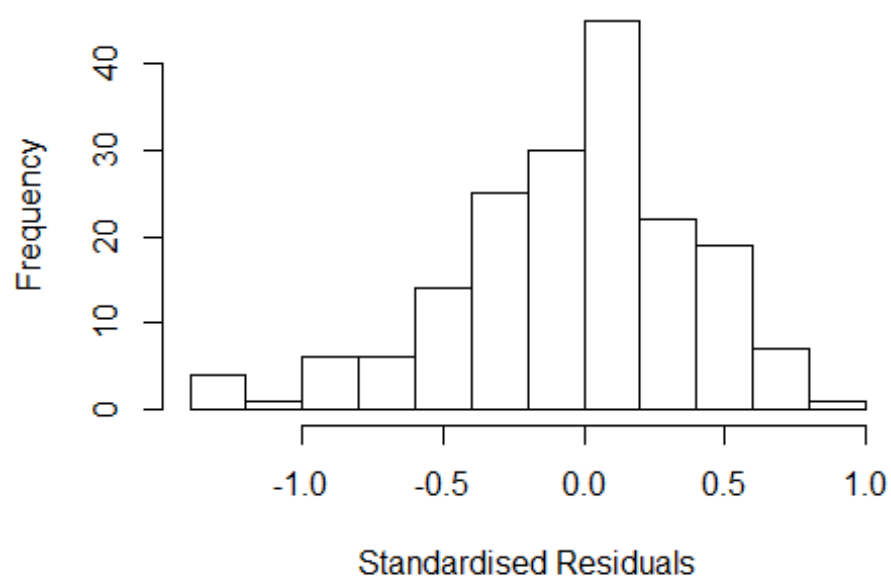
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2.5, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.47602664488645"
## [1] "Male last author team size 2018 geometric mean: 2.46621207433047"

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

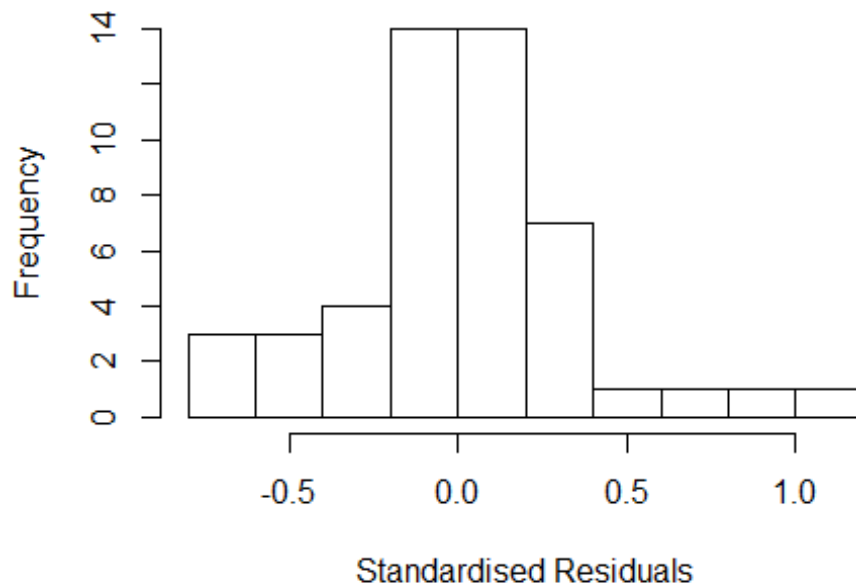
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5.5, 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 1.725e+14  1      1.313e+07
## LastAuthorFemale  1.573e+01  1       3.966e+00
## UniqueAuthors    4.648e+32  4       1.212e+04
## Year              3.228e+46 12       8.667e+01
```

Residuals from first and last author and team size



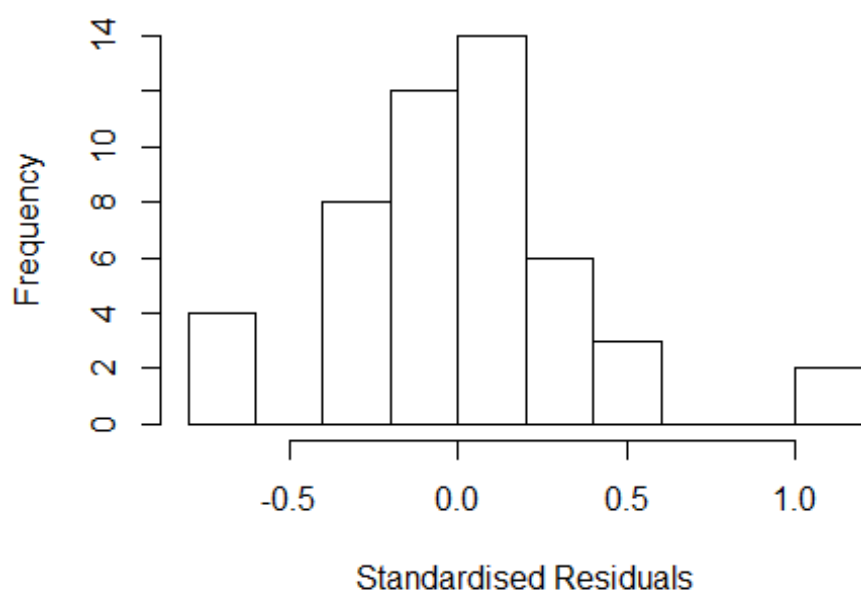
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.707 -0.183 0.005 0.193 1.159
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.139 0.579 0.24 0.81177
## FirstAuthorFemale1 0.130 0.166 0.78 0.43945
## LastAuthorFemale1 -0.639 1.098 -0.58 0.56484
## UniqueAuthors2 0.239 0.679 0.35 0.72700
## UniqueAuthors3 0.122 0.628 0.19 0.84749
## UniqueAuthors4 0.204 0.683 0.30 0.76671
## UniqueAuthors5 0.327 0.657 0.50 0.62243
## Year1999 0.854 0.325 2.63 0.01344 *
## Year2000 1.089 0.274 3.97 0.00041 ***
## Year2001 1.063 0.278 3.83 0.00062 ***
```

```

## Year2002          0.845      0.283      2.98  0.00561 **
## Year2003          0.767      0.266      2.88  0.00722 **
## Year2004          1.087      0.166      6.55  3.1e-07 ***
## Year2005          0.893      0.269      3.32  0.00237 **
## Year2008          1.139      0.284      4.01  0.00038 ***
## Year2009          0.930      0.382      2.43  0.02122 *
## Year2010          0.807      0.348      2.32  0.02742 *
## Year2011          1.006      0.281      3.58  0.00119 **
## Year2012          0.838      0.240      3.49  0.00151 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.361, Adjusted R-squared:  -0.0232
## Convergence in 43 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 41 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.187  0.891  0.962   0.894   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      2.04e-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 -4.094e+15  1      NaN
## LastAuthorFemale  8.627e+00  1      2.937
## Year              -3.530e+16 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
## -0.71017 -0.16295 0.00281 0.19529 1.19877
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.261 0.161 1.62 0.11545
## FirstAuthorFemale1 0.130 0.161 0.81 0.42601
## LastAuthorFemale1 -0.847 0.811 -1.05 0.30330
## Year1999 0.971 0.161 6.01 8.3e-07 ***
## Year2000 1.294 0.161 8.01 2.4e-09 ***
## Year2001 1.156 0.213 5.42 5.0e-06 ***
## Year2002 0.942 0.273 3.46 0.00149 **
## Year2003 0.886 0.139 6.40 2.6e-07 ***
## Year2004 1.087 0.161 6.73 9.8e-08 ***
## Year2005 0.976 0.162 6.04 7.6e-07 ***
## Year2008 1.160 0.274 4.24 0.00016 ***
## Year2009 1.094 0.252 4.34 0.00012 ***
```



```

## Year2010          0.936      0.206      4.55  6.5e-05 ***
## Year2011          1.167      0.298      3.92  0.00041 ***
## Year2012          0.860      0.169      5.08  1.4e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.366, Adjusted R-squared:  0.105
## Convergence in 31 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 37 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.180  0.890  0.957  0.882  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      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"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale -1.138e+25  1          NaN
## Year              -1.138e+25 12          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
## -1.04e+00 -1.54e-01 -8.88e-16  2.21e-01  1.13e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)

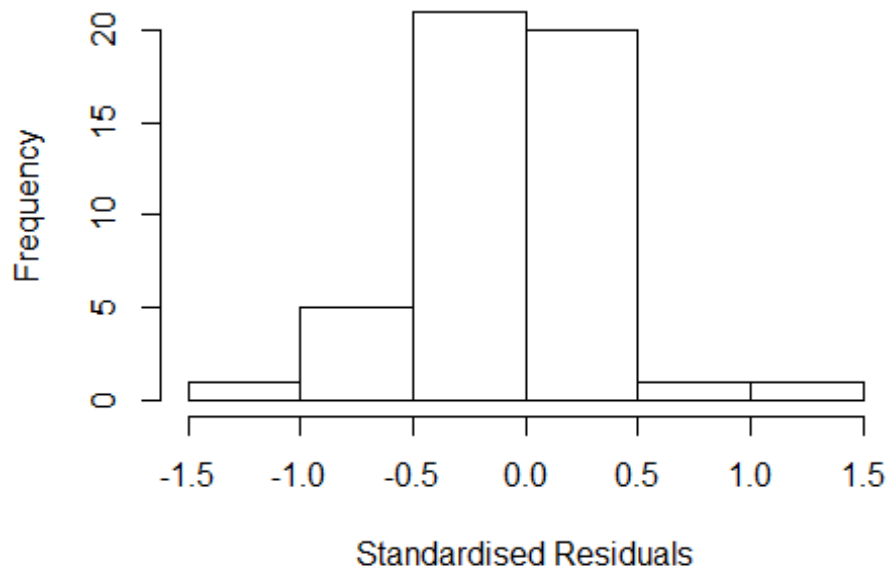
```

```

## (Intercept)      0.4156      0.3508      1.18      0.2440
## FirstAuthorFemale1 -0.0246      0.3508     -0.07      0.9444
## Year1999          0.8164      0.3508      2.33      0.0258 *
## Year2000          1.1394      0.3508      3.25      0.0026 **
## Year2001          0.8806      0.3261      2.70      0.0106 *
## Year2002          0.7805      0.4525      1.72      0.0934 .
## Year2003          0.7856      0.2491      3.15      0.0033 **
## Year2004          0.9324      0.3508      2.66      0.0118 *
## Year2005          0.8214      0.3508      2.34      0.0250 *
## Year2008          0.9798      0.9128      1.07      0.2905
## Year2009          1.0109      0.3220      3.14      0.0034 **
## Year2010          0.7942      0.3831      2.07      0.0456 *
## Year2011          1.0384      0.3184      3.26      0.0025 **
## Year2012          0.7588      0.2499      3.04      0.0045 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.355
## Multiple R-squared:  0.206, Adjusted R-squared:  -0.0891
## Convergence in 31 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 40 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.291  0.880  0.957  0.888  0.985  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           2.04e-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"
## 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

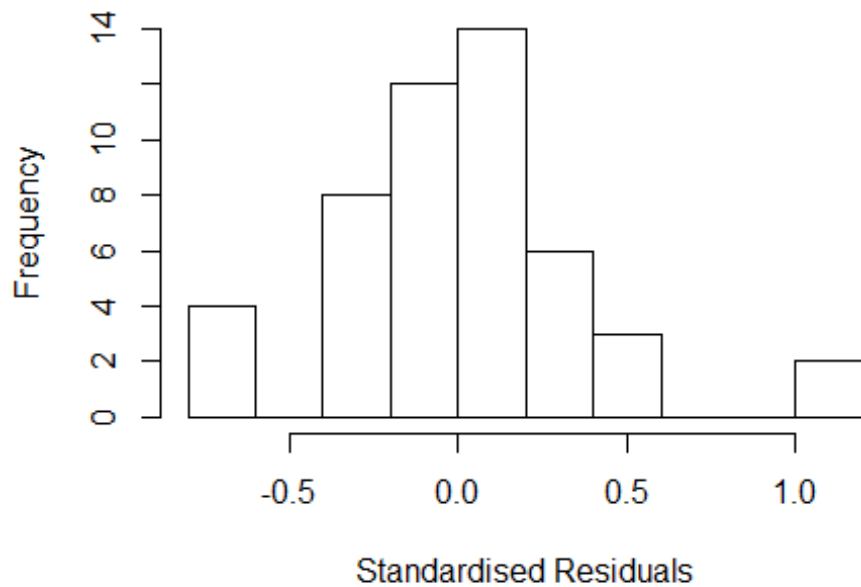
```

Residuals from first author



##	GVIF	Df	$GVIF^{1/(2 \cdot Df)}$
## 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
## -7.13e-01 -1.53e-01 1.11e-16 1.87e-01 1.15e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.39100 0.00000 Inf < 2e-16 ***
## LastAuthorFemale1 -0.74664 0.78609 -0.95 0.34872
## Year1999 0.84100 0.00000 Inf < 2e-16 ***
## Year2000 1.16400 0.00000 Inf < 2e-16 ***
## Year2001 1.05260 0.17255 6.10 5.7e-07 ***
## Year2002 0.81351 0.21857 3.72 0.00069 ***
## Year2003 0.80221 0.11500 6.98 4.1e-08 ***
## Year2004 0.95700 0.00000 Inf < 2e-16 ***
## Year2005 0.84600 0.00354 239.28 < 2e-16 ***
## Year2008 1.03790 0.23705 4.38 0.00010 ***
## Year2009 1.01021 0.29306 3.45 0.00149 **
## Year2010 0.81771 0.14413 5.67 2.1e-06 ***
## Year2011 1.05224 0.21165 4.97 1.7e-05 ***
## Year2012 0.77444 0.10837 7.15 2.5e-08 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.334
## Multiple R-squared: 0.357, Adjusted R-squared: 0.119
## Convergence in 31 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 38 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.210 0.876 0.948 0.882 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 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

```

```

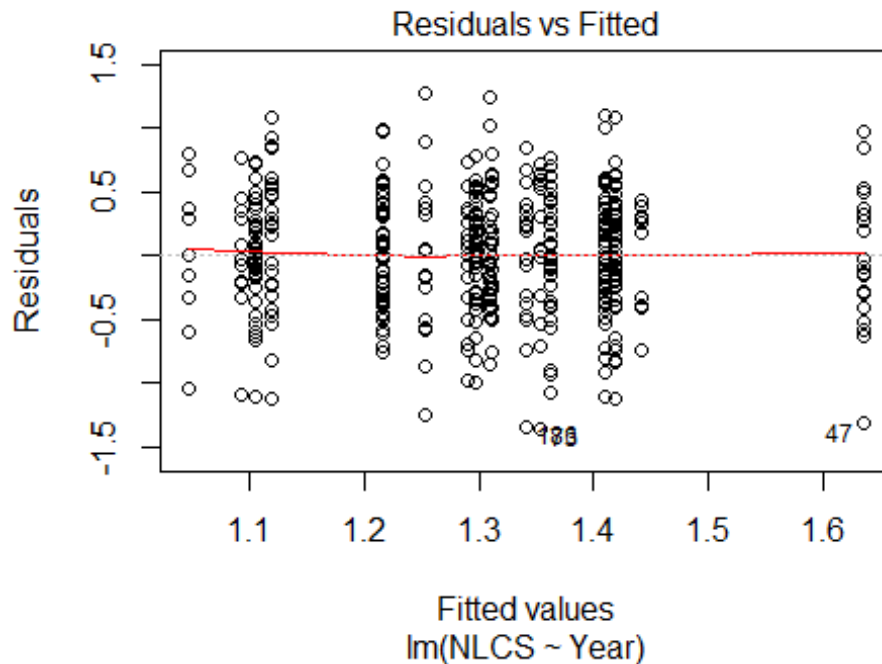
##          0          1000          0
##          psi          subsampling          cov
##          "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 49"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3506"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2006 2007 2008 2009 2010 2011
##    9    5    2    2    2    3    5    5    1    3    1    1    5    2    5
## 2012
##    2
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2006 2007 2008 2009 2010 2011
##    4    1    0    1    0    2    1    3    0    1    1    1    3    0    2
## 2012
##    1
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2006 2007 2008 2009 2010 2011
##    3    1    0    1    0    2    1    3    0    0    1    1    3    0    1
## 2012
##    1
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 5"
## [1] "Male first author team size 2018 geometric mean: 2.64575131106459"
##
## 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: 5"
## [1] "Male last author team size 2018 geometric mean: 2.64575131106459"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1, 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: 18"
## [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
##    29    29    23    32    37    30    46    44    32    62    64    65    44    23    15
## 2011 2012
##    12    11
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    17    20    14    19    27    17    37    35    25    53    53    53    41    20    12
## 2011 2012
##    10     9
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    16    19    13    16    22    15    35    32    22    49    50    47    38    16    11
## 2011 2012
##     8     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

```



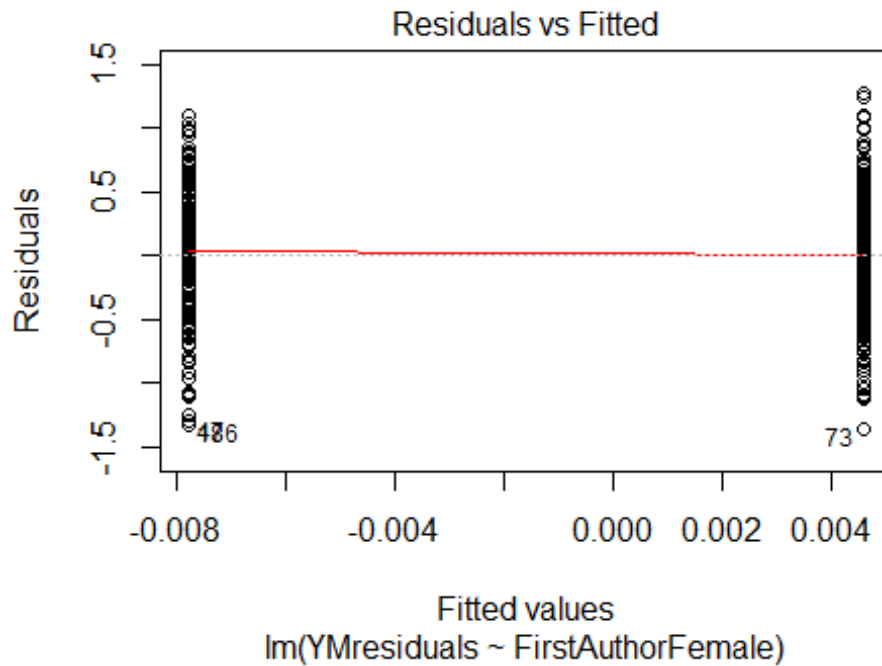
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.3, df = 1, p-value = 0.07

## [1] "Female first author team size 2018 geometric mean: 2.82101301234021"
## [1] "Male first author team size 2018 geometric mean: 1.68179283050743"

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

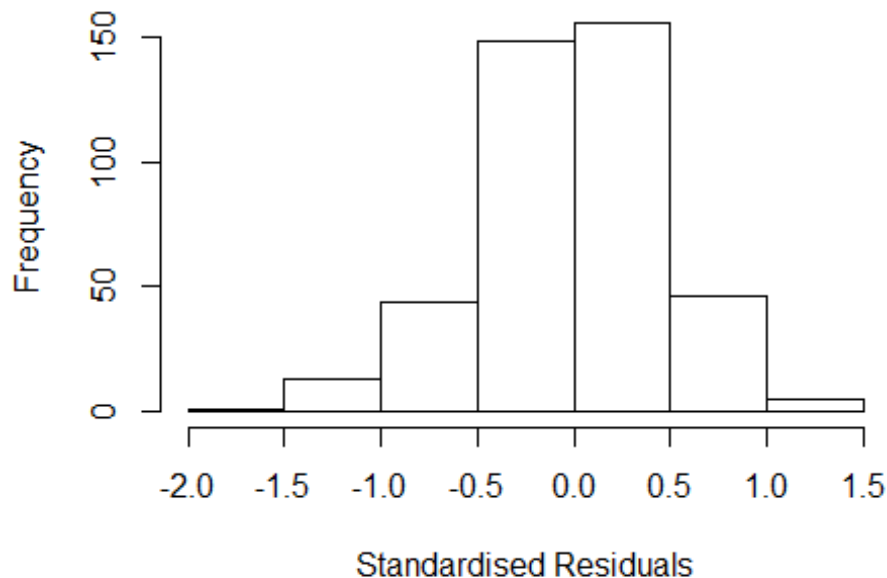
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.68575127330659"
## [1] "Male last author team size 2018 geometric mean: 1.5874010519682"

## Warning in wilcox.test.default(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 = 0.3
## 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 1.538 1          1.240
## LastAuthorFemale  1.594 1          1.262
## UniqueAuthors     2.020 4          1.092
## Year               2.520 16         1.029
```


Residuals from first and last author and team size



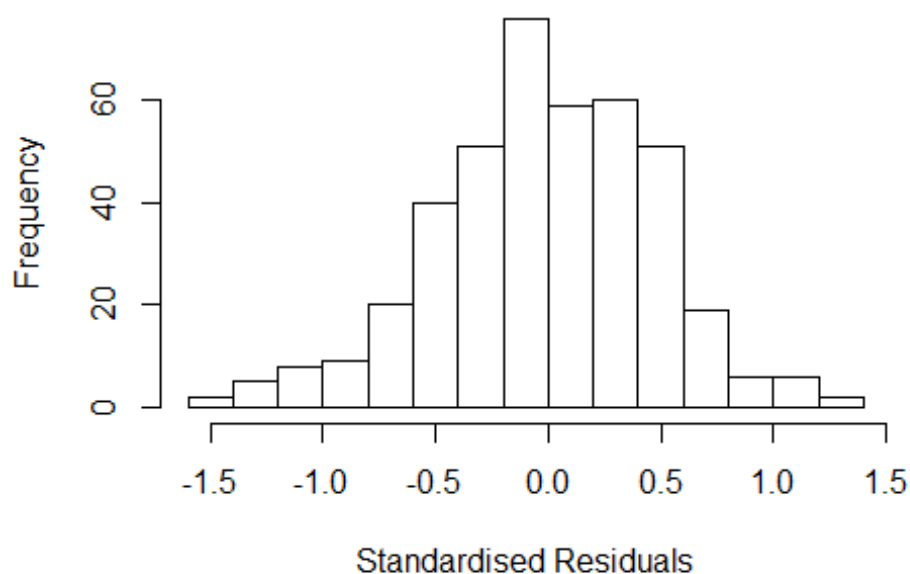
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.59173 -0.30072 0.00208 0.32317 1.35163
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19317 0.21004 5.68 2.6e-08 ***
## FirstAuthorFemale1 -0.04721 0.05933 -0.80 0.427
## LastAuthorFemale1 0.02941 0.05692 0.52 0.606
## UniqueAuthors2 0.07684 0.07426 1.03 0.301
## UniqueAuthors3 0.16376 0.06989 2.34 0.020 *
## UniqueAuthors4 0.14850 0.08984 1.65 0.099 .
## UniqueAuthors5 0.29565 0.06654 4.44 1.2e-05 ***
## Year1997 0.41224 0.24529 1.68 0.094 .
## Year1998 0.16707 0.27222 0.61 0.540
## Year1999 0.12788 0.22480 0.57 0.570
```

```

## Year2000      -0.00404    0.26503   -0.02    0.988
## Year2001      0.23480    0.23742    0.99    0.323
## Year2002      0.02037    0.21268    0.10    0.924
## Year2003      0.08421    0.21616    0.39    0.697
## Year2004     -0.00433    0.22360   -0.02    0.985
## Year2005      0.10054    0.21067    0.48    0.633
## Year2006      0.11122    0.21217    0.52    0.600
## Year2007     -0.11219    0.21118   -0.53    0.596
## Year2008     -0.15224    0.21000   -0.72    0.469
## Year2009      0.04463    0.23420    0.19    0.849
## Year2010     -0.18635    0.23410   -0.80    0.426
## Year2011      0.14691    0.24251    0.61    0.545
## Year2012     -0.26172    0.46547   -0.56    0.574
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.463
## Multiple R-squared:  0.112, Adjusted R-squared:  0.0616
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 35 weights are ~= 1. The remaining 379 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.213  0.881  0.949  0.904  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.42e-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 1.431 1      1.196
## LastAuthorFemale  1.534 1      1.239
## Year              1.361 16      1.010

```

Residuals from first and last author



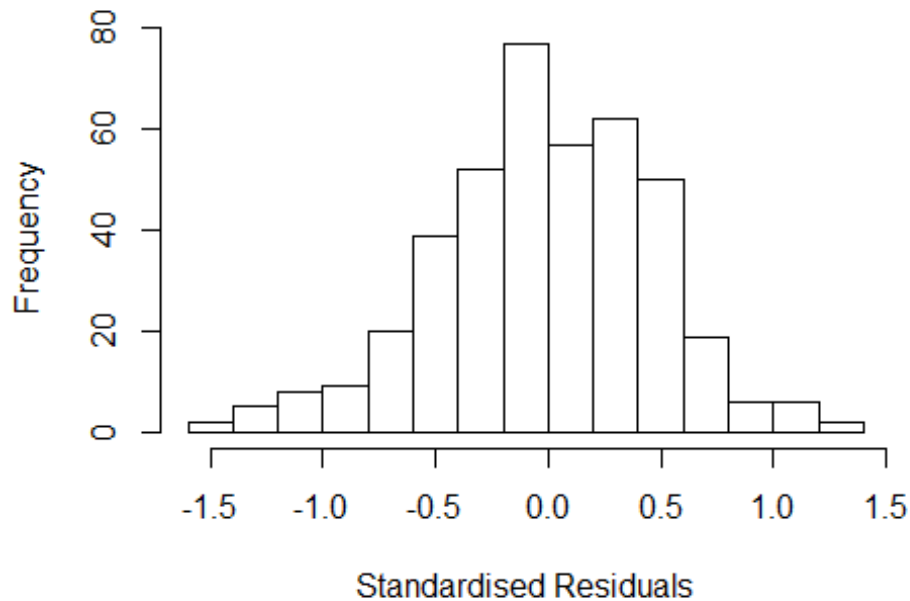
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4988 -0.3193 -0.0112  0.3214  1.2755
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2898     0.1935   6.67 8.9e-11 ***
## FirstAuthorFemale1 -0.0361     0.0591  -0.61   0.54
## LastAuthorFemale1  0.0178     0.0573   0.31   0.76
## Year1997          0.3870     0.2373   1.63   0.10
## Year1998          0.1542     0.2739   0.56   0.57
## Year1999          0.1309     0.2157   0.61   0.54
## Year2000         -0.0141     0.2590  -0.05   0.96
## Year2001          0.2090     0.2268   0.92   0.36
## Year2002          0.0451     0.2034   0.22   0.82
## Year2003          0.1251     0.2085   0.60   0.55
## Year2004          0.0168     0.2176   0.08   0.94
## Year2005          0.1378     0.2004   0.69   0.49
```

```

## Year2006          0.1405      0.2014      0.70      0.49
## Year2007          -0.0563     0.2040     -0.28     0.78
## Year2008          -0.1270     0.2002     -0.63     0.53
## Year2009           0.0613     0.2258      0.27     0.79
## Year2010          -0.2252     0.2222     -1.01     0.31
## Year2011           0.1856     0.2536      0.73     0.46
## Year2012          -0.3578     0.4480     -0.80     0.42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0738, Adjusted R-squared:  0.0316
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 369 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.295  0.880   0.946   0.902   0.983   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.42e-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 1.146 1      1.070
## Year              1.146 16      1.004

```

Residuals from first author



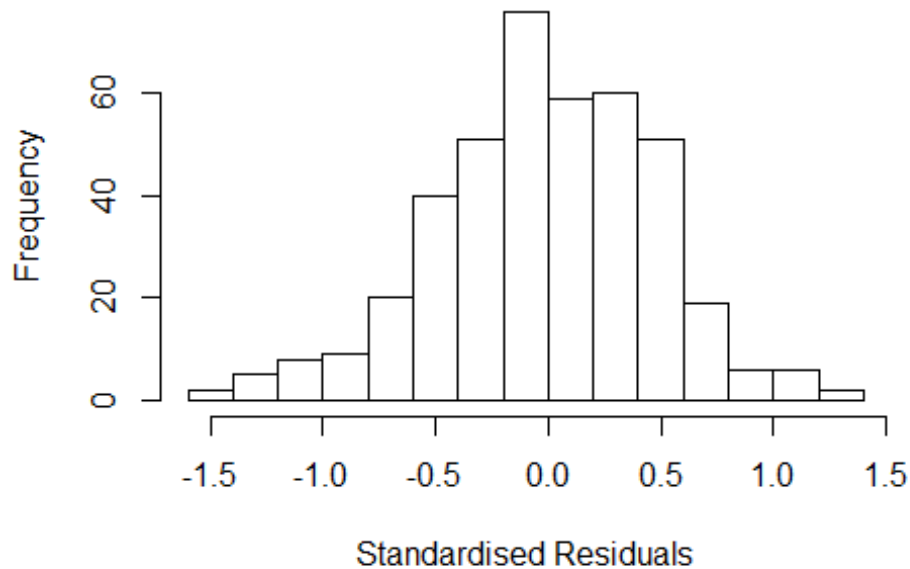
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.50673 -0.32460 -0.00827  0.32185  1.26755
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.29271    0.19433   6.65 9.6e-11 ***
## FirstAuthorFemale1 -0.02900    0.05308  -0.55  0.59
## Year1997         0.39436    0.23238   1.70  0.09 .
## Year1998         0.15683    0.27438   0.57  0.57
## Year1999         0.13275    0.21446   0.62  0.54
## Year2000        -0.00829    0.25775  -0.03  0.97
## Year2001         0.21402    0.22499   0.95  0.34
## Year2002         0.04740    0.20193   0.23  0.81
## Year2003         0.12805    0.20722   0.62  0.54
## Year2004         0.01475    0.21821   0.07  0.95
## Year2005         0.13911    0.19948   0.70  0.49
## Year2006         0.14309    0.19993   0.72  0.47
```

```

## Year2007      -0.05522    0.20285   -0.27    0.79
## Year2008      -0.12384    0.19859   -0.62    0.53
## Year2009       0.06691    0.22332    0.30    0.76
## Year2010      -0.21866    0.21836   -1.00    0.32
## Year2011       0.19271    0.25008    0.77    0.44
## Year2012      -0.35048    0.44258   -0.79    0.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0736, Adjusted R-squared:  0.0338
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 373 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.288  0.879   0.946   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      2.42e-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 1.235 1          1.111
## Year              1.235 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
## -1.4882 -0.3231 -0.0123 0.3200 1.2550
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26823 0.18480 6.86 2.6e-11 ***
## LastAuthorFemale1 0.00375 0.05153 0.07 0.942
## Year1997 0.39753 0.23308 1.71 0.089 .
## Year1998 0.15524 0.27242 0.57 0.569
## Year1999 0.12754 0.21299 0.60 0.550
## Year2000 -0.00458 0.25689 -0.02 0.986
## Year2001 0.22000 0.22395 0.98 0.327
## Year2002 0.04925 0.20017 0.25 0.806
## Year2003 0.13294 0.20480 0.65 0.517
## Year2004 0.02364 0.21450 0.11 0.912
## Year2005 0.14524 0.19608 0.74 0.459
## Year2006 0.14618 0.19787 0.74 0.460
```

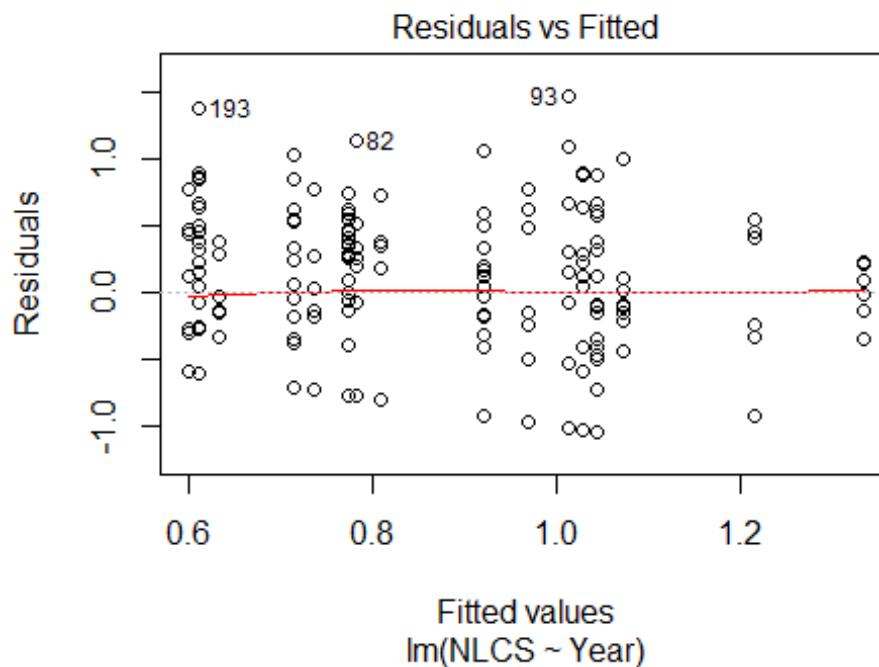
```

## Year2007          -0.04840      0.19962    -0.24      0.809
## Year2008          -0.11710      0.19632    -0.60      0.551
## Year2009           0.06528      0.22486     0.29      0.772
## Year2010          -0.21729      0.21806    -1.00      0.320
## Year2011           0.19259      0.24936     0.77      0.440
## Year2012          -0.35730      0.45223    -0.79      0.430
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.474
## Multiple R-squared:  0.0726, Adjusted R-squared:  0.0328
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 368 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.303  0.875   0.946   0.902   0.983   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.42e-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: 414"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3601"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8    8    6    6    9    7   13    6   10   11   12    9   11   19   39
## 2011 2012
##   26   26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8    7    6    6    8    6   11    6    9   10   11    8    9   15   32
## 2011 2012

```



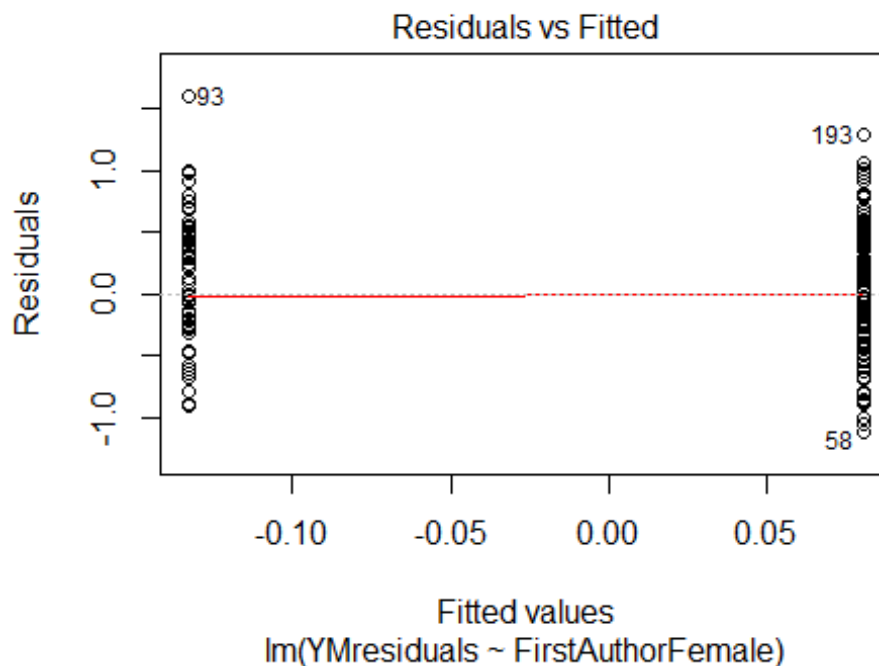
```
## 22 25
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 7 6 6 8 6 11 5 9 10 11 6 8 13 30
## 2011 2012
## 20 23
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 17, df = 16, p-value = 0.4
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.96, df = 1, p-value = 0.3
## [1] "Female first author team size 2018 geometric mean: 2.16093273781394"
## [1] "Male first author team size 2018 geometric mean: 2.70529322256312"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

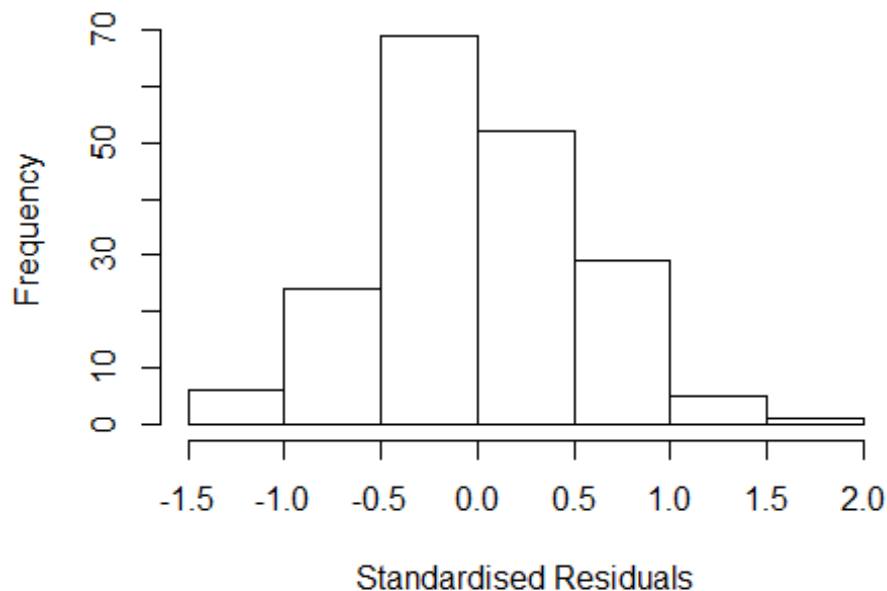
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 34, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.42973635498699"
## [1] "Male last author team size 2018 geometric mean: 2.13982638786733"

## Warning in wilcox.test.default(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 = 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 2.451 1      1.566
## LastAuthorFemale 3.105 1      1.762
## UniqueAuthors    6.397 4      1.261
## Year             13.478 16     1.085
```

Residuals from first and last author and team size



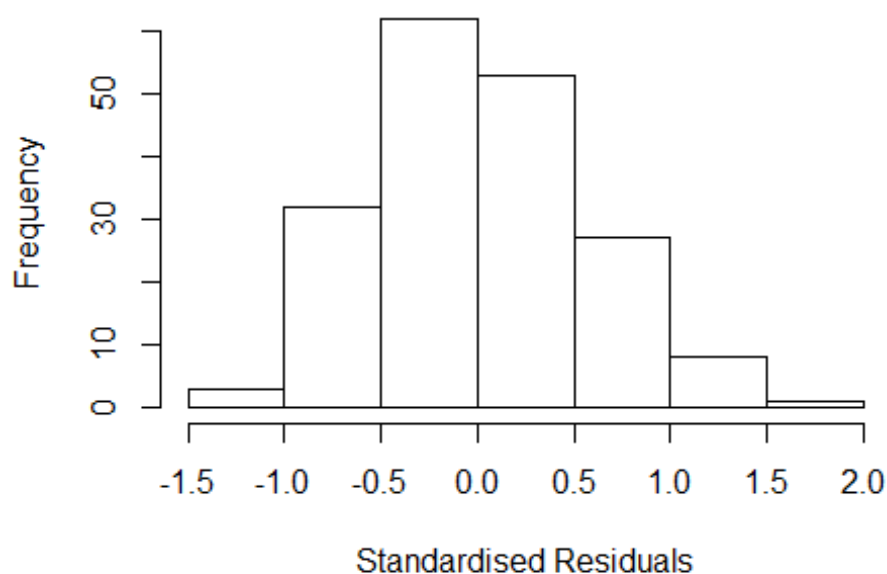
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1814 -0.3263 -0.0212 0.3845 1.7311
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6866 0.2461 2.79 0.00591 **
## FirstAuthorFemale1 0.1145 0.1194 0.96 0.33911
## LastAuthorFemale1 0.1607 0.1237 1.30 0.19590
## UniqueAuthors2 0.3796 0.1122 3.38 0.00090 ***
## UniqueAuthors3 0.4776 0.1516 3.15 0.00194 **
## UniqueAuthors4 0.5040 0.1432 3.52 0.00056 ***
## UniqueAuthors5 0.4588 0.1628 2.82 0.00543 **
## Year1997 -0.1427 0.3311 -0.43 0.66719
## Year1998 -0.3893 0.2476 -1.57 0.11784
## Year1999 -0.0522 0.3764 -0.14 0.88994
```

```

## Year2000          0.0749      0.3120      0.24  0.81054
## Year2001         -0.5337      0.2666     -2.00  0.04695 *
## Year2002         -0.0784      0.3022     -0.26  0.79552
## Year2003          0.0859      0.2401      0.36  0.72091
## Year2004         -0.3942      0.3230     -1.22  0.22410
## Year2005         -0.3223      0.4350     -0.74  0.45976
## Year2006         -0.2031      0.2939     -0.69  0.49054
## Year2007         -0.6618      0.3129     -2.12  0.03594 *
## Year2008         -0.2355      0.2626     -0.90  0.37109
## Year2009         -0.3490      0.2616     -1.33  0.18404
## Year2010         -0.4175      0.2383     -1.75  0.08164 .
## Year2011         -0.5185      0.2608     -1.99  0.04846 *
## Year2012         -0.2958      0.2502     -1.18  0.23892
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.268, Adjusted R-squared:  0.169
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 171 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.257  0.860  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      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 1.962 1      1.401
## LastAuthorFemale  2.607 1      1.615
## Year              2.086 16      1.023

```

Residuals from first and last author



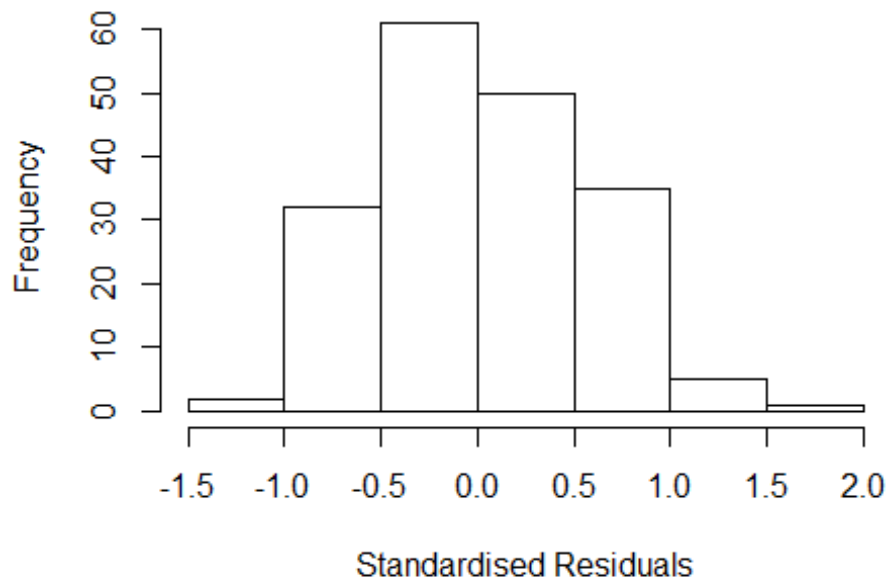
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1103 -0.4057 -0.0216 0.4177 1.6971
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8747 0.2211 3.96 0.00011 ***
## FirstAuthorFemale1 0.1811 0.1235 1.47 0.14443
## LastAuthorFemale1 0.1541 0.1332 1.16 0.24906
## Year1997 -0.0996 0.3521 -0.28 0.77760
## Year1998 -0.3214 0.2653 -1.21 0.22731
## Year1999 -0.1682 0.3291 -0.51 0.60997
## Year2000 0.1638 0.2895 0.57 0.57234
## Year2001 -0.4116 0.2423 -1.70 0.09121 .
## Year2002 -0.0535 0.2956 -0.18 0.85656
## Year2003 0.2270 0.2352 0.97 0.33574
## Year2004 -0.2967 0.2994 -0.99 0.32314
## Year2005 -0.0968 0.4152 -0.23 0.81588
```

```

## Year2006          -0.0879      0.2599   -0.34   0.73569
## Year2007          -0.4985      0.3211   -1.55   0.12248
## Year2008          -0.0133      0.2308   -0.06   0.95405
## Year2009          -0.1638      0.2519   -0.65   0.51642
## Year2010          -0.4691      0.2298   -2.04   0.04284 *
## Year2011          -0.4193      0.2440   -1.72   0.08760 .
## Year2012          -0.3492      0.2333   -1.50   0.13634
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.543
## Multiple R-squared:  0.17,   Adjusted R-squared:  0.0805
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 172 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.308  0.851  0.938  0.904  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 1.077 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
## -1.1278 -0.4490 -0.0411 0.4467 1.6199
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.91956 0.21661 4.25 3.6e-05 ***
## FirstAuthorFemale1 0.26434 0.09244 2.86 0.0048 **
## Year1997 -0.12736 0.34436 -0.37 0.7120
## Year1998 -0.33930 0.27444 -1.24 0.2181
## Year1999 -0.20035 0.34084 -0.59 0.5574
## Year2000 0.16893 0.27866 0.61 0.5452
## Year2001 -0.46327 0.23612 -1.96 0.0514 .
## Year2002 -0.05611 0.29171 -0.19 0.8477
## Year2003 0.28586 0.24157 1.18 0.2384
## Year2004 -0.31894 0.29002 -1.10 0.2730
## Year2005 -0.06442 0.42288 -0.15 0.8791
## Year2006 -0.10206 0.26249 -0.39 0.6979
```

```

## Year2007          -0.51600    0.30457   -1.69    0.0921 .
## Year2008          -0.00871    0.22746   -0.04    0.9695
## Year2009          -0.19830    0.24801   -0.80    0.4251
## Year2010          -0.47053    0.23202   -2.03    0.0441 *
## Year2011          -0.43623    0.24326   -1.79    0.0747 .
## Year2012          -0.35234    0.23310   -1.51    0.1325
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.592
## Multiple R-squared:  0.155, Adjusted R-squared:  0.0692
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 172 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.435  0.872  0.946  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      5.38e-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 1.689 1          1.300
## Year              1.689 16          1.017

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.10188 -0.45434  0.00389  0.39592  1.66928

```



```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.93263    0.22044   4.23 3.8e-05 ***
## LastAuthorFemale1 0.24664    0.10265   2.40  0.017 *
## Year1997       -0.07739    0.33934  -0.23  0.820
## Year1998       -0.32110    0.26178  -1.23  0.222
## Year1999       -0.17756    0.31900  -0.56  0.579
## Year2000        0.15438    0.29719   0.52  0.604
## Year2001       -0.37759    0.23824  -1.58  0.115
## Year2002       -0.05121    0.30117  -0.17  0.865
## Year2003        0.18368    0.22023   0.83  0.405
## Year2004       -0.28384    0.31339  -0.91  0.366
## Year2005       -0.12691    0.39744  -0.32  0.750
## Year2006       -0.05616    0.25734  -0.22  0.828
## Year2007       -0.46007    0.34165  -1.35  0.180
## Year2008       -0.00296    0.23625  -0.01  0.990
## Year2009       -0.12821    0.25692  -0.50  0.618
## Year2010       -0.47829    0.23196  -2.06  0.041 *
## Year2011       -0.41416    0.24761  -1.67  0.096 .
## Year2012       -0.34230    0.23376  -1.46  0.145
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.566
## Multiple R-squared:  0.155, Adjusted R-squared:  0.0696
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 172 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.365  0.871  0.942  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      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 3602"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1    2    4   10    6    6    1    4    9    9    3    6   14   10   12
## 2012
##    20
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    0    0    8    5    6    1    3    7    8    3    5   13   10   11
## 2012
##    20
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    0    0    7    3    6    1    3    7    7    3    5   11    8    8
## 2012
##    17
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4.09534502215844"
## [1] "Male first author team size 2018 geometric mean: 3.07567546368897"

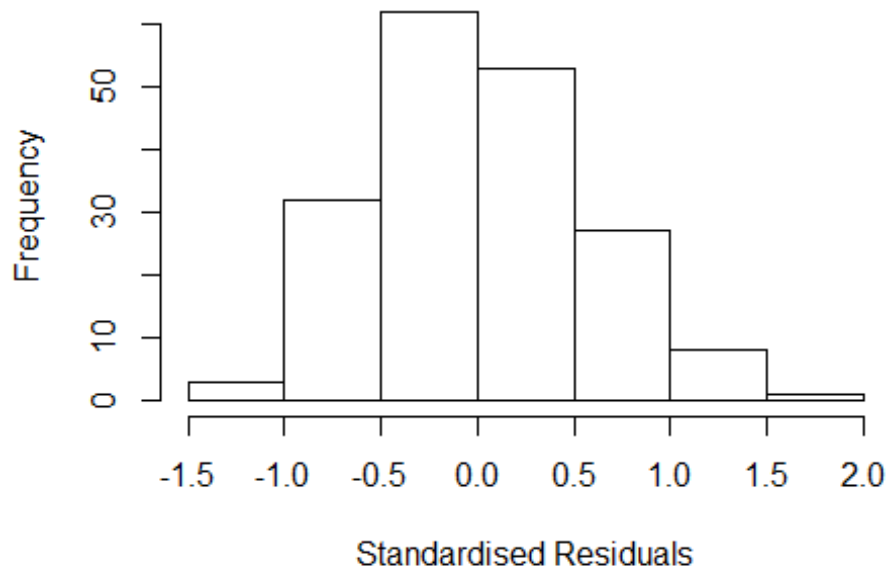
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 23, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.63424118566428"
## [1] "Male last author team size 2018 geometric mean: 3.37577529551882"

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

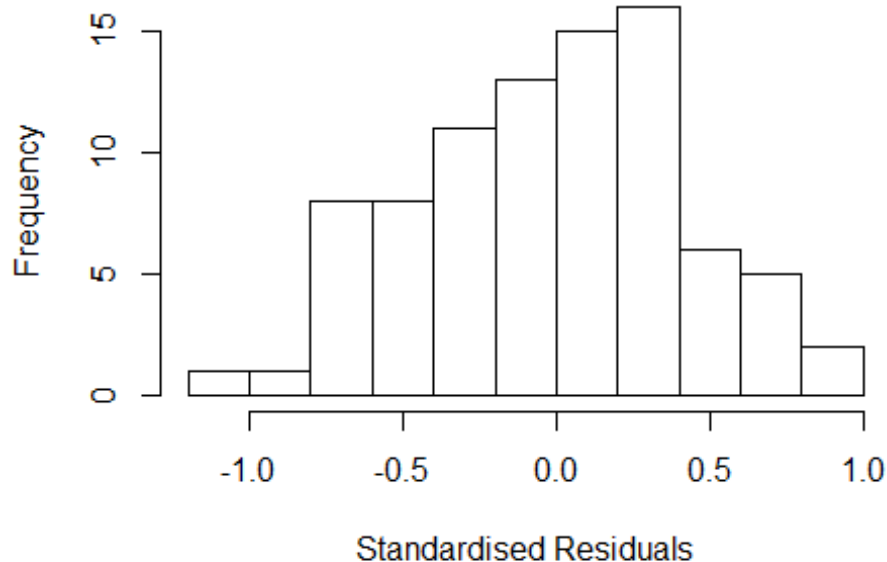
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15, 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 14.61 1          3.822
## LastAuthorFemale  16.70 1          4.087
## UniqueAuthors    81.43 4          1.733
## Year             487.72 12          1.294
```

Residuals from first and last author and team size



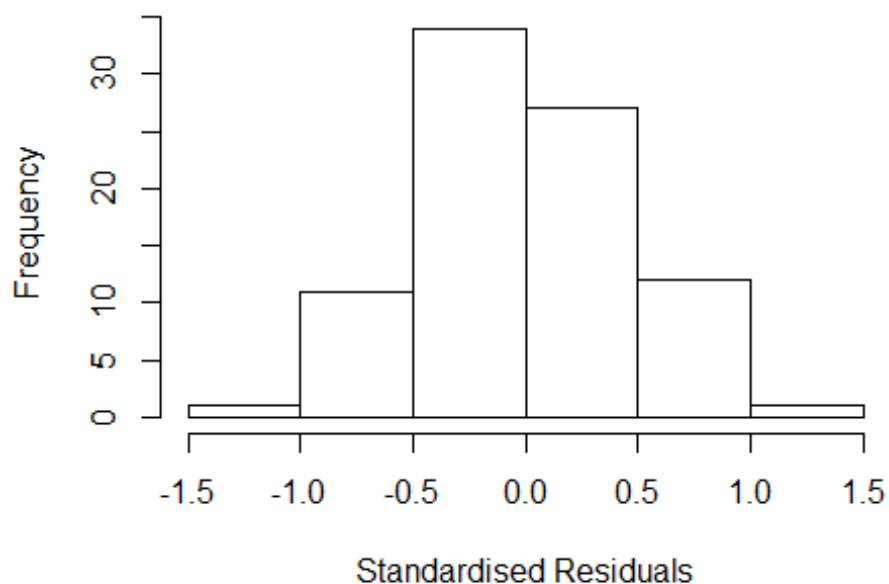
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1557 -0.3355 0.0255 0.2919 0.9907
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6133 0.4343 1.41 0.1626
## FirstAuthorFemale1 -0.0896 0.1609 -0.56 0.5794
## LastAuthorFemale1 -0.3200 0.1763 -1.81 0.0740 .
## UniqueAuthors2 0.2381 0.1606 1.48 0.1428
## UniqueAuthors3 0.5729 0.2082 2.75 0.0076 **
## UniqueAuthors4 0.5844 0.1533 3.81 0.0003 ***
## UniqueAuthors5 0.6711 0.2314 2.90 0.0050 **
## Year2001 0.1818 0.5259 0.35 0.7307
## Year2002 0.4170 0.4808 0.87 0.3889
## Year2003 0.9635 0.4118 2.34 0.0223 *
```

```

## Year2004          0.6294      0.4983      1.26      0.2110
## Year2005          0.5416      0.4854      1.12      0.2686
## Year2006          0.3440      0.4819      0.71      0.4778
## Year2007          0.6303      0.4873      1.29      0.2003
## Year2008          0.2081      0.4199      0.50      0.6218
## Year2009          0.3655      0.4538      0.81      0.4234
## Year2010          0.4680      0.4540      1.03      0.3064
## Year2011          0.1170      0.4168      0.28      0.7798
## Year2012          0.0591      0.4304      0.14      0.8912
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.394, Adjusted R-squared:  0.232
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 79 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.440  0.865  0.950   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          1.16e-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 10.812 1          3.288
## LastAuthorFemale  8.913 1          2.986
## Year              26.292 12          1.146

```

Residuals from first and last author



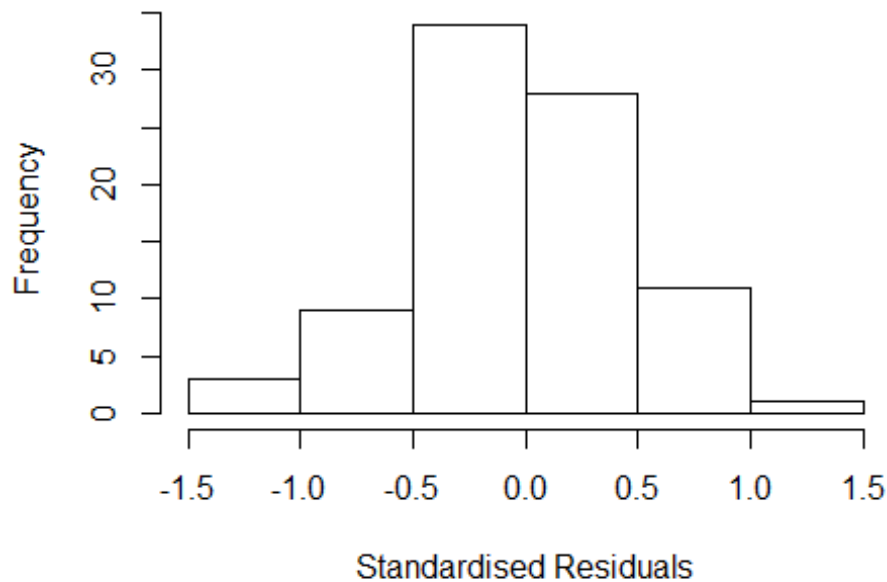
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2557 -0.3097 -0.0338 0.3673 1.3073
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9544 0.2565 3.72 0.00039 ***
## FirstAuthorFemale1 0.0386 0.1738 0.22 0.82498
## LastAuthorFemale1 -0.2374 0.1630 -1.46 0.14967
## Year2001 0.3589 0.5299 0.68 0.50046
## Year2002 0.5437 0.3177 1.71 0.09142 .
## Year2003 0.9845 0.2896 3.40 0.00111 **
## Year2004 0.6316 0.4159 1.52 0.13330
## Year2005 0.5801 0.4004 1.45 0.15178
## Year2006 0.1564 0.4477 0.35 0.72796
## Year2007 0.6379 0.4309 1.48 0.14318
## Year2008 0.2294 0.3037 0.76 0.45246
## Year2009 0.3014 0.3070 0.98 0.32953
```

```

## Year2010          0.5421      0.2842      1.91  0.06054 .
## Year2011          0.2035      0.2771      0.73  0.46510
## Year2012         -0.0841      0.2638     -0.32  0.75086
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.492
## Multiple R-squared:  0.242, Adjusted R-squared:  0.092
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 76 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.460  0.874  0.943  0.905  0.976  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.16e-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 10.35 1          3.218
## Year              10.35 12          1.102

```

Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22657 -0.28590 -0.00813 0.37855 1.39095
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.842298 0.238563 3.53 0.00073 ***
## FirstAuthorFemale1 -0.014176 0.178389 -0.08 0.93688
## Year2001 0.354935 0.481761 0.74 0.46367
## Year2002 0.640772 0.317590 2.02 0.04736 *
## Year2003 0.911878 0.293190 3.11 0.00268 **
## Year2004 0.756372 0.422747 1.79 0.07779 .
## Year2005 0.694327 0.386200 1.80 0.07639 .
## Year2006 0.184756 0.490526 0.38 0.70754
## Year2007 0.552684 0.454568 1.22 0.22802
## Year2008 0.360599 0.285913 1.26 0.21130
## Year2009 0.384273 0.301086 1.28 0.20595
## Year2010 0.648266 0.269505 2.41 0.01873 *
```

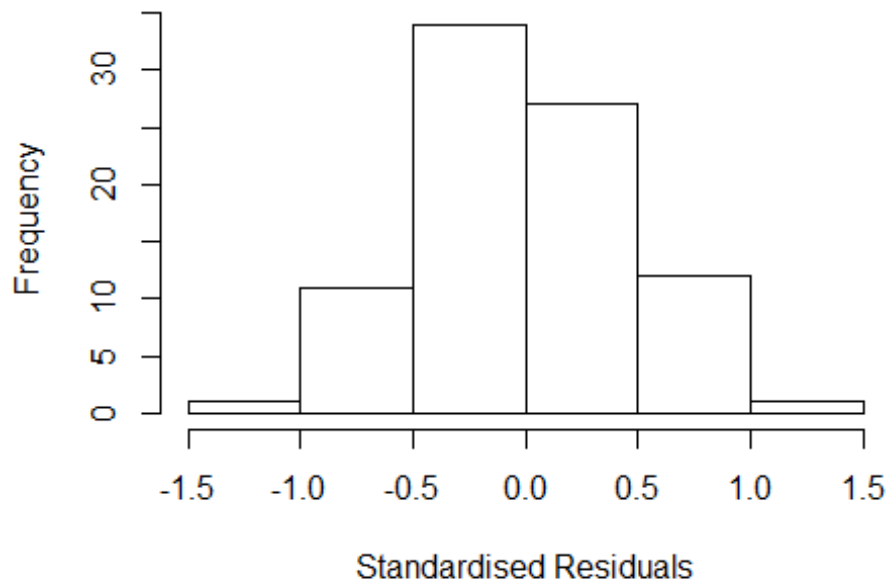


```

## Year2011          0.254275    0.276538    0.92  0.36091
## Year2012          0.000832    0.257542    0.00  0.99743
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.502
## Multiple R-squared:  0.224, Adjusted R-squared:  0.0838
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 75 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.422  0.878  0.940  0.904  0.980  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          1.16e-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 8.257 1          2.874
## Year            8.257 12          1.092

```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ 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 -0.2928 -0.0388 0.3783 1.2996
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9513 0.2513 3.79 0.00031 ***
## LastAuthorFemale1 -0.2302 0.1625 -1.42 0.16082
## Year2001 0.3476 0.5137 0.68 0.50084
## Year2002 0.5446 0.3120 1.75 0.08518 .
## Year2003 1.0189 0.2315 4.40 3.7e-05 ***
## Year2004 0.6459 0.4283 1.51 0.13590
## Year2005 0.5874 0.3903 1.50 0.13670
## Year2006 0.1671 0.4288 0.39 0.69797
## Year2007 0.6607 0.4031 1.64 0.10559
## Year2008 0.2463 0.2953 0.83 0.40712
## Year2009 0.3184 0.2859 1.11 0.26906
## Year2010 0.5617 0.2737 2.05 0.04376 *
```

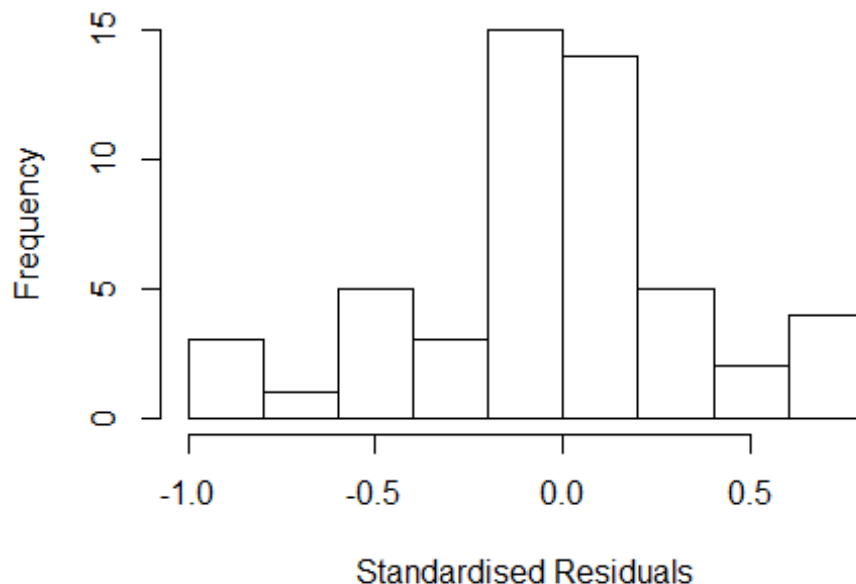
```

## Year2011          0.2085      0.2700      0.77  0.44255
## Year2012          -0.0706      0.2597     -0.27  0.78638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.512
## Multiple R-squared:  0.239, Adjusted R-squared:  0.102
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 76 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.499  0.886  0.943   0.911  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           1.16e-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: 86"
## [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
##    4    1    1    2    3    3    7    2    5    6    4    5    9    7    7
## 2011 2012
##    9    3
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    1    1    1    3    1    1    2    3    4    1    5    8    7    6
## 2011 2012
##    8    3
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    1    1    0    3    1    0    1    3    2    1    5    8    7    5

```

```
## 2011 2012
##      7      3
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 7.11378660898013"
## [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 3.738e+14 1      19333764
## LastAuthorFemale -4.279e+13 1           NaN
## UniqueAuthors    -9.411e+44 4           NaN
## Year              -3.678e+60 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
```

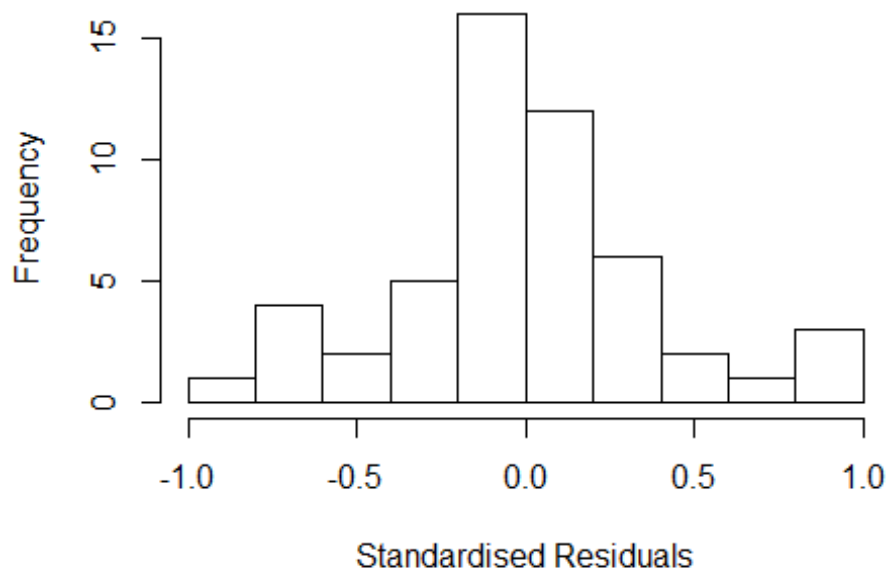
```

## -9.18e-01 -1.74e-01 1.22e-15 1.73e-01 7.94e-01
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.32090    0.20667    6.39 4.0e-07 ***
## FirstAuthorFemale1 -0.01569    0.24874   -0.06 0.95012
## LastAuthorFemale1 -0.44189    0.29433   -1.50 0.14339
## UniqueAuthors2     0.22598    0.25274    0.89 0.37815
## UniqueAuthors3    -0.04027    0.22550   -0.18 0.85944
## UniqueAuthors4     0.02362    0.20504    0.12 0.90903
## UniqueAuthors5     0.03664    0.14538    0.25 0.80270
## Year1997          -0.82090    0.20667   -3.97 0.00039 ***
## Year1998           0.92894    0.17205    5.40 6.8e-06 ***
## Year2000           -0.07978    0.17765   -0.45 0.65651
## Year2001           1.07225    0.27688    3.87 0.00052 ***
## Year2003           0.41605    0.23802    1.75 0.09036 .
## Year2004           -0.36722    0.30038   -1.22 0.23073
## Year2005           0.37082    0.23832    1.56 0.12986
## Year2006           -0.43319    0.27776   -1.56 0.12901
## Year2007           -0.01874    0.26426   -0.07 0.94393
## Year2008           0.04880    0.13664    0.36 0.72341
## Year2009           0.00762    0.22258    0.03 0.97293
## Year2010           0.06448    0.66989    0.10 0.92394
## Year2011           -0.18719    0.27512   -0.68 0.50130
## Year2012           -0.19438    0.36895   -0.53 0.60204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.375, Adjusted R-squared:  -0.0279
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 42 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.597  0.882  0.973  0.907  0.989  0.999
## Algorithmic parameters:
##           tuning.chi              bb          tuning.psi          refine.tol
##           1.55e+00             5.00e-01          4.69e+00          1.00e-07
##           rel.tol             solve.tol          eps.outlier          eps.x
##           1.00e-07             1.00e-07          1.92e-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 8.653e+14 1      29415329
## LastAuthorFemale 1.127e+15 1      33575295
## Year             -2.094e+30 14      NaN
```

Residuals from first and last author



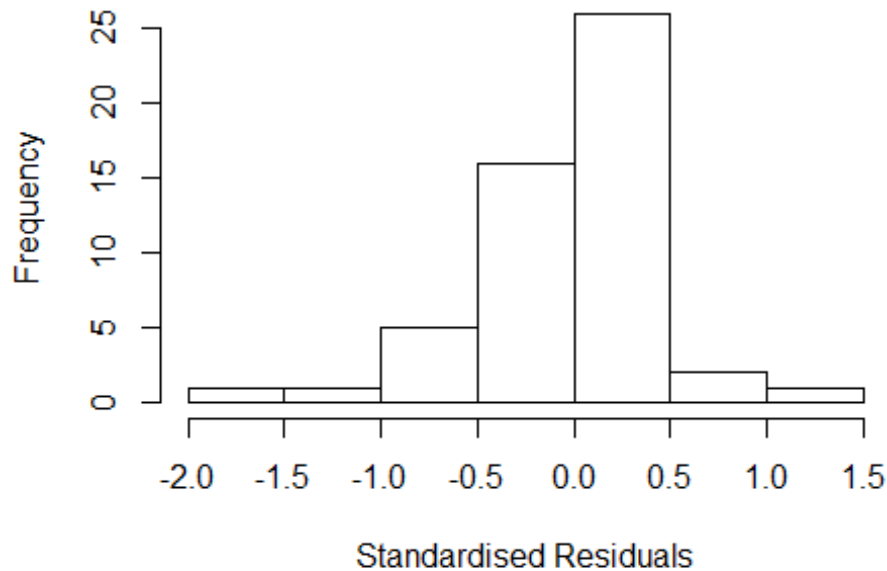
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.971 -0.168  0.000  0.171  0.970
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3286     0.0946   14.05 5.9e-16 ***
## FirstAuthorFemale1  0.0236     0.4072    0.06 0.95407
## LastAuthorFemale1 -0.5004     0.3369   -1.49 0.14635
```

```

## Year1997          -0.8286      0.0946    -8.76  2.4e-10 ***
## Year1998           0.9002      0.1737     5.18  9.3e-06 ***
## Year2000           0.0779      0.1872     0.42  0.67996
## Year2001           1.0828      0.2705     4.00  0.00031 ***
## Year2003           0.4512      0.1737     2.60  0.01365 *
## Year2004          -0.2073      0.4728    -0.44  0.66382
## Year2005           0.4752      0.1853     2.56  0.01477 *
## Year2006          -0.2542      0.4621    -0.55  0.58575
## Year2007           0.0649      0.4163     0.16  0.87698
## Year2008           0.1188      0.1252     0.95  0.34917
## Year2009           0.0169      0.2335     0.07  0.94258
## Year2010           0.0359      0.6841     0.05  0.95847
## Year2011          -0.0777      0.2427    -0.32  0.75080
## Year2012          -0.1827      0.3863    -0.47  0.63931
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.362, Adjusted R-squared:  0.07
## Convergence in 29 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 42 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.443  0.814  0.953  0.876  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           1.92e-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 -1.202e+15  1           NaN
## Year              -1.202e+15 14           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
## -1.5463 -0.1495 0.0158 0.1446 1.0648
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1866 0.1705 6.96 3.7e-08 ***
## FirstAuthorFemale1 -0.3416 0.1593 -2.14 0.03884 *
## Year1997 -0.6866 0.1705 -4.03 0.00028 ***
## Year1998 0.9070 0.1561 5.81 1.3e-06 ***
## Year2000 0.1367 0.2170 0.63 0.53265
## Year2001 0.7244 0.1705 4.25 0.00014 ***
## Year2003 0.4580 0.1561 2.93 0.00580 **
## Year2004 0.3597 0.1598 2.25 0.03056 *
## Year2005 0.4820 0.1689 2.85 0.00712 **
## Year2006 0.2530 0.1561 1.62 0.11390
## Year2007 0.1886 0.2068 0.91 0.36768
## Year2008 0.0886 0.1687 0.53 0.60255
```



```

## Year2009          0.1362      0.2148      0.63  0.53001
## Year2010          0.3678      0.5203      0.71  0.48414
## Year2011          0.1082      0.2306      0.47  0.64181
## Year2012         -0.1160      0.3866     -0.30  0.76596
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.354, Adjusted R-squared:  0.0848
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 38 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0271 0.7940 0.9280 0.8440 0.9880 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      1.92e-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 1.642e+15 1      4.052e+07
## Year              1.642e+15 14      3.495e+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 ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q    Median      3Q      Max
## -9.73e-01 -1.55e-01  3.47e-16  1.62e-01  9.51e-01
##
## Coefficients:

```

```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3255    0.0910   14.57 < 2e-16 ***
## LastAuthorFemale1 -0.4821    0.1643   -2.93  0.0058 **
## Year1997         -0.8255    0.0910   -9.07  7.8e-11 ***
## Year1998          0.9086    0.0977    9.30  4.1e-11 ***
## Year2000          0.0842    0.1669    0.50  0.6167
## Year2001          1.0676    0.0977   10.93  5.5e-13 ***
## Year2003          0.4596    0.0977    4.70  3.7e-05 ***
## Year2004         -0.1875    0.3918   -0.48  0.6352
## Year2005          0.4836    0.1170    4.13  0.0002 ***
## Year2006         -0.2275    0.0910   -2.50  0.0171 *
## Year2007          0.0687    0.2198    0.31  0.7564
## Year2008          0.1295    0.1135    1.14  0.2616
## Year2009          0.0282    0.1481    0.19  0.8499
## Year2010          0.0257    0.5918    0.04  0.9656
## Year2011         -0.0676    0.2585   -0.26  0.7953
## Year2012         -0.1633    0.3786   -0.43  0.6689
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.364, Adjusted R-squared:  0.0991
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 42 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.396  0.795  0.949  0.864  0.986  0.998
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.92e-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: 52"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3604"
## [1] "#####"

```

```

## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1998 1999 2000 2001 2005 2006 2007 2008 2009 2010 2011 2012
##    1    4    3    1    2    2    2    3    7    1   16   15
##
## 1998 1999 2000 2001 2005 2006 2007 2008 2009 2010 2011 2012
##    1    3    2    1    0    1    1    2    5    0    1    0
##
## 1998 1999 2000 2001 2005 2006 2007 2008 2009 2010 2011 2012
##    1    3    2    1    0    1    1    2    5    0    1    0
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 6.6332495807108"
## [1] "Male first author team size 2018 geometric mean: 7.09034881362097"

## Warning in wilcox.test.default(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: 10.3020383076056"
## [1] "Male last author team size 2018 geometric mean: 4.12128529980856"

## Warning in wilcox.test.default(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"
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
## [1] "Regression 3: First author gender, Year as factors"
## [1] "Regression 4: Last author gender, Year as factors"
## [1] "Sample size for the above analysis: 17"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3605"
## [1] "#####"

```

```

## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    5    2    4    7    3    4    7    8   15   12   14   22   19    7
## 2011 2012
##    10   16
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    2    1    3    1    1    3    6    6   10    9   12   17   16    4
## 2011 2012
##    7    10
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    2    1    3    1    0    3    6    6   10    7   10   13   13    4
## 2011 2012
##    6    8
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3.80730787743176"
## [1] "Male first author team size 2018 geometric mean: 3.68704102068099"

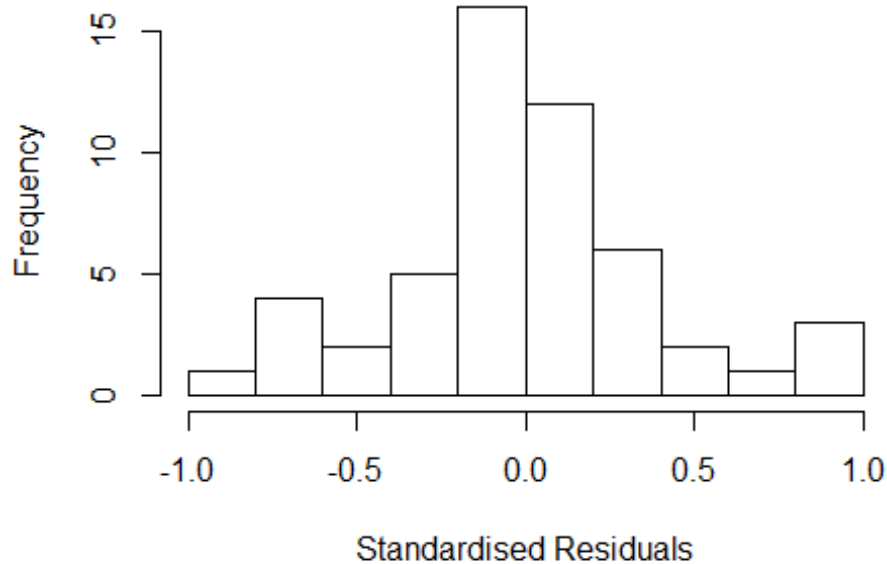
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 41, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.91486764116886"
## [1] "Male last author team size 2018 geometric mean: 3.63838790933338"

## 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 = 48, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##               GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.630e+00  1          1.905
## LastAuthorFemale  2.888e+00  1          1.699
## UniqueAuthors    6.491e+15  4          94.741
## Year              1.383e+16 15          3.452
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
```

```

##      Min      1Q      Median      3Q      Max
## -0.84868 -0.30666  0.00775  0.26315  1.66367
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.11e+00   4.50e-08  2.48e+07  <2e-16 ***
## FirstAuthorFemale1 -1.89e-01   1.39e-01 -1.35e+00   0.1803
## LastAuthorFemale1  4.09e-03   1.25e-01   3.00e-02   0.9739
## UniqueAuthors2    2.53e-01   2.30e-01   1.10e+00   0.2769
## UniqueAuthors3    4.44e-01   2.12e-01   2.10e+00   0.0394 *
## UniqueAuthors4    5.53e-01   1.83e-01   3.03e+00   0.0034 **
## UniqueAuthors5    4.46e-01   1.81e-01   2.47e+00   0.0160 *
## Year1997          -5.27e-02   3.59e-01  -1.50e-01   0.8837
## Year1998           5.80e-01   2.30e-01   2.52e+00   0.0140 *
## Year1999          -7.55e-01   2.32e-01  -3.26e+00   0.0017 **
## Year2000           6.70e-01   4.75e-08   1.41e+07  <2e-16 ***
## Year2002           2.67e-02   4.22e-01   6.00e-02   0.9497
## Year2003          -2.23e-01   2.41e-01  -9.30e-01   0.3577
## Year2004          -5.87e-01   2.14e-01  -2.74e+00   0.0077 **
## Year2005          -1.31e-01   1.59e-01  -8.30e-01   0.4106
## Year2006           1.15e-01   1.91e-01   6.00e-01   0.5494
## Year2007          -2.67e-01   2.38e-01  -1.12e+00   0.2668
## Year2008          -4.44e-01   2.27e-01  -1.96e+00   0.0544 .
## Year2009          -3.35e-01   1.81e-01  -1.85e+00   0.0682 .
## Year2010          -5.68e-01   1.87e-01  -3.04e+00   0.0033 **
## Year2011          -3.86e-02   2.15e-01  -1.80e-01   0.8579
## Year2012          -1.43e-01   3.34e-01  -4.30e-01   0.6695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.419
## Multiple R-squared:  0.341, Adjusted R-squared:  0.149
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 84 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0785 0.8900 0.9480 0.9020 0.9810 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.06e-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

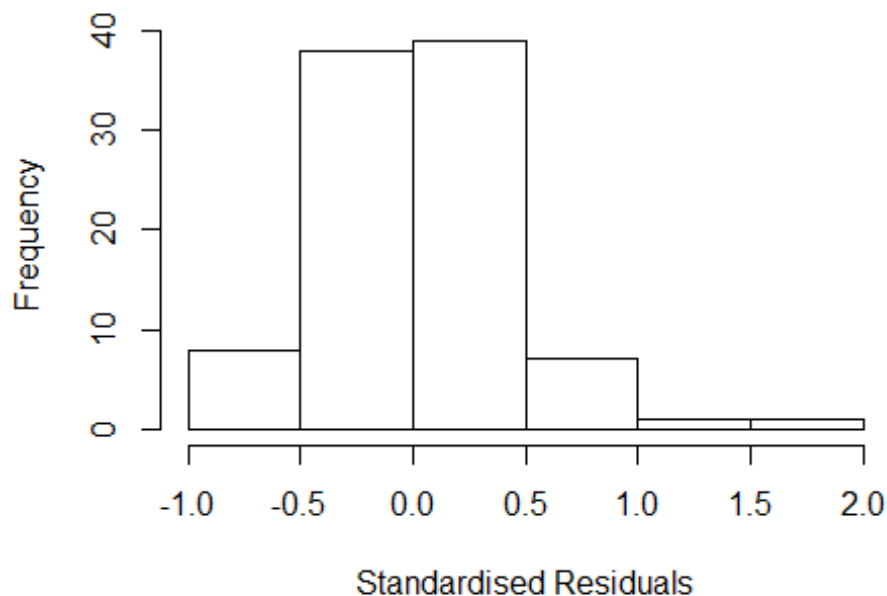
```

```
##           "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 lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful
```

Residuals from first and last author and team size



```
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale NaN 1 NaN
## LastAuthorFemale NaN 1 NaN
## Year NaN 15 NaN

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

```

## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.13781 -0.28571  0.00634  0.27281  1.20043
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11500    0.00000      Inf <2e-16 ***
## FirstAuthorFemale1 -0.13535    0.12687    -1.07   0.289
## LastAuthorFemale1  0.03361    0.12852     0.26   0.794
## Year1997         0.17050    0.13328     1.28   0.205
## Year1998         0.83300    0.00000      Inf <2e-16 ***
## Year1999        -0.50087    0.27270    -1.84   0.070 .
## Year2000         0.67000    0.00000      Inf <2e-16 ***
## Year2002         0.29579    0.39727     0.74   0.459
## Year2003         0.02367    0.22738     0.10   0.917
## Year2004        -0.28430    0.18988    -1.50   0.138
## Year2005         0.13695    0.14701     0.93   0.355
## Year2006         0.37439    0.21493     1.74   0.086 .
## Year2007         0.12456    0.24645     0.51   0.615
## Year2008        -0.06368    0.13935    -0.46   0.649
## Year2009         0.00998    0.14755     0.07   0.946
## Year2010        -0.10977    0.13735    -0.80   0.427
## Year2011         0.31843    0.17447     1.83   0.072 .
## Year2012         0.07107    0.18105     0.39   0.696
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.483
## Multiple R-squared:  0.189, Adjusted R-squared:  0.00738
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 86 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.517  0.884  0.965   0.917   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.06e-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"

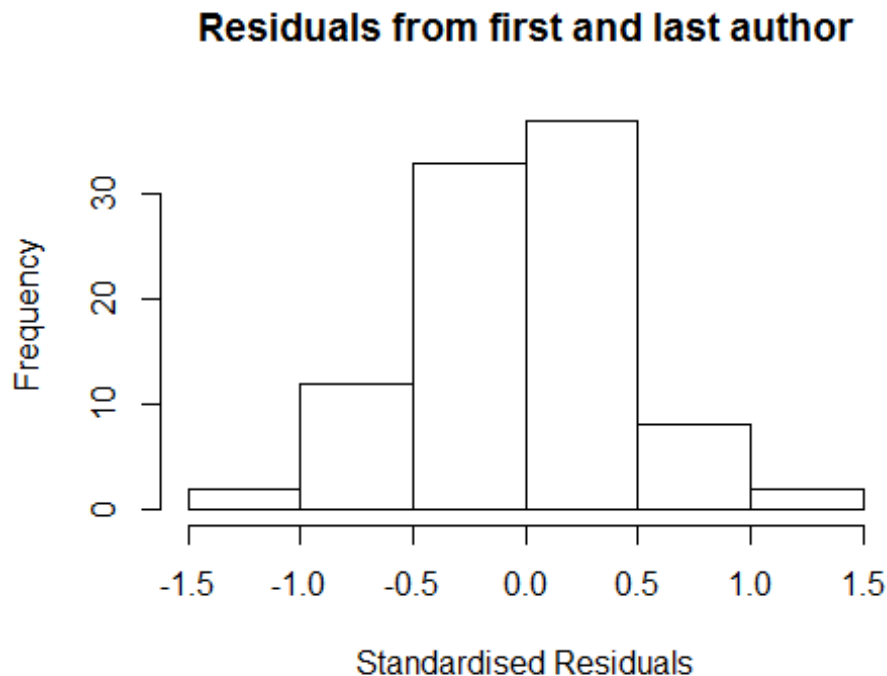
```



```
## 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
```



```
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale NaN 1          NaN
## Year             NaN 15          NaN

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.12e+00 -2.82e-01 -1.11e-16  2.75e-01  1.22e+00
##
## Coefficients:
```

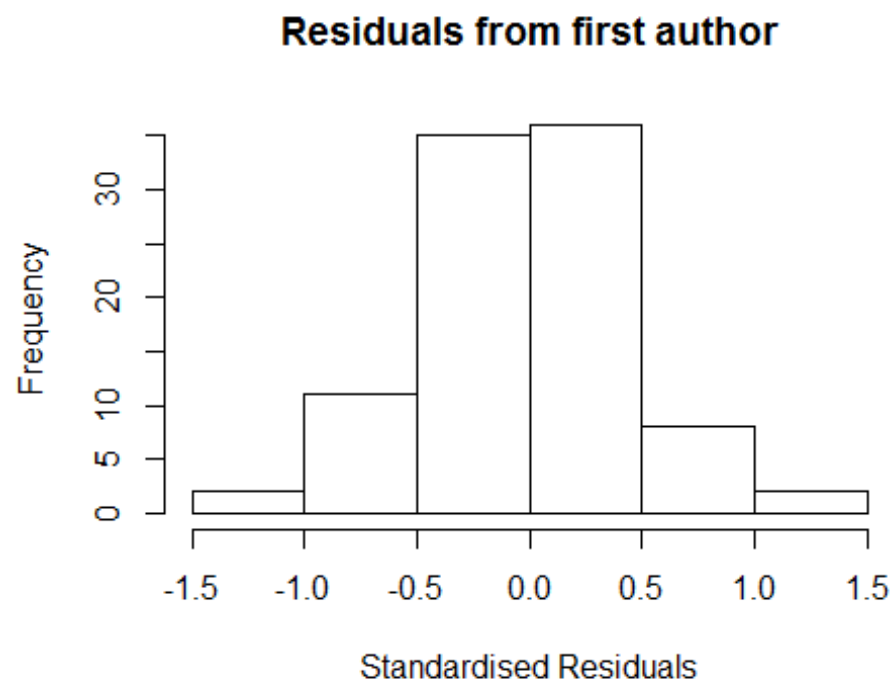
```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1150    0.0000    Inf    <2e-16 ***
## FirstAuthorFemale1 -0.1252    0.1285   -0.97    0.333
## Year1997          0.1705    0.1333    1.28    0.205
## Year1998          0.8330    0.0000    Inf    <2e-16 ***
## Year1999         -0.5008    0.2728   -1.84    0.070 .
## Year2000          0.6700    0.0000    Inf    <2e-16 ***
## Year2002          0.3003    0.4102    0.73    0.466
## Year2003          0.0227    0.2313    0.10    0.922
## Year2004         -0.2669    0.1847   -1.45    0.152
## Year2005          0.1505    0.1419    1.06    0.292
## Year2006          0.3778    0.2109    1.79    0.077 .
## Year2007          0.1263    0.2414    0.52    0.602
## Year2008         -0.0559    0.1403   -0.40    0.691
## Year2009          0.0231    0.1294    0.18    0.859
## Year2010         -0.0981    0.1207   -0.81    0.419
## Year2011          0.3299    0.1731    1.91    0.060 .
## Year2012          0.0866    0.1669    0.52    0.605
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.482
## Multiple R-squared:  0.189, Adjusted R-squared:  0.0201
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 86 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.505  0.882  0.965  0.916  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.06e-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

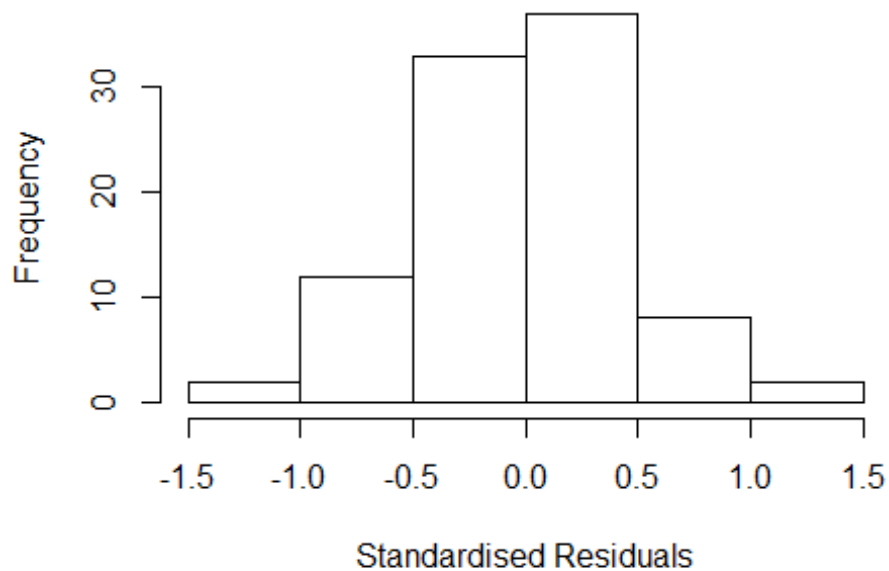
```

```
## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful
```



```
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale  NaN  1          NaN
## Year              NaN 15          NaN
```

Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19039 -0.29082  0.00158  0.30222  1.16444
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11500    0.00000      Inf <2e-16 ***
## LastAuthorFemale1 -0.00678    0.13217   -0.05  0.959
## Year1997         0.17050    0.13323    1.28  0.204
## Year1998         0.83300    0.00000      Inf <2e-16 ***
## Year1999        -0.50123    0.27239   -1.84  0.070 .
## Year2000         0.67000    0.00000      Inf <2e-16 ***
## Year2002         0.27506    0.35680    0.77  0.443
## Year2003        -0.00060    0.24225    0.00  0.998
## Year2004        -0.28266    0.20028   -1.41  0.162
## Year2005         0.09455    0.12965    0.73  0.468
## Year2006         0.34301    0.20366    1.68  0.096 .
## Year2007         0.08217    0.26387    0.31  0.756
```

```

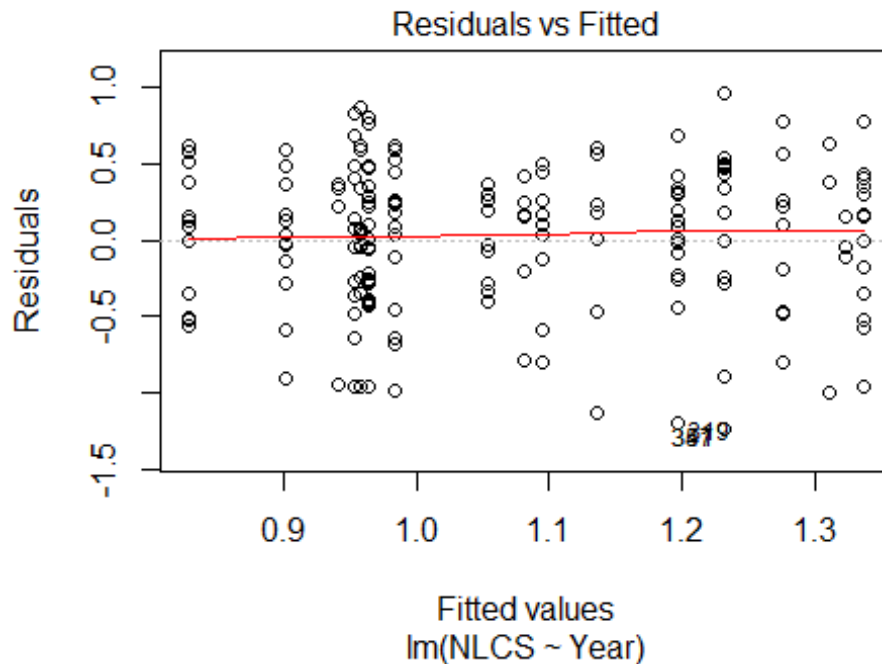
## Year2008          -0.07894      0.14209    -0.56      0.580
## Year2009          -0.04508      0.13785    -0.33      0.745
## Year2010          -0.15704      0.14970    -1.05      0.297
## Year2011           0.26311      0.16596     1.59      0.117
## Year2012          -0.01771      0.15833    -0.11      0.911
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.176, Adjusted R-squared:  0.00502
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 80 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.529  0.878  0.957   0.910  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      1.06e-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: 94"
## [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]"
## < 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 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
##    19    14     9    15    43    25    26    20    15    16     5    21    24    25    19
## 2011 2012
##    24    17
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##     9     9     6     4     7     3    13    13     9    13     3    13    18    17    14
## 2011 2012
##    21    13
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##     9     6     6     2     4     2     9    13     7     9     3    13    16    15    14
## 2011 2012
##    17    12
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.5

```



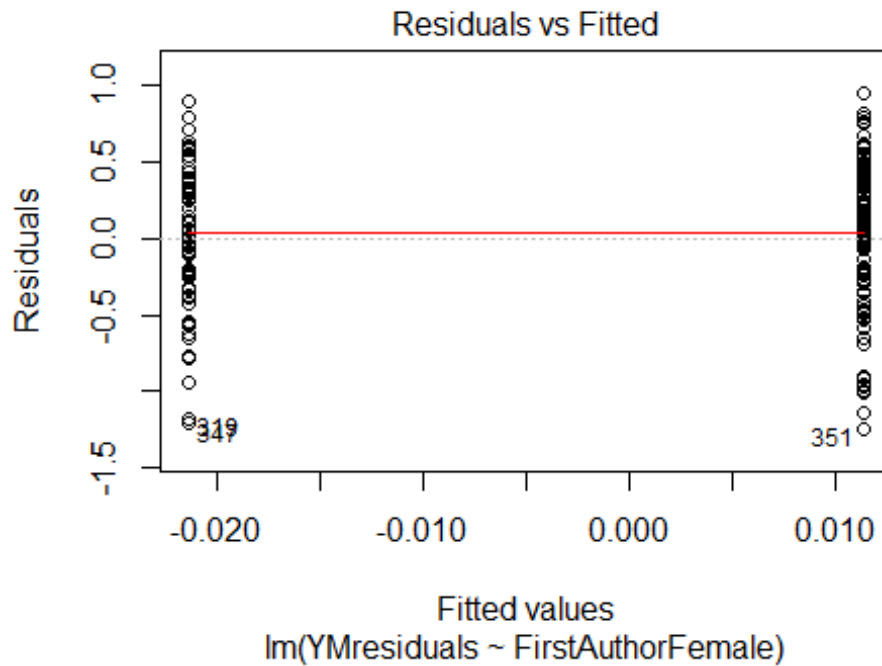
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.099, df = 1, p-value = 0.8

## [1] "Female first author team size 2018 geometric mean: 2.03393700979443"
## [1] "Male first author team size 2018 geometric mean: 2.74945927399721"

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

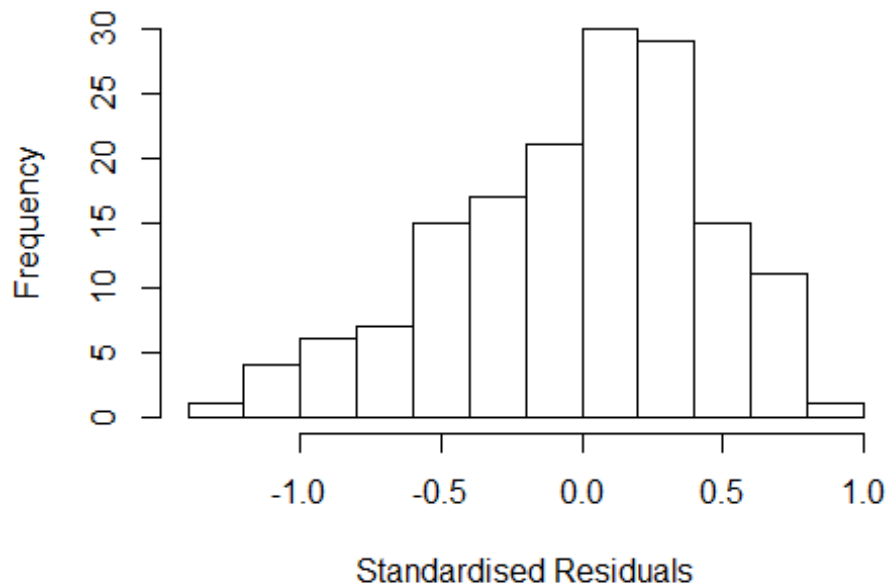
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.76632373444518"
## [1] "Male last author team size 2018 geometric mean: 2.10397901101729"

## Warning in wilcox.test.default(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 = 0.7
## 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 1.436 1      1.198
## LastAuthorFemale  1.570 1      1.253
## UniqueAuthors    6.640 4      1.267
## Year             10.830 16     1.077
```


Residuals from first and last author and team size



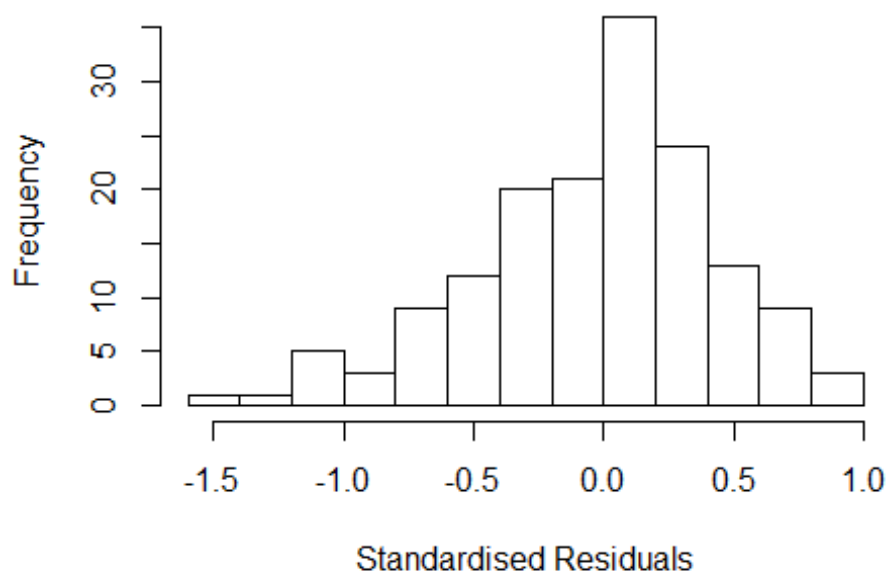
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3897 -0.3224 0.0656 0.2733 0.8023
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.91798 0.17853 5.14 9.4e-07 ***
## FirstAuthorFemale1 -0.08419 0.08611 -0.98 0.330
## LastAuthorFemale1 -0.05523 0.08700 -0.63 0.527
## UniqueAuthors2 0.35889 0.16955 2.12 0.036 *
## UniqueAuthors3 0.23189 0.16134 1.44 0.153
## UniqueAuthors4 0.29663 0.14847 2.00 0.048 *
## UniqueAuthors5 0.38251 0.14764 2.59 0.011 *
## Year1997 0.16109 0.25341 0.64 0.526
## Year1998 0.00661 0.18907 0.03 0.972
## Year1999 -0.39092 1.07329 -0.36 0.716
```

```

## Year2000      0.25391      0.20449      1.24      0.217
## Year2001      0.34692      0.15253      2.27      0.025 *
## Year2002      0.30342      0.23790      1.28      0.204
## Year2003     -0.22980      0.19838     -1.16      0.249
## Year2004     -0.15561      0.19458     -0.80      0.425
## Year2005     -0.38751      0.18876     -2.05      0.042 *
## Year2006      0.24941      0.54346      0.46      0.647
## Year2007     -0.35142      0.18237     -1.93      0.056 .
## Year2008     -0.11818      0.16653     -0.71      0.479
## Year2009      0.01187      0.15754      0.08      0.940
## Year2010      0.17509      0.22110      0.79      0.430
## Year2011     -0.17646      0.17655     -1.00      0.319
## Year2012     -0.09707      0.18928     -0.51      0.609
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.185, Adjusted R-squared:  0.0516
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 146 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.398  0.903  0.960  0.920  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.37e-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 1.339 1      1.157
## LastAuthorFemale  1.498 1      1.224
## Year              1.850 16      1.019

```

Residuals from first and last author



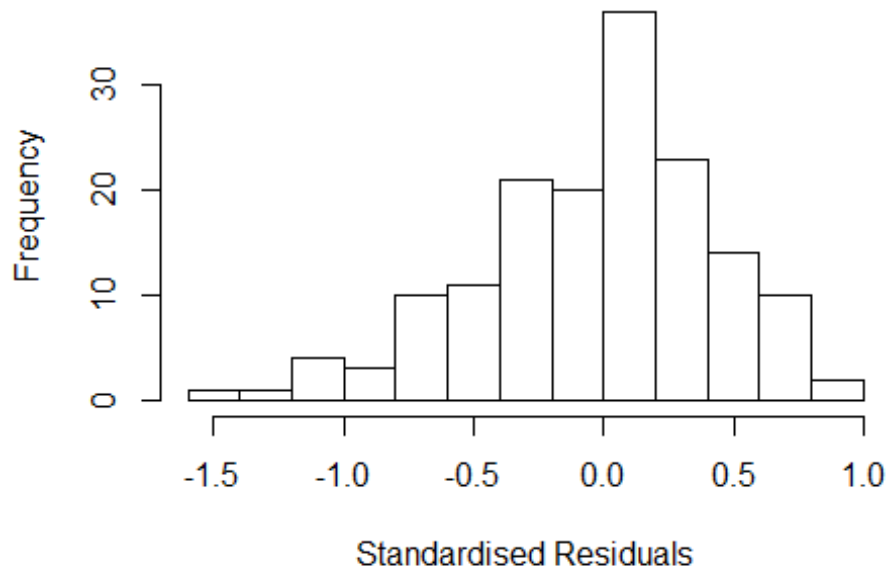
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4102 -0.3100 0.0549 0.3068 0.8250
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1985 0.1449 8.27 9.9e-14 ***
## FirstAuthorFemale1 -0.0914 0.0859 -1.06 0.289
## LastAuthorFemale1 -0.0407 0.0889 -0.46 0.648
## Year1997 0.0860 0.2855 0.30 0.764
## Year1998 -0.0695 0.2130 -0.33 0.745
## Year1999 -0.5555 0.7344 -0.76 0.451
## Year2000 0.3189 0.1857 1.72 0.088 .
## Year2001 0.2110 0.1478 1.43 0.156
## Year2002 0.2380 0.2256 1.05 0.293
## Year2003 -0.1958 0.2116 -0.93 0.356
## Year2004 -0.1276 0.1935 -0.66 0.510
## Year2005 -0.3644 0.2031 -1.79 0.075 .
```

```

## Year2006          0.2901      0.5629      0.52      0.607
## Year2007         -0.3354      0.1924     -1.74      0.084 .
## Year2008         -0.0875      0.1797     -0.49      0.627
## Year2009          0.0475      0.1660      0.29      0.775
## Year2010          0.2117      0.2235      0.95      0.345
## Year2011         -0.1581      0.1801     -0.88      0.381
## Year2012         -0.0941      0.1943     -0.48      0.629
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.155, Adjusted R-squared:  0.0448
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 147 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.354  0.885   0.955   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      6.37e-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 1.294 1      1.138
## Year              1.294 16      1.008

```

Residuals from first author



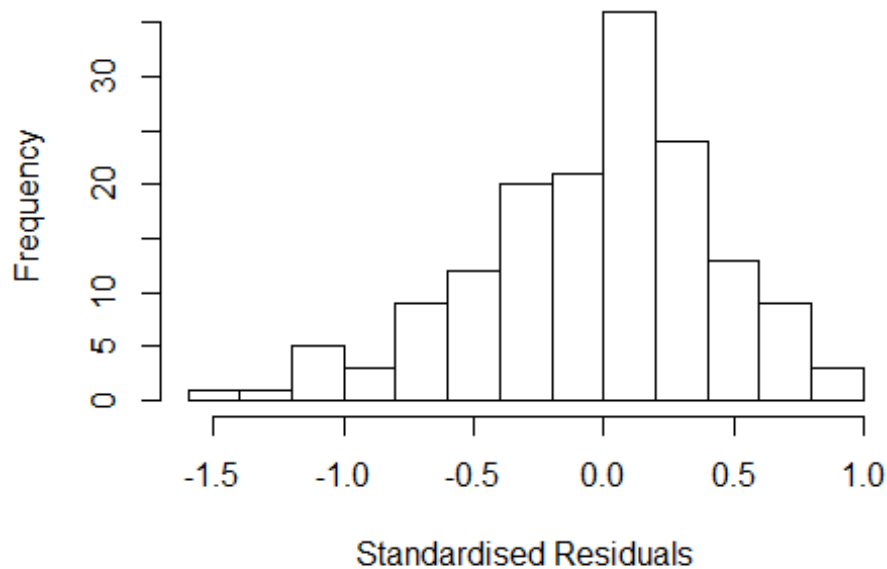
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4040 -0.3218 0.0453 0.3130 0.8258
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1784 0.1451 8.12 2.2e-13 ***
## FirstAuthorFemale1 -0.0973 0.0857 -1.14 0.258
## Year1997 0.0923 0.2834 0.33 0.745
## Year1998 -0.0589 0.2138 -0.28 0.783
## Year1999 -0.5354 0.7530 -0.71 0.478
## Year2000 0.3407 0.1850 1.84 0.068 .
## Year2001 0.2138 0.1531 1.40 0.165
## Year2002 0.2433 0.2266 1.07 0.285
## Year2003 -0.1937 0.2110 -0.92 0.360
## Year2004 -0.1043 0.1932 -0.54 0.590
## Year2005 -0.3414 0.2032 -1.68 0.095 .
## Year2006 0.3231 0.5622 0.57 0.566
```

```

## Year2007          -0.3200      0.1959   -1.63    0.105
## Year2008          -0.0749      0.1827   -0.41    0.682
## Year2009           0.0583      0.1698    0.34    0.732
## Year2010           0.2257      0.2267    1.00    0.321
## Year2011          -0.1457      0.1823   -0.80    0.425
## Year2012          -0.0789      0.1995   -0.40    0.693
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.464
## Multiple R-squared:  0.156, Adjusted R-squared:  0.0533
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 145 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.339  0.877   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      6.37e-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 1.431 1          1.196
## Year              1.431 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
## -1.3776 -0.3273 0.0458 0.3218 0.8554
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1524 0.1443 7.99 4.8e-13 ***
## LastAuthorFemale1 -0.0563 0.0904 -0.62 0.534
## Year1997 0.1100 0.2913 0.38 0.706
## Year1998 -0.0269 0.2165 -0.12 0.901
## Year1999 -0.5094 0.7286 -0.70 0.486
## Year2000 0.3406 0.1915 1.78 0.077 .
## Year2001 0.2193 0.1590 1.38 0.170
## Year2002 0.2732 0.2154 1.27 0.207
## Year2003 -0.1645 0.2142 -0.77 0.444
## Year2004 -0.1329 0.1879 -0.71 0.481
## Year2005 -0.3572 0.2009 -1.78 0.078 .
## Year2006 0.3322 0.5607 0.59 0.554
```

```

## Year2007          -0.3154      0.1916    -1.65      0.102
## Year2008          -0.0755      0.1812    -0.42      0.678
## Year2009           0.0591      0.1677      0.35      0.725
## Year2010           0.2252      0.2340      0.96      0.338
## Year2011          -0.1316      0.1797    -0.73      0.465
## Year2012          -0.0889      0.1920    -0.46      0.644
## ---
## 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:  0.0432
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 143 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.383  0.894  0.952  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.37e-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: 157"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3608"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
## < table of extent 0 >
## < table of extent 0 >
## < table of extent 0 >
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: NaN"
## [1] "Male first author team size 2018 geometric mean: NaN"
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"

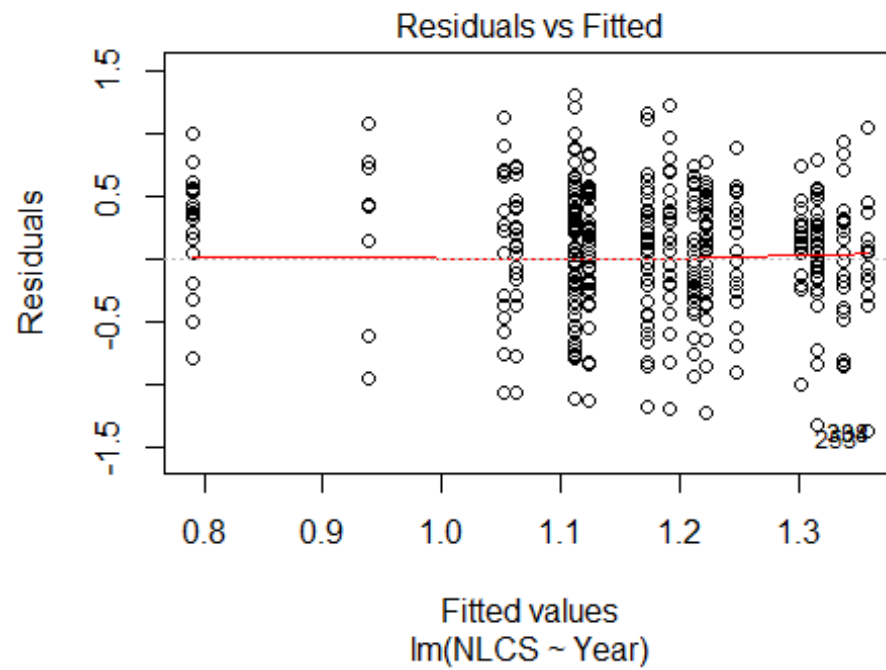
```



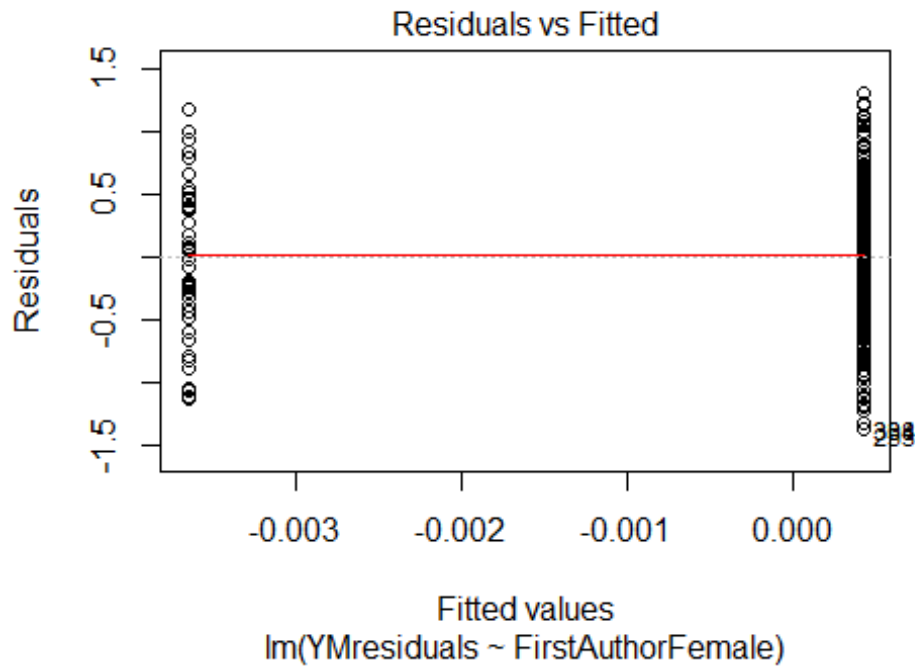
```

## [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 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
##    34    20    22    24    21    44    27    15    15    23    39    55    26    38    49
## 2011 2012
##    47    45
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    31    18    20    22    18    31    25    14    12    19    34    51    20    33    42
## 2011 2012
##    44    38
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    29    17    20    20    18    27    24    14    12    18    32    47    19    30    37
## 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 = 16, df = 16, p-value = 0.5

```

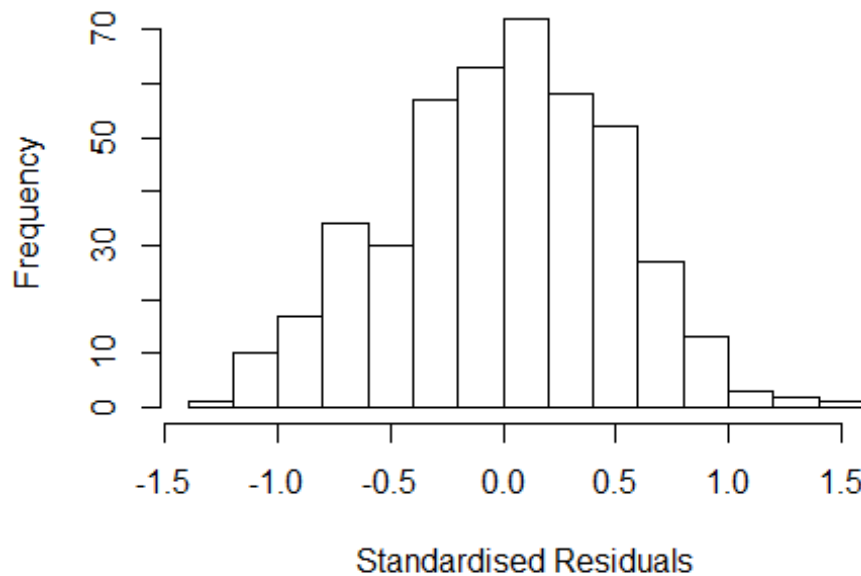


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.8, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.51434696283621"
## [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 1.160 1          1.077
## LastAuthorFemale  1.337 1          1.156
## UniqueAuthors    2.919 4          1.143
## Year              3.572 16          1.041
```

Residuals from first and last author and team size



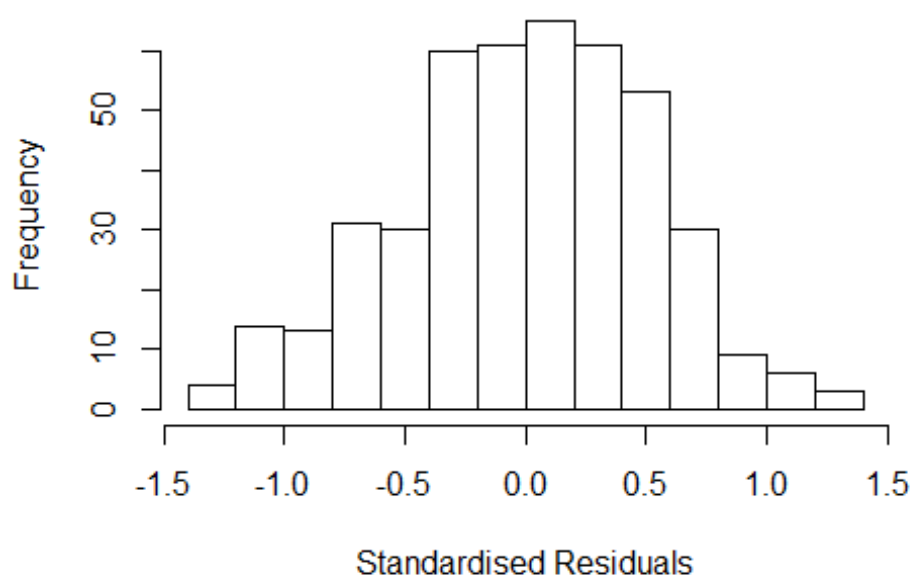
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3057 -0.3191 0.0119 0.3503 1.4519
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01858 0.16694 6.10 2.4e-09 ***
## FirstAuthorFemale1 -0.01255 0.10580 -0.12 0.906
## LastAuthorFemale1 0.01169 0.06770 0.17 0.863
## UniqueAuthors2 0.16035 0.09526 1.68 0.093 .
## UniqueAuthors3 0.10111 0.09266 1.09 0.276
## UniqueAuthors4 0.27957 0.10115 2.76 0.006 **
## UniqueAuthors5 0.40883 0.09877 4.14 4.2e-05 ***
## Year1997 0.12318 0.17656 0.70 0.486
## Year1998 0.23037 0.15904 1.45 0.148
## Year1999 -0.06871 0.14894 -0.46 0.645
```

```

## Year2000      -0.11046    0.21872   -0.51    0.614
## Year2001      0.02946    0.13820    0.21    0.831
## Year2002     -0.39779    0.15780   -2.52    0.012 *
## Year2003      0.16743    0.15902    1.05    0.293
## Year2004     -0.27326    0.32339   -0.84    0.399
## Year2005     -0.10407    0.15516   -0.67    0.503
## Year2006      0.12761    0.13014    0.98    0.327
## Year2007     -0.10043    0.12523   -0.80    0.423
## Year2008      0.04914    0.15421    0.32    0.750
## Year2009     -0.04766    0.15434   -0.31    0.758
## Year2010     -0.03438    0.14361   -0.24    0.811
## Year2011     -0.00725    0.13292   -0.05    0.957
## Year2012     -0.08879    0.14411   -0.62    0.538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.503
## Multiple R-squared:  0.122, Adjusted R-squared:  0.0759
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 401 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.385  0.866  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      2.27e-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 1.125 1      1.061
## LastAuthorFemale  1.194 1      1.093
## Year              1.342 16      1.009

```

Residuals from first and last author



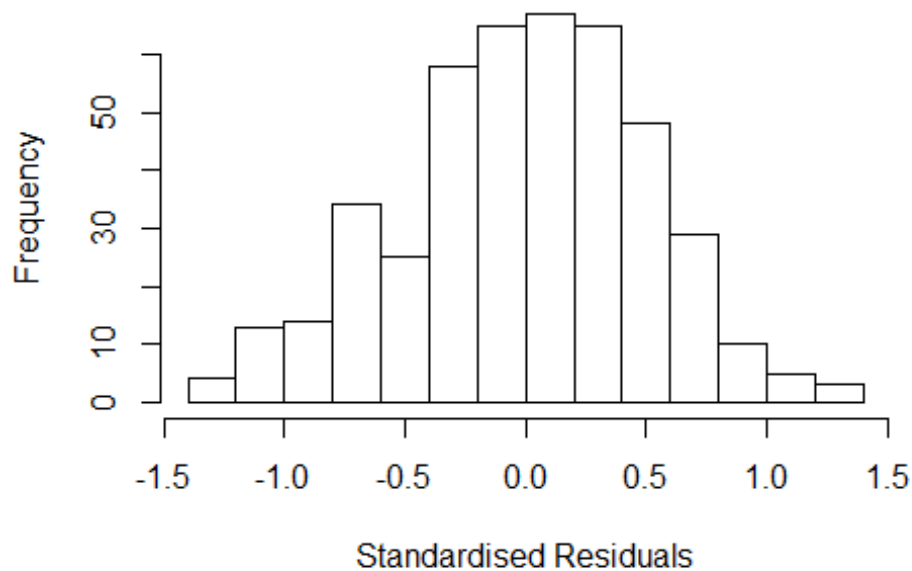
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3763 -0.3421 0.0233 0.3552 1.2761
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1632 0.1428 8.14 4.4e-15 ***
## FirstAuthorFemale1 -0.0196 0.1050 -0.19 0.852
## LastAuthorFemale1 -0.0542 0.0679 -0.80 0.425
## Year1997 0.1298 0.1706 0.76 0.447
## Year1998 0.2531 0.1467 1.73 0.085 .
## Year1999 -0.0552 0.1550 -0.36 0.722
## Year2000 -0.0567 0.2117 -0.27 0.789
## Year2001 0.1204 0.1355 0.89 0.375
## Year2002 -0.3445 0.1604 -2.15 0.032 *
## Year2003 0.2869 0.1490 1.93 0.055 .
## Year2004 -0.1430 0.3172 -0.45 0.652
## Year2005 0.0459 0.1483 0.31 0.757
```

```

## Year2006          0.2415      0.1253      1.93      0.055 .
## Year2007          0.0301      0.1173      0.26      0.797
## Year2008          0.1911      0.1513      1.26      0.207
## Year2009          0.0663      0.1500      0.44      0.659
## Year2010          0.0774      0.1385      0.56      0.576
## Year2011          0.1315      0.1169      1.12      0.261
## Year2012          0.0565      0.1397      0.40      0.686
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.0663, Adjusted R-squared:  0.0264
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 35 weights are ~= 1. The remaining 405 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.456  0.871  0.952  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      2.27e-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 1.127 1      1.062
## 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
## -1.3848 -0.3398 0.0235 0.3581 1.2707
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1347 0.1407 8.06 7.7e-15 ***
## FirstAuthorFemale1 -0.0299 0.1064 -0.28 0.779
## Year1997 0.1150 0.1675 0.69 0.493
## Year1998 0.2386 0.1440 1.66 0.098 .
## Year1999 -0.0669 0.1538 -0.43 0.664
## Year2000 -0.0733 0.2126 -0.34 0.731
## Year2001 0.1119 0.1346 0.83 0.406
## Year2002 -0.3476 0.1652 -2.10 0.036 *
## Year2003 0.2800 0.1454 1.93 0.055 .
## Year2004 -0.1346 0.3306 -0.41 0.684
## Year2005 0.0370 0.1472 0.25 0.801
## Year2006 0.2405 0.1256 1.91 0.056 .
```

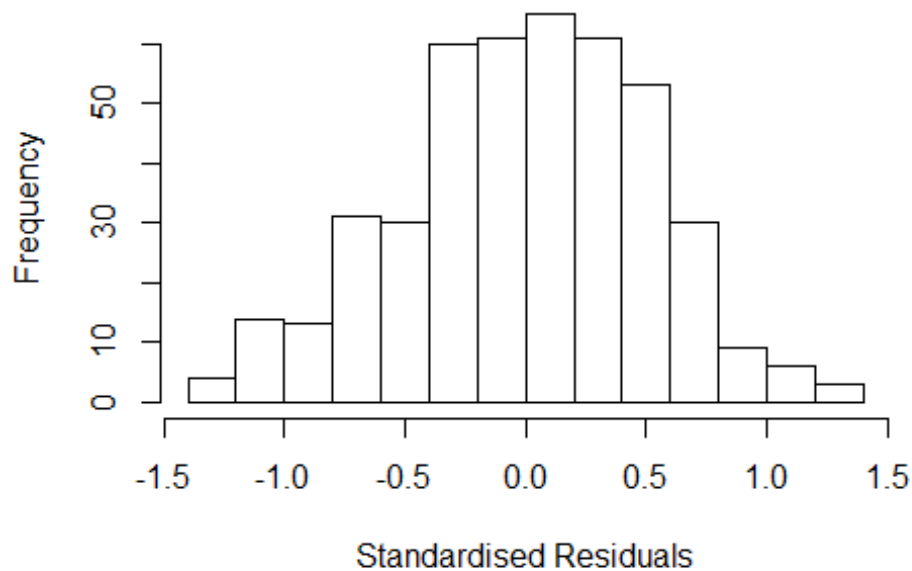


```

## Year2007          0.0318      0.1179      0.27      0.787
## Year2008          0.1849      0.1537      1.20      0.230
## Year2009          0.0636      0.1508      0.42      0.673
## Year2010          0.0752      0.1398      0.54      0.591
## Year2011          0.1202      0.1153      1.04      0.298
## Year2012          0.0465      0.1386      0.34      0.737
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.504
## Multiple R-squared:  0.0653, Adjusted R-squared:  0.0277
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 409 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.431  0.862  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      2.27e-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 1.189 1      1.090
## Year      1.189 16      1.005

```

Residuals from last author



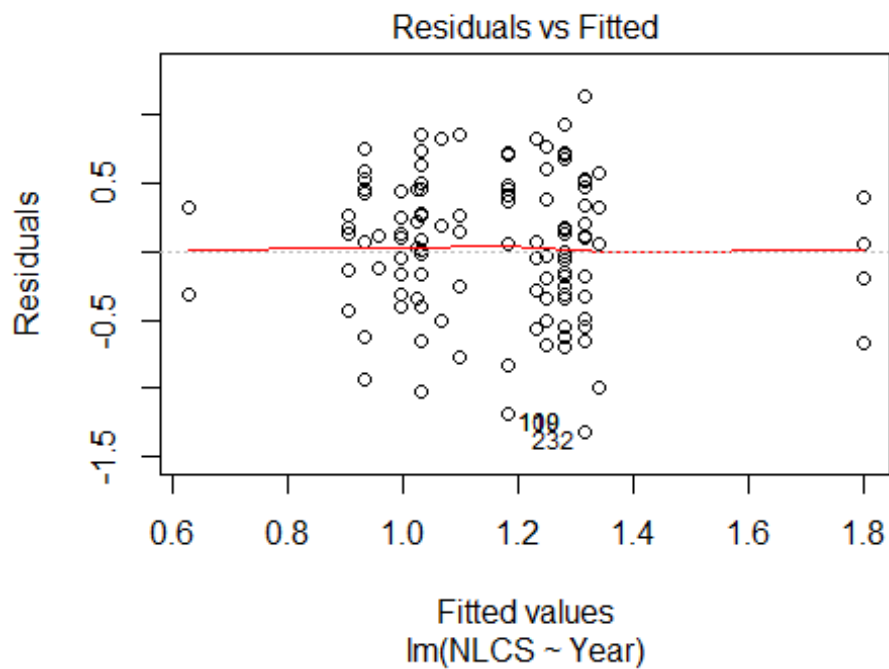
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3802 -0.3441 0.0219 0.3531 1.2751
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1453 0.1027 11.15 <2e-16 ***
## LastAuthorFemale1 -0.0565 0.0689 -0.82 0.413
## Year1997 0.1326 0.1708 0.78 0.438
## Year1998 0.2581 0.1463 1.76 0.078 .
## Year1999 -0.0525 0.1551 -0.34 0.735
## Year2000 -0.0573 0.2132 -0.27 0.788
## Year2001 0.1207 0.1359 0.89 0.375
## Year2002 -0.3426 0.1627 -2.11 0.036 *
## Year2003 0.2913 0.1489 1.96 0.051 .
## Year2004 -0.1412 0.3266 -0.43 0.666
## Year2005 0.0509 0.1491 0.34 0.733
## Year2006 0.2469 0.1246 1.98 0.048 *
```

```

## Year2007          0.0325      0.1177      0.28      0.783
## Year2008          0.1940      0.1520      1.28      0.203
## Year2009          0.0686      0.1497      0.46      0.647
## Year2010          0.0789      0.1390      0.57      0.570
## Year2011          0.1341      0.1171      1.15      0.253
## Year2012          0.0581      0.1401      0.41      0.679
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.504
## Multiple R-squared:  0.0671, Adjusted R-squared:  0.0295
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 409 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.434  0.866  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      2.27e-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: 440"
## [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
##    7    5    2    6    9    9   11    9    7    3    7   14   11   21   23
## 2011 2012
##   26   22
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    4    2    5    7    5    8    8    5    2    6   10    5   11   12
## 2011 2012

```

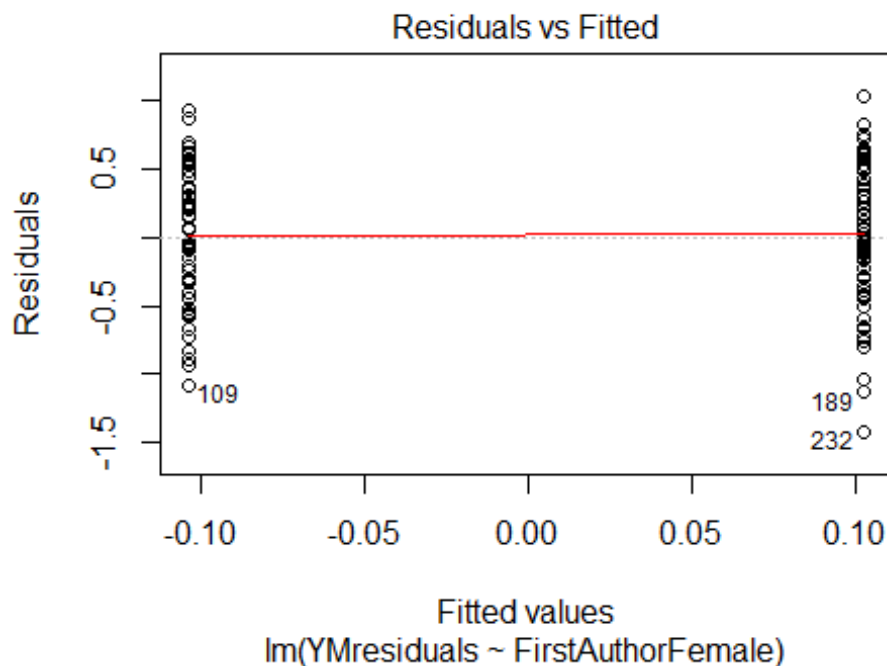
```
## 19 15
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5 4 2 5 7 5 6 6 4 1 5 9 5 8 11
## 2011 2012
## 18 14
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.6
```



```
##
## 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: 4.03054144421362"
## [1] "Male first author team size 2018 geometric mean: 4.82028452835046"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.7224194364084"
## [1] "Male last author team size 2018 geometric mean: 4.64775708972681"

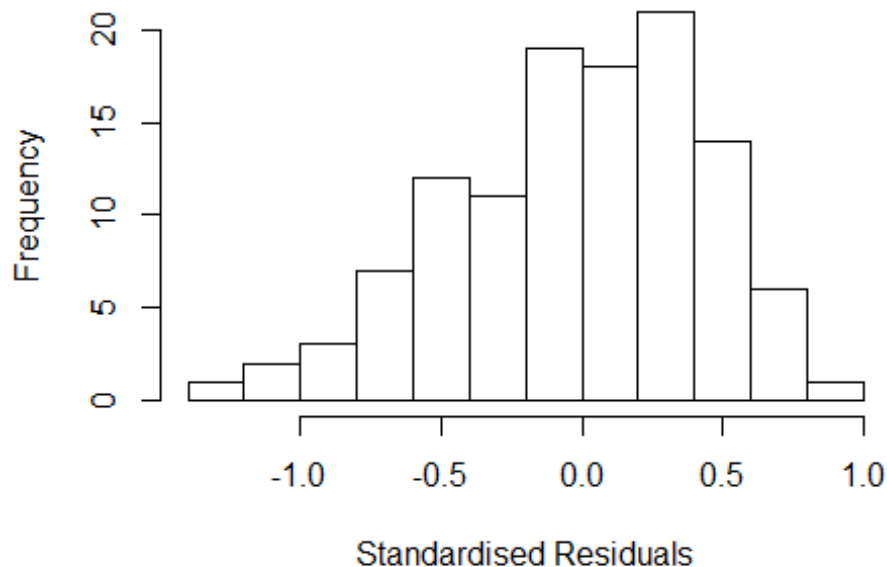
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6.5, p-value = 0.3
## 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	4.710	1	2.170
LastAuthorFemale	3.242	1	1.801
UniqueAuthors	209.271	4	1.950
Year	573.544	16	1.220

Residuals from first and last author and team size



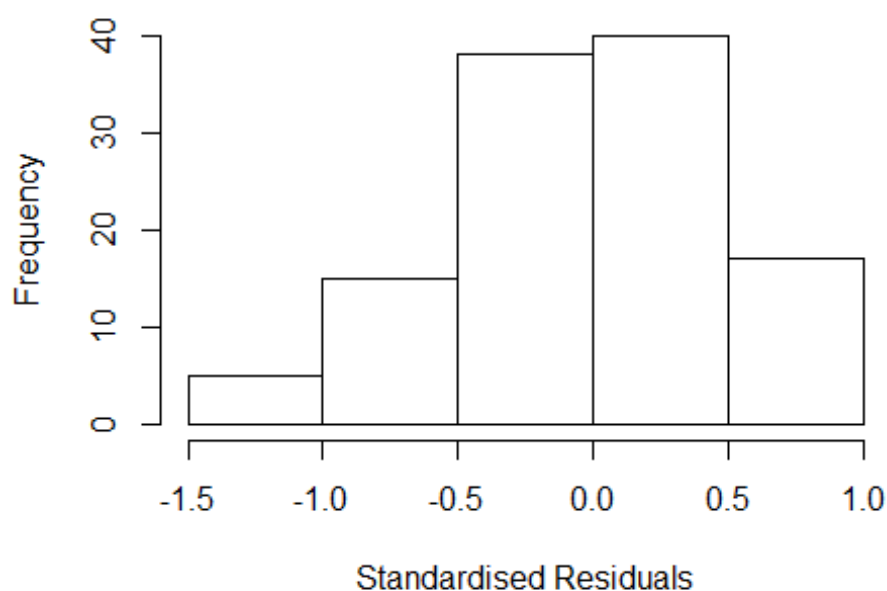
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2717 -0.3167 0.0627 0.3232 0.8629
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.653 0.259 6.37 7.2e-09 ***
## FirstAuthorFemale1 0.382 0.109 3.49 0.00073 ***
## LastAuthorFemale1 -0.379 0.110 -3.45 0.00084 ***
## UniqueAuthors2 0.270 0.214 1.26 0.21082
## UniqueAuthors3 0.451 0.178 2.53 0.01313 *
## UniqueAuthors4 0.491 0.152 3.23 0.00171 **
## UniqueAuthors5 0.553 0.189 2.93 0.00432 **
## Year1997 -0.959 0.355 -2.70 0.00820 **
## Year1998 -1.500 0.269 -5.57 2.5e-07 ***
## Year1999 -0.632 0.388 -1.63 0.10650
```

```

## Year2000          -0.589      0.257    -2.29   0.02420 *
## Year2001          -1.078      0.268    -4.03   0.00012 ***
## Year2002          -0.851      0.277    -3.07   0.00285 **
## Year2003          -0.805      0.269    -2.99   0.00355 **
## Year2004          -1.305      0.245    -5.32   7.3e-07 ***
## Year2005          -1.030      0.287    -3.59   0.00053 ***
## Year2006          -1.063      0.318    -3.34   0.00122 **
## Year2007          -1.165      0.344    -3.38   0.00106 **
## Year2008          -0.777      0.307    -2.53   0.01322 *
## Year2009          -1.152      0.420    -2.74   0.00735 **
## Year2010          -0.994      0.291    -3.41   0.00095 ***
## Year2011          -0.964      0.273    -3.53   0.00065 ***
## Year2012          -0.937      0.331    -2.83   0.00570 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.518
## Multiple R-squared:  0.355, Adjusted R-squared:  0.201
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 108 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.526  0.895   0.961   0.928   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 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 4.361 1           2.088
## LastAuthorFemale  3.624 1           1.904
## Year              9.839 16           1.074

```

Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + 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 -2.98e-01  4.44e-16  3.21e-01  9.01e-01
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.843      0.160   11.48 < 2e-16 ***
## FirstAuthorFemale1  0.438      0.108    4.06 0.00010 ***
## LastAuthorFemale1 -0.382      0.115   -3.32 0.00128 **
## Year1997          -0.689      0.270   -2.55 0.01227 *
## Year1998          -1.245      0.147   -8.45 3.2e-13 ***
## Year1999          -0.679      0.263   -2.58 0.01144 *
## Year2000          -0.649      0.155   -4.20 6.1e-05 ***
## Year2001          -1.064      0.239   -4.45 2.3e-05 ***
## Year2002          -0.660      0.201   -3.28 0.00146 **
## Year2003          -0.912      0.190   -4.79 6.0e-06 ***
## Year2004          -1.168      0.231   -5.05 2.1e-06 ***
## Year2005          -0.770      0.160   -4.80 5.9e-06 ***
```

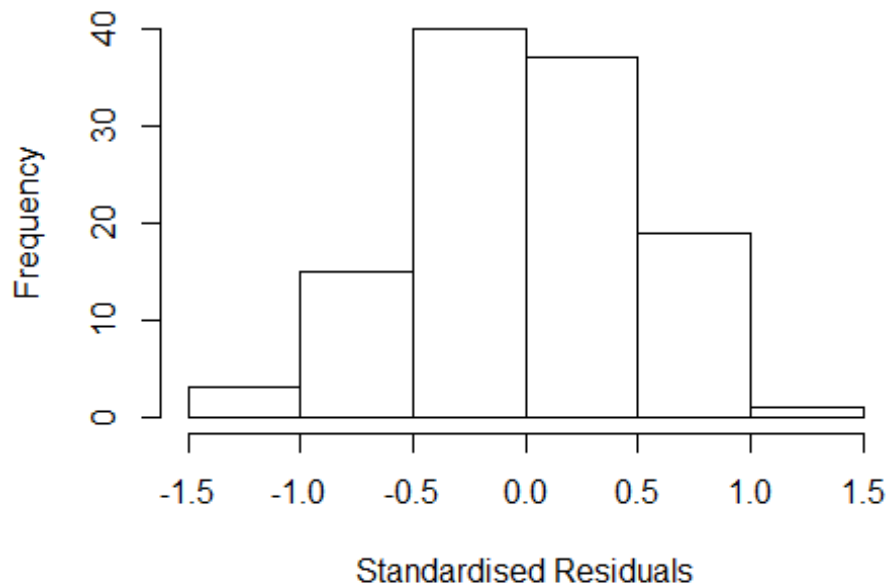


```

## Year2006          -0.822      0.280   -2.93  0.00422 **
## Year2007          -0.938      0.245   -3.83  0.00023 ***
## Year2008          -0.554      0.277   -2.00  0.04829 *
## Year2009          -0.881      0.310   -2.84  0.00548 **
## Year2010          -0.784      0.205   -3.83  0.00023 ***
## Year2011          -0.821      0.199   -4.13  7.7e-05 ***
## Year2012          -0.675      0.231   -2.92  0.00438 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.293, Adjusted R-squared:  0.161
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 104 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.439  0.893  0.960  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      8.70e-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 4.863 1      2.205
## Year              4.863 16      1.051

```

Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.460 -0.351 0.000 0.363 1.132
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.575 0.177 8.91 3.0e-14 ***
## FirstAuthorFemale1 0.383 0.125 3.05 0.0029 **
## Year1997 -0.610 0.294 -2.08 0.0404 *
## Year1998 -1.141 0.181 -6.32 8.1e-09 ***
## Year1999 -0.643 0.344 -1.87 0.0646 .
## Year2000 -0.515 0.195 -2.65 0.0095 **
## Year2001 -0.828 0.189 -4.39 2.9e-05 ***
## Year2002 -0.585 0.214 -2.73 0.0075 **
## Year2003 -0.644 0.204 -3.15 0.0021 **
## Year2004 -0.949 0.187 -5.08 1.8e-06 ***
## Year2005 -0.502 0.177 -2.84 0.0054 **
## Year2006 -0.694 0.289 -2.40 0.0184 *
```

```

## Year2007          -0.656      0.284   -2.31   0.0231 *
## Year2008          -0.271      0.277   -0.98   0.3307
## Year2009          -0.721      0.393   -1.83   0.0698 .
## Year2010          -0.568      0.231   -2.46   0.0157 *
## Year2011          -0.694      0.211   -3.30   0.0014 **
## Year2012          -0.498      0.280   -1.78   0.0782 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.51
## Multiple R-squared:  0.22,   Adjusted R-squared:  0.0835
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 102 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.392  0.873   0.943   0.904   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      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 3.457 1          1.859
## Year            3.457 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 ~ 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 -0.3404  0.0699  0.3426  1.0186

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.045      0.219   9.33 3.7e-15 ***
## LastAuthorFemale1 -0.304      0.130  -2.34 0.02160 *
## Year1997          -0.829      0.378  -2.19 0.03080 *
## Year1998          -1.268      0.235  -5.39 5.0e-07 ***
## Year1999          -0.646      0.271  -2.38 0.01936 *
## Year2000          -0.640      0.240  -2.66 0.00902 **
## Year2001          -1.080      0.266  -4.05 0.00010 ***
## Year2002          -0.838      0.271  -3.09 0.00262 **
## Year2003          -1.114      0.242  -4.61 1.2e-05 ***
## Year2004          -1.069      0.289  -3.70 0.00035 ***
## Year2005          -0.972      0.219  -4.44 2.4e-05 ***
## Year2006          -0.874      0.349  -2.50 0.01401 *
## Year2007          -0.776      0.391  -1.99 0.04982 *
## Year2008          -0.642      0.323  -1.99 0.04949 *
## Year2009          -0.865      0.362  -2.39 0.01876 *
## Year2010          -0.707      0.273  -2.59 0.01115 *
## Year2011          -0.778      0.260  -3.00 0.00348 **
## Year2012          -0.612      0.252  -2.43 0.01688 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.513
## Multiple R-squared:  0.181, Adjusted R-squared:  0.0379
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~ = 1. The remaining 107 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.519  0.872  0.956  0.906  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      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] ""

```

```

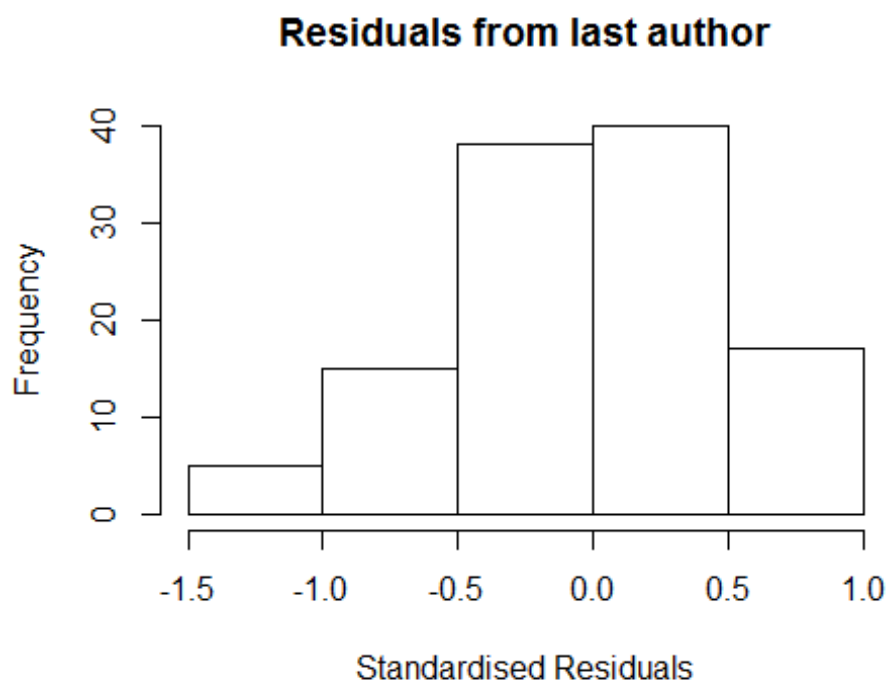
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3611"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 7 5 11 14 13 15 14 10 20 24 32 40 41 56
## 2011 2012
## 52 36
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 11 5 3 8 1 2 14 12 8 16 22 32 36 39 52
## 2011 2012
## 47 34
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 9 5 3 7 1 2 14 10 8 14 16 29 36 36 47
## 2011 2012
## 43 30
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4.74229011352561"
## [1] "Male first author team size 2018 geometric mean: 3.34124942092595"

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

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 470, p-value = 0.008
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.26483283370182"
## [1] "Male last author team size 2018 geometric mean: 4.24005475394138"

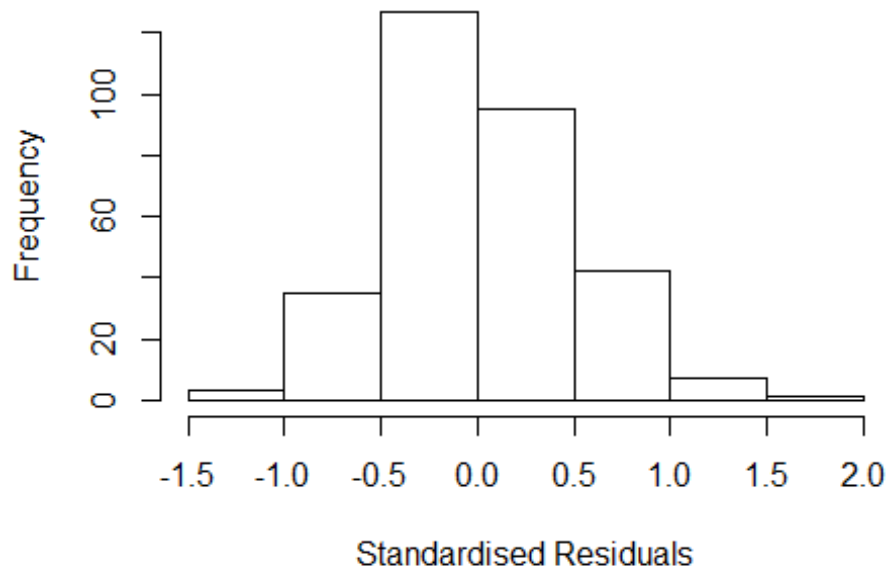
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 380, p-value = 0.6
## 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 2.920 1      1.709
## LastAuthorFemale  2.135 1      1.461
## UniqueAuthors    12.738 4      1.374
## Year              27.974 16     1.110
```

Residuals from first and last author and team size



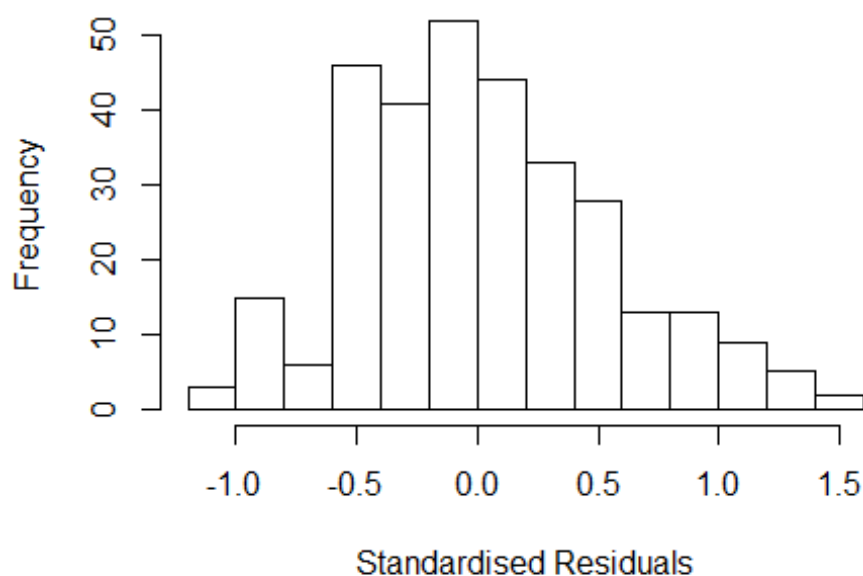
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1830 -0.3524 -0.0274 0.3636 1.6172
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.2520 0.1535 1.64 0.10182
## FirstAuthorFemale1 0.0434 0.0625 0.69 0.48838
## LastAuthorFemale1 -0.1052 0.0654 -1.61 0.10874
## UniqueAuthors2 0.2515 0.0978 2.57 0.01059 *
## UniqueAuthors3 0.2254 0.0996 2.26 0.02443 *
## UniqueAuthors4 0.2071 0.0955 2.17 0.03091 *
## UniqueAuthors5 0.5156 0.1148 4.49 1e-05 ***
## Year1997 -0.2181 0.2375 -0.92 0.35927
## Year1998 0.5477 0.1585 3.46 0.00063 ***
## Year1999 -0.0547 0.2200 -0.25 0.80373
```

```

## Year2000          0.4438      0.1444      3.07  0.00232 **
## Year2001          0.0684      0.4895      0.14  0.88901
## Year2002          0.2621      0.1743      1.50  0.13375
## Year2003          0.1092      0.2444      0.45  0.65524
## Year2004         -0.0822      0.2525     -0.33  0.74502
## Year2005          0.1709      0.1995      0.86  0.39223
## Year2006          0.0870      0.1694      0.51  0.60816
## Year2007         -0.0459      0.1457     -0.32  0.75297
## Year2008          0.0395      0.1620      0.24  0.80745
## Year2009          0.4484      0.1581      2.83  0.00491 **
## Year2010          0.2895      0.1430      2.02  0.04387 *
## Year2011          0.4773      0.1581      3.02  0.00277 **
## Year2012          0.4229      0.1696      2.49  0.01321 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.511
## Multiple R-squared:  0.217, Adjusted R-squared:  0.157
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 283 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.295  0.884  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      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 3.212  1      1.792
## LastAuthorFemale  2.197  1      1.482
## Year              4.675 16      1.049

```


Residuals from first and last author



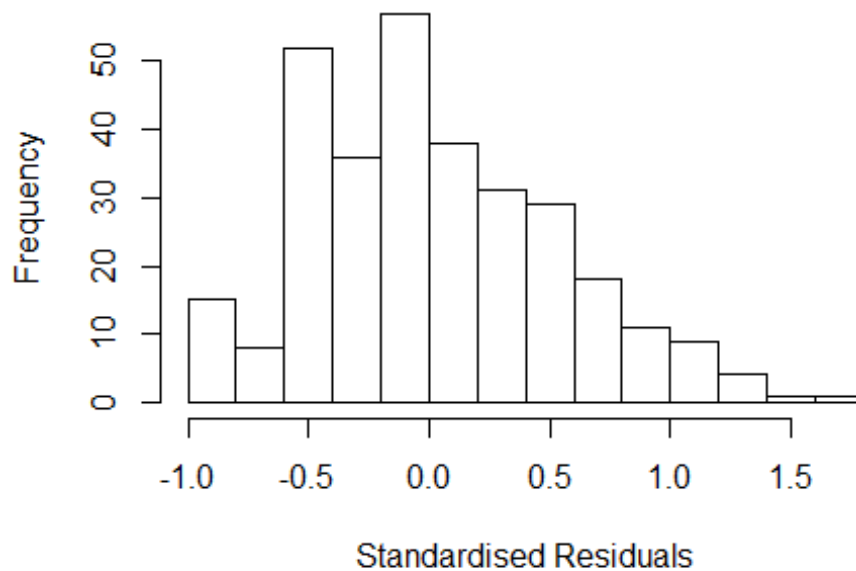
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0571 -0.3709 -0.0417 0.3562 1.5743
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.48834 0.15771 3.10 0.0021 **
## FirstAuthorFemale1 0.03611 0.06526 0.55 0.5805
## LastAuthorFemale1 -0.09406 0.06812 -1.38 0.1684
## Year1997 -0.21756 0.24589 -0.88 0.3770
## Year1998 0.54287 0.17403 3.12 0.0020 **
## Year1999 -0.09384 0.24002 -0.39 0.6961
## Year2000 0.41062 0.15557 2.64 0.0088 **
## Year2001 0.03916 0.49382 0.08 0.9368
## Year2002 0.20293 0.18086 1.12 0.2628
## Year2003 0.14582 0.27109 0.54 0.5911
## Year2004 -0.01350 0.26755 -0.05 0.9598
## Year2005 0.12229 0.20154 0.61 0.5445
```

```

## Year2006          0.04515      0.19805      0.23      0.8198
## Year2007          0.00341      0.17094      0.02      0.9841
## Year2008          0.02857      0.17878      0.16      0.8732
## Year2009          0.41691      0.18427      2.26      0.0244 *
## Year2010          0.31453      0.16530      1.90      0.0581 .
## Year2011          0.49824      0.18136      2.75      0.0064 **
## Year2012          0.53269      0.18420      2.89      0.0041 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.513
## Multiple R-squared:  0.147, Adjusted R-squared:  0.0938
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 284 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.326  0.889   0.947   0.906   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      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 3.007 1      1.734
## Year              3.007 16      1.035

```

Residuals from first author



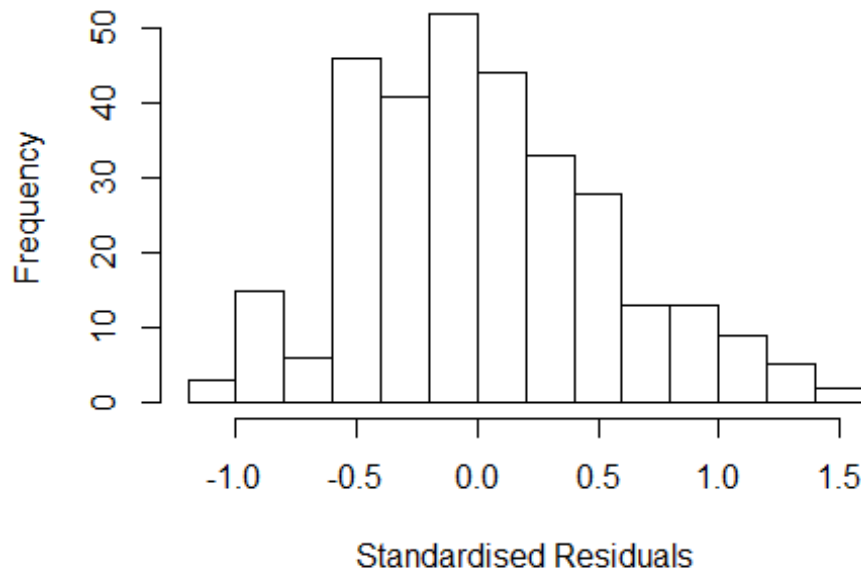
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.9868 -0.3873 -0.0457 0.3617 1.6393
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4357 0.1535 2.84 0.0048 **
## FirstAuthorFemale1 0.0164 0.0634 0.26 0.7967
## Year1997 -0.1708 0.2509 -0.68 0.4966
## Year1998 0.5015 0.1713 2.93 0.0037 **
## Year1999 -0.0973 0.2413 -0.40 0.6870
## Year2000 0.3890 0.1553 2.50 0.0128 *
## Year2001 0.0918 0.4917 0.19 0.8520
## Year2002 0.2071 0.1832 1.13 0.2593
## Year2003 0.1507 0.2696 0.56 0.5766
## Year2004 0.0452 0.2535 0.18 0.8587
## Year2005 0.1369 0.2099 0.65 0.5147
## Year2006 0.0493 0.2004 0.25 0.8058
```

```

## Year2007          0.0261      0.1719      0.15      0.8793
## Year2008          0.0288      0.1782      0.16      0.8717
## Year2009          0.4290      0.1833      2.34      0.0199 *
## Year2010          0.3257      0.1650      1.97      0.0493 *
## Year2011          0.5057      0.1821      2.78      0.0058 **
## Year2012          0.5348      0.1841      2.90      0.0040 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.14,   Adjusted R-squared:  0.0903
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 283 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.290  0.891  0.944  0.905  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.23e-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      2  1      1.414
## Year      2 16      1.022

```

Residuals from last author



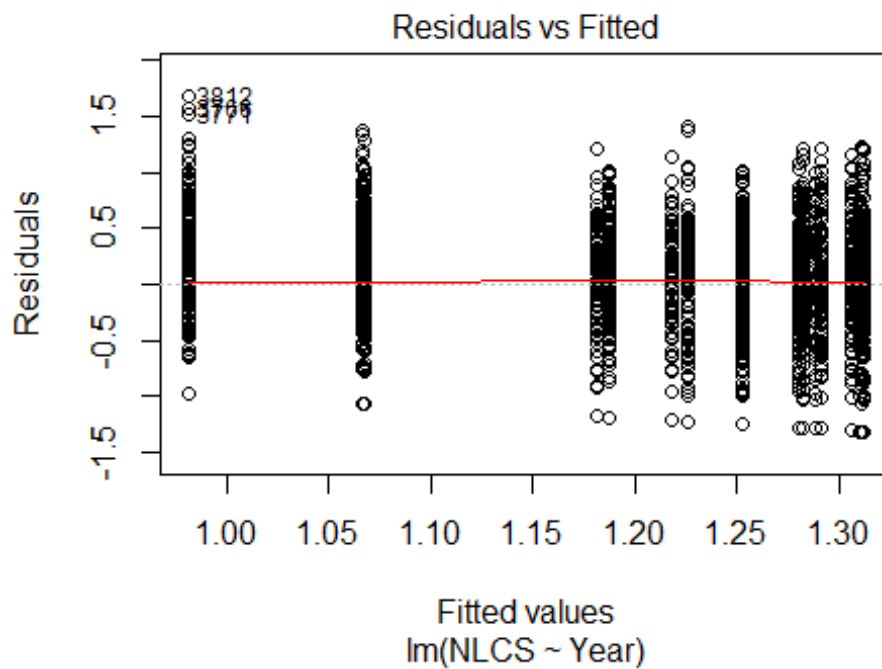
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0361 -0.3607 -0.0472  0.3594  1.5980
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.49934    0.15566   3.21  0.0015 **
## LastAuthorFemale1 -0.08616    0.06600  -1.31  0.1927
## Year1997        -0.20267    0.24672  -0.82  0.4121
## Year1998         0.52396    0.16929   3.10  0.0022 **
## Year1999        -0.09792    0.24231  -0.40  0.6864
## Year2000         0.42782    0.15121   2.83  0.0050 **
## Year2001         0.02816    0.49420   0.06  0.9546
## Year2002         0.21018    0.17803   1.18  0.2387
## Year2003         0.14706    0.26758   0.55  0.5830
## Year2004        -0.00873    0.26440  -0.03  0.9737
## Year2005         0.12859    0.20180   0.64  0.5245
## Year2006         0.05124    0.19697   0.26  0.7949
```

```

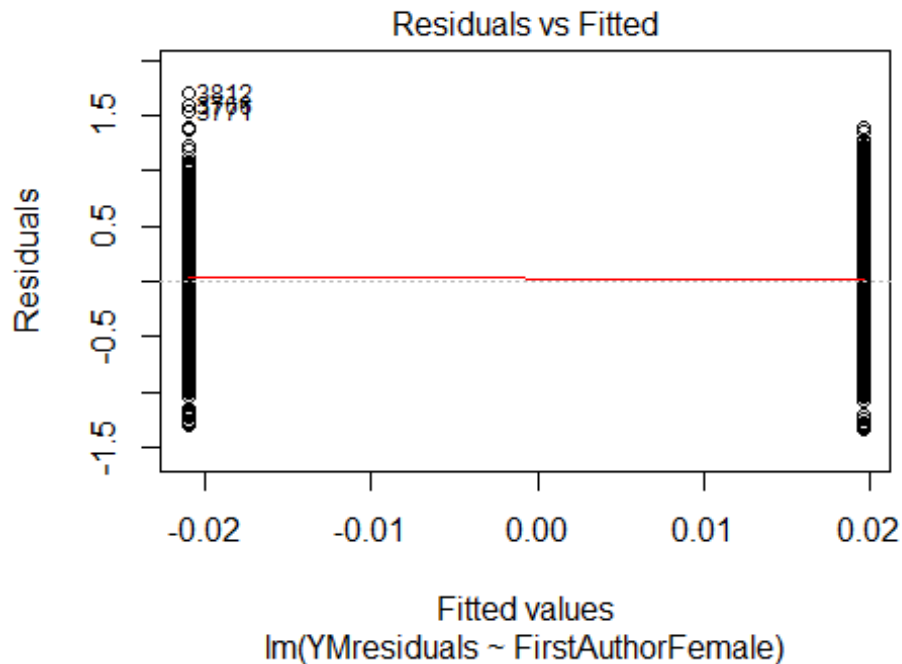
## Year2007      0.00700      0.16971      0.04      0.9671
## Year2008      0.03302      0.17764      0.19      0.8526
## Year2009      0.42195      0.18292      2.31      0.0218 *
## Year2010      0.32074      0.16391      1.96      0.0513 .
## Year2011      0.49967      0.18091      2.76      0.0061 **
## Year2012      0.53671      0.18307      2.93      0.0036 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.511
## Multiple R-squared:  0.145, Adjusted R-squared:  0.0957
## 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.
##   0.307  0.890  0.945  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      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 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
## 127 125 127 130 163 203 169 138 161 165 184 211 193 271 306
## 2011 2012
## 292 324
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 87 85 81 92 111 103 144 107 120 126 141 170 158 227 253
## 2011 2012

```

```
## 238 276
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 79 82 75 84 98 95 131 99 106 114 137 153 140 203 227
## 2011 2012
## 216 251
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 53, df = 16, p-value = 9e-06
```

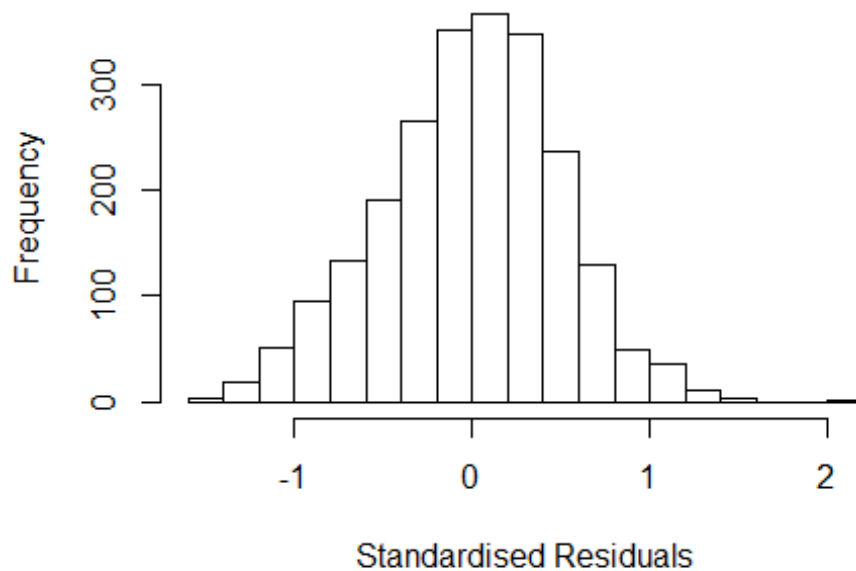


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0031, df = 1, p-value = 1
```



```
## [1] "Female first author team size 2018 geometric mean: 3.49040683676945"
## [1] "Male first author team size 2018 geometric mean: 3.68537068123838"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9900, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.38162524327757"
## [1] "Male last author team size 2018 geometric mean: 3.7465573615057"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.115 1          1.056
## LastAuthorFemale  1.130 1          1.063
## UniqueAuthors     1.226 4          1.026
## Year              1.280 16         1.008
```


Residuals from first and last author and team size



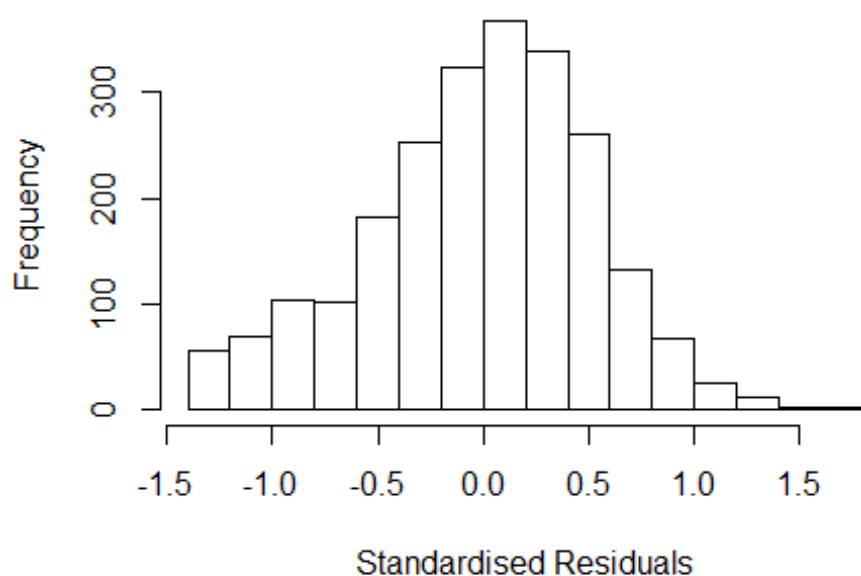
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.486 -0.329 0.019 0.344 2.064
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.98687 0.07048 14.00 < 2e-16 ***
## FirstAuthorFemale1 0.03396 0.02267 1.50 0.1343
## LastAuthorFemale1 0.00487 0.02366 0.21 0.8370
## UniqueAuthors2 0.29841 0.04609 6.47 1.2e-10 ***
## UniqueAuthors3 0.40155 0.04506 8.91 < 2e-16 ***
## UniqueAuthors4 0.45710 0.04539 10.07 < 2e-16 ***
## UniqueAuthors5 0.55548 0.04382 12.68 < 2e-16 ***
## Year1997 -0.01008 0.08767 -0.11 0.9085
## Year1998 0.00312 0.08777 0.04 0.9716
## Year1999 -0.15169 0.07931 -1.91 0.0559 .
```

```

## Year2000      -0.14319    0.08695   -1.65    0.0997 .
## Year2001      -0.06363    0.08527   -0.75    0.4556
## Year2002      -0.13604    0.08074   -1.69    0.0921 .
## Year2003      -0.08110    0.07749   -1.05    0.2954
## Year2004      -0.09176    0.08249   -1.11    0.2661
## Year2005      -0.09720    0.07684   -1.26    0.2060
## Year2006      -0.13109    0.07432   -1.76    0.0779 .
## Year2007      -0.13351    0.07350   -1.82    0.0694 .
## Year2008      -0.16826    0.07414   -2.27    0.0233 *
## Year2009      -0.31428    0.07310   -4.30    1.8e-05 ***
## Year2010      -0.20715    0.07113   -2.91    0.0036 **
## Year2011      -0.31619    0.07584   -4.17    3.2e-05 ***
## Year2012      -0.39373    0.07693   -5.12    3.4e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.489
## Multiple R-squared:  0.144, Adjusted R-squared:  0.136
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 2105 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0357 0.8660 0.9480 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.37e-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 1.102 1 1.050
## LastAuthorFemale 1.098 1 1.048
## Year 1.085 16 1.003

```

Residuals from first and last author



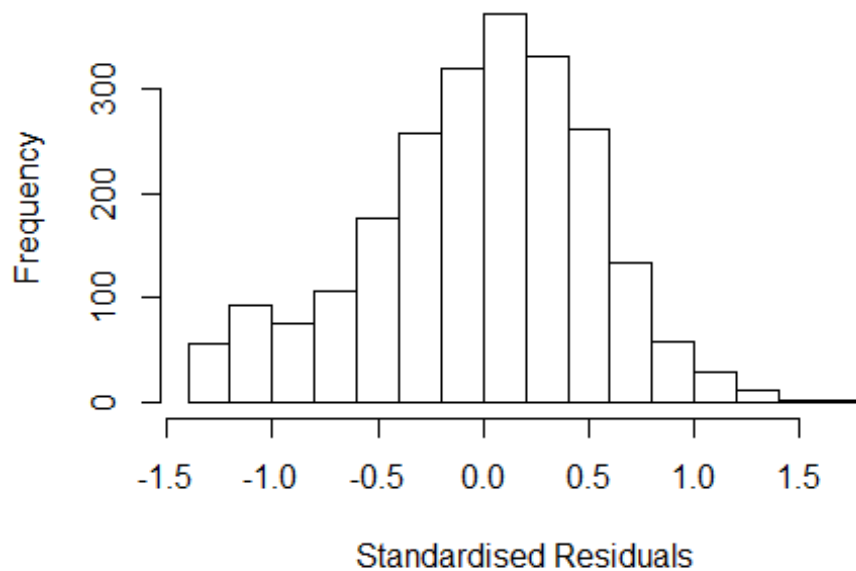
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.346 -0.345  0.033  0.353  1.691
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.26053    0.06860   18.38 < 2e-16 ***
## FirstAuthorFemale1 0.05444    0.02363    2.30 0.02135 *
## LastAuthorFemale1 -0.02394    0.02477   -0.97 0.33389
## Year1997         0.02996    0.09152    0.33 0.74347
## Year1998         0.03802    0.09305    0.41 0.68284
## Year1999        -0.02451    0.08675   -0.28 0.77754
## Year2000        -0.08633    0.08824   -0.98 0.32801
## Year2001         0.05532    0.09179    0.60 0.54679
## Year2002        -0.02174    0.08633   -0.25 0.80118
## Year2003         0.04569    0.08214    0.56 0.57816
## Year2004         0.03202    0.08614    0.37 0.71013
## Year2005         0.02581    0.08045    0.32 0.74835
```

```

## Year2006          -0.00204      0.07996    -0.03  0.97967
## Year2007          -0.01815      0.07800    -0.23  0.81599
## Year2008          -0.03751      0.07845    -0.48  0.63263
## Year2009          -0.20403      0.07878    -2.59  0.00966 **
## Year2010          -0.07698      0.07596    -1.01  0.31102
## Year2011          -0.21654      0.08159    -2.65  0.00801 **
## Year2012          -0.29412      0.08241    -3.57  0.00037 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.0458, Adjusted R-squared:  0.0383
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 177 weights are ~= 1. The remaining 2113 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.232  0.865   0.947   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      4.37e-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 1.045 1      1.022
## Year              1.045 16      1.001

```

Residuals from first author



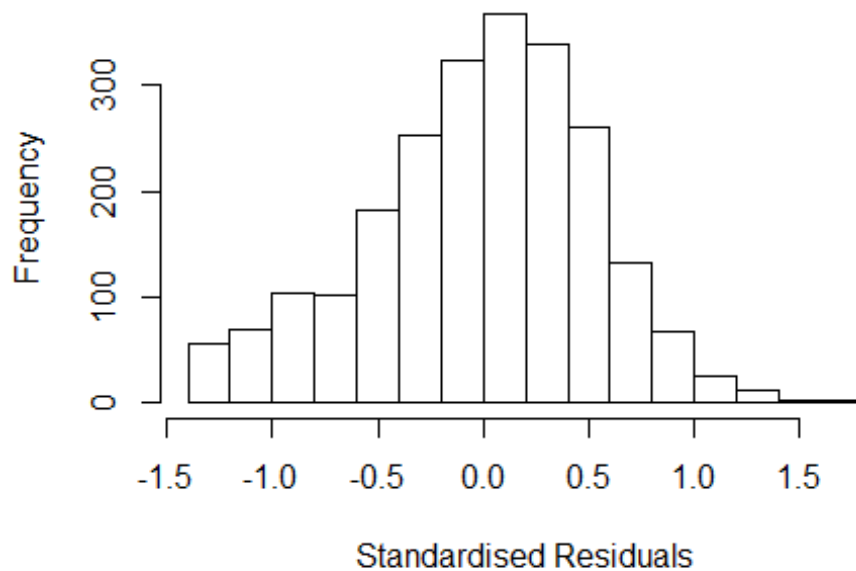
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3579 -0.3469 0.0347 0.3526 1.6974
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25725 0.06881 18.27 < 2e-16 ***
## FirstAuthorFemale1 0.04728 0.02309 2.05 0.04073 *
## Year1997 0.02721 0.09137 0.30 0.76591
## Year1998 0.03724 0.09291 0.40 0.68856
## Year1999 -0.02650 0.08685 -0.31 0.76033
## Year2000 -0.08755 0.08824 -0.99 0.32121
## Year2001 0.05339 0.09163 0.58 0.56021
## Year2002 -0.02393 0.08642 -0.28 0.78183
## Year2003 0.04507 0.08220 0.55 0.58352
## Year2004 0.03280 0.08635 0.38 0.70412
## Year2005 0.02385 0.08048 0.30 0.76703
## Year2006 -0.00338 0.08003 -0.04 0.96627
```

```

## Year2007          -0.02126    0.07798   -0.27  0.78520
## Year2008          -0.04073    0.07852   -0.52  0.60403
## Year2009          -0.20820    0.07873   -2.64  0.00824 **
## Year2010          -0.08028    0.07595   -1.06  0.29058
## Year2011          -0.21981    0.08163   -2.69  0.00714 **
## Year2012          -0.29763    0.08250   -3.61  0.00032 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.0454, Adjusted R-squared:  0.0383
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 178 weights are ~= 1. The remaining 2112 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.228  0.866  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      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 1.044 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
## -1.3355 -0.3427 0.0329 0.3518 1.6685
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27730 0.06813 18.75 < 2e-16 ***
## LastAuthorFemale1 -0.00689 0.02418 -0.29 0.77566
## Year1997 0.03230 0.09150 0.35 0.72414
## Year1998 0.04372 0.09287 0.47 0.63782
## Year1999 -0.02464 0.08662 -0.28 0.77608
## Year2000 -0.08513 0.08819 -0.97 0.33447
## Year2001 0.05819 0.09181 0.63 0.52624
## Year2002 -0.02321 0.08616 -0.27 0.78762
## Year2003 0.05237 0.08216 0.64 0.52392
## Year2004 0.03968 0.08575 0.46 0.64358
## Year2005 0.02934 0.08043 0.36 0.71534
## Year2006 0.00671 0.07961 0.08 0.93287
```

```

## Year2007          -0.01431      0.07806   -0.18  0.85452
## Year2008          -0.02936      0.07832   -0.37  0.70785
## Year2009          -0.19804      0.07856   -2.52  0.01177 *
## Year2010          -0.06956      0.07583   -0.92  0.35906
## Year2011          -0.21141      0.08158   -2.59  0.00962 **
## Year2012          -0.28882      0.08233   -3.51  0.00046 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.0438, Adjusted R-squared:  0.0366
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 182 weights are ~= 1. The remaining 2108 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.245  0.862  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      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: 2290"
## [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 2003 2005 2008 2009 2012
##    2    2    1    1    1    1    1    3    2    2
##
## 1996 1997 1998 1999 2000 2003 2005 2008 2009 2012
##    1    2    0    0    1    1    1    3    1    1
##
## 1996 1997 1998 1999 2000 2003 2005 2008 2009 2012
##    1    2    0    0    1    1    1    3    1    1

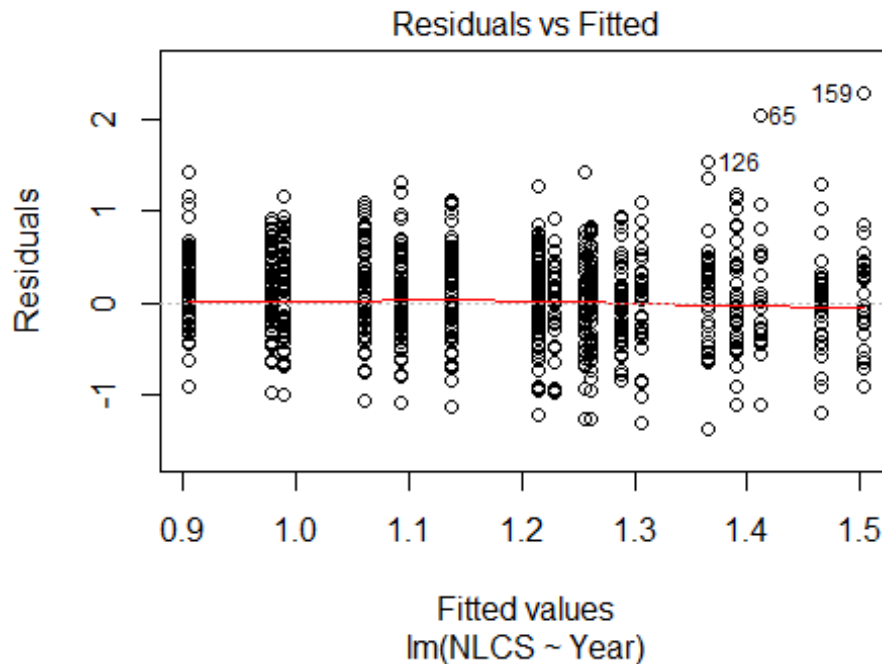
```



```

## [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: 6"
## [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: 11"
## [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
##    57   78   76   76   67   74   72   81   95  112  124  157  159  148  144
## 2011 2012
##   141  152
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    23   34   34   32   35   30   38   59   49   62   83   92  110  101   85
## 2011 2012
##    79   94
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    18   29   24   27   29   23   31   44   43   47   72   80   97   83   68
## 2011 2012
##    68   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 = 28, df = 16, p-value = 0.03

```



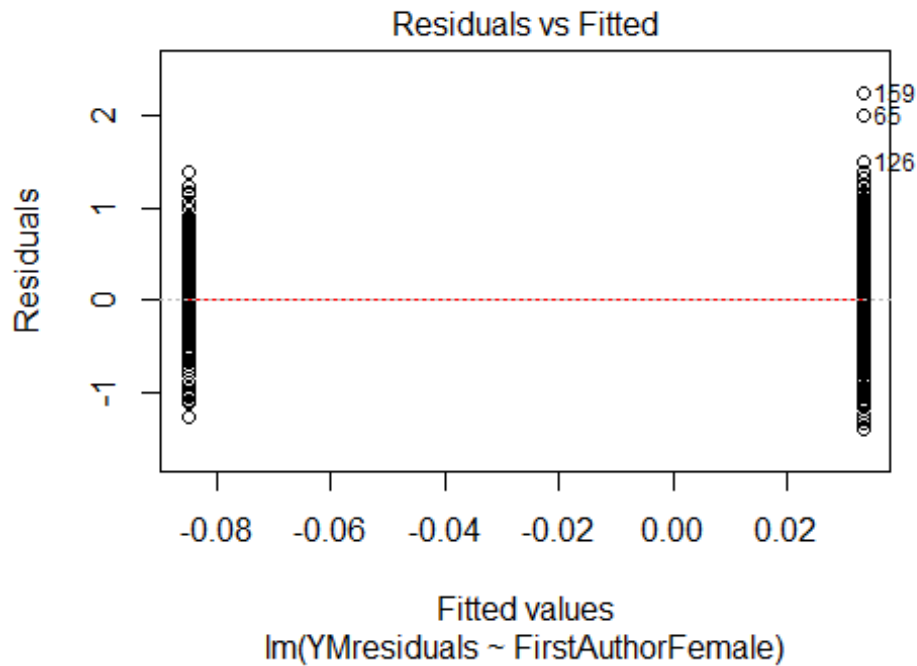
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0096, df = 1, p-value = 0.9

## [1] "Female first author team size 2018 geometric mean: 4.74227003185324"
## [1] "Male first author team size 2018 geometric mean: 4.17302015498945"

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

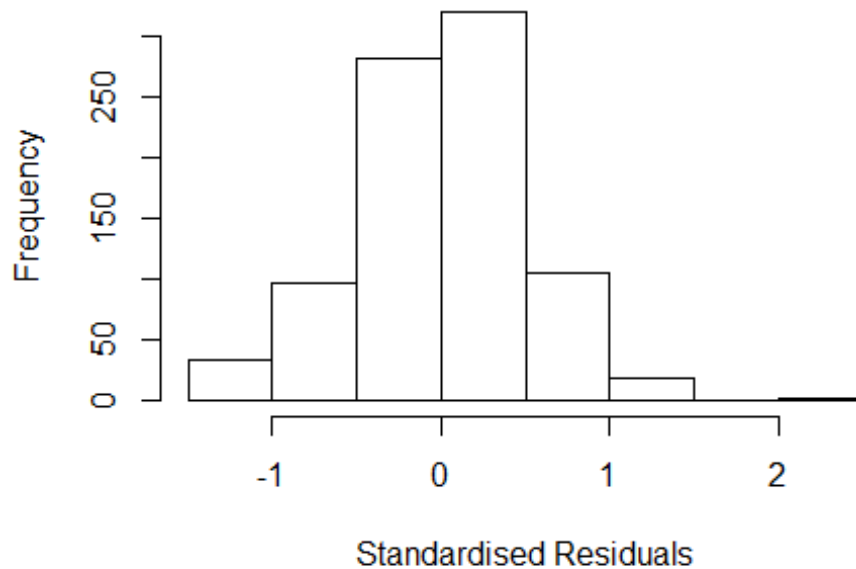
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 860, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.60318271229449"
## [1] "Male last author team size 2018 geometric mean: 4.29601530067501"

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



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 780, 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 1.169 1          1.081
## LastAuthorFemale  1.241 1          1.114
## UniqueAuthors    1.409 4          1.044
## Year             1.568 16          1.014
```

Residuals from first and last author and team size



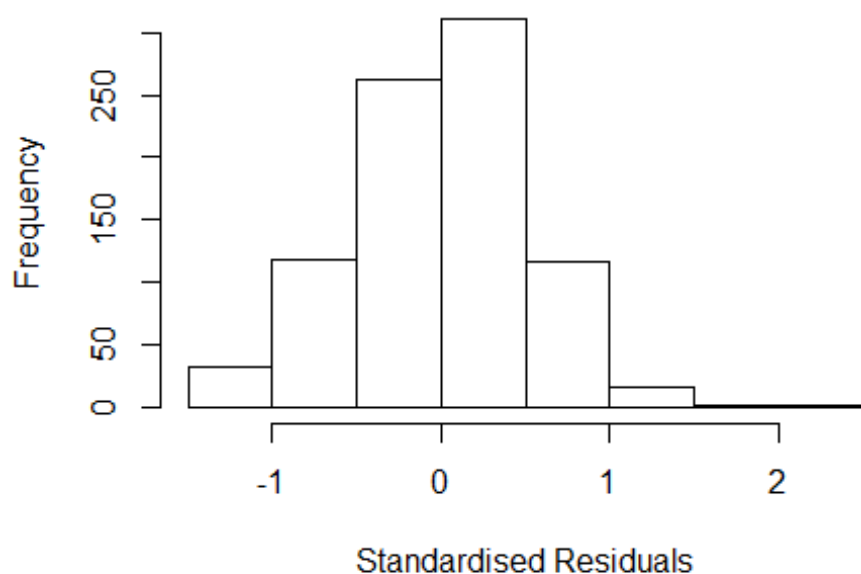
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4533 -0.3337 0.0271 0.3395 2.3485
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0746 0.1662 6.47 1.7e-10 ***
## FirstAuthorFemale1 -0.0998 0.0401 -2.49 0.0130 *
## LastAuthorFemale1 -0.1075 0.0481 -2.23 0.0257 *
## UniqueAuthors2 0.2554 0.0751 3.40 0.0007 ***
## UniqueAuthors3 0.3478 0.0674 5.16 3.1e-07 ***
## UniqueAuthors4 0.3678 0.0730 5.04 5.7e-07 ***
## UniqueAuthors5 0.4784 0.0640 7.47 2.0e-13 ***
## Year1997 -0.0997 0.1852 -0.54 0.5907
## Year1998 0.0120 0.1766 0.07 0.9457
## Year1999 -0.0138 0.1865 -0.07 0.9411
```

```

## Year2000          -0.1076      0.1826   -0.59   0.5556
## Year2001          -0.0289      0.1826   -0.16   0.8745
## Year2002          -0.1745      0.1777   -0.98   0.3263
## Year2003          -0.1054      0.1681   -0.63   0.5310
## Year2004          -0.0830      0.1650   -0.50   0.6151
## Year2005          -0.1998      0.1652   -1.21   0.2269
## Year2006          -0.2309      0.1662   -1.39   0.1650
## Year2007          -0.1578      0.1608   -0.98   0.3267
## Year2008          -0.4594      0.1595   -2.88   0.0041 **
## Year2009          -0.2226      0.1639   -1.36   0.1746
## Year2010          -0.4203      0.1628   -2.58   0.0100 **
## Year2011          -0.2731      0.1669   -1.64   0.1021
## Year2012          -0.4006      0.1609   -2.49   0.0130 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.49
## Multiple R-squared:  0.167, Adjusted R-squared:  0.145
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 51 is an outlier with |weight| = 0 ( < 0.00012);
## 69 weights are ~= 1. The remaining 788 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0215 0.8780 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          1.17e-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 1.105 1          1.051
## LastAuthorFemale 1.196 1          1.094
## Year 1.188 16          1.005

```

Residuals from first and last author



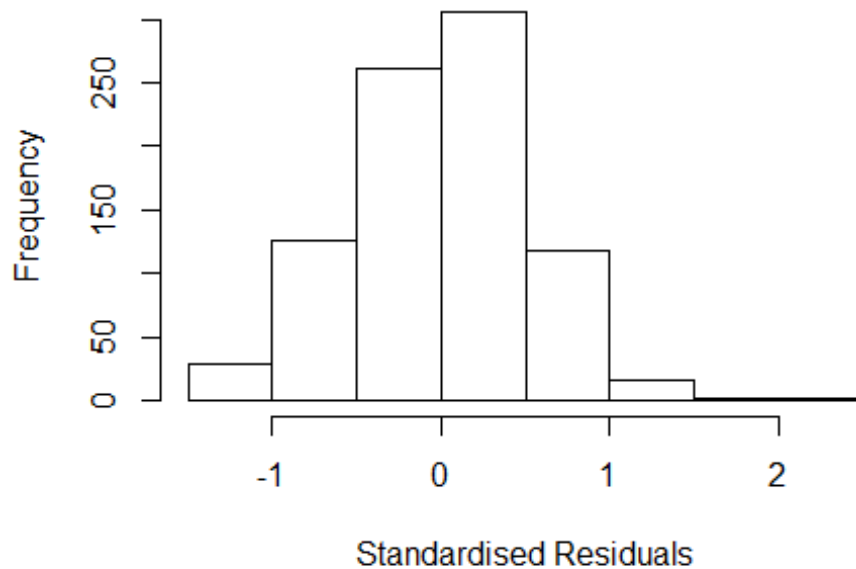
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.398 -0.337 0.026 0.363 2.336
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4367 0.1515 9.48 <2e-16 ***
## FirstAuthorFemale1 -0.0774 0.0408 -1.90 0.0582 .
## LastAuthorFemale1 -0.1466 0.0498 -2.94 0.0033 **
## Year1997 -0.1073 0.1828 -0.59 0.5572
## Year1998 0.0102 0.1787 0.06 0.9547
## Year1999 -0.0383 0.1875 -0.20 0.8383
## Year2000 -0.1238 0.1824 -0.68 0.4974
## Year2001 -0.0569 0.1814 -0.31 0.7537
## Year2002 -0.1778 0.1726 -1.03 0.3034
## Year2003 -0.1634 0.1652 -0.99 0.3231
## Year2004 -0.0893 0.1636 -0.55 0.5853
## Year2005 -0.1871 0.1645 -1.14 0.2556
```

```

## Year2006          -0.2571      0.1667    -1.54    0.1234
## Year2007          -0.1660      0.1607    -1.03    0.3017
## Year2008          -0.5007      0.1600    -3.13    0.0018 **
## Year2009          -0.2555      0.1656    -1.54    0.1233
## Year2010          -0.4472      0.1634    -2.74    0.0063 **
## Year2011          -0.3105      0.1697    -1.83    0.0676 .
## Year2012          -0.4102      0.1593    -2.57    0.0102 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.511
## Multiple R-squared:  0.105, Adjusted R-squared:  0.0858
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 787 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0023 0.8740 0.9480 0.9040 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.17e-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 1.043 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
## -1.3735 -0.3406 0.0277 0.3764 2.3532
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3990 0.1498 9.34 <2e-16 ***
## FirstAuthorFemale1 -0.1104 0.0405 -2.72 0.0066 **
## Year1997 -0.0994 0.1839 -0.54 0.5891
## Year1998 0.0308 0.1815 0.17 0.8651
## Year1999 -0.0255 0.1868 -0.14 0.8914
## Year2000 -0.0897 0.1832 -0.49 0.6246
## Year2001 -0.0255 0.1778 -0.14 0.8858
## Year2002 -0.1700 0.1743 -0.98 0.3296
## Year2003 -0.1382 0.1646 -0.84 0.4012
## Year2004 -0.0606 0.1624 -0.37 0.7090
## Year2005 -0.1585 0.1632 -0.97 0.3320
## Year2006 -0.2324 0.1658 -1.40 0.1615
```

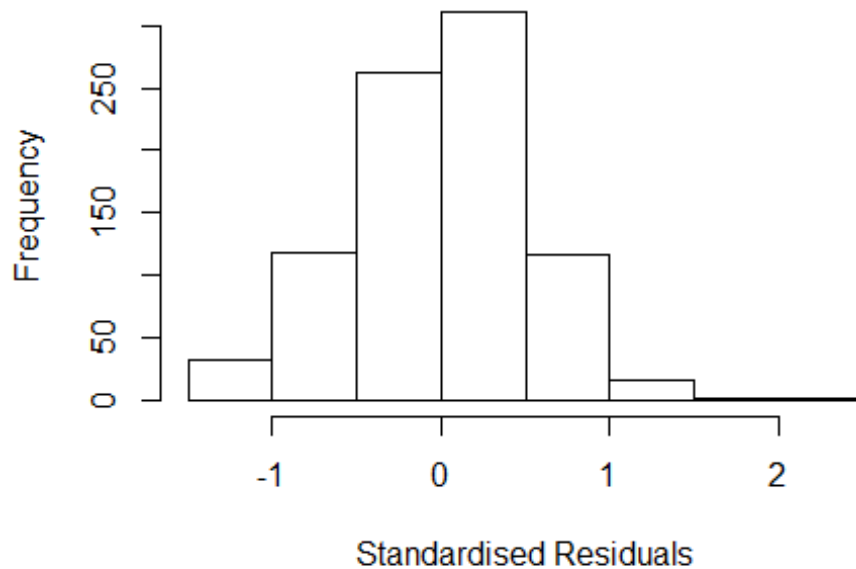


```

## Year2007          -0.1417      0.1594   -0.89    0.3743
## Year2008          -0.4880      0.1601   -3.05    0.0024 **
## Year2009          -0.2368      0.1652   -1.43    0.1522
## Year2010          -0.4242      0.1631   -2.60    0.0094 **
## Year2011          -0.2994      0.1696   -1.76    0.0780 .
## Year2012          -0.4069      0.1600   -2.54    0.0112 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.513
## Multiple R-squared:  0.0929, Adjusted R-squared:  0.0746
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 783 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0018 0.8740 0.9460 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      1.17e-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 1.128 1      1.062
## Year      1.128 16      1.004

```

Residuals from last author



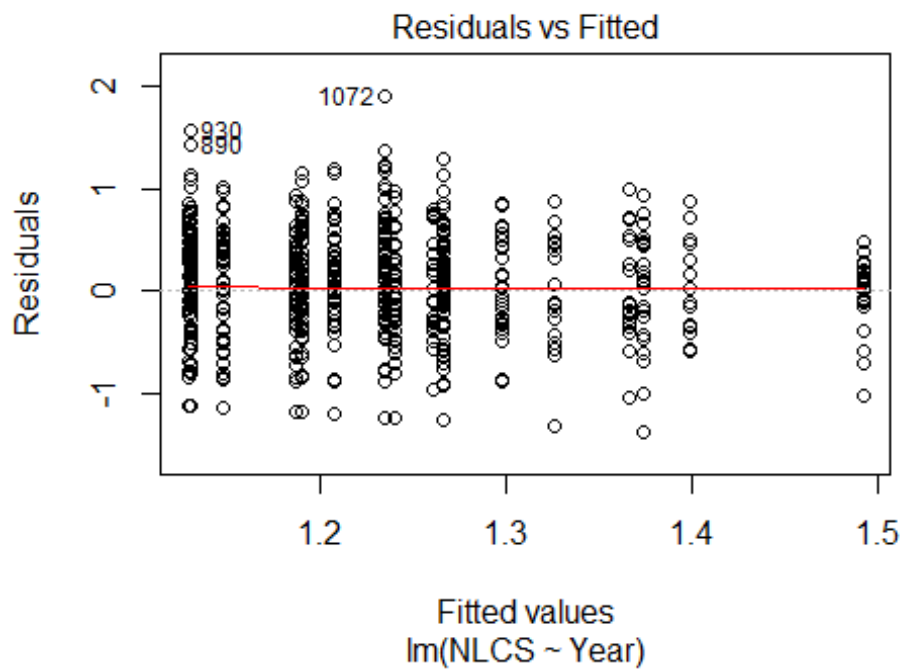
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3924 -0.3336 0.0283 0.3589 2.3391
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4218 0.1517 9.37 < 2e-16 ***
## LastAuthorFemale1 -0.1700 0.0490 -3.47 0.00054 ***
## Year1997 -0.1035 0.1837 -0.56 0.57348
## Year1998 0.0221 0.1771 0.12 0.90092
## Year1999 -0.0294 0.1881 -0.16 0.87576
## Year2000 -0.1258 0.1817 -0.69 0.48892
## Year2001 -0.0551 0.1824 -0.30 0.76257
## Year2002 -0.1716 0.1720 -1.00 0.31878
## Year2003 -0.1685 0.1649 -1.02 0.30727
## Year2004 -0.0949 0.1641 -0.58 0.56302
## Year2005 -0.1949 0.1639 -1.19 0.23470
## Year2006 -0.2603 0.1667 -1.56 0.11894
```

```

## Year2007          -0.1762      0.1601    -1.10   0.27156
## Year2008          -0.5092      0.1602    -3.18   0.00153 **
## Year2009          -0.2563      0.1660    -1.54   0.12291
## Year2010          -0.4548      0.1636    -2.78   0.00556 **
## Year2011          -0.3209      0.1698    -1.89   0.05912 .
## Year2012          -0.4137      0.1594    -2.60   0.00959 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.511
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0829
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 792 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.002  0.871   0.949   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      1.17e-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: 858"
## [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
##   22  38  31  33  33  33  52  35  34  40  63  61  67  64  89
## 2011 2012
##   95  87
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18  29  25  22  18  16  44  29  33  28  51  58  57  55  76
## 2011 2012

```

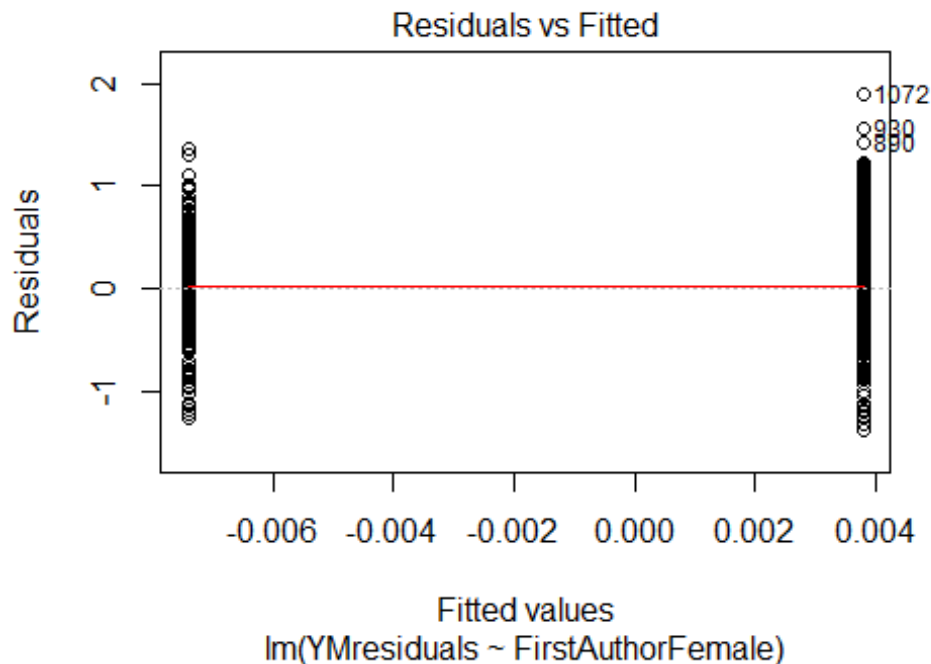
```
##      84      77
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    17    27    23    19    17    16    42    26    32    25    44    53    51    49    69
## 2011 2012
##    80    64
## [1] "Heteroscedasticity checks, confirming that there are problems with
##      these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 31, df = 16, p-value = 0.01
```



```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.042, df = 1, p-value = 0.8
##
## [1] "Female first author team size 2018 geometric mean: 3.21916225606056"
## [1] "Male first author team size 2018 geometric mean: 2.87563058645459"
##
## Warning in wilcox.test.default(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 = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.16603486959347"
## [1] "Male last author team size 2018 geometric mean: 3.06064971804522"

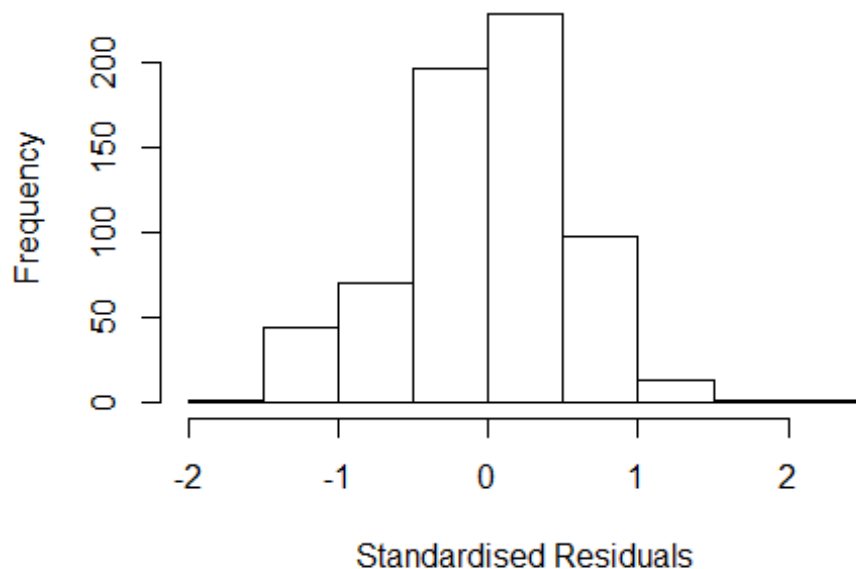
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 410, 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	1.245	1	1.116
LastAuthorFemale	1.316	1	1.147
UniqueAuthors	1.899	4	1.083
Year	2.303	16	1.026

Residuals from first and last author and team size



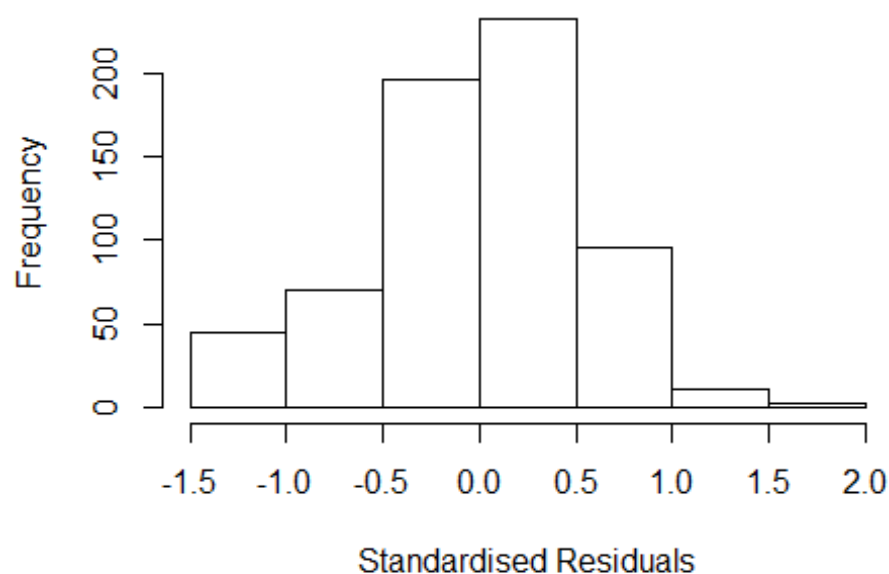
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5393 -0.3546 0.0168 0.3571 2.0726
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25933 0.14589 8.63 <2e-16 ***
## FirstAuthorFemale1 -0.02727 0.05085 -0.54 0.592
## LastAuthorFemale1 0.04464 0.04805 0.93 0.353
## UniqueAuthors2 0.16700 0.07057 2.37 0.018 *
## UniqueAuthors3 0.12073 0.07465 1.62 0.106
## UniqueAuthors4 0.14435 0.08937 1.62 0.107
## UniqueAuthors5 0.14307 0.09399 1.52 0.128
## Year1997 0.09563 0.16995 0.56 0.574
## Year1998 -0.15689 0.17907 -0.88 0.381
## Year1999 -0.18766 0.17599 -1.07 0.287
```

```

## Year2000      -0.11022    0.19272   -0.57    0.568
## Year2001      -0.00395    0.18694   -0.02    0.983
## Year2002      -0.14273    0.15616   -0.91    0.361
## Year2003      -0.09271    0.16686   -0.56    0.579
## Year2004      -0.18508    0.20403   -0.91    0.365
## Year2005       0.11992    0.15030    0.80    0.425
## Year2006      -0.18393    0.17087   -1.08    0.282
## Year2007      -0.08794    0.16140   -0.54    0.586
## Year2008      -0.19676    0.15346   -1.28    0.200
## Year2009      -0.17659    0.15630   -1.13    0.259
## Year2010      -0.20132    0.15779   -1.28    0.202
## Year2011      -0.25708    0.15569   -1.65    0.099 .
## Year2012      -0.21433    0.16255   -1.32    0.188
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.512
## Multiple R-squared:  0.0437, Adjusted R-squared:  0.0103
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 592 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0639 0.8540 0.9440 0.8930 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.53e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.178 1          1.086
## LastAuthorFemale 1.174 1          1.083
## Year      1.334 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
## -1.4775 -0.3600 0.0289 0.3483 1.9319
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.35560 0.14185 9.56 <2e-16 ***
## FirstAuthorFemale1 -0.01351 0.05110 -0.26 0.79
## LastAuthorFemale1 0.04582 0.04627 0.99 0.32
## Year1997 0.08957 0.17648 0.51 0.61
## Year1998 -0.15276 0.18059 -0.85 0.40
## Year1999 -0.17435 0.17353 -1.00 0.32
## Year2000 -0.06459 0.18953 -0.34 0.73
## Year2001 0.00722 0.18402 0.04 0.97
## Year2002 -0.13448 0.15765 -0.85 0.39
## Year2003 -0.07215 0.17022 -0.42 0.67
## Year2004 -0.17012 0.20904 -0.81 0.42
## Year2005 0.14211 0.15184 0.94 0.35
```

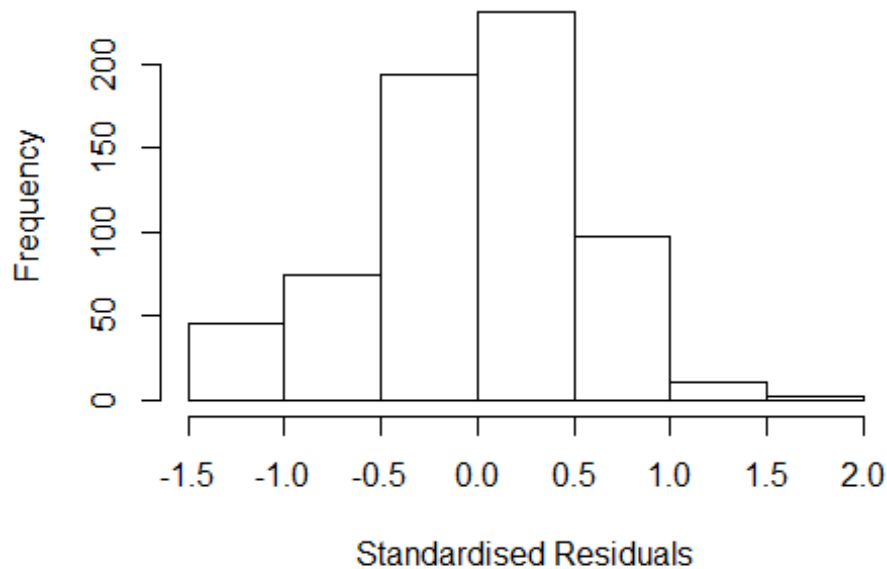


```

## Year2006      -0.17822    0.17414   -1.02    0.31
## Year2007      -0.08692    0.16345   -0.53    0.60
## Year2008      -0.18367    0.15398   -1.19    0.23
## Year2009      -0.16469    0.15662   -1.05    0.29
## Year2010      -0.18183    0.15916   -1.14    0.25
## Year2011      -0.23992    0.15806   -1.52    0.13
## Year2012      -0.18479    0.16481   -1.12    0.26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.509
## Multiple R-squared:  0.032, Adjusted R-squared:  0.00457
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 598 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.118  0.856  0.949  0.892  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.53e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.175 1      1.084
## Year              1.175 16      1.005

```

Residuals from first author



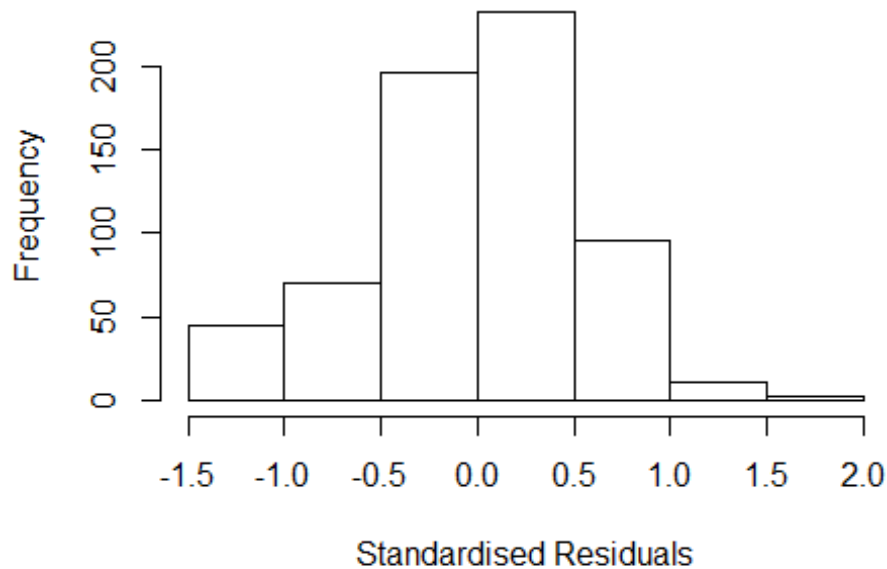
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4607 -0.3519 0.0424 0.3495 1.9470
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37939 0.13800 10.00 <2e-16 ***
## FirstAuthorFemale1 -0.00170 0.05154 -0.03 0.97
## Year1997 0.08305 0.17619 0.47 0.64
## Year1998 -0.15614 0.18092 -0.86 0.39
## Year1999 -0.18339 0.17171 -1.07 0.29
## Year2000 -0.07072 0.19002 -0.37 0.71
## Year2001 0.00233 0.18210 0.01 0.99
## Year2002 -0.14517 0.15578 -0.93 0.35
## Year2003 -0.07996 0.16903 -0.47 0.64
## Year2004 -0.18492 0.20611 -0.90 0.37
## Year2005 0.13795 0.15037 0.92 0.36
## Year2006 -0.18905 0.17248 -1.10 0.27
```

```

## Year2007      -0.08824    0.16233   -0.54    0.59
## Year2008      -0.19504    0.15169   -1.29    0.20
## Year2009      -0.16912    0.15497   -1.09    0.28
## Year2010      -0.18394    0.15796   -1.16    0.24
## Year2011      -0.24744    0.15663   -1.58    0.11
## Year2012      -0.18968    0.16380   -1.16    0.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.509
## Multiple R-squared:  0.0304, Adjusted R-squared:  0.00447
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 608 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.111  0.856  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      1.53e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.162 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
## -1.4830 -0.3566 0.0276 0.3490 1.9276
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.35214 0.14095 9.59 <2e-16 ***
## LastAuthorFemale1 0.04283 0.04658 0.92 0.36
## Year1997 0.08806 0.17622 0.50 0.62
## Year1998 -0.15781 0.17801 -0.89 0.38
## Year1999 -0.17600 0.17310 -1.02 0.31
## Year2000 -0.06460 0.18941 -0.34 0.73
## Year2001 0.00318 0.18262 0.02 0.99
## Year2002 -0.13858 0.15654 -0.89 0.38
## Year2003 -0.07570 0.16933 -0.45 0.65
## Year2004 -0.17351 0.20862 -0.83 0.41
## Year2005 0.13796 0.15064 0.92 0.36
## Year2006 -0.18016 0.17412 -1.03 0.30
```

```

## Year2007          -0.09227      0.16138    -0.57      0.57
## Year2008          -0.18701      0.15324    -1.22      0.22
## Year2009          -0.16935      0.15583    -1.09      0.28
## Year2010          -0.18692      0.15732    -1.19      0.24
## Year2011          -0.24602      0.15493    -1.59      0.11
## Year2012          -0.18756      0.16423    -1.14      0.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.508
## Multiple R-squared:  0.0319, Adjusted R-squared:  0.00606
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 599 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.119  0.856  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      1.53e-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: 654"

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	2990		59.26	20.588	36.923	16.3348	11.765	24.231	12.4661	
## 1100	9002		63.72	21.368	43.619	22.2520	9.402	30.394	20.9927	
## 1101	1661		42.26	30.000	48.000	18.0000	18.000	36.000	18.0000	
## 1102	9022		34.46	16.038	31.544	15.5059	19.811	26.174	6.3632	
## 1103	12470		45.69	21.860	55.484	33.6234	17.674	30.000	12.3256	
## 1104	11565		55.60	23.024	48.571	25.5474	15.464	26.984	11.5202	
## 1105	22518		62.85	23.647	44.578	20.9310	16.433	24.632	8.1990	
## 1106	8523		35.57	37.500	59.091	21.5909	18.750	44.444	25.6944	
## 1107	6558		50.84	4.494	31.875	27.3806	6.742	24.375	17.6334	
## 1108	3429		29.48	23.810	16.000	-7.8095	23.810	20.000	-3.8095	
## 1109	4316		51.76	28.713	47.170	18.4569	17.822	29.245	11.4235	
## 1110	11408		43.15	28.704	38.122	9.4178	15.278	23.204	7.9266	
## 1111	4595		32.97	15.789	32.927	17.1374	5.263	23.171	17.9076	
## 1200	742		83.83	75.000	48.148	-26.8519	50.000	51.852	1.8519	
## 1201	7895		75.47	34.463	50.606	16.1428	30.508	47.576	17.0673	
## 1202	5537		84.54	21.127	44.595	23.4678	18.310	43.243	24.9334	
## 1203	4839		79.48	45.161	53.261	8.0996	41.935	56.522	14.5863	
## 1204	1158		74.70	22.222	35.938	13.7153	22.222	26.562	4.3403	
## 1205	450		79.33	33.333	10.000	-23.3333	33.333	10.000	-23.3333	
## 1206	184		78.80	0.000	63.636	63.6364	0.000	63.636	63.6364	
## 1207	2318		82.36	25.000	37.719	12.7193	20.000	45.614	25.6140	
## 1208	3950		86.84	38.298	52.201	13.9034	31.915	50.943	19.0285	
## 1209	157		85.35	50.000	66.667	16.6667	0.000	66.667	66.6667	
## 1210	639		88.26	33.333	54.286	20.9524	33.333	57.143	23.8095	
## 1211	5069		87.10	22.115	34.892	12.7767	20.192	33.813	13.6206	
## 1212	2600		87.38	22.222	28.966	6.7433	22.222	27.586	5.3640	
## 1213	2055		85.35	61.111	59.712	-1.3989	50.000	60.432	10.4317	
## 1300	10843		62.03	19.492	42.748	23.2566	11.864	29.389	17.5249	
## 1301	87		54.02	NaN	22.222	NaN	NaN	0.000	NaN	
## 1302	1175		72.68	40.909	66.667	25.7576	31.818	63.333	31.5152	
## 1303	27293		55.24	30.791	33.643	2.8513	14.101	17.169	3.0687	
## 1304	8568		51.73	25.581	42.130	16.5482	18.605	22.685	4.0805	
## 1305	7865		46.83	23.636	37.288	13.6518	16.364	15.819	-0.5444	
## 1306	7810		57.03	32.203	46.226	14.0230	25.424	30.189	4.7650	
## 1307	16470		55.68	36.927	44.554	7.6278	19.954	19.802	-0.1521	
## 1308	4872		55.64	36.792	52.525	15.7328	12.264	22.222	9.9581	
## 1309	4084		60.24	36.697	51.724	15.0269	23.853	28.448	4.5951	
## 1310	6265		56.01	34.821	59.281	24.4600	25.000	31.737	6.7365	
## 1311	15463		49.64	31.118	44.660	13.5424	22.356	25.243	2.8862	
## 1312	24103		54.64	33.982	41.788	7.8060	17.876	22.263	4.3867	
## 1313	4592		57.01	26.562	37.952	11.3893	10.938	24.699	13.7613	
## 1314	15149		55.73	29.032	47.884	18.8513	13.710	24.603	10.8935	
## 1315	2745		50.86	29.630	40.816	11.1867	5.556	26.531	20.9751	
## 1400	1972		70.18	35.897	39.286	3.3883	30.769	26.190	-4.5788	
## 1401	723		69.57	25.000	42.857	17.8571	25.000	51.429	26.4286	

## 1402	1256	70.54	14.286	29.091	14.8052	21.429	20.000	-1.4286
## 1403	3414	71.91	31.667	39.286	7.6190	25.000	33.571	8.5714
## 1404	763	59.11	21.053	21.429	0.3759	21.053	21.429	0.3759
## 1405	2486	72.77	24.051	35.870	11.8189	29.114	34.783	5.6687
## 1406	1740	69.89	10.526	43.836	33.3093	7.895	34.247	26.3518
## 1407	2220	79.86	33.824	59.756	25.9326	35.294	58.537	23.2425
## 1408	4870	65.46	21.569	31.053	9.4840	20.588	36.316	15.7276
## 1409	1445	79.17	26.667	46.552	19.8851	26.667	43.103	16.4368
## 1410	519	69.94	10.000	47.368	37.3684	10.000	47.368	37.3684
## 1500	12310	35.52	14.184	33.193	19.0089	5.674	17.647	11.9733
## 1501	439	39.41	25.000	36.364	11.3636	0.000	18.182	18.1818
## 1502	4574	46.74	17.143	32.456	15.3133	25.714	20.175	-5.5388
## 1503	6344	52.18	16.111	31.677	15.5659	7.778	19.255	11.4769
## 1504	167	61.68	0.000	75.000	75.0000	0.000	50.000	50.0000
## 1505	2528	45.89	14.815	26.829	12.0145	5.556	17.073	11.5176
## 1506	694	37.03	0.000	41.667	41.6667	50.000	0.000	-50.0000
## 1507	1477	30.74	0.000	10.606	10.6061	0.000	7.576	7.5758
## 1508	1023	38.51	23.077	25.641	2.5641	7.692	12.821	5.1282
## 1600	20979	45.25	20.222	33.740	13.5182	7.756	16.057	8.3007
## 1601	413	48.18	0.000	52.941	52.9412	0.000	29.412	29.4118
## 1602	6068	48.73	24.176	34.074	9.8982	17.582	28.148	10.5657
## 1603	3706	33.22	22.222	31.579	9.3567	11.111	28.947	17.8363
## 1604	4971	56.53	19.718	29.565	9.8469	8.451	23.478	15.0276
## 1605	13838	52.77	14.589	31.381	16.7919	8.753	17.992	9.2383
## 1606	14246	47.18	13.619	24.415	10.7960	5.837	18.729	12.8925
## 1607	4595	46.25	18.182	41.964	23.7825	9.091	35.714	26.6234
## 1700	2666	52.10	11.538	18.750	7.2115	15.385	14.286	-1.0989
## 1701	272	56.99	NaN	25.000	NaN	NaN	12.500	NaN
## 1702	3279	51.78	11.321	14.286	2.9650	11.321	14.286	2.9650
## 1703	2902	52.76	17.143	20.312	3.1696	8.571	9.375	0.8036
## 1704	1759	44.17	11.538	10.000	-1.5385	7.692	5.000	-2.6923
## 1705	4183	44.94	7.143	27.660	20.5167	7.143	8.511	1.3678
## 1706	11756	45.17	8.333	25.269	16.9355	6.481	21.505	15.0239
## 1707	1676	48.21	0.000	28.125	28.1250	7.407	15.625	8.2176
## 1708	2743	42.62	13.953	12.500	-1.4535	16.279	10.417	-5.8624
## 1709	1642	55.91	12.500	33.871	21.3710	25.000	37.097	12.0968
## 1710	3970	53.73	11.765	27.619	15.8543	13.235	23.810	10.5742
## 1711	2928	41.29	2.083	16.176	14.0931	8.333	8.824	0.4902
## 1712	7077	46.53	10.417	16.154	5.7372	10.417	9.231	-1.1859
## 1800	610	65.57	22.222	44.444	22.2222	27.778	37.037	9.2593
## 1801	20	60.00	NaN	100.000	NaN	NaN	50.000	NaN
## 1802	1227	59.25	22.727	33.333	10.6061	13.636	25.641	12.0047
## 1803	3079	51.64	12.069	30.693	18.6241	15.517	22.772	7.2550
## 1804	2687	49.46	11.111	19.608	8.4967	7.407	15.686	8.2789
## 1900	4881	48.15	13.846	23.529	9.6833	10.769	17.647	6.8778
## 1901	3401	43.99	5.797	31.481	25.6844	7.246	14.815	7.5684
## 1902	5031	40.87	7.143	37.008	29.8650	2.857	23.622	20.7649
## 1903	1065	47.79	5.263	24.390	19.1271	0.000	21.951	21.9512
## 1904	4652	57.67	10.526	38.356	27.8298	6.579	26.027	19.4484
## 1905	716	50.00	13.333	30.435	17.1014	13.333	17.391	4.0580

## 1906	6255	46.35	13.125	36.082	22.9575	10.625	17.526	6.9008
## 1907	5033	47.37	13.889	26.471	12.5817	10.185	19.118	8.9325
## 1908	4654	45.29	16.162	32.468	16.3059	9.091	22.078	12.9870
## 1909	4100	34.32	4.255	23.967	19.7116	2.128	13.223	11.0955
## 1910	4094	50.32	9.174	45.679	36.5047	8.257	29.630	21.3727
## 1911	2036	52.75	13.235	28.571	15.3361	8.824	30.952	22.1289
## 1912	4486	39.70	8.219	29.310	21.0912	2.740	18.966	16.2258
## 1913	524	48.66	14.286	20.000	5.7143	0.000	20.000	20.0000
## 2000	999	70.07	14.286	25.000	10.7143	14.286	20.000	5.7143
## 2001	489	72.60	25.000	24.242	-0.7576	50.000	9.091	-40.9091
## 2002	8774	68.38	13.962	23.175	9.2123	15.849	21.270	5.4208
## 2003	2819	65.48	11.290	20.661	9.3708	16.129	17.355	1.2263
## 2100	1747	49.91	NaN	18.421	NaN	NaN	14.474	NaN
## 2101	269	49.81	NaN	0.000	NaN	NaN	6.250	NaN
## 2102	6699	34.66	7.407	18.293	10.8853	5.556	11.585	6.0298
## 2103	4021	34.29	7.692	20.968	13.2754	3.846	13.710	9.8635
## 2104	1768	28.28	7.692	19.565	11.8729	7.692	15.217	7.5251
## 2105	5389	39.12	17.647	23.789	6.1415	11.765	14.097	2.3322
## 2200	3891	42.48	13.433	23.636	10.2035	7.463	10.909	3.4464
## 2201	811	54.50	0.000	25.000	25.0000	0.000	22.500	22.5000
## 2202	3050	33.84	3.704	1.667	-2.0370	0.000	3.333	3.3333
## 2203	1264	41.30	0.000	15.909	15.9091	0.000	9.091	9.0909
## 2204	5886	49.61	17.241	30.263	13.0218	24.138	21.053	-3.0853
## 2205	8481	42.74	13.393	18.605	5.2118	7.143	10.963	3.8206
## 2206	1080	27.78	25.000	4.167	-20.8333	25.000	4.167	-20.8333
## 2207	5753	37.44	5.769	13.793	8.0239	1.923	17.241	15.3183
## 2208	21701	37.09	9.314	11.413	2.0993	6.373	10.326	3.9535
## 2209	7302	34.87	20.000	18.943	-1.0573	0.000	14.978	14.9780
## 2210	15046	29.62	10.714	11.796	1.0820	4.762	10.188	5.4258
## 2211	10161	29.37	20.755	13.617	-7.1377	7.547	12.340	4.7933
## 2212	1125	47.47	5.882	18.605	12.7223	11.765	13.953	2.1888
## 2213	2849	47.28	36.364	27.928	-8.4357	18.182	18.018	-0.1638
## 2214	924	34.42	22.222	33.333	11.1111	44.444	33.333	-11.1111
## 2215	3839	45.45	11.429	17.576	6.1472	2.857	9.091	6.2338
## 2216	495	58.99	100.000	22.222	-77.7778	0.000	16.667	16.6667
## 2300	7972	50.63	18.841	33.624	14.7839	15.942	28.384	12.4423
## 2301	1338	75.41	31.818	34.615	2.7972	31.818	24.359	-7.4592
## 2302	1633	50.64	23.810	45.614	21.8045	9.524	17.544	8.0201
## 2303	14235	60.74	22.273	42.180	19.9074	18.636	24.171	5.5343
## 2304	8569	51.11	29.412	45.296	15.8844	14.706	29.617	14.9108
## 2305	4671	43.72	22.222	33.939	11.7172	12.963	16.970	4.0067
## 2306	3234	58.53	8.108	38.525	30.4165	21.622	25.410	3.7882
## 2307	4812	54.66	30.000	51.295	21.2953	15.714	35.233	19.5189
## 2308	6652	64.40	19.048	39.726	20.6784	17.460	28.425	10.9643
## 2309	3520	68.18	13.514	43.396	29.8827	13.514	30.818	17.3041
## 2310	5675	47.08	20.000	42.308	22.3077	11.429	25.000	13.5714
## 2311	4094	42.77	18.919	37.255	18.3360	18.919	21.569	2.6497
## 2312	8812	44.89	16.058	36.752	20.6937	5.839	19.658	13.8187
## 2400	2378	59.34	10.000	35.789	25.7895	0.000	36.842	36.8421
## 2401	29	68.97	NaN	50.000	NaN	NaN	66.667	NaN

##	2402	3428	47.02	18.919	46.154	27.2349	16.216	19.231	3.0146
##	2403	9948	53.59	34.454	54.167	19.7129	23.109	35.000	11.8908
##	2404	6509	52.00	26.562	46.795	20.2324	19.531	26.282	6.7508
##	2405	1456	51.85	27.027	37.143	10.1158	24.324	20.000	-4.3243
##	2406	2626	55.75	54.167	50.633	-3.5338	14.583	24.051	9.4673
##	2500	13928	32.98	10.870	18.063	7.1933	8.696	12.565	3.8698
##	2501	490	35.10	33.333	15.385	-17.9487	33.333	15.385	-17.9487
##	2502	3174	46.47	25.000	26.923	1.9231	12.500	26.923	14.4231
##	2503	3822	39.48	15.000	19.444	4.4444	2.500	12.500	10.0000
##	2504	9829	32.51	7.143	17.293	10.1504	8.929	12.030	3.1015
##	2505	11884	34.65	23.188	32.632	9.4432	5.797	15.789	9.9924
##	2506	4643	33.32	19.118	24.096	4.9787	8.824	12.048	3.2247
##	2507	5418	34.27	20.896	22.000	1.1045	7.463	14.000	6.5373
##	2508	6919	32.10	20.755	21.186	0.4317	5.660	19.492	13.8311
##	2600	4474	55.61	7.080	18.705	11.6254	7.965	15.827	7.8627
##	2601	473	60.04	5.882	30.769	24.8869	0.000	23.077	23.0769
##	2602	2156	54.96	11.905	12.500	0.5952	9.524	12.500	2.9762
##	2603	1676	46.36	10.345	17.391	7.0465	10.345	13.043	2.6987
##	2604	9616	43.07	8.750	16.034	7.2838	6.875	13.080	6.2052
##	2605	2237	44.39	4.545	10.448	5.9023	4.545	2.985	-1.5604
##	2606	984	47.87	7.143	5.769	-1.3736	14.286	11.538	-2.7473
##	2607	1803	52.91	12.195	17.021	4.8262	14.634	25.532	10.8978
##	2608	1330	56.54	20.833	17.949	-2.8846	12.500	20.513	8.0128
##	2609	344	67.44	23.077	0.000	-23.0769	23.077	0.000	-23.0769
##	2610	2172	45.81	6.452	5.128	-1.3234	3.226	2.564	-0.6617
##	2611	4978	41.86	6.667	17.647	10.9804	6.667	13.971	7.3039
##	2612	927	36.35	7.692	16.000	8.3077	0.000	12.000	12.0000
##	2613	5523	48.74	7.865	15.672	7.8065	6.742	11.940	5.1987
##	2614	2882	53.75	12.727	21.127	8.3995	14.545	16.901	2.3560
##	2700	23390	65.95	31.538	49.751	18.2128	25.000	42.289	17.2886
##	2701	5598	62.75	44.872	58.755	13.8831	43.590	42.023	-1.5664
##	2702	1160	61.12	44.118	55.556	11.4379	26.471	38.095	11.6246
##	2703	4416	61.96	21.875	46.617	24.7415	29.167	39.098	9.9311
##	2704	861	56.33	43.750	63.636	19.8864	18.750	36.364	17.6136
##	2705	12902	64.38	18.341	37.562	19.2216	10.480	28.856	18.3754
##	2706	2870	65.47	28.205	47.368	19.1633	14.103	34.211	20.1080
##	2707	701	67.05	16.667	50.000	33.3333	16.667	40.000	23.3333
##	2708	1885	62.39	13.793	57.009	43.2162	17.241	38.318	21.0764
##	2709	33	66.67	NaN	NaN	NaN	NaN	NaN	NaN
##	2710	551	53.18	42.857	60.000	17.1429	57.143	30.000	-27.1429
##	2711	2350	71.87	28.571	38.889	10.3175	14.286	36.420	22.1340
##	2712	6347	56.06	44.444	60.648	16.2037	25.000	32.407	7.4074
##	2713	4033	60.72	38.889	63.587	24.6981	22.222	51.087	28.8647
##	2714	1472	74.32	36.842	67.470	30.6278	31.579	59.036	27.4572
##	2715	3670	56.13	17.308	41.892	24.5842	15.385	35.135	19.7505
##	2716	3650	48.60	39.286	55.046	15.7602	37.500	29.358	-8.1422
##	2717	3775	74.60	47.887	68.156	20.2691	42.254	55.307	13.0537
##	2718	1927	64.09	0.000	59.589	59.5890	0.000	53.425	53.4247
##	2719	6140	73.62	25.926	61.722	35.7966	33.333	52.871	19.5375
##	2720	4448	53.46	20.833	53.383	32.5501	29.167	32.331	3.1642

##	2721	1307	58.76	18.182	46.429	28.2468	15.152	32.143	16.9913
##	2722	1169	54.75	45.833	39.623	-6.2107	16.667	20.755	4.0881
##	2723	5227	56.27	28.205	57.672	29.4668	23.077	42.328	19.2511
##	2724	2968	58.52	21.569	57.843	36.2745	13.725	34.314	20.5882
##	2725	6026	59.44	29.167	47.511	18.3446	26.389	32.579	6.1903
##	2726	3460	60.32	25.000	44.444	19.4444	28.333	30.556	2.2222
##	2727	2156	64.10	24.390	50.000	25.6098	19.512	36.765	17.2525
##	2728	13183	61.72	30.165	53.931	23.7652	22.727	32.907	10.1795
##	2729	5534	55.28	38.298	65.737	27.4392	36.170	49.402	13.2322
##	2730	11675	57.10	27.928	49.633	21.7053	28.829	34.719	5.8900
##	2731	4087	60.63	25.000	44.961	19.9612	18.421	35.659	17.2379
##	2732	8318	64.56	22.124	40.365	18.2407	19.469	26.823	7.3539
##	2733	2800	65.29	25.000	36.893	11.8932	22.500	27.184	4.6845
##	2734	3980	63.77	30.579	47.799	17.2202	28.099	45.283	17.1838
##	2735	10518	63.92	38.514	67.012	28.4989	29.730	48.548	18.8180
##	2736	5464	62.23	38.272	53.403	15.1315	29.630	39.267	9.6374
##	2737	7446	59.91	26.344	44.915	18.5712	16.129	24.153	8.0235
##	2738	14635	70.82	39.924	65.015	25.0906	43.346	47.959	4.6132
##	2739	17363	67.17	45.089	65.792	20.7032	39.732	49.601	9.8688
##	2740	5928	59.46	27.891	50.000	22.1088	10.204	34.314	24.1096
##	2741	11357	52.79	20.611	38.043	17.4328	23.664	29.076	5.4120
##	2742	4715	71.50	62.295	64.683	2.3875	42.623	53.968	11.3453
##	2743	1761	47.08	14.286	56.250	41.9643	28.571	39.062	10.4911
##	2744	1	100.00	NaN	NaN	NaN	NaN	NaN	NaN
##	2745	2182	58.02	36.585	51.471	14.8852	26.829	51.471	24.6413
##	2746	12417	60.93	18.391	38.619	20.2286	11.494	27.239	15.7446
##	2747	2181	49.34	25.000	50.820	25.8197	12.500	29.508	17.0082
##	2748	3323	64.10	12.903	32.468	19.5643	16.129	20.130	4.0008
##	2800	10926	56.07	32.161	53.636	21.4756	19.598	31.212	11.6141
##	2801	703	68.99	100.000	43.137	-56.8627	100.000	21.569	-78.4314
##	2802	4093	67.48	38.835	58.042	19.2070	28.155	45.455	17.2992
##	2803	2009	68.34	36.111	67.391	31.2802	25.000	41.304	16.3043
##	2804	4208	55.30	29.487	48.951	19.4639	19.231	26.573	7.3427
##	2805	3920	66.17	31.746	54.144	22.3976	34.921	42.541	7.6208
##	2806	1454	67.88	45.455	66.129	20.6745	21.212	43.548	22.3363
##	2807	743	54.91	0.000	70.833	70.8333	60.000	54.167	-5.8333
##	2808	7370	60.18	30.928	55.452	24.5239	17.526	30.841	13.3153
##	2809	2569	62.36	41.818	50.633	8.8147	27.273	27.848	0.5754
##	2900	4023	75.57	86.667	80.690	-5.9770	88.000	73.103	-14.8966
##	2901	326	75.77	66.667	74.074	7.4074	33.333	62.963	29.6296
##	2902	1483	75.46	50.000	73.611	23.6111	43.750	65.278	21.5278
##	2903	169	70.41	100.000	45.455	-54.5455	100.000	45.455	-54.5455
##	2904	126	72.22	NaN	72.727	NaN	NaN	63.636	NaN
##	2905	991	80.32	64.706	86.957	22.2506	47.059	71.739	24.6803
##	2906	275	74.18	100.000	33.333	-66.6667	100.000	66.667	-33.3333
##	2907	375	69.87	33.333	30.000	-3.3333	33.333	30.000	-3.3333
##	2908	241	73.86	NaN	66.667	NaN	NaN	53.333	NaN
##	2909	1745	77.94	54.167	74.390	20.2236	50.000	59.756	9.7561
##	2910	929	80.41	66.667	59.615	-7.0513	66.667	78.846	12.1795
##	2911	760	75.00	77.778	76.000	-1.7778	44.444	80.000	35.5556

##	2912	354	71.75	50.000	66.667	16.6667	75.000	55.556	-19.4444
##	2913	264	74.62	100.000	93.333	-6.6667	100.000	86.667	-13.3333
##	2914	229	77.73	100.000	100.000	0.0000	100.000	85.714	-14.2857
##	2915	10	90.00	NaN	0.000	NaN	NaN	0.000	NaN
##	2916	4081	61.19	62.745	69.022	6.2766	50.980	47.826	-3.1543
##	2917	486	70.78	100.000	90.000	-10.0000	100.000	45.000	-55.0000
##	2918	0	NA	NA	NA	NA	NA	NA	NA
##	2919	469	78.46	100.000	80.000	-20.0000	100.000	70.000	-30.0000
##	2920	44	65.91	NaN	100.000	NaN	NaN	100.000	NaN
##	2921	884	70.93	92.857	77.500	-15.3571	92.857	67.500	-25.3571
##	2922	195	87.69	NaN	61.538	NaN	NaN	76.923	NaN
##	2923	187	67.91	NaN	85.714	NaN	NaN	85.714	NaN
##	3000	590	60.17	40.000	45.000	5.0000	20.000	50.000	30.0000
##	3001	115	70.43	0.000	57.143	57.1429	0.000	28.571	28.5714
##	3002	3592	57.24	14.865	41.053	26.1878	6.757	17.895	11.1380
##	3003	3357	56.03	29.167	48.544	19.3770	27.083	50.485	23.4021
##	3004	9130	55.77	29.545	45.327	15.7816	23.636	32.243	8.6066
##	3005	4106	53.99	34.211	51.200	16.9895	13.158	32.000	18.8421
##	3100	9664	38.35	30.000	16.216	-13.7838	20.000	13.063	-6.9369
##	3101	3760	34.89	19.231	9.375	-9.8558	15.385	6.250	-9.1346
##	3102	2378	49.03	18.644	22.222	3.5782	15.254	15.556	0.3013
##	3103	2729	43.28	7.547	24.390	16.8431	3.774	21.951	18.1776
##	3104	19913	31.30	9.722	15.407	5.6848	3.472	13.081	9.6092
##	3105	3912	38.19	2.941	16.667	13.7255	2.941	15.333	12.3922
##	3106	4655	39.33	11.111	7.018	-4.0936	8.889	8.772	-0.1170
##	3107	8186	38.42	8.955	14.423	5.4679	7.463	12.019	4.5565
##	3108	2322	48.92	11.111	31.250	20.1389	16.667	21.875	5.2083
##	3109	2677	44.90	8.571	6.522	-2.0497	2.857	4.348	1.4907
##	3110	3057	29.11	18.182	36.170	17.9884	4.545	19.149	14.6035
##	3200	6852	75.70	32.258	54.114	21.8559	29.677	45.253	15.5757
##	3201	1023	70.77	41.176	57.143	15.9664	47.059	57.143	10.0840
##	3202	4028	73.51	29.870	68.750	38.8799	29.870	50.000	20.1299
##	3203	6166	75.33	34.959	63.345	28.3858	39.837	47.687	7.8494
##	3204	7880	74.73	51.381	74.286	22.9045	40.884	55.065	14.1810
##	3205	6694	73.30	35.176	54.545	19.3696	32.161	44.269	12.1080
##	3206	2544	68.32	37.313	65.421	28.1071	34.328	44.860	10.5315
##	3207	5280	78.62	34.831	61.818	26.9867	32.584	52.364	19.7794
##	3300	3105	81.03	19.149	50.920	31.7713	23.404	49.693	26.2890
##	3301	4438	80.42	35.294	63.158	27.8638	31.765	58.772	27.0072
##	3302	1158	74.35	42.857	40.984	-1.8735	28.571	37.705	9.1335
##	3303	3302	75.56	24.706	45.802	21.0956	20.000	41.221	21.2214
##	3304	15761	77.98	43.966	64.436	20.4702	43.103	59.843	16.7391
##	3305	10659	75.19	21.839	42.762	20.9226	22.989	38.307	15.3188
##	3306	6330	75.85	52.055	65.934	13.8793	43.836	63.736	19.9006
##	3307	1255	67.89	44.444	35.294	-9.1503	44.444	37.255	-7.1895
##	3308	5946	80.12	33.735	46.749	13.0143	27.711	47.678	19.9672
##	3309	3121	76.77	44.737	59.091	14.3541	44.737	57.273	12.5359
##	3310	5672	79.80	43.363	57.477	14.1138	43.363	60.280	16.9175
##	3311	751	71.64	42.857	47.727	4.8701	42.857	31.818	-11.0390
##	3312	14440	82.04	35.636	49.868	14.2321	32.364	47.368	15.0048

```

## 3313 1324      59.74  54.545  34.177 -20.3682  27.273  18.987  -8.2854
## 3314 3444      79.73  46.341  60.591  14.2497  46.341  56.158   9.8162
## 3315 2092      80.69  59.091  51.852  -7.2391  59.091  46.667 -12.4242
## 3316 5662      82.82  30.769  52.676  21.9068  30.769  52.958  22.1885
## 3317 1612      73.82  50.000  62.617  12.6168  45.000  54.206   9.2056
## 3318 2061      83.70  73.077  79.310   6.2334  53.846  79.310  25.4642
## 3319 1184      77.79  57.143  69.565  12.4224  38.095  55.072  16.9772
## 3320 3958      82.77  17.778  36.036  18.2583  22.222  34.685  12.4625
## 3321 2264      78.00  21.739  48.276  26.5367  21.739  47.414  25.6747
## 3322 1565      80.83  41.667  40.000  -1.6667  45.833  29.333 -16.5000
## 3400 5215      59.90  45.238  62.651  17.4125  38.095  45.783   7.6879
## 3401  161      19.88    NaN  66.667    NaN    NaN  66.667    NaN
## 3402  341      17.89 100.000  75.000 -25.0000  75.000   0.000 -75.0000
## 3403 1845      23.04   0.000  75.758  75.7576  33.333  45.455  12.1212
## 3404  386      37.56 100.000  66.667 -33.3333 100.000  33.333 -66.6667
## 3500 1469      50.85  20.000  60.000  40.0000  12.000  48.000  36.0000
## 3501   68      36.76    NaN   0.000    NaN    NaN   0.000    NaN
## 3502   0        NA     NA     NA     NA     NA     NA     NA
## 3503   0        NaN    NaN    NaN    NaN    NaN    NaN    NaN
## 3504  480      57.50  31.250  54.545  23.2955  18.750  27.273   8.5227
## 3505  162      54.94    NaN  33.333    NaN    NaN  50.000    NaN
## 3506   82      36.59   0.000  33.333  33.3333   0.000  33.333  33.3333
## 3600  717      69.32  75.000  60.000 -15.0000  62.500  70.000   7.5000
## 3601  381      80.05  57.143  66.667   9.5238  57.143  66.667   9.5238
## 3602  265      80.38    NaN  38.462    NaN    NaN  23.077    NaN
## 3603  102      68.63  25.000 100.000  75.0000  50.000   0.000 -50.0000
## 3604  120      26.67    NaN  28.571    NaN    NaN  57.143    NaN
## 3605  329      62.92   0.000  23.810  23.8095   0.000  28.571  28.5714
## 3606   0        NaN    NaN    NaN    NaN    NaN    NaN    NaN
## 3607  460      51.74  66.667  53.846 -12.8205  66.667  38.462 -28.2051
## 3608   0        NaN    NaN    NaN    NaN    NaN    NaN    NaN
## 3609  808      79.08  93.103 100.000   6.8966  72.414  75.000   2.5862
## 3610  284      58.80  60.000  70.000  10.0000  80.000  40.000 -40.0000
## 3611  714      78.71  44.444  69.091  24.6465  66.667  63.636  -3.0303
## 3612 5356      72.31  40.506  62.081  21.5742  29.114  48.993  19.8794
## 3613   27      66.67   0.000   0.000   0.0000 100.000   0.000 -100.0000
## 3614 2675      49.72  27.778  44.872  17.0940  22.222  41.026  18.8034
## 3615   0        NA     NA     NA     NA     NA     NA     NA
## 3616 1367      75.71  41.176  74.576  33.3998  64.706  64.407  -0.2991

```

```

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         768  24.740  20.588  25.000   4.4118  15.365  11.765  12.500
## 1100        2541  36.600  21.368  49.323  27.9555  21.173   9.402  24.952
## 1101         558  29.928  30.000  31.250   1.2500  18.817  18.000   6.250

```

## 1102	1998	29.179	16.038	35.294	19.2564	21.321	19.811	26.471
## 1103	3933	38.291	21.860	42.809	20.9489	23.494	17.674	26.756
## 1104	4450	37.438	23.024	44.959	21.9351	19.371	15.464	23.978
## 1105	9545	34.814	23.647	41.102	17.4549	21.173	16.433	25.488
## 1106	1871	42.972	37.500	48.889	11.3889	27.953	18.750	35.556
## 1107	2273	24.725	4.494	31.343	26.8489	16.806	6.742	23.383
## 1108	669	25.411	23.810	31.481	7.6720	24.066	23.810	18.519
## 1109	1552	33.892	28.713	41.818	13.1053	21.456	17.822	27.273
## 1110	3558	36.818	28.704	43.644	14.9404	21.922	15.278	24.576
## 1111	948	31.435	15.789	42.697	26.9072	15.612	5.263	20.225
## 1200	294	30.952	75.000	39.583	-35.4167	29.592	50.000	35.417
## 1201	3728	44.984	34.463	47.658	13.1951	39.619	30.508	40.220
## 1202	2738	32.286	21.127	35.915	14.7887	32.031	18.310	33.803
## 1203	2422	50.619	45.161	42.500	-2.6613	47.275	41.935	40.000
## 1204	494	31.579	22.222	28.846	6.6239	26.518	22.222	17.308
## 1205	221	27.602	33.333	21.875	-11.4583	27.149	33.333	21.875
## 1206	78	46.154	0.000	25.000	25.0000	44.872	0.000	25.000
## 1207	1201	37.469	25.000	42.636	17.6357	35.554	20.000	42.636
## 1208	2084	43.330	38.298	43.367	5.0695	43.090	31.915	44.388
## 1209	84	45.238	50.000	33.333	-16.6667	39.286	0.000	33.333
## 1210	301	39.535	33.333	31.579	-1.7544	40.199	33.333	44.737
## 1211	2606	23.446	22.115	26.378	4.2626	23.676	20.192	29.134
## 1212	1342	24.441	22.222	23.009	0.7866	24.739	22.222	21.239
## 1213	792	49.369	61.111	48.113	-12.9979	48.737	50.000	47.170
## 1300	2857	36.682	19.492	47.535	28.0437	20.756	11.864	25.176
## 1301	13	30.769	NaN	25.000	NaN	0.000	NaN	0.000
## 1302	595	46.218	40.909	63.265	22.3562	35.126	31.818	34.694
## 1303	11342	34.095	30.791	35.146	4.3545	16.329	14.101	15.782
## 1304	2981	33.613	25.581	32.780	7.1987	17.679	18.605	16.183
## 1305	2456	35.708	23.636	36.444	12.8081	19.137	16.364	25.333
## 1306	3153	45.354	32.203	48.000	15.7966	27.244	25.424	30.182
## 1307	6863	41.556	36.927	45.868	8.9412	20.122	19.954	20.661
## 1308	1984	36.694	36.792	38.346	1.5534	18.397	12.264	15.789
## 1309	1716	46.037	36.697	58.678	21.9804	25.117	23.853	23.967
## 1310	2388	47.571	34.821	57.592	22.7702	28.894	25.000	38.743
## 1311	5161	41.387	31.118	48.848	17.7301	23.426	22.356	24.194
## 1312	9471	39.056	33.982	43.741	9.7587	19.702	17.876	22.446
## 1313	1668	34.592	26.562	38.125	11.5625	18.825	10.938	18.750
## 1314	6013	36.687	29.032	49.749	20.7165	19.608	13.710	23.869
## 1315	1054	29.127	29.630	37.500	7.8704	15.560	5.556	16.667
## 1400	951	28.707	35.897	30.864	-5.0332	27.550	30.769	20.988
## 1401	329	36.170	25.000	25.926	0.9259	34.043	25.000	29.630
## 1402	571	21.016	14.286	43.590	29.3040	23.643	21.429	41.026
## 1403	1662	34.296	31.667	41.727	10.0600	30.806	25.000	34.532
## 1404	319	30.721	21.053	46.154	25.1012	29.154	21.053	30.769
## 1405	1316	31.763	24.051	30.682	6.6312	30.395	29.114	27.273
## 1406	837	36.201	10.526	42.373	31.8466	34.170	7.895	40.678
## 1407	1241	39.565	33.824	45.545	11.7210	39.162	35.294	37.624
## 1408	2117	31.271	21.569	38.012	16.4431	29.287	20.588	35.088
## 1409	706	45.892	26.667	58.333	31.6667	41.218	26.667	51.667

## 1410	242	23.140	10.000	28.571	18.5714	28.512	10.000	35.714
## 1500	2527	20.538	14.184	19.841	5.6569	9.300	5.674	14.286
## 1501	108	23.148	25.000	NaN	NaN	7.407	0.000	NaN
## 1502	1307	30.834	17.143	25.595	8.4524	20.505	25.714	19.643
## 1503	2373	21.239	16.111	27.778	11.6667	9.945	7.778	16.049
## 1504	79	26.582	0.000	42.857	42.8571	27.848	0.000	28.571
## 1505	911	21.844	14.815	33.898	19.0835	11.636	5.556	13.559
## 1506	173	26.012	0.000	40.000	40.0000	11.561	50.000	20.000
## 1507	216	8.796	0.000	16.667	16.6667	8.796	0.000	12.500
## 1508	221	21.267	23.077	21.429	-1.6484	13.575	7.692	50.000
## 1600	6304	23.461	20.222	28.322	8.1001	11.405	7.756	16.084
## 1601	118	33.051	0.000	57.143	57.1429	21.186	0.000	71.429
## 1602	2037	28.473	24.176	33.918	9.7423	13.353	17.582	15.789
## 1603	808	27.847	22.222	27.957	5.7348	14.604	11.111	11.828
## 1604	2081	25.517	19.718	26.744	7.0259	11.004	8.451	14.535
## 1605	5424	23.636	14.589	24.599	10.0101	10.804	8.753	12.299
## 1606	4914	21.713	13.619	23.757	10.1382	10.765	5.837	12.707
## 1607	1495	27.559	18.182	35.088	16.9059	13.579	9.091	18.421
## 1700	864	19.097	11.538	14.130	2.5920	15.046	15.385	11.957
## 1701	82	24.390	NaN	12.500	NaN	20.732	NaN	25.000
## 1702	1203	15.628	11.321	15.152	3.8308	14.214	11.321	12.121
## 1703	1006	14.115	17.143	13.187	-3.9560	12.425	8.571	6.593
## 1704	524	9.924	11.538	14.000	2.4615	7.634	7.692	6.000
## 1705	1361	14.401	7.143	23.585	16.4420	13.887	7.143	19.811
## 1706	3295	15.296	8.333	17.751	9.4181	13.445	6.481	12.722
## 1707	561	10.160	0.000	17.544	17.5439	8.378	7.407	7.018
## 1708	836	11.962	13.953	10.345	-3.6087	11.124	16.279	12.069
## 1709	544	29.779	12.500	35.088	22.5877	26.471	25.000	28.070
## 1710	1458	22.702	11.765	28.972	17.2073	20.850	13.235	19.626
## 1711	801	10.612	2.083	15.789	13.7061	7.116	8.333	9.211
## 1712	2243	12.885	10.417	15.873	5.4563	10.834	10.417	9.524
## 1800	250	32.400	22.222	34.483	12.2605	25.200	27.778	27.586
## 1801	2	0.000	NaN	NaN	NaN	50.000	NaN	NaN
## 1802	488	23.566	22.727	35.484	12.7566	18.033	13.636	16.129
## 1803	1026	17.544	12.069	22.973	10.9040	14.620	15.517	14.865
## 1804	982	16.497	11.111	27.273	16.1616	13.340	7.407	15.152
## 1900	1633	21.494	13.846	31.429	17.5824	14.942	10.769	23.810
## 1901	1066	21.670	5.797	29.167	23.3696	15.478	7.246	20.833
## 1902	1348	20.846	7.143	29.008	21.8648	12.389	2.857	16.031
## 1903	311	16.399	5.263	21.212	15.9490	11.254	0.000	9.091
## 1904	1733	27.063	10.526	36.301	25.7751	19.446	6.579	20.548
## 1905	185	24.865	13.333	31.579	18.2456	20.000	13.333	21.053
## 1906	2083	20.451	13.125	34.586	21.4615	13.154	10.625	14.286
## 1907	1603	17.779	13.889	27.193	13.3041	13.475	10.185	14.912
## 1908	1490	19.530	16.162	31.783	15.6213	12.483	9.091	14.729
## 1909	774	11.499	4.255	6.897	2.6412	8.398	2.128	13.793
## 1910	1501	24.184	9.174	33.636	24.4621	13.991	8.257	17.273
## 1911	774	21.189	13.235	35.294	22.0588	11.628	8.824	21.569
## 1912	1271	14.477	8.219	21.111	12.8919	10.071	2.740	10.000
## 1913	152	24.342	14.286	40.000	25.7143	19.079	0.000	26.667

## 2000	391	25.831	14.286	38.776	24.4898	22.762	14.286	28.571
## 2001	206	25.728	25.000	31.579	6.5789	26.699	50.000	42.105
## 2002	4133	20.324	13.962	25.667	11.7044	19.356	15.849	20.667
## 2003	1199	16.013	11.290	19.780	8.4899	15.763	16.129	15.385
## 2100	429	19.580	NaN	21.667	NaN	11.189	NaN	8.333
## 2101	51	13.725	NaN	0.000	NaN	7.843	NaN	0.000
## 2102	1237	10.428	7.407	11.504	4.0970	7.114	5.556	7.080
## 2103	723	13.416	7.692	9.722	2.0299	6.777	3.846	8.333
## 2104	270	9.630	7.692	13.793	6.1008	4.444	7.692	3.448
## 2105	911	18.880	17.647	14.789	-2.8583	13.392	11.765	14.085
## 2200	1132	16.696	13.433	22.105	8.6724	15.106	7.463	15.789
## 2201	270	20.370	0.000	32.000	32.0000	13.333	0.000	24.000
## 2202	595	12.437	3.704	18.182	14.4781	7.899	0.000	5.455
## 2203	253	17.787	0.000	19.231	19.2308	9.881	0.000	7.692
## 2204	1573	30.324	17.241	27.350	10.1090	20.470	24.138	22.650
## 2205	1965	13.588	13.393	14.216	0.8228	9.975	7.143	10.784
## 2206	198	8.586	25.000	16.667	-8.3333	7.576	25.000	5.556
## 2207	1229	9.520	5.769	14.530	8.7607	6.672	1.923	4.274
## 2208	5105	10.147	9.314	11.111	1.7974	7.796	6.373	11.556
## 2209	1412	16.006	20.000	20.000	0.0000	9.703	0.000	13.333
## 2210	2314	10.328	10.714	12.542	1.8281	7.001	4.762	10.169
## 2211	1598	13.016	20.755	13.402	-7.3527	10.325	7.547	11.340
## 2212	309	10.356	5.882	14.815	8.9325	8.091	11.765	11.111
## 2213	696	25.144	36.364	31.818	-4.5455	19.109	18.182	20.455
## 2214	218	20.183	22.222	36.842	14.6199	13.761	44.444	15.789
## 2215	817	12.607	11.429	11.927	0.4980	8.323	2.857	7.339
## 2216	129	31.008	100.000	23.529	-76.4706	30.233	0.000	29.412
## 2300	2686	30.901	18.841	34.804	15.9633	20.774	15.942	25.490
## 2301	624	32.853	31.818	32.727	0.9091	31.891	31.818	29.091
## 2302	505	27.723	23.810	36.538	12.7289	16.832	9.524	28.846
## 2303	5770	33.865	22.273	42.218	19.9455	21.837	18.636	27.728
## 2304	2774	35.076	29.412	39.326	9.9141	18.133	14.706	22.472
## 2305	1150	31.130	22.222	31.387	9.1646	16.435	12.963	18.248
## 2306	1220	28.525	8.108	33.600	25.4919	20.082	21.622	20.000
## 2307	1645	42.128	30.000	53.147	23.1469	23.465	15.714	22.378
## 2308	2466	31.144	19.048	36.759	17.7113	25.101	17.460	29.249
## 2309	1362	34.435	13.514	40.909	27.3956	24.082	13.514	23.485
## 2310	1464	35.997	20.000	41.026	21.0256	20.902	11.429	23.718
## 2311	911	32.492	18.919	29.730	10.8108	19.868	18.919	20.721
## 2312	2462	28.107	16.058	32.512	16.4539	16.450	5.839	20.690
## 2400	720	35.278	10.000	43.182	33.1818	21.389	0.000	27.273
## 2401	7	28.571	NaN	50.000	NaN	0.000	NaN	0.000
## 2402	1156	39.187	18.919	49.462	30.5434	18.599	16.216	31.183
## 2403	3863	44.887	34.454	49.660	15.2061	23.376	23.109	27.891
## 2404	2328	42.612	26.562	49.754	23.1912	20.704	19.531	24.631
## 2405	509	38.507	27.027	44.681	17.6538	21.807	24.324	25.532
## 2406	993	46.425	54.167	48.421	-5.7456	17.321	14.583	21.053
## 2500	2574	17.521	10.870	17.778	6.9082	11.150	8.696	13.333
## 2501	129	23.256	33.333	0.000	-33.3333	8.527	33.333	100.000
## 2502	891	34.680	25.000	29.167	4.1667	25.253	12.500	21.667

## 2503	995	21.307	15.000	25.000	10.0000	12.663	2.500	13.889
## 2504	1987	14.746	7.143	12.755	5.6122	9.411	8.929	14.286
## 2505	2880	23.785	23.188	25.781	2.5928	11.076	5.797	13.672
## 2506	1015	20.197	19.118	25.287	6.1697	10.049	8.824	11.494
## 2507	1190	25.378	20.896	30.435	9.5393	12.101	7.463	12.174
## 2508	1397	21.976	20.755	24.088	3.3329	10.379	5.660	13.139
## 2600	1589	12.649	7.080	14.085	7.0049	13.027	7.965	13.380
## 2601	193	14.508	5.882	18.182	12.2995	11.399	0.000	18.182
## 2602	815	10.307	11.905	5.797	-6.1077	8.221	9.524	7.246
## 2603	489	12.679	10.345	17.073	6.7283	11.247	10.345	14.634
## 2604	2751	12.904	8.750	13.433	4.6828	10.251	6.875	10.075
## 2605	636	13.522	4.545	12.500	7.9545	9.591	4.545	7.143
## 2606	236	10.593	7.143	11.765	4.6218	11.017	14.286	0.000
## 2607	647	12.519	12.195	18.000	5.8049	12.365	14.634	20.000
## 2608	481	11.019	20.833	13.333	-7.5000	10.603	12.500	8.889
## 2609	177	10.169	23.077	15.385	-7.6923	9.040	23.077	15.385
## 2610	747	8.835	6.452	7.547	1.0956	6.827	3.226	5.660
## 2611	1260	14.127	6.667	19.286	12.6190	10.238	6.667	8.571
## 2612	223	11.211	7.692	22.222	14.5299	7.623	0.000	16.667
## 2613	1845	16.423	7.865	19.075	11.2100	13.604	6.742	10.983
## 2614	1110	12.432	12.727	6.452	-6.2757	11.982	14.545	11.828
## 2700	9346	45.228	31.538	49.796	18.2570	35.759	25.000	36.299
## 2701	2015	53.846	44.872	53.922	9.0498	41.836	43.590	41.667
## 2702	447	41.163	44.118	53.571	9.4538	26.622	26.471	28.571
## 2703	1861	35.519	21.875	43.750	21.8750	25.954	29.167	31.250
## 2704	333	40.841	43.750	25.714	-18.0357	26.126	18.750	20.000
## 2705	5708	25.893	18.341	30.617	12.2767	15.872	10.480	16.790
## 2706	1274	31.947	28.205	43.678	15.4730	22.449	14.103	35.632
## 2707	288	41.319	16.667	37.931	21.2644	37.153	16.667	27.586
## 2708	698	36.963	13.793	46.667	32.8736	27.077	17.241	40.000
## 2709	13	69.231	NaN	100.000	NaN	61.538	NaN	100.000
## 2710	199	56.784	42.857	80.000	37.1429	34.673	57.143	26.667
## 2711	956	30.230	28.571	35.165	6.5934	22.490	14.286	32.967
## 2712	2113	50.450	44.444	60.748	16.3032	31.661	25.000	38.785
## 2713	1436	46.379	38.889	54.598	15.7088	32.939	22.222	43.678
## 2714	655	51.603	36.842	50.000	13.1579	41.221	31.579	40.000
## 2715	1428	31.022	17.308	39.344	22.0366	20.168	15.385	23.770
## 2716	1111	47.255	39.286	55.882	16.5966	34.653	37.500	34.314
## 2717	1794	60.424	47.887	67.320	19.4329	50.836	42.254	50.327
## 2718	605	48.430	0.000	65.934	65.9341	40.331	0.000	48.352
## 2719	2240	52.991	25.926	57.879	31.9529	43.839	33.333	53.030
## 2720	1558	43.838	20.833	50.355	29.5213	27.856	29.167	30.496
## 2721	474	33.122	18.182	45.000	26.8182	18.354	15.152	22.500
## 2722	372	41.935	45.833	52.778	6.9444	23.118	16.667	22.222
## 2723	1867	46.331	28.205	52.318	24.1128	26.406	23.077	32.450
## 2724	1067	41.799	21.569	57.547	35.9785	27.366	13.725	38.679
## 2725	2171	44.404	29.167	48.178	19.0115	24.873	26.389	31.579
## 2726	1325	40.151	25.000	55.000	30.0000	28.075	28.333	31.667
## 2727	895	36.089	24.390	37.662	13.2721	25.140	19.512	29.870
## 2728	5037	39.885	30.165	50.000	19.8347	26.325	22.727	29.665

## 2729	1782	58.474	38.298	74.684	36.3857	39.731	36.170	50.633
## 2730	4100	44.902	27.928	48.492	20.5645	29.537	28.829	33.668
## 2731	1765	34.164	25.000	39.860	14.8601	26.232	18.421	25.175
## 2732	3221	31.357	22.124	32.581	10.4568	19.776	19.469	20.645
## 2733	1182	31.641	25.000	31.481	6.4815	19.966	22.500	17.593
## 2734	1583	42.956	30.579	50.000	19.4215	33.354	28.099	26.543
## 2735	4246	53.650	38.514	64.921	26.4080	40.438	29.730	49.476
## 2736	2171	47.121	38.272	56.604	18.3322	29.480	29.630	37.736
## 2737	3060	34.771	26.344	42.975	16.6311	19.608	16.129	22.314
## 2738	6217	51.182	39.924	54.839	14.9148	41.306	43.346	42.704
## 2739	6818	55.133	45.089	63.731	18.6413	44.823	39.732	48.316
## 2740	2339	34.245	27.891	43.085	15.1939	19.282	10.204	23.936
## 2741	3798	29.989	20.611	31.579	10.9683	19.300	23.664	24.931
## 2742	2011	62.705	62.295	61.404	-0.8916	49.279	42.623	47.807
## 2743	509	53.635	14.286	69.565	55.2795	29.862	28.571	34.783
## 2744	1	0.000	NaN	NaN	NaN	0.000	NaN	NaN
## 2745	821	53.837	36.585	61.446	24.8604	35.688	26.829	43.373
## 2746	4794	23.488	18.391	27.027	8.6362	16.083	11.494	17.199
## 2747	761	34.823	25.000	40.000	15.0000	20.237	12.500	27.273
## 2748	1313	19.573	12.903	19.118	6.2144	13.557	16.129	15.441
## 2800	4255	37.579	32.161	44.364	12.2028	20.094	19.598	25.091
## 2801	205	40.488	100.000	31.579	-68.4211	26.829	100.000	18.421
## 2802	1785	47.395	38.835	50.000	11.1650	32.605	28.155	38.710
## 2803	830	45.783	36.111	52.041	15.9297	25.060	25.000	31.633
## 2804	1540	41.299	29.487	45.912	16.4248	23.052	19.231	23.270
## 2805	1534	45.632	31.746	43.651	11.9048	32.595	34.921	26.190
## 2806	621	46.860	45.455	51.020	5.5659	36.554	21.212	46.939
## 2807	253	57.312	0.000	70.000	70.0000	45.059	60.000	40.000
## 2808	2619	39.939	30.928	46.371	15.4431	25.544	17.526	29.032
## 2809	1062	38.418	41.818	48.193	6.3746	29.849	27.273	26.506
## 2900	2205	80.227	86.667	79.605	-7.0614	75.102	88.000	73.026
## 2901	119	79.832	66.667	72.222	5.5556	68.908	33.333	61.111
## 2902	665	61.504	50.000	80.247	30.2469	51.729	43.750	60.494
## 2903	60	56.667	100.000	50.000	-50.0000	58.333	100.000	33.333
## 2904	42	45.238	NaN	75.000	NaN	45.238	NaN	75.000
## 2905	506	73.123	64.706	69.388	4.6819	66.601	47.059	57.143
## 2906	136	87.500	100.000	82.609	-17.3913	80.882	100.000	73.913
## 2907	148	41.216	33.333	38.462	5.1282	40.541	33.333	38.462
## 2908	99	81.818	NaN	85.714	NaN	79.798	NaN	71.429
## 2909	881	62.883	54.167	73.239	19.0728	55.505	50.000	57.746
## 2910	427	57.611	66.667	53.030	-13.6364	53.396	66.667	57.576
## 2911	399	70.426	77.778	95.652	17.8744	62.907	44.444	69.565
## 2912	154	82.468	50.000	70.000	20.0000	77.922	75.000	60.000
## 2913	109	88.991	100.000	94.118	-5.8824	76.147	100.000	76.471
## 2914	91	89.011	100.000	90.000	-10.0000	76.923	100.000	80.000
## 2915	2	50.000	NaN	0.000	NaN	50.000	NaN	0.000
## 2916	1375	62.473	62.745	63.057	0.3122	47.345	50.980	52.866
## 2917	193	80.829	100.000	70.270	-29.7297	67.358	100.000	51.351
## 2918	0	NA	NA	NA	NA	NA	NA	NA
## 2919	210	88.095	100.000	93.548	-6.4516	76.190	100.000	70.968

## 2920	19	68.421	NaN	NaN	NaN	57.895	NaN	NaN
## 2921	360	68.056	92.857	60.465	-32.3920	61.667	92.857	46.512
## 2922	94	79.787	NaN	56.250	NaN	72.340	NaN	68.750
## 2923	86	93.023	NaN	100.000	NaN	89.535	NaN	85.714
## 3000	229	41.921	40.000	47.368	7.3684	25.764	20.000	26.316
## 3001	24	58.333	0.000	66.667	66.6667	33.333	0.000	33.333
## 3002	1497	26.453	14.865	32.039	17.1740	12.024	6.757	12.621
## 3003	1279	37.060	29.167	30.392	1.2255	28.772	27.083	39.216
## 3004	3678	39.859	29.545	46.290	16.7443	22.023	23.636	25.442
## 3005	1496	45.521	34.211	51.020	16.8099	25.936	13.158	20.408
## 3100	2417	12.453	30.000	13.084	-16.9159	9.433	20.000	12.150
## 3101	899	10.790	19.231	11.111	-8.1197	6.563	15.385	11.111
## 3102	808	22.277	18.644	27.848	9.2040	16.955	15.254	20.253
## 3103	848	12.146	7.547	11.475	3.9282	9.552	3.774	11.475
## 3104	3954	14.896	9.722	15.957	6.2352	8.978	3.472	13.298
## 3105	869	17.722	2.941	24.390	21.4491	7.595	2.941	12.195
## 3106	1324	10.574	11.111	17.647	6.5359	7.779	8.889	9.412
## 3107	1943	11.065	8.955	13.080	4.1249	10.190	7.463	10.127
## 3108	776	28.866	11.111	33.803	22.6917	18.428	16.667	25.352
## 3109	894	8.277	8.571	5.882	-2.6891	7.383	2.857	3.529
## 3110	566	25.972	18.182	27.869	9.6870	17.138	4.545	11.475
## 3200	3356	43.832	32.258	52.941	20.6831	38.886	29.677	44.118
## 3201	489	45.194	41.176	43.902	2.7260	44.172	47.059	36.585
## 3202	1714	47.666	29.870	55.429	25.5584	43.524	29.870	49.143
## 3203	2820	53.085	34.959	61.511	26.5514	45.106	39.837	51.439
## 3204	3651	61.873	51.381	67.164	15.7830	52.944	40.884	55.821
## 3205	3222	44.662	35.176	50.579	15.4033	37.182	32.161	37.838
## 3206	1147	48.649	37.313	51.923	14.6096	37.489	34.328	45.192
## 3207	2386	51.928	34.831	61.667	26.8352	46.982	32.584	50.833
## 3300	1479	39.689	19.149	45.625	26.4761	38.540	23.404	45.000
## 3301	2177	52.595	35.294	64.593	29.2992	50.390	31.765	60.766
## 3302	527	31.499	42.857	31.373	-11.4846	28.083	28.571	21.569
## 3303	1601	39.413	24.706	46.296	21.5904	38.726	20.000	37.037
## 3304	7506	53.251	43.966	58.430	14.4643	51.412	43.103	56.371
## 3305	5086	34.231	21.839	40.278	18.4387	32.422	22.989	39.120
## 3306	2762	60.210	52.055	65.698	13.6429	53.621	43.836	59.884
## 3307	508	38.976	44.444	40.816	-3.6281	30.512	44.444	26.531
## 3308	2689	39.792	33.735	47.855	14.1198	37.114	27.711	39.934
## 3309	1504	50.332	44.737	50.327	5.5900	50.399	44.737	49.020
## 3310	2890	51.869	43.363	43.220	-0.1425	49.412	43.363	42.797
## 3311	276	33.696	42.857	45.714	2.8571	31.159	42.857	40.000
## 3312	7126	41.819	35.636	49.635	13.9987	40.065	32.364	43.504
## 3313	407	21.376	54.545	21.622	-32.9238	18.673	27.273	27.027
## 3314	1475	50.305	46.341	60.366	14.0244	48.068	46.341	56.098
## 3315	857	45.858	59.091	49.153	-9.9384	44.924	59.091	46.610
## 3316	2381	47.123	30.769	47.603	16.8335	46.325	30.769	45.548
## 3317	605	58.017	50.000	75.806	25.8065	52.397	45.000	64.516
## 3318	971	77.240	73.077	80.992	7.9148	74.871	53.846	76.860
## 3319	530	61.887	57.143	67.308	10.1648	56.981	38.095	71.154
## 3320	1919	28.088	17.778	29.630	11.8519	29.182	22.222	26.852

## 3321	1148	31.620	21.739	46.914	25.1744	32.491	21.739	44.444
## 3322	794	31.612	41.667	30.000	-11.6667	32.872	45.833	31.250
## 3400	2219	50.518	45.238	59.477	14.2390	35.602	38.095	41.176
## 3401	12	50.000	NaN	0.000	NaN	50.000	NaN	100.000
## 3402	34	47.059	100.000	50.000	-50.0000	38.235	75.000	50.000
## 3403	257	51.751	0.000	66.667	66.6667	31.518	33.333	33.333
## 3404	96	50.000	100.000	71.429	-28.5714	45.833	100.000	85.714
## 3500	549	29.508	20.000	50.000	30.0000	22.587	12.000	33.333
## 3501	6	0.000	NaN	0.000	NaN	33.333	NaN	66.667
## 3502	0	NA	NA	NA	NA	NA	NA	NA
## 3503	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
## 3504	180	26.667	31.250	23.077	-8.1731	15.556	18.750	23.077
## 3505	49	24.490	NaN	33.333	NaN	8.163	NaN	0.000
## 3506	18	33.333	0.000	100.000	100.0000	5.556	0.000	0.000
## 3600	414	61.836	75.000	80.000	5.0000	56.763	62.500	80.000
## 3601	186	62.903	57.143	73.913	16.7702	54.839	57.143	69.565
## 3602	86	27.907	NaN	35.294	NaN	23.256	NaN	17.647
## 3603	52	57.692	25.000	33.333	8.3333	55.769	50.000	33.333
## 3604	17	29.412	NaN	NaN	NaN	47.059	NaN	NaN
## 3605	94	41.489	0.000	87.500	87.5000	41.489	0.000	75.000
## 3606	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
## 3607	157	36.943	66.667	50.000	-16.6667	26.752	66.667	16.667
## 3608	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
## 3609	440	90.227	93.103	97.222	4.1188	82.273	72.414	88.889
## 3610	115	52.174	60.000	71.429	11.4286	33.913	80.000	35.714
## 3611	310	56.774	44.444	56.667	12.2222	56.774	66.667	63.333
## 3612	2290	51.747	40.506	56.972	16.4658	40.655	29.114	46.614
## 3613	11	54.545	0.000	0.000	0.0000	27.273	100.000	100.000
## 3614	858	30.303	27.778	32.000	4.2222	19.930	22.222	28.000
## 3615	0	NA	NA	NA	NA	NA	NA	NA
## 3616	654	66.055	41.176	60.938	19.7610	55.657	64.706	57.812
##	ChgN 1CtyMF96 1CtyMF14							
## 1000	0.7353	34	64					
## 1100	15.5499	117	517					
## 1101	-11.7500	50	16					
## 1102	6.6593	106	170					
## 1103	9.0814	215	299					
## 1104	8.5143	291	367					
## 1105	9.0551	499	871					
## 1106	16.8056	96	180					
## 1107	16.6415	89	201					
## 1108	-5.2910	42	54					
## 1109	9.4509	101	110					
## 1110	9.2985	216	236					
## 1111	14.9616	38	89					
## 1200	-14.5833	4	48					
## 1201	9.7119	177	363					
## 1202	15.4930	71	284					
## 1203	-1.9355	93	200					
## 1204	-4.9145	9	52					

## 1205	-11.4583	3	32
## 1206	25.0000	2	4
## 1207	22.6357	40	129
## 1208	12.4729	47	196
## 1209	33.3333	2	3
## 1210	11.4035	12	38
## 1211	8.9416	104	254
## 1212	-0.9833	54	113
## 1213	-2.8302	18	106
## 1300	13.3116	118	568
## 1301	NaN	0	4
## 1302	2.8757	22	49
## 1303	1.6818	695	754
## 1304	-2.4221	129	241
## 1305	8.9697	110	225
## 1306	4.7581	118	275
## 1307	0.7070	436	484
## 1308	3.5253	106	133
## 1309	0.1137	109	121
## 1310	13.7435	112	191
## 1311	1.8371	331	434
## 1312	4.5699	565	695
## 1313	7.8125	64	160
## 1314	10.1597	372	398
## 1315	11.1111	54	72
## 1400	-9.7816	39	81
## 1401	4.6296	16	27
## 1402	19.5971	28	39
## 1403	9.5324	60	139
## 1404	9.7166	19	26
## 1405	-1.8412	79	88
## 1406	32.7832	38	59
## 1407	2.3296	68	101
## 1408	14.4995	102	171
## 1409	25.0000	30	60
## 1410	25.7143	10	14
## 1500	8.6120	141	252
## 1501	NaN	4	0
## 1502	-6.0714	35	168
## 1503	8.2716	180	162
## 1504	28.5714	2	7
## 1505	8.0038	54	59
## 1506	-30.0000	2	10
## 1507	12.5000	7	24
## 1508	42.3077	13	14
## 1600	8.3277	361	572
## 1601	71.4286	1	7
## 1602	-1.7929	91	171
## 1603	0.7168	27	93
## 1604	6.0842	142	172

## 1605	3.5461	377	374
## 1606	6.8706	257	362
## 1607	9.3301	55	114
## 1700	-3.4281	26	92
## 1701	NaN	0	8
## 1702	0.8005	53	99
## 1703	-1.9780	35	91
## 1704	-1.6923	26	50
## 1705	12.6685	42	106
## 1706	6.2404	108	338
## 1707	-0.3899	27	57
## 1708	-4.2101	43	58
## 1709	3.0702	8	57
## 1710	6.3909	68	107
## 1711	0.8772	48	76
## 1712	-0.8929	96	189
## 1800	-0.1916	18	29
## 1801	NaN	0	0
## 1802	2.4927	22	31
## 1803	-0.6524	58	74
## 1804	7.7441	54	66
## 1900	13.0403	130	105
## 1901	13.5870	69	72
## 1902	13.1734	70	131
## 1903	9.0909	19	33
## 1904	13.9690	76	146
## 1905	7.7193	15	19
## 1906	3.6607	160	133
## 1907	4.7271	108	114
## 1908	5.6378	99	129
## 1909	11.6654	47	58
## 1910	9.0158	109	110
## 1911	12.7451	68	51
## 1912	7.2603	73	90
## 1913	26.6667	7	15
## 2000	14.2857	7	49
## 2001	-7.8947	4	19
## 2002	4.8176	265	300
## 2003	-0.7444	62	91
## 2100	NaN	0	60
## 2101	NaN	0	4
## 2102	1.5241	54	113
## 2103	4.4872	26	72
## 2104	-4.2440	13	29
## 2105	2.3198	17	142
## 2200	8.3268	134	95
## 2201	24.0000	13	25
## 2202	5.4545	27	55
## 2203	7.6923	7	26
## 2204	-1.4884	29	234

## 2205	3.6415	112	204
## 2206	-19.4444	8	18
## 2207	2.3504	52	117
## 2208	5.1830	204	450
## 2209	13.3333	5	135
## 2210	5.4076	84	295
## 2211	3.7930	53	194
## 2212	-0.6536	17	27
## 2213	2.2727	22	88
## 2214	-28.6550	9	19
## 2215	4.4823	35	109
## 2216	29.4118	1	17
## 2300	9.5482	138	204
## 2301	-2.7273	22	55
## 2302	19.3223	21	52
## 2303	9.0917	220	559
## 2304	7.7660	102	267
## 2305	5.2852	54	137
## 2306	-1.6216	37	125
## 2307	6.6633	70	143
## 2308	11.7887	63	253
## 2309	9.9713	37	132
## 2310	12.2894	70	156
## 2311	1.8018	37	111
## 2312	14.8502	137	203
## 2400	27.2727	20	88
## 2401	NaN	0	2
## 2402	14.9666	74	93
## 2403	4.7819	238	294
## 2404	5.0993	128	203
## 2405	1.2076	37	47
## 2406	6.4693	48	95
## 2500	4.6377	92	315
## 2501	66.6667	6	1
## 2502	9.1667	24	120
## 2503	11.3889	40	108
## 2504	5.3571	56	196
## 2505	7.8748	138	256
## 2506	2.6707	68	87
## 2507	4.7112	67	115
## 2508	7.4783	53	137
## 2600	5.4157	113	142
## 2601	18.1818	17	11
## 2602	-2.2774	42	69
## 2603	4.2893	29	41
## 2604	3.1996	160	268
## 2605	2.5974	22	56
## 2606	-14.2857	14	17
## 2607	5.3659	41	50
## 2608	-3.6111	24	45

## 2609	-7.6923	13	13
## 2610	2.4346	31	53
## 2611	1.9048	45	140
## 2612	16.6667	13	18
## 2613	4.2411	89	173
## 2614	-2.7175	55	93
## 2700	11.2986	260	978
## 2701	-1.9231	78	204
## 2702	2.1008	34	28
## 2703	2.0833	96	144
## 2704	1.2500	16	35
## 2705	6.3098	229	405
## 2706	21.5296	78	87
## 2707	10.9195	6	29
## 2708	22.7586	29	75
## 2709	NaN	0	1
## 2710	-30.4762	7	15
## 2711	18.6813	14	91
## 2712	13.7850	72	214
## 2713	21.4559	54	174
## 2714	8.4211	38	90
## 2715	8.3859	52	122
## 2716	-3.1863	56	102
## 2717	8.0733	71	153
## 2718	48.3516	4	91
## 2719	19.6970	27	330
## 2720	1.3298	48	141
## 2721	7.3485	33	40
## 2722	5.5556	24	36
## 2723	9.3734	117	151
## 2724	24.9538	51	106
## 2725	5.1901	72	247
## 2726	3.3333	60	120
## 2727	10.3579	41	77
## 2728	6.9378	242	418
## 2729	14.4627	47	158
## 2730	4.8395	111	398
## 2731	6.7538	76	143
## 2732	1.1761	113	310
## 2733	-4.9074	40	108
## 2734	-1.5560	121	162
## 2735	19.7467	148	382
## 2736	8.1062	81	212
## 2737	6.1850	186	242
## 2738	-0.6425	263	651
## 2739	8.5839	224	772
## 2740	13.7321	147	188
## 2741	1.2666	131	361
## 2742	5.1841	61	228
## 2743	6.2112	7	46

## 2744	NaN	0	0
## 2745	16.5442	41	83
## 2746	5.7048	174	407
## 2747	14.7727	24	55
## 2748	-0.6879	31	136
## 2800	5.4929	199	275
## 2801	-81.5789	1	38
## 2802	10.5543	103	124
## 2803	6.6327	36	98
## 2804	4.0397	78	159
## 2805	-8.7302	63	126
## 2806	25.7267	33	49
## 2807	-20.0000	5	20
## 2808	11.5065	97	248
## 2809	-0.7667	55	83
## 2900	-14.9737	75	152
## 2901	27.7778	3	18
## 2902	16.7438	16	81
## 2903	-66.6667	1	6
## 2904	NaN	0	4
## 2905	10.0840	17	49
## 2906	-26.0870	1	23
## 2907	5.1282	3	13
## 2908	NaN	0	7
## 2909	7.7465	24	71
## 2910	-9.0909	6	66
## 2911	25.1208	9	23
## 2912	-15.0000	4	10
## 2913	-23.5294	2	17
## 2914	-20.0000	2	10
## 2915	NaN	0	1
## 2916	1.8858	51	157
## 2917	-48.6486	3	37
## 2918	NA	NA	NA
## 2919	-29.0323	1	31
## 2920	NaN	0	0
## 2921	-46.3455	14	43
## 2922	NaN	0	16
## 2923	NaN	0	7
## 3000	6.3158	15	19
## 3001	33.3333	1	3
## 3002	5.8646	74	103
## 3003	12.1324	48	102
## 3004	1.8053	220	283
## 3005	7.2503	76	98
## 3100	-7.8505	10	214
## 3101	-4.2735	26	54
## 3102	4.9989	59	79
## 3103	7.7018	53	61
## 3104	9.8257	144	376

## 3105	9.2539	34	82
## 3106	0.5229	45	85
## 3107	2.6639	67	237
## 3108	8.6854	18	71
## 3109	0.6723	35	85
## 3110	6.9300	22	61
## 3200	14.4402	155	272
## 3201	-10.4735	34	41
## 3202	19.2727	77	175
## 3203	11.6015	123	278
## 3204	14.9369	181	335
## 3205	5.6770	199	259
## 3206	10.8639	67	104
## 3207	18.2491	89	240
## 3300	21.5957	47	160
## 3301	29.0008	85	209
## 3302	-7.0028	14	51
## 3303	17.0370	85	108
## 3304	13.2672	232	777
## 3305	16.1319	174	432
## 3306	16.0481	73	344
## 3307	-17.9138	18	49
## 3308	12.2232	83	303
## 3309	4.2828	76	153
## 3310	-0.5662	113	236
## 3311	-2.8571	7	35
## 3312	11.1400	275	685
## 3313	-0.2457	11	37
## 3314	9.7561	41	164
## 3315	-12.4807	22	118
## 3316	14.7787	39	292
## 3317	19.5161	20	62
## 3318	23.0134	26	121
## 3319	33.0586	21	52
## 3320	4.6296	45	216
## 3321	22.7053	46	81
## 3322	-14.5833	24	80
## 3400	3.0812	126	153
## 3401	NaN	0	1
## 3402	-25.0000	4	2
## 3403	0.0000	3	21
## 3404	-14.2857	3	7
## 3500	21.3333	25	30
## 3501	NaN	0	3
## 3502	NA	NA	NA
## 3503	NaN	0	0
## 3504	4.3269	16	13
## 3505	NaN	0	3
## 3506	0.0000	3	1
## 3600	17.5000	16	5

```
## 3601 12.4224      7      23
## 3602      NaN      0      17
## 3603 -16.6667      4       3
## 3604      NaN      0       0
## 3605 75.0000      1       8
## 3606      NaN      0       0
## 3607 -50.0000      9      12
## 3608      NaN      0       0
## 3609 16.4751     29      36
## 3610 -44.2857      5      14
## 3611 -3.3333      9      30
## 3612 17.4996     79     251
## 3613  0.0000      1       1
## 3614  5.7778     18      75
## 3615      NA      NA      NA
## 3616 -6.8934     17      64
```

```
print("GenderTeamSize in 2018")
```

```
## [1] "GenderTeamSize in 2018"
```

```
print(GenderTeamSize)
```

```
##      Articles FirstF FirstM   FirstP LastF LastM   LastP
## 1000      260  4.371  4.151 4.901e-01  3.944  4.327 4.012e-01
## 1100      431  4.563  3.992 2.018e-02  4.366  4.174 4.950e-01
## 1101       25  2.641  2.495 6.773e-01  2.244  2.764 4.188e-01
## 1102      149  2.951  3.148 5.263e-01  2.701  3.233 5.633e-02
## 1103      310  3.356  2.780 2.858e-02  3.240  3.023 5.704e-01
## 1104      315  3.336  3.571 2.417e-01  3.320  3.506 2.701e-01
## 1105      747  3.078  2.922 2.602e-01  2.879  3.028 1.953e-01
## 1106      198  3.880  3.612 6.796e-01  3.435  4.058 3.131e-02
## 1107      160  3.576  3.001 1.406e-02  3.585  3.051 6.068e-02
## 1108       25  3.663  3.190 5.030e-01  3.366  3.236 9.712e-01
## 1109      106  2.863  2.766 8.505e-01  2.620  2.894 4.081e-01
## 1110      181  3.357  3.345 9.929e-01  3.433  3.325 5.557e-01
## 1111       82  3.373  3.463 8.087e-01  3.512  3.409 6.370e-01
## 1200       27  1.581  1.292 5.048e-01  1.607  1.249 2.965e-01
## 1201      330  2.042  1.838 1.246e-01  1.977  1.903 7.473e-01
## 1202      296  1.174  1.117 2.101e-01  1.142  1.143 9.277e-01
## 1203      184  1.622  1.617 9.626e-01  1.682  1.542 3.508e-01
## 1204       64  2.087  1.503 4.972e-02  1.894  1.623 4.296e-01
## 1205       10  1.000  1.000      NaN  1.000  1.000      NaN
## 1206       11  1.170  1.189 8.885e-01  1.104  1.316 6.742e-01
## 1207      114  1.757  1.447 5.536e-02  2.015  1.254 3.439e-05
## 1208      159  1.074  1.000 5.717e-03  1.058  1.018 1.625e-01
## 1209        9  1.000  1.000      NaN  1.000  1.000      NaN
## 1210       35  1.196  1.425 1.953e-01  1.196  1.441 3.282e-01
## 1211      278  1.207  1.073 1.278e-02  1.166  1.094 1.601e-01
## 1212      145  1.104  1.054 2.241e-01  1.075  1.065 9.923e-01
## 1213      139  1.166  1.082 7.600e-02  1.173  1.070 2.869e-02
```

## 1300	524	4.821	4.193	1.825e-02	4.541	4.414	6.593e-01
## 1301	9	3.162	3.513	1.000e+00	NA	NA	NA
## 1302	30	4.338	4.488	9.292e-01	3.600	6.175	1.034e-02
## 1303	431	4.331	3.833	1.945e-02	3.808	4.034	4.085e-01
## 1304	216	3.852	3.480	9.894e-02	3.300	3.735	4.969e-02
## 1305	177	4.294	3.586	1.298e-01	4.372	3.742	2.020e-01
## 1306	212	6.319	6.291	9.901e-01	6.328	6.293	8.227e-01
## 1307	303	4.338	4.116	5.901e-01	3.961	4.278	2.736e-01
## 1308	99	4.760	5.496	3.481e-01	4.683	5.221	8.488e-01
## 1309	116	3.869	3.892	8.867e-01	3.665	3.969	3.099e-01
## 1310	167	4.476	4.510	7.142e-01	4.254	4.604	4.024e-01
## 1311	412	4.474	3.925	4.126e-02	4.057	4.197	7.062e-01
## 1312	548	4.160	4.042	5.885e-01	3.803	4.177	1.288e-01
## 1313	166	5.483	5.134	4.692e-01	5.039	5.339	5.653e-01
## 1314	378	3.979	3.615	2.524e-01	3.613	3.843	1.925e-01
## 1315	49	3.464	3.549	9.016e-01	2.813	3.808	2.070e-01
## 1400	84	2.067	1.935	6.533e-01	1.664	2.114	8.648e-02
## 1401	35	2.246	2.119	8.103e-01	2.661	1.752	4.698e-02
## 1402	55	1.895	1.803	8.227e-01	1.849	1.825	8.857e-01
## 1403	140	2.172	1.903	1.435e-01	2.024	1.995	9.447e-01
## 1404	14	4.217	2.022	3.844e-02	2.520	2.326	1.000e+00
## 1405	92	2.071	1.956	6.204e-01	1.912	2.043	4.688e-01
## 1406	73	2.115	1.967	4.868e-01	1.816	2.153	1.868e-01
## 1407	82	2.049	1.789	2.901e-01	1.944	1.935	7.791e-01
## 1408	190	1.839	2.054	1.864e-01	2.053	1.947	5.533e-01
## 1409	58	2.357	1.919	1.424e-01	2.398	1.918	1.229e-01
## 1410	19	1.636	1.320	6.103e-01	1.423	1.496	9.262e-01
## 1500	238	3.596	3.040	8.520e-03	3.428	3.170	5.439e-01
## 1501	11	3.130	3.557	5.560e-01	4.472	3.194	1.782e-01
## 1502	114	4.207	3.934	5.551e-01	4.821	3.840	6.357e-02
## 1503	161	3.796	3.563	2.476e-01	4.335	3.486	1.803e-02
## 1504	4	3.476	4.000	1.000e+00	2.828	4.583	6.667e-01
## 1505	41	4.390	3.499	1.545e-01	3.590	3.746	7.507e-01
## 1506	12	3.366	2.246	2.645e-01	NA	NA	NA
## 1507	66	2.380	2.646	4.585e-01	2.221	2.652	5.537e-01
## 1508	39	3.149	3.217	5.600e-01	2.993	3.232	7.116e-01
## 1600	492	3.950	3.486	2.320e-02	3.963	3.577	6.989e-02
## 1601	17	4.077	2.847	1.868e-01	3.949	3.252	4.217e-01
## 1602	135	3.275	3.428	4.609e-01	3.004	3.533	4.280e-02
## 1603	38	3.185	3.806	2.026e-01	3.124	3.811	1.650e-01
## 1604	115	3.834	3.242	1.453e-01	4.130	3.211	1.982e-02
## 1605	239	3.795	3.455	1.340e-01	4.163	3.438	1.516e-02
## 1606	299	3.727	2.804	5.350e-05	3.623	2.879	4.319e-03
## 1607	112	3.803	3.631	5.315e-01	3.563	3.782	4.704e-01
## 1700	112	2.700	2.569	6.143e-01	2.840	2.554	3.459e-01
## 1701	8	2.000	1.698	8.609e-01	1.000	1.919	3.588e-01
## 1702	14	2.000	2.199	9.218e-01	3.162	2.038	3.260e-01
## 1703	64	2.508	2.159	3.009e-01	1.587	2.305	8.619e-02
## 1704	40	2.913	3.208	6.279e-01	2.449	3.221	4.270e-01
## 1705	47	2.622	2.548	8.348e-01	2.449	2.580	6.510e-01

## 1706	186	2.597	2.600	5.145e-01	2.142	2.741	1.146e-02
## 1707	32	2.587	2.873	4.444e-01	2.491	2.848	5.161e-01
## 1708	48	3.238	2.611	2.736e-01	2.862	2.662	8.208e-01
## 1709	62	2.096	2.162	9.755e-01	2.165	2.124	9.101e-01
## 1710	105	2.286	2.388	6.804e-01	1.941	2.508	1.454e-02
## 1711	68	2.728	2.472	6.077e-01	2.749	2.490	6.759e-01
## 1712	130	2.876	2.689	7.113e-01	2.618	2.729	8.188e-01
## 1800	27	2.641	1.977	2.084e-01	2.392	2.168	7.957e-01
## 1801	NA	NA	NA	NA	NA	NA	NA
## 1802	39	3.174	1.697	1.551e-03	2.731	1.907	1.229e-01
## 1803	101	2.162	2.171	7.782e-01	1.971	2.230	4.191e-01
## 1804	51	1.966	1.608	2.193e-01	1.622	1.682	8.900e-01
## 1900	136	3.047	2.768	5.209e-01	2.690	2.862	5.598e-01
## 1901	54	2.993	2.761	6.015e-01	3.510	2.728	2.061e-01
## 1902	127	2.848	2.983	9.735e-01	2.843	2.961	9.120e-01
## 1903	41	2.244	2.852	4.794e-01	1.986	2.930	9.623e-02
## 1904	146	2.287	2.377	5.064e-01	2.190	2.398	5.118e-01
## 1905	23	2.965	2.328	2.715e-01	2.632	2.481	5.315e-01
## 1906	97	2.938	2.970	8.915e-01	2.963	2.958	8.865e-01
## 1907	136	2.640	2.790	7.046e-01	2.713	2.758	8.694e-01
## 1908	77	2.679	2.970	2.377e-01	2.752	2.907	6.721e-01
## 1909	121	2.647	2.833	6.274e-01	2.466	2.840	5.046e-01
## 1910	81	3.315	3.024	5.730e-01	3.412	3.051	4.153e-01
## 1911	42	2.389	2.488	8.747e-01	2.595	2.401	6.948e-01
## 1912	58	2.028	2.529	8.327e-02	1.969	2.476	2.126e-01
## 1913	15	4.380	3.137	3.791e-01	5.192	3.006	1.069e-01
## 2000	40	1.597	1.412	4.879e-01	1.364	1.480	7.145e-01
## 2001	33	2.782	1.495	1.111e-02	3.175	1.636	1.047e-01
## 2002	315	1.806	1.557	2.521e-02	1.759	1.574	2.044e-01
## 2003	121	1.888	1.515	4.321e-02	1.826	1.539	1.630e-01
## 2100	76	3.783	2.492	1.916e-02	3.677	2.553	7.734e-02
## 2101	NA	NA	NA	NA	NA	NA	NA
## 2102	164	3.239	2.745	3.373e-02	3.329	2.770	7.952e-02
## 2103	124	3.049	2.859	3.154e-01	2.554	2.957	2.079e-01
## 2104	46	3.302	2.447	8.033e-02	2.119	2.691	2.465e-01
## 2105	227	3.047	2.802	2.144e-01	2.628	2.898	4.538e-01
## 2200	55	2.683	2.709	9.425e-01	4.036	2.573	5.780e-02
## 2201	40	2.024	2.892	1.948e-01	1.537	3.097	5.035e-03
## 2202	60	2.000	2.365	6.313e-01	3.464	2.328	2.147e-01
## 2203	44	2.208	2.697	1.379e-01	2.632	2.611	9.823e-01
## 2204	228	4.031	4.006	5.589e-01	3.693	4.104	4.911e-01
## 2205	301	2.770	2.557	2.475e-01	2.528	2.603	5.534e-01
## 2206	24	2.000	3.006	2.920e-01	3.000	2.954	1.000e+00
## 2207	58	2.823	2.645	9.905e-01	2.530	2.698	4.408e-01
## 2208	184	3.126	2.821	8.588e-02	2.783	2.863	7.841e-01
## 2209	227	2.731	2.861	5.500e-01	2.705	2.860	5.898e-01
## 2210	373	2.824	2.705	6.024e-01	2.920	2.697	3.050e-01
## 2211	235	3.171	2.750	9.920e-02	2.934	2.786	6.019e-01
## 2212	43	3.105	1.941	3.615e-02	2.804	2.024	2.252e-01
## 2213	111	3.098	2.593	2.040e-01	2.624	2.747	8.654e-01

## 2214	24	1.542	1.993	1.833e-01	1.542	1.993	1.833e-01
## 2215	165	2.691	2.604	7.029e-01	2.815	2.600	5.039e-01
## 2216	36	2.087	1.683	3.826e-01	1.782	1.762	1.000e+00
## 2300	229	3.204	2.906	2.250e-01	3.500	2.826	2.356e-02
## 2301	78	2.264	1.660	2.514e-02	2.198	1.748	1.358e-01
## 2302	57	3.061	3.289	8.453e-01	2.252	3.426	2.169e-01
## 2303	422	3.143	3.046	8.023e-01	2.952	3.131	2.691e-01
## 2304	287	4.121	3.639	4.553e-02	3.724	3.904	2.745e-01
## 2305	165	4.173	3.381	1.473e-02	4.112	3.540	3.724e-01
## 2306	122	2.816	2.559	3.626e-01	2.584	2.680	9.713e-01
## 2307	193	4.338	3.883	1.628e-01	4.188	4.068	9.358e-01
## 2308	292	2.535	2.519	8.315e-01	2.545	2.517	9.887e-01
## 2309	159	3.036	2.992	9.604e-01	3.069	2.986	9.970e-01
## 2310	260	4.005	3.577	7.401e-02	3.894	3.706	5.482e-01
## 2311	153	4.440	3.473	3.304e-03	4.166	3.712	4.618e-01
## 2312	234	3.397	2.938	3.387e-02	3.203	3.074	6.364e-01
## 2400	95	4.648	3.570	2.223e-02	4.061	3.846	8.000e-01
## 2401	6	5.313	5.944	1.000e+00	3.807	12.247	1.333e-01
## 2402	52	3.756	3.855	3.141e-01	3.970	3.771	7.661e-01
## 2403	240	5.360	4.876	5.875e-02	5.109	5.145	8.076e-01
## 2404	156	4.545	4.238	6.271e-01	4.088	4.487	3.426e-01
## 2405	35	5.170	4.380	2.823e-01	4.406	4.723	9.833e-01
## 2406	79	4.599	5.420	3.468e-01	4.848	5.033	5.508e-01
## 2500	382	3.538	3.254	1.368e-01	3.135	3.329	4.019e-01
## 2501	13	4.899	3.094	3.149e-01	4.899	3.094	3.149e-01
## 2502	78	5.014	4.356	2.189e-01	4.769	4.437	4.910e-01
## 2503	72	3.293	2.917	3.076e-01	3.292	2.946	6.665e-01
## 2504	133	3.216	2.855	3.047e-01	3.702	2.820	4.711e-02
## 2505	95	3.261	3.247	9.350e-01	3.378	3.229	5.858e-01
## 2506	83	2.648	2.834	3.809e-01	2.608	2.813	5.408e-01
## 2507	100	3.166	3.280	3.977e-01	4.060	3.140	6.562e-02
## 2508	118	3.378	3.191	6.138e-01	3.560	3.155	3.919e-01
## 2600	139	1.682	1.393	5.605e-02	1.551	1.423	3.985e-01
## 2601	13	1.414	1.080	1.706e-01	1.260	1.149	7.290e-01
## 2602	56	1.104	1.304	2.753e-01	1.170	1.293	4.132e-01
## 2603	46	1.364	1.429	7.298e-01	1.122	1.468	1.073e-01
## 2604	237	1.863	1.937	5.530e-01	1.899	1.929	9.478e-01
## 2605	67	3.120	2.216	1.087e-01	2.449	2.292	7.851e-01
## 2606	52	1.817	2.128	6.495e-01	2.221	2.095	8.921e-01
## 2607	47	1.251	1.695	1.268e-01	1.658	1.593	8.230e-01
## 2608	39	1.292	1.496	3.892e-01	1.297	1.501	4.021e-01
## 2609	NA	NA	NA	NA	NA	NA	NA
## 2610	39	1.414	1.557	9.434e-01	1.000	1.567	4.282e-01
## 2611	136	2.543	2.294	7.377e-01	2.783	2.271	2.253e-01
## 2612	25	1.414	1.682	4.346e-01	1.260	1.695	2.461e-01
## 2613	134	2.196	1.880	1.778e-01	1.815	1.942	7.763e-01
## 2614	71	1.940	1.688	2.341e-01	1.861	1.715	6.757e-01
## 2700	402	3.914	2.960	4.469e-04	3.571	3.283	3.696e-01
## 2701	257	4.711	4.644	8.636e-01	4.703	4.669	8.971e-01
## 2702	63	4.520	2.760	1.612e-02	4.545	3.161	5.766e-02

## 2703	133	4.116	4.826	2.126e-01	4.067	4.768	4.166e-01
## 2704	22	5.614	5.662	1.000e+00	4.472	6.424	2.875e-01
## 2705	402	5.661	5.564	6.837e-01	5.268	5.741	3.134e-01
## 2706	114	5.497	4.661	1.470e-01	5.209	4.954	8.925e-01
## 2707	30	3.783	2.955	4.759e-01	3.273	3.391	7.482e-01
## 2708	107	3.667	3.525	9.157e-01	3.562	3.633	7.195e-01
## 2709	NA	NA	NA	NA	NA	NA	NA
## 2710	10	3.175	2.213	5.038e-01	2.520	2.852	8.117e-01
## 2711	162	4.266	4.683	2.972e-01	4.205	4.704	1.637e-01
## 2712	216	5.124	4.774	8.440e-01	4.898	5.025	5.179e-01
## 2713	184	4.592	4.245	4.944e-01	4.563	4.360	7.124e-01
## 2714	83	4.197	3.578	2.935e-01	3.672	4.483	1.139e-01
## 2715	74	6.215	5.886	6.191e-01	7.139	5.491	3.048e-02
## 2716	109	5.309	4.995	8.017e-01	4.640	5.401	2.483e-01
## 2717	179	4.452	4.513	8.759e-01	4.250	4.760	1.859e-01
## 2718	146	4.986	3.989	1.814e-02	4.756	4.337	2.448e-01
## 2719	418	3.823	2.807	3.202e-05	3.460	3.328	7.216e-01
## 2720	133	5.167	5.808	5.835e-01	4.560	5.946	1.316e-01
## 2721	28	7.936	6.886	5.767e-01	6.926	7.568	7.847e-01
## 2722	53	4.226	2.302	2.491e-03	3.363	2.824	6.404e-01
## 2723	189	5.481	4.845	7.712e-02	5.444	5.032	1.921e-01
## 2724	102	4.957	4.285	6.400e-01	4.953	4.516	4.430e-01
## 2725	221	5.761	5.400	4.329e-01	5.455	5.625	5.885e-01
## 2726	108	5.710	4.933	4.033e-01	5.023	5.375	3.910e-01
## 2727	68	5.151	4.832	7.810e-01	5.596	4.667	2.678e-01
## 2728	547	5.003	4.467	1.329e-02	4.955	4.651	1.785e-01
## 2729	251	3.734	3.596	6.952e-01	3.701	3.672	9.958e-01
## 2730	409	6.041	6.063	6.388e-01	6.120	6.016	7.088e-01
## 2731	129	3.806	3.422	3.860e-01	3.587	3.591	7.609e-01
## 2732	384	3.873	3.929	7.794e-01	3.535	4.052	1.877e-02
## 2733	103	4.417	4.393	8.491e-01	4.076	4.530	3.620e-01
## 2734	159	3.448	2.843	8.274e-02	3.252	3.011	3.633e-01
## 2735	482	4.085	3.999	7.161e-01	3.778	4.337	4.391e-02
## 2736	191	5.088	4.166	1.719e-02	4.678	4.608	6.868e-01
## 2737	236	4.384	4.771	2.331e-01	4.617	4.585	7.057e-01
## 2738	686	4.124	4.308	7.142e-02	4.055	4.314	6.647e-02
## 2739	877	4.070	3.742	3.888e-02	3.828	4.083	2.589e-01
## 2740	204	6.587	5.279	4.357e-02	6.497	5.606	2.750e-01
## 2741	368	5.179	4.819	1.283e-01	4.793	5.020	5.732e-01
## 2742	252	3.780	4.155	3.678e-01	3.794	4.047	7.330e-01
## 2743	64	4.632	4.581	6.367e-01	4.825	4.476	7.330e-01
## 2744	NA	NA	NA	NA	NA	NA	NA
## 2745	68	5.919	5.222	1.971e-01	5.121	6.088	1.907e-01
## 2746	536	4.743	4.594	8.916e-01	4.541	4.693	5.636e-01
## 2747	61	6.674	6.365	9.365e-01	6.892	6.371	8.739e-01
## 2748	154	5.326	6.055	1.398e-01	4.757	6.108	6.891e-03
## 2800	330	3.866	3.672	5.184e-01	3.621	3.847	3.691e-01
## 2801	51	4.434	4.565	9.692e-01	3.723	4.751	3.413e-01
## 2802	143	3.980	3.542	3.121e-01	4.036	3.596	2.675e-01
## 2803	92	4.890	5.145	9.731e-01	4.842	5.065	8.162e-01

## 2804	143	4.628	3.721	1.408e-02	4.349	4.067	5.108e-01
## 2805	181	3.956	3.407	9.082e-02	3.607	3.759	5.614e-01
## 2806	62	4.484	2.869	1.085e-02	4.211	3.599	4.898e-01
## 2807	24	6.016	2.995	5.359e-03	5.151	4.637	9.070e-01
## 2808	321	4.957	4.623	2.202e-01	4.996	4.722	2.879e-01
## 2809	79	4.072	3.432	2.413e-01	4.468	3.495	1.547e-01
## 2900	145	3.233	2.392	2.546e-02	3.055	3.036	9.748e-01
## 2901	27	3.303	2.908	6.953e-01	2.885	3.803	3.098e-01
## 2902	72	2.833	3.841	1.143e-01	2.687	3.944	4.604e-02
## 2903	11	2.551	1.674	1.053e-01	3.308	1.348	4.543e-02
## 2904	11	4.376	2.289	3.505e-01	4.293	2.783	5.645e-01
## 2905	46	3.648	3.397	6.532e-01	3.481	3.975	4.703e-01
## 2906	9	2.289	3.004	6.924e-01	3.286	1.913	5.978e-01
## 2907	20	4.036	4.821	2.798e-01	3.659	5.028	5.586e-02
## 2908	15	2.421	1.431	1.290e-01	2.328	1.739	4.028e-01
## 2909	82	3.631	3.087	3.972e-01	3.501	3.456	8.553e-01
## 2910	52	1.642	2.397	3.134e-02	1.969	1.718	5.663e-01
## 2911	25	2.562	2.854	6.491e-01	2.635	2.605	9.723e-01
## 2912	18	3.113	1.698	9.414e-02	2.961	2.104	4.961e-01
## 2913	15	2.919	4.000	3.992e-01	3.073	2.449	3.768e-01
## 2914	NA	NA	NA	NA	NA	NA	NA
## 2915	NA	NA	NA	NA	NA	NA	NA
## 2916	184	4.748	4.396	6.072e-01	4.243	5.029	3.774e-02
## 2917	20	4.113	3.162	6.097e-01	3.147	4.881	4.647e-01
## 2918	NA	NA	NA	NA	NA	NA	NA
## 2919	20	2.344	3.130	4.355e-01	2.329	2.884	4.696e-01
## 2920	NA	NA	NA	NA	NA	NA	NA
## 2921	40	3.277	4.417	2.172e-01	3.414	3.703	8.487e-01
## 2922	13	1.364	1.783	3.752e-01	1.534	1.442	9.258e-01
## 2923	7	2.994	2.000	6.139e-01	2.667	4.000	8.008e-01
## 3000	20	2.706	2.611	5.863e-01	2.763	2.548	9.691e-01
## 3001	14	6.790	4.579	2.969e-01	7.200	5.236	3.914e-01
## 3002	95	4.387	4.778	5.642e-01	3.898	4.786	2.260e-01
## 3003	103	4.480	3.898	5.720e-02	4.282	4.060	3.260e-01
## 3004	214	4.584	4.372	6.214e-01	4.389	4.504	7.471e-01
## 3005	125	4.237	3.589	1.142e-01	3.865	3.928	7.390e-01
## 3100	222	3.797	2.484	3.216e-03	3.004	2.613	2.870e-01
## 3101	64	2.449	2.238	4.471e-01	2.000	2.275	5.950e-01
## 3102	45	2.792	2.697	6.922e-01	2.034	2.867	4.971e-02
## 3103	41	1.762	1.956	5.221e-01	1.817	1.933	8.916e-01
## 3104	344	2.857	2.697	4.859e-01	2.688	2.726	8.143e-01
## 3105	150	3.410	3.097	4.398e-01	3.712	3.054	1.283e-01
## 3106	57	2.632	2.150	3.953e-01	2.702	2.137	7.454e-01
## 3107	208	3.320	3.070	4.458e-01	3.165	3.097	6.945e-01
## 3108	64	5.142	3.619	1.274e-01	4.768	3.855	4.276e-01
## 3109	46	1.817	1.860	9.252e-01	1.414	1.881	5.700e-01
## 3110	47	3.861	3.136	1.547e-01	2.794	3.537	8.989e-02
## 3200	316	2.771	2.389	2.502e-02	2.722	2.483	2.004e-01
## 3201	35	2.343	2.415	6.948e-01	2.600	2.102	3.309e-01
## 3202	192	3.010	2.179	4.225e-04	2.804	2.640	4.340e-01

## 3203	281	3.369	3.884	9.137e-03	3.327	3.765	2.964e-02
## 3204	385	3.220	2.771	9.809e-02	3.048	3.161	3.047e-01
## 3205	253	3.235	2.618	2.385e-02	3.030	2.868	9.576e-01
## 3206	107	3.962	3.809	8.571e-01	4.139	3.730	6.144e-01
## 3207	275	2.677	2.094	2.660e-03	2.430	2.445	9.814e-01
## 3300	163	1.841	1.407	2.485e-03	1.830	1.424	5.165e-03
## 3301	228	2.414	1.735	3.477e-04	2.341	1.878	1.582e-02
## 3302	61	1.644	1.540	6.008e-01	1.575	1.585	9.727e-01
## 3303	131	1.750	1.560	2.761e-01	1.554	1.711	3.563e-01
## 3304	762	2.205	1.947	1.021e-02	2.121	2.093	7.404e-01
## 3305	449	1.761	1.702	4.163e-01	1.710	1.737	4.939e-01
## 3306	364	2.906	2.666	3.412e-01	2.809	2.845	8.777e-01
## 3307	51	2.959	3.268	7.111e-01	2.566	3.568	5.242e-02
## 3308	323	1.823	1.594	2.214e-02	1.893	1.536	2.935e-03
## 3309	110	1.646	1.939	1.697e-01	1.633	1.946	1.249e-01
## 3310	214	1.646	1.678	6.809e-01	1.690	1.613	7.176e-01
## 3311	44	2.769	2.576	7.346e-01	2.307	2.853	3.432e-01
## 3312	760	1.668	1.449	5.972e-04	1.629	1.490	2.814e-02
## 3313	79	2.668	2.359	5.706e-01	2.936	2.361	2.807e-01
## 3314	203	1.390	1.211	3.217e-02	1.309	1.326	9.791e-01
## 3315	135	1.781	1.329	1.183e-03	1.693	1.430	5.964e-02
## 3316	355	1.251	1.186	1.717e-01	1.259	1.176	1.028e-01
## 3317	107	1.749	1.340	1.496e-02	1.551	1.622	9.145e-01
## 3318	116	1.750	2.000	6.595e-01	1.761	1.950	7.158e-01
## 3319	69	2.824	2.840	7.306e-01	2.728	2.959	4.735e-01
## 3320	222	1.162	1.157	8.502e-01	1.150	1.164	6.003e-01
## 3321	116	1.833	1.592	1.973e-01	1.806	1.618	4.130e-01
## 3322	75	1.721	1.581	3.282e-01	1.495	1.699	3.944e-01
## 3400	166	3.784	3.370	2.923e-01	3.247	3.976	1.512e-02
## 3401	3	4.472	7.000	6.667e-01	5.292	5.000	1.000e+00
## 3402	4	5.944	6.000	1.000e+00	NA	NA	NA
## 3403	33	4.481	5.555	4.559e-02	4.175	5.230	1.035e-01
## 3404	9	4.702	3.302	8.960e-01	3.780	4.394	6.010e-01
## 3500	25	3.258	3.104	5.743e-01	2.884	3.512	3.243e-01
## 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	11	3.302	3.154	7.813e-01	2.289	3.681	6.049e-01
## 3505	6	2.449	3.201	6.386e-01	3.476	2.466	8.248e-01
## 3506	3	5.000	2.646	1.000e+00	5.000	2.646	1.000e+00
## 3600	10	2.821	1.682	3.253e-01	2.686	1.587	3.500e-01
## 3601	18	2.161	2.705	8.433e-01	2.430	2.140	8.048e-01
## 3602	13	4.095	3.076	7.121e-01	3.634	3.376	1.000e+00
## 3603	NA	NA	NA	NA	NA	NA	NA
## 3604	7	6.633	7.090	8.451e-01	10.302	4.121	2.118e-01
## 3605	21	3.807	3.687	9.664e-01	3.915	3.638	8.428e-01
## 3606	NA	NA	NA	NA	NA	NA	NA
## 3607	13	2.034	2.749	4.651e-01	2.766	2.104	6.533e-01
## 3608	NA	NA	NA	NA	NA	NA	NA
## 3609	NA	NA	NA	NA	NA	NA	NA


```
## 3610      10  4.031  4.820 4.642e-01  3.722  4.648 2.538e-01
## 3611      55  4.742  3.341 7.621e-03  4.265  4.240 6.331e-01
## 3612     298  3.490  3.685 4.618e-01  3.382  3.747 2.627e-01
## 3613      NA      NA      NA      NA      NA      NA      NA
## 3614      78  4.742  4.173 2.820e-01  4.603  4.296 6.442e-01
## 3615      NA      NA      NA      NA      NA      NA      NA
## 3616      59  3.219  2.876 3.426e-01  3.166  3.061 8.718e-01
```

```
print(RegCoef)
```

```
##          FFA1          FLA1          2          3          4          5+
## 1000 -0.0544622 -0.0600169  0.9263875  0.905906  1.101689  1.069707
## 1100 -0.0213526  0.0039657  0.1694596  0.142881  0.218343  0.258535
## 1101  0.0434481 -0.0711997  0.3446914  0.229586  0.414163  0.301753
## 1102  0.0752420  0.0017017  0.1372358  0.182617  0.244975  0.288072
## 1103  0.0057876 -0.0102997  0.1146182  0.120675  0.203867  0.180128
## 1104 -0.0069111 -0.0054626  0.1198487  0.109436  0.168323  0.215220
## 1105  0.0168187 -0.0237885  0.1351490  0.148310  0.187299  0.240120
## 1106  0.0398604 -0.0026127  0.4469824  0.455695  0.460740  0.467389
## 1107 -0.0355588 -0.0065392  0.1815001  0.225546  0.263949  0.339733
## 1108  0.0955990  0.0295515 -0.0308670  0.033551  0.017572 -0.054893
## 1109  0.0960223 -0.0134520  0.2909296  0.288757  0.376282  0.490163
## 1110  0.0562777  0.0229171  0.1895009  0.206040  0.256490  0.342433
## 1111 -0.0076584 -0.0355751  0.1504794  0.127336  0.202625  0.285673
## 1200  0.1682733  0.0817420  0.2903260 -0.125934  1.093386  1.123597
## 1201  0.0139239  0.0132673  0.2110651  0.254595  0.211896  0.306448
## 1202  0.1035959 -0.0293018  0.1832460  0.578785  0.845017  0.945127
## 1203 -0.0185411  0.0673981  0.4692103  0.505344  0.578441  0.525990
## 1204  0.1700028 -0.0872801  0.1191221  0.443873  0.514417  0.336463
## 1205  0.2501177 -0.1231255  0.6273279 -0.193977 -0.459573 -0.352401
## 1206 -0.5389165  0.6689816  0.4063011  1.555362 -0.395123  0.624502
## 1207  0.1338966  0.0890343  0.2916608  0.490547  0.469738  0.540005
## 1208 -0.0263491  0.1604988  0.1496888  0.902978  2.514638  0.853008
## 1209 -0.2317596 -0.0934081  0.5321586  0.677219  0.531001  0.915903
## 1210 -0.2126326  0.0394795  0.8745647  1.336776  1.929714  1.534017
## 1211 -0.1085089  0.1466169  0.1978192  0.133850  0.508801  0.341132
## 1212 -0.0158866  0.0937419  0.1696086  0.191962 -0.581814 -0.294661
## 1213  0.3718791 -0.2076817 -0.1431129  0.630887  0.766856 -0.009292
## 1300 -0.0513072 -0.0118569  0.0817973  0.010983  0.135115  0.226482
## 1301      NA      NA      NA      NA      NA      NA
## 1302  0.0438648 -0.0439451  0.4938005  0.664794  0.748047  0.804913
## 1303 -0.0148063 -0.0105427  0.1889859  0.209607  0.235889  0.277241
## 1304  0.0008831  0.0223581  0.1744340  0.223722  0.213977  0.238959
## 1305  0.0065259  0.0596061  0.3385737  0.355295  0.405476  0.476062
## 1306 -0.0222477 -0.0036933  0.3473683  0.367811  0.426304  0.536764
## 1307 -0.0126082 -0.0072042  0.2132661  0.243486  0.271069  0.349559
## 1308  0.0074012  0.0472992  0.2465697  0.244371  0.259812  0.330763
## 1309 -0.0113239  0.0263763  0.0859505  0.081825  0.095474  0.165359
## 1310 -0.0136066 -0.0223743  0.2182190  0.223395  0.227679  0.286234
## 1311 -0.0138191 -0.0157776  0.0885114  0.118786  0.110126  0.180441
```

## 1312	0.0008066	-0.0000411	0.1962439	0.230099	0.247461	0.331906
## 1313	-0.0145977	0.0459117	0.3187391	0.341450	0.394521	0.437247
## 1314	0.0022279	-0.0194155	0.1454089	0.172266	0.173653	0.236483
## 1315	0.0578486	-0.0138620	0.1309913	0.125702	0.190013	0.335517
## 1400	0.0495046	-0.0262025	0.2115714	0.372947	0.330955	0.145625
## 1401	0.0051412	-0.0508030	0.2450669	0.293287	0.300703	-0.223160
## 1402	0.0174098	-0.0559269	0.1748683	0.281853	0.114262	0.066749
## 1403	-0.0319679	0.0920007	0.2612441	0.220335	0.165635	0.299508
## 1404	0.3098446	-0.0328438	0.2580104	0.363706	0.434773	0.269901
## 1405	-0.0026697	0.0076202	0.0469771	0.023792	0.135488	0.148651
## 1406	-0.0419416	0.0682932	0.1444458	0.256594	0.200214	-0.046929
## 1407	0.0441182	0.0362304	0.0979087	0.109350	0.233885	0.051543
## 1408	-0.0346686	0.0428775	0.1621978	0.156211	0.179496	0.340358
## 1409	-0.0172573	0.0268443	0.0616901	0.175075	0.394165	0.215510
## 1410	-0.0414962	0.1370651	0.4116375	0.331816	0.351364	0.299522
## 1500	-0.0548435	-0.0250087	0.3231035	0.357192	0.430941	0.430120
## 1501	0.2037031	-0.2573061	0.4908983	0.662000	0.664862	0.127127
## 1502	-0.0273936	0.0611039	0.3509061	0.339583	0.449606	0.428898
## 1503	-0.0540432	-0.0117853	0.1376892	0.158942	0.208142	0.206572
## 1504	NA	NA	NA	NA	NA	NA
## 1505	-0.0745625	0.0043349	0.1884864	0.153159	0.238137	0.230298
## 1506	0.1069920	-0.1011819	-0.1539335	-0.116135	-0.050564	0.071020
## 1507	-0.1092220	0.1203286	0.5421097	0.557399	0.752804	0.786208
## 1508	0.0949376	0.0406416	0.5033249	0.577998	0.586191	0.636640
## 1600	-0.0114431	0.0165100	0.2308399	0.229852	0.283970	0.329097
## 1601	0.1076414	0.0458010	0.2298087	0.161456	0.437795	0.497281
## 1602	-0.0171529	-0.0350805	0.3261414	0.360795	0.361491	0.401874
## 1603	-0.0322791	0.0204174	0.3241060	0.362209	0.418828	0.483179
## 1604	0.0067792	0.0046797	0.0409179	0.075241	0.096053	0.137959
## 1605	-0.0434432	-0.0102009	0.0531343	0.082934	0.111511	0.106620
## 1606	0.0088778	-0.0102726	0.1125917	0.141415	0.170893	0.192522
## 1607	-0.0240059	0.0114850	-0.0072073	0.035739	0.043603	0.066381
## 1700	-0.0797775	-0.0211598	0.0930568	0.431274	0.307621	0.391118
## 1701	0.0862527	-0.4483401	-0.1206257	0.151644	0.057406	0.262147
## 1702	0.0315960	-0.1041870	0.1178274	0.201046	0.231458	0.305670
## 1703	-0.0350263	-0.0900570	0.1903965	0.258429	0.315879	0.458076
## 1704	-0.0348796	0.0112390	0.1843536	0.257578	0.309234	0.109270
## 1705	0.0485792	-0.0729560	0.1836269	0.279969	0.140025	0.386810
## 1706	0.0170839	-0.0387627	0.1867331	0.243171	0.195966	0.329116
## 1707	-0.1176098	0.0967074	0.1310087	0.049636	0.255653	0.159395
## 1708	-0.0902143	0.0589742	0.1180414	0.270935	0.167526	0.293020
## 1709	0.0743094	0.0345287	0.1315818	0.372068	0.260002	0.300397
## 1710	0.0848570	0.0270400	0.2059745	0.230352	0.252062	0.274314
## 1711	-0.0025890	0.0897982	0.1393048	0.089862	0.314366	0.197678
## 1712	-0.0732241	0.0443494	0.1533747	0.244648	0.193650	0.299462
## 1800	0.0284270	0.1672745	0.0931567	0.200872	0.399449	0.257805
## 1801	NA	NA	NA	NA	NA	NA
## 1802	0.1026860	-0.0015998	0.2844373	0.390998	0.307921	-0.333457
## 1803	-0.0511192	0.0457473	0.2075350	0.363177	0.203044	0.443251
## 1804	-0.1663591	0.0830030	0.1633489	0.179381	0.088364	0.030834

## 1900	-0.0176437	0.0155278	0.2781628	0.316248	0.357848	0.454515
## 1901	-0.0710578	-0.1893644	0.2047197	0.188397	0.247459	0.339154
## 1902	0.0424339	0.0342156	-0.0234286	-0.002033	0.095921	0.185185
## 1903	0.0026530	0.0695306	0.1221148	0.208257	0.234364	0.300026
## 1904	0.0312636	-0.1196438	0.2176770	0.186635	0.171043	0.267975
## 1905	0.0119918	0.0080660	-0.1415292	-0.087680	0.017434	0.028226
## 1906	-0.0085210	0.0176706	-0.0316302	-0.008609	0.064192	0.057573
## 1907	-0.0220443	-0.0290852	0.0861976	0.136733	0.201842	0.227985
## 1908	0.0212803	-0.0365977	0.0339488	0.033430	0.139372	0.135611
## 1909	0.0541758	-0.0886538	0.3581955	0.303688	0.266381	0.222400
## 1910	0.0084929	0.0148491	0.1405392	0.166721	0.247002	0.265746
## 1911	-0.0021329	-0.0663778	0.0821424	0.123716	0.186628	0.260807
## 1912	0.0860323	-0.0833254	0.0602177	0.086074	0.133113	0.239576
## 1913	0.0269614	-0.1654217	0.0006127	0.142296	0.236423	0.303301
## 2000	0.1201712	0.0243180	0.1821039	0.199038	0.705947	0.079117
## 2001	-0.2243202	0.0808844	0.2315588	0.458653	0.706629	0.167188
## 2002	0.0632641	0.0123726	0.1468948	0.180710	0.150582	0.089944
## 2003	0.1689715	-0.1307545	0.1922278	0.230091	0.072861	0.220779
## 2100	NA	NA	NA	NA	NA	NA
## 2101	0.4666780	-0.4245987	0.7577199	0.851739	0.718743	0.958406
## 2102	-0.0492943	-0.0149495	0.4151361	0.553828	0.439856	0.520576
## 2103	-0.1333803	0.0757893	0.3885744	0.626706	0.452791	0.410542
## 2104	0.0780333	0.0301237	0.0063082	-0.059371	-0.051346	0.140564
## 2105	-0.0949962	0.0979993	0.2193826	0.343586	0.346061	0.301213
## 2200	0.0529053	0.1345957	0.2015431	0.358831	0.201218	0.370612
## 2201	0.1690904	-0.1647542	0.4222896	0.584305	0.640121	0.717554
## 2202	-0.0445792	0.1432166	0.3567924	0.437346	0.314952	0.641296
## 2203	-0.0337966	-0.0017341	0.6737418	0.702526	0.683068	0.517245
## 2204	-0.0208385	0.0219026	0.5375988	0.501040	0.575162	0.553785
## 2205	0.0050711	-0.0396572	0.1138034	0.129497	0.058246	0.125282
## 2206	-0.0645107	0.1634783	0.1498179	0.194132	0.147107	0.438030
## 2207	-0.0380017	0.0464011	0.1201463	0.253875	0.260241	0.249922
## 2208	0.0016370	-0.0006078	0.1791603	0.222215	0.168135	0.241471
## 2209	0.0202056	-0.0317849	0.3025437	0.301755	0.400855	0.322372
## 2210	0.0168578	-0.0634705	0.3243388	0.362472	0.351208	0.573599
## 2211	0.0647838	0.0718448	0.2381258	0.252001	0.268908	0.422642
## 2212	0.0610000	-0.2064894	0.3463282	0.545090	0.465442	0.525374
## 2213	0.0180853	0.0950820	0.2984293	0.389594	0.352919	0.516416
## 2214	0.2008389	0.1570245	0.4242927	0.464988	0.147695	0.404350
## 2215	-0.0746236	0.0011226	0.2781065	0.380622	0.341980	0.463249
## 2216	-0.0009215	0.0952232	0.1479996	0.452766	0.900013	1.081669
## 2300	0.0456822	0.0249778	0.1250105	0.116661	0.168663	0.233683
## 2301	0.1551818	0.0119885	0.0952753	0.016278	-0.104417	0.024907
## 2302	0.0119827	0.0494533	0.1788328	0.228769	0.186945	0.272506
## 2303	0.0111682	-0.0111707	0.2989906	0.309304	0.339243	0.430984
## 2304	-0.0079557	0.0105043	0.2184896	0.233758	0.218437	0.320189
## 2305	-0.0410071	-0.1197936	0.2954041	0.269112	0.260284	0.340366
## 2306	-0.0524505	-0.0792555	0.2138444	0.246689	0.243382	0.313657
## 2307	-0.0031737	0.0121398	0.2098336	0.202408	0.273356	0.313864
## 2308	-0.0101575	-0.0000641	0.1698043	0.217899	0.173925	0.273444

##	2309	-0.0175640	0.0654180	0.2774822	0.288117	0.311279	0.260307
##	2310	-0.0139915	0.0366539	0.1755593	0.235977	0.214524	0.298545
##	2311	-0.0159017	-0.0661409	0.3434015	0.374622	0.313476	0.433994
##	2312	-0.0332501	-0.0013187	0.2750864	0.287393	0.283267	0.329849
##	2400	-0.0326143	0.0305013	0.0915808	-0.005429	0.110626	0.098719
##	2401	NA	NA	NA	NA	NA	NA
##	2402	0.0490255	0.0385920	0.3375639	0.364863	0.383926	0.459952
##	2403	-0.0081289	0.0002471	0.1350908	0.166392	0.220465	0.295112
##	2404	0.0165288	-0.0109648	0.2180299	0.276188	0.274802	0.317025
##	2405	0.0733962	-0.0121164	0.1197721	0.209272	0.197623	0.379831
##	2406	-0.0273197	-0.0005909	0.0513844	0.043195	0.049556	0.154861
##	2500	0.0343236	-0.0018759	0.1615654	0.087503	0.212543	0.375506
##	2501	-0.0592150	-0.1381852	0.3403796	0.234879	0.356817	0.604088
##	2502	-0.0033477	0.0917315	0.1304944	0.140949	0.215759	0.184766
##	2503	-0.0163571	0.0562895	0.2184301	0.238040	0.255057	0.253211
##	2504	0.0147932	0.0162004	0.0658129	0.120301	0.137403	0.205092
##	2505	0.0045744	0.0174992	0.1739505	0.165000	0.216912	0.269391
##	2506	0.0298021	-0.0768831	0.5151085	0.530858	0.565397	0.639742
##	2507	-0.0310586	0.0014480	0.2419890	0.267039	0.296786	0.363303
##	2508	0.0211923	0.0098236	0.2103665	0.220972	0.293028	0.317180
##	2600	-0.0403226	-0.0149449	0.1404137	0.278216	0.443154	0.537282
##	2601	0.3548505	-0.4022240	0.1912872	0.618883	0.139694	0.078754
##	2602	0.0279262	-0.1044325	0.1974533	0.122425	0.202841	0.898192
##	2603	0.1379536	-0.1550829	0.0564901	0.123095	0.086544	-0.324874
##	2604	-0.0239689	0.0318674	0.1666799	0.235231	0.174358	0.297657
##	2605	-0.0092782	-0.0902104	0.0671366	0.134427	0.179833	0.376772
##	2606	-0.3076871	0.2201133	0.2360965	0.329934	0.373873	0.425716
##	2607	0.0787137	-0.1245345	-0.0172747	0.046133	0.037102	-0.218509
##	2608	-0.1633982	0.0477918	0.2005366	0.247464	0.275882	0.216709
##	2609	0.2500250	-0.3165790	0.3367669	0.430555	2.173717	-0.085669
##	2610	0.0359796	-0.1015309	0.1068089	0.088338	0.269156	0.210628
##	2611	0.1120068	-0.0752042	0.1549161	0.267195	0.219039	0.250018
##	2612	-0.1801466	0.2114748	0.2030913	0.305856	-0.167757	-0.101304
##	2613	-0.0308395	-0.0018220	0.1343431	0.182456	0.254402	0.375325
##	2614	-0.0112121	-0.0684177	0.0355404	0.089493	0.139257	0.156385
##	2700	0.0081213	0.0183150	0.5288680	0.686995	0.854172	1.202136
##	2701	-0.0115873	-0.0478013	0.1666336	0.158385	0.199331	0.281196
##	2702	-0.0639956	0.0643284	0.0805012	0.160766	0.193124	0.229706
##	2703	-0.0133560	0.1131394	0.3250071	0.436294	0.520173	0.585241
##	2704	0.0106475	0.0876990	0.3169279	0.445749	0.454999	0.583548
##	2705	0.0514829	0.0605412	0.2395299	0.337977	0.361819	0.594086
##	2706	0.0135058	0.0097676	0.4487487	0.681967	0.759578	0.859369
##	2707	0.0106713	0.0291052	0.0814282	0.349913	0.261157	0.423234
##	2708	0.0171268	0.0023177	0.3367333	0.264887	0.302975	0.434217
##	2709	NA	NA	NA	NA	NA	NA
##	2710	-0.0252362	-0.1064577	0.1967532	0.218201	0.200903	0.336656
##	2711	-0.0073657	0.0218726	0.4737095	0.702882	0.869559	1.033760
##	2712	-0.0384098	-0.0480003	0.3157442	0.328238	0.360923	0.478290
##	2713	0.0322015	0.0023373	0.1507192	0.197211	0.199721	0.227498
##	2714	0.0593060	0.1522808	0.2652779	0.289678	0.447121	0.673861

## 2715	0.0189428	0.0345256	0.1342070	0.243480	0.252985	0.408651
## 2716	-0.0022465	-0.0577019	0.1625952	0.161810	0.157625	0.238332
## 2717	0.0620220	0.0365309	0.2640241	0.389620	0.429543	0.508426
## 2718	0.0393305	-0.0165273	0.2438674	0.214507	0.373558	0.428905
## 2719	0.0527615	-0.0216170	0.1693868	0.292061	0.394414	0.385787
## 2720	0.0323528	0.0527276	0.1279742	0.273863	0.271967	0.413050
## 2721	-0.0300060	-0.0150653	0.0678097	0.126361	0.074658	0.251364
## 2722	-0.0766835	-0.0506646	0.3224224	0.336995	0.279406	0.353858
## 2723	-0.0273334	-0.0048817	0.1912438	0.197699	0.305514	0.398880
## 2724	-0.1244562	-0.0865115	0.4103713	0.510833	0.553862	0.725004
## 2725	-0.0263821	-0.0745857	0.3372751	0.390871	0.458382	0.524304
## 2726	0.0120822	-0.0720015	0.3317931	0.418935	0.533661	0.632784
## 2727	-0.0080017	0.0689679	0.1698234	0.240878	0.241724	0.399815
## 2728	0.0313371	0.0709361	0.2205962	0.250974	0.309488	0.388768
## 2729	0.0856429	-0.0154111	0.3621551	0.369561	0.412205	0.548860
## 2730	0.0208634	0.0356496	0.3382960	0.416863	0.488254	0.603014
## 2731	0.0689610	0.0062824	0.2327385	0.200931	0.297443	0.402289
## 2732	0.0437108	0.0256121	0.2857837	0.361507	0.403585	0.457731
## 2733	0.0401997	0.1083110	0.2094581	0.254511	0.342911	0.468469
## 2734	-0.0029571	0.0230876	0.1840167	0.192804	0.193985	0.297762
## 2735	0.0489064	0.0232571	0.2901083	0.318716	0.414584	0.525553
## 2736	-0.0041001	-0.1106490	0.3056088	0.354778	0.405528	0.504862
## 2737	0.0037216	-0.0210904	0.1384834	0.160469	0.157063	0.274180
## 2738	0.0132240	-0.0361253	0.2944589	0.342523	0.371191	0.433406
## 2739	-0.0019805	-0.0165468	0.1473709	0.206150	0.230891	0.288890
## 2740	0.0031152	0.0458269	0.4069320	0.459628	0.548438	0.705359
## 2741	0.0010701	-0.0034684	0.1146691	0.194106	0.278404	0.462167
## 2742	-0.0192695	-0.0440136	0.3453878	0.423301	0.484010	0.633093
## 2743	0.0907257	0.0230789	0.2252578	0.252392	0.248583	0.294400
## 2744	NA	NA	NA	NA	NA	NA
## 2745	-0.0295487	0.0755582	0.0646632	0.149721	0.245248	0.325712
## 2746	0.0555190	0.0076470	0.2497609	0.291603	0.390861	0.536492
## 2747	0.0119818	-0.0166752	0.2171059	0.150147	0.178620	0.324812
## 2748	0.0341394	0.0340345	0.3439782	0.452256	0.523070	0.714243
## 2800	-0.0055653	0.0133269	0.2026081	0.215749	0.239675	0.323038
## 2801	NA	NA	NA	NA	NA	NA
## 2802	0.0242803	0.0238420	0.0738297	0.079614	0.078104	0.146033
## 2803	0.0634400	-0.0202183	0.2925681	0.333532	0.391424	0.389347
## 2804	0.0184443	-0.0044703	0.0018605	0.008911	0.033988	0.076021
## 2805	0.0210577	0.0277665	0.0714983	0.052339	0.037169	0.125639
## 2806	0.0203940	0.0359600	0.2047697	0.153939	0.174340	0.242692
## 2807	0.0577546	0.0864848	0.0242395	0.004373	0.067912	0.112805
## 2808	0.0273011	0.0794379	0.2955885	0.341941	0.333642	0.423497
## 2809	0.0142959	-0.0135509	0.2438602	0.263499	0.299869	0.358032
## 2900	-0.0988139	0.0192710	0.4036005	0.432906	0.522583	0.604653
## 2901	-0.1972060	-0.0101001	0.0289684	0.300537	0.067025	0.012491
## 2902	-0.1313855	-0.0202008	0.3093613	0.602372	0.720032	0.868467
## 2903	NA	NA	NA	NA	NA	NA
## 2904	-0.6306154	0.0826777	0.1069392	-0.757164	1.214279	0.575507
## 2905	0.0503284	0.1208414	0.0710501	0.239716	0.059297	0.065711

## 2906	0.4375313	-0.2987667	0.5337373	0.970747	0.973464	0.945400
## 2907	0.1312022	0.0863449	0.6288527	0.818396	0.848161	1.283099
## 2908	-0.1266107	0.1417761	0.5135823	0.893804	0.812209	1.462590
## 2909	0.1186623	0.0990648	0.1575369	0.407706	0.390337	0.404450
## 2910	0.1078310	0.0531176	0.2636984	0.269460	0.445141	0.463087
## 2911	0.1027062	0.1987984	0.2024299	0.440124	0.750116	0.600171
## 2912	-0.1194181	0.0554607	0.3902510	0.585640	0.780283	0.524551
## 2913	0.8198653	-0.2039379	0.0134421	0.120181	0.333240	0.597142
## 2914	0.3988109	0.1875935	0.2643990	0.583389	0.604669	0.420896
## 2915	NA	NA	NA	NA	NA	NA
## 2916	-0.0377791	-0.0935266	0.3360723	0.338986	0.351377	0.407017
## 2917	0.4155625	0.0613393	0.2604911	0.448386	0.301524	0.337282
## 2918	NA	NA	NA	NA	NA	NA
## 2919	0.4449449	-0.0810010	-0.0407730	0.365539	0.446238	0.706103
## 2920	NA	NA	NA	NA	NA	NA
## 2921	0.1229968	0.1755881	0.2199820	0.330077	0.560051	0.433951
## 2922	0.2505801	0.0009692	0.1480245	0.358160	0.085022	0.474569
## 2923	0.0030172	0.2602092	-0.0170396	0.377699	0.184051	0.294145
## 3000	-0.0045059	0.0734419	0.0770340	0.224934	0.234139	0.240028
## 3001	NA	NA	NA	NA	NA	NA
## 3002	-0.0051268	0.0169857	0.1580245	0.156199	0.193464	0.247949
## 3003	-0.1267872	-0.0136837	0.6710778	0.730959	0.743870	0.824141
## 3004	0.0151078	-0.0138354	0.1903067	0.223505	0.253564	0.306871
## 3005	-0.0207639	0.0387806	0.0814524	0.097615	0.142612	0.157447
## 3100	0.0281004	0.0606505	0.2418625	0.248858	0.380425	0.400272
## 3101	-0.0904676	0.0517404	0.1974202	0.273283	0.201373	0.438075
## 3102	-0.0473649	-0.0523212	0.3517938	0.390049	0.381375	0.455726
## 3103	0.0872678	-0.1138419	0.1750744	0.245081	0.290627	0.342151
## 3104	-0.0151096	0.0431894	0.0829971	0.052848	0.088131	0.186125
## 3105	-0.1018386	0.0321566	0.2998275	0.269915	0.302377	0.288030
## 3106	-0.0786091	0.0162853	0.1264863	0.190017	0.320170	0.090307
## 3107	-0.0288132	-0.0668906	0.1030012	0.135016	0.267897	0.352164
## 3108	-0.0316068	0.0188025	0.2355793	0.391141	0.336543	0.524995
## 3109	0.1341431	-0.0653487	0.1051459	0.113483	0.205191	0.296549
## 3110	0.0688772	0.0433709	0.1160730	0.059131	0.089517	0.160250
## 3200	-0.0179347	-0.0323433	0.2218302	0.268465	0.292479	0.333055
## 3201	0.0299509	-0.0541746	-0.0380086	0.116162	-0.023343	0.039837
## 3202	-0.0241522	-0.0224567	0.2061221	0.263667	0.214062	0.264348
## 3203	-0.0184108	-0.0209759	0.3390304	0.340453	0.387570	0.424657
## 3204	0.0359575	0.0139502	0.1889095	0.239273	0.225037	0.286808
## 3205	0.0276234	0.0146120	0.0408961	0.069657	0.072886	0.144197
## 3206	0.0462578	0.0395249	0.1871071	0.202783	0.118493	0.245428
## 3207	-0.0304009	-0.0358940	0.2599729	0.276995	0.295822	0.265679
## 3300	0.1141669	-0.1101865	0.1704515	0.304390	0.201077	0.210507
## 3301	0.0578594	0.0365556	0.2106037	0.290065	0.330733	0.359951
## 3302	0.1455966	-0.1503704	0.1671665	0.482662	0.539253	0.389043
## 3303	0.0642033	-0.0201118	0.1236239	0.195579	0.017736	0.142959
## 3304	0.0361899	0.0478939	0.2255675	0.315790	0.389577	0.432615
## 3305	0.0481094	0.0474796	0.1274462	0.170921	0.228618	0.357046
## 3306	-0.0057722	0.0406391	0.1232816	0.264969	0.249962	0.272926

##	3307	0.0223525	0.1121208	0.1755311	0.229549	0.169247	0.258653
##	3308	-0.0225239	-0.0318614	0.2782373	0.431080	0.434067	0.513915
##	3309	0.0909050	-0.0547096	0.3625474	0.459993	0.440884	0.298628
##	3310	0.0087959	0.0333154	0.4340008	0.454014	0.495916	0.395173
##	3311	-0.1430219	0.0924014	0.2803612	0.392804	-0.004115	0.270566
##	3312	0.0260275	-0.0154439	0.1806263	0.239450	0.244604	0.240499
##	3313	-0.0932657	0.0771491	0.3734101	0.475988	0.226227	0.467383
##	3314	0.0101027	-0.0405599	0.2574636	0.231808	0.211149	0.223439
##	3315	0.1233647	-0.0386167	0.2808603	0.337376	0.354772	0.309594
##	3316	0.0078623	0.0641463	0.1930387	0.255947	0.054999	0.151543
##	3317	0.1597923	0.1106821	0.1452907	0.118338	0.277685	0.376739
##	3318	-0.0178562	-0.0164956	0.2393408	0.233723	0.305506	0.474929
##	3319	-0.0423775	-0.0596477	0.1901451	0.215763	0.222855	0.215487
##	3320	0.0721696	0.0427957	0.0934557	0.303248	0.323552	0.559064
##	3321	0.0009468	0.0041650	0.0657181	0.007993	0.052276	0.224119
##	3322	-0.1142659	0.0255608	0.0532499	-0.083292	0.157865	0.018460
##	3400	0.0449614	0.0349842	0.2675629	0.349287	0.461658	0.537862
##	3401	NA	NA	NA	NA	NA	NA
##	3402	NA	NA	NA	NA	NA	NA
##	3403	-0.1053361	-0.0245948	0.1825847	0.120337	0.263252	0.293897
##	3404	-0.1355391	-0.1850393	0.2780029	0.362137	0.336304	0.163606
##	3500	0.0048031	-0.0350310	0.4321740	0.519578	0.539720	0.634108
##	3501	NA	NA	NA	NA	NA	NA
##	3502	NA	NA	NA	NA	NA	NA
##	3503	NA	NA	NA	NA	NA	NA
##	3504	-0.0117409	0.0264555	0.2395960	0.332161	0.400955	0.536100
##	3505	0.1300895	-0.6392538	0.2391790	0.121793	0.204495	0.326703
##	3506	NA	NA	NA	NA	NA	NA
##	3600	-0.0472087	0.0294082	0.0768381	0.163759	0.148503	0.295651
##	3601	0.1145046	0.1606967	0.3796149	0.477623	0.504035	0.458773
##	3602	-0.0896015	-0.3200143	0.2381475	0.572891	0.584383	0.671100
##	3603	-0.0156852	-0.4418871	0.2259760	-0.040266	0.023621	0.036637
##	3604	NA	NA	NA	NA	NA	NA
##	3605	-0.1886981	0.0040934	0.2525253	0.444388	0.553215	0.446361
##	3606	NA	NA	NA	NA	NA	NA
##	3607	-0.0841940	-0.0552307	0.3588916	0.231891	0.296633	0.382507
##	3608	NA	NA	NA	NA	NA	NA
##	3609	-0.0125467	0.0116921	0.1603537	0.101106	0.279574	0.408832
##	3610	0.3819129	-0.3788309	0.2696526	0.450522	0.491443	0.552916
##	3611	0.0433602	-0.1052358	0.2515091	0.225404	0.207096	0.515592
##	3612	0.0339560	0.0048688	0.2984065	0.401553	0.457102	0.555481
##	3613	NA	NA	NA	NA	NA	NA
##	3614	-0.0997549	-0.1074799	0.2554001	0.347847	0.367795	0.478363
##	3615	NA	NA	NA	NA	NA	NA
##	3616	-0.0272737	0.0446374	0.1670004	0.120733	0.144350	0.143073
##		FFA2	FLA2	FFA3	FLA4		
##	1000	0.0062575	-0.0541274	0.0013751	-0.0532006		
##	1100	-0.0044535	0.0087036	-0.0035386	0.0080461		
##	1101	0.0947113	-0.0969216	0.0776383	-0.0680109		
##	1102	0.0842790	-0.0061098	0.0835022	0.0070484		

## 1103	0.0130081	-0.0150740	0.0106134	-0.0123603
## 1104	0.0023739	-0.0106608	0.0012846	-0.0103022
## 1105	0.0267564	-0.0275974	0.0227241	-0.0222774
## 1106	0.0597812	-0.0175382	0.0574847	-0.0076971
## 1107	-0.0213068	-0.0024135	-0.0216895	-0.0068973
## 1108	0.1022408	0.0320651	0.1036924	0.0370255
## 1109	0.1222320	-0.0108263	0.1207035	0.0118997
## 1110	0.0696985	0.0165732	0.0718122	0.0287381
## 1111	0.0174457	-0.0267248	0.0158856	-0.0249792
## 1200	-0.0164887	0.3026467	0.2409063	0.2882206
## 1201	0.0398696	-0.0115450	0.0344318	0.0079669
## 1202	0.1341501	-0.0463942	0.0926928	0.0737455
## 1203	0.0744353	0.0135645	0.0839523	0.0655914
## 1204	0.2421207	-0.1670758	0.1511992	-0.0157123
## 1205	0.3444378	-0.2255920	0.1275471	0.1069040
## 1206	-0.2119806	0.3599961	0.0582406	0.1965422
## 1207	0.2493516	0.0899147	0.3100928	0.2650146
## 1208	0.0585705	0.0810383	0.1366697	0.1379447
## 1209	-0.0504223	-0.3221059	-0.2895559	-0.3548699
## 1210	-0.3563169	0.2403429	-0.1405579	-0.0711654
## 1211	-0.1052470	0.1547166	0.0376787	0.0576608
## 1212	0.0375635	0.0488761	0.0827677	0.0832119
## 1213	0.3629246	-0.1969662	0.1837178	0.1357983
## 1300	-0.0458833	-0.0130675	-0.0473447	-0.0201941
## 1301	NA	NA	NA	NA
## 1302	0.1060951	-0.0397666	0.0935469	-0.0036593
## 1303	-0.0071021	-0.0063447	-0.0074696	-0.0070139
## 1304	0.0087173	0.0264457	0.0107817	0.0275028
## 1305	0.0243910	0.0549713	0.0288187	0.0578950
## 1306	-0.0173709	-0.0005320	-0.0174070	-0.0021742
## 1307	-0.0072056	-0.0050840	-0.0074908	-0.0056883
## 1308	0.0108488	0.0450295	0.0141153	0.0462102
## 1309	-0.0084250	0.0278403	-0.0056857	0.0267466
## 1310	-0.0061287	-0.0278236	-0.0095409	-0.0287340
## 1311	-0.0073513	-0.0158208	-0.0092665	-0.0170182
## 1312	0.0083611	0.0043873	0.0086344	0.0051696
## 1313	-0.0095384	0.0365135	-0.0063612	0.0352507
## 1314	0.0061482	-0.0224919	0.0047287	-0.0219209
## 1315	0.0824369	-0.0216982	0.0800449	-0.0064381
## 1400	0.0541460	-0.0290789	0.0372219	0.0029325
## 1401	0.0359500	-0.0604519	-0.0054462	-0.0361074
## 1402	-0.0014063	-0.0357420	-0.0209372	-0.0364721
## 1403	-0.0108154	0.0832127	0.0298040	0.0776778
## 1404	0.2704677	-0.0660929	0.2354736	0.0794586
## 1405	-0.0040044	0.0141433	0.0028788	0.0121241
## 1406	-0.0364506	0.0686177	-0.0062604	0.0522692
## 1407	0.0460696	0.0356611	0.0639790	0.0587726
## 1408	-0.0262484	0.0410323	-0.0068618	0.0280782
## 1409	-0.0024932	0.0227077	0.0106805	0.0212694
## 1410	-0.1228077	0.1682689	-0.0171313	0.0972597

##	1500	-0.0150723	-0.0082325	-0.0155110	-0.0098467
##	1501	0.2546026	-0.2615024	0.2741294	-0.2649093
##	1502	-0.0117858	0.0615790	-0.0017501	0.0591454
##	1503	-0.0518734	-0.0115174	-0.0524302	-0.0162983
##	1504	0.1296887	-0.0699459	0.1023622	-0.0199261
##	1505	-0.0764069	-0.0048280	-0.0769502	-0.0189476
##	1506	0.0926984	-0.1099923	0.0819122	-0.0944563
##	1507	-0.0796432	0.1705393	-0.0321408	0.1465074
##	1508	0.1379024	0.0264183	0.1403673	0.0524258
##	1600	0.0035548	0.0197880	0.0045950	0.0201276
##	1601	0.0875006	0.0878190	0.0955841	0.1001536
##	1602	0.0008244	-0.0318576	-0.0002109	-0.0318080
##	1603	0.0031546	0.0427240	0.0045411	0.0429435
##	1604	0.0097104	0.0061477	0.0099470	0.0068525
##	1605	-0.0410994	-0.0076480	-0.0414465	-0.0111374
##	1606	0.0194773	-0.0057749	0.0192346	-0.0042793
##	1607	-0.0191369	0.0102908	-0.0194067	0.0111586
##	1700	-0.0485557	-0.0179639	-0.0530850	-0.0330809
##	1701	0.0549895	-0.4902083	-0.1743453	-0.4620078
##	1702	0.0297379	-0.1089960	-0.0171549	-0.0952689
##	1703	-0.0157544	-0.0832741	-0.0427142	-0.0892910
##	1704	-0.0210459	0.0060660	-0.0197438	-0.0007361
##	1705	0.0430867	-0.0866996	0.0165625	-0.0733503
##	1706	0.0189741	-0.0507304	0.0009952	-0.0431154
##	1707	-0.1232708	0.0924699	-0.1028130	0.0577133
##	1708	-0.0945141	0.0610467	-0.0777807	0.0341210
##	1709	0.0896032	0.0269234	0.1013365	0.0688641
##	1710	0.0783803	0.0126417	0.0848290	0.0549866
##	1711	-0.0028323	0.1060817	0.0112034	0.1053570
##	1712	-0.0673035	0.0369728	-0.0561775	0.0134056
##	1800	0.0482494	0.1440251	0.1046146	0.1666838
##	1801	NA	NA	NA	NA
##	1802	0.1414208	-0.0547911	0.1233993	0.0035407
##	1803	-0.0235141	0.0509821	-0.0100845	0.0438790
##	1804	-0.1373968	0.0806882	-0.1041338	0.0143348
##	1900	-0.0039853	-0.0003187	-0.0040583	-0.0015405
##	1901	-0.0541805	-0.2073072	-0.0961233	-0.2232670
##	1902	0.0434331	0.0537815	0.0514153	0.0636732
##	1903	0.0306344	0.0910331	0.0326077	0.0916993
##	1904	0.0505739	-0.1472802	0.0172121	-0.1313233
##	1905	0.0405812	-0.0279802	0.0368668	-0.0247970
##	1906	-0.0111839	0.0224995	-0.0080340	0.0202540
##	1907	-0.0130684	-0.0164030	-0.0159888	-0.0193911
##	1908	0.0240542	-0.0345173	0.0183525	-0.0289019
##	1909	0.0976731	-0.0788269	0.0862613	-0.0583088
##	1910	0.0423159	0.0141062	0.0444812	0.0241540
##	1911	0.0318296	-0.0851607	0.0115991	-0.0733650
##	1912	0.1123804	-0.0726677	0.1039528	-0.0530703
##	1913	0.0191581	-0.1425836	0.0026761	-0.1400325
##	2000	0.1407617	-0.0030154	0.1387478	0.1030578

##	2001	-0.1726162	0.0242567	-0.1598894	-0.0686982
##	2002	0.0713360	0.0123858	0.0788141	0.0560198
##	2003	0.1784685	-0.1252607	0.1123984	-0.0290182
##	2100	NA	NA	NA	NA
##	2101	0.6603351	0.2099061	0.6763804	0.0924572
##	2102	-0.0013965	-0.0016791	-0.0016500	-0.0019815
##	2103	-0.0724930	0.0935479	-0.0592429	0.0739456
##	2104	0.0881413	0.0743826	0.0961446	0.0986588
##	2105	-0.0847610	0.0788111	-0.0689644	0.0564225
##	2200	0.0529694	0.1355002	0.1003744	0.1561260
##	2201	0.2493324	-0.3258380	0.1366962	-0.2061559
##	2202	-0.0238509	0.0155913	-0.0187133	0.0042040
##	2203	0.0146832	-0.0821077	0.0048080	-0.0791841
##	2204	-0.0136983	0.0129755	-0.0117262	0.0102709
##	2205	0.0090501	-0.0421435	0.0048277	-0.0409334
##	2206	-0.0655051	0.1613892	-0.0292696	0.1442402
##	2207	-0.0141619	0.0307479	-0.0073988	0.0259015
##	2208	0.0114492	-0.0026702	0.0111003	-0.0007065
##	2209	0.0559997	-0.0274022	0.0523082	-0.0150903
##	2210	0.0659182	-0.0542828	0.0608001	-0.0450783
##	2211	0.1109130	0.1043327	0.1246652	0.1212195
##	2212	0.1174231	-0.1781846	0.0858761	-0.1483886
##	2213	0.0526549	0.0834455	0.0775383	0.1026691
##	2214	0.2491757	0.1349724	0.2841182	0.2223839
##	2215	-0.0270834	0.0094317	-0.0263653	0.0064583
##	2216	0.0886668	-0.0708702	0.0446205	-0.0157590
##	2300	0.0567544	0.0158560	0.0602025	0.0317731
##	2301	0.1569987	0.0154825	0.1670242	0.1200230
##	2302	0.0282747	0.0486879	0.0318731	0.0517371
##	2303	0.0359607	-0.0202753	0.0327135	-0.0124350
##	2304	0.0033374	0.0029650	0.0035610	0.0033475
##	2305	-0.0293570	-0.1290623	-0.0384476	-0.1323289
##	2306	-0.0337820	-0.0789813	-0.0465045	-0.0863073
##	2307	0.0092334	0.0085000	0.0097773	0.0093206
##	2308	0.0077685	-0.0089919	0.0045254	-0.0058292
##	2309	0.0051287	0.0623759	0.0153962	0.0634910
##	2310	-0.0027372	0.0244214	-0.0010799	0.0241663
##	2311	-0.0074984	-0.0720852	-0.0147063	-0.0731071
##	2312	-0.0107690	-0.0066048	-0.0115356	-0.0084759
##	2400	-0.0276465	0.0247069	-0.0259821	0.0222636
##	2401	NA	NA	NA	NA
##	2402	0.0637472	0.0352597	0.0672910	0.0449737
##	2403	-0.0016796	-0.0048323	-0.0019739	-0.0049701
##	2404	0.0222223	-0.0136071	0.0220337	-0.0131804
##	2405	0.1200689	-0.0099475	0.1196788	-0.0012265
##	2406	-0.0353598	-0.0088695	-0.0354672	-0.0095201
##	2500	0.0586612	0.0171551	0.0599214	0.0233757
##	2501	0.0016437	-0.2340822	-0.0410828	-0.2332856
##	2502	-0.0006841	0.0971958	0.0126608	0.0969877
##	2503	-0.0031685	0.0666976	0.0052121	0.0660549

##	2504	0.0240478	0.0124153	0.0249996	0.0154436
##	2505	0.0154268	0.0268229	0.0166201	0.0280964
##	2506	0.0657327	-0.0859594	0.0573016	-0.0747412
##	2507	-0.0165330	0.0176561	-0.0147575	0.0146673
##	2508	0.0313358	0.0184120	0.0322160	0.0212627
##	2600	-0.0272053	0.0030210	-0.0251694	-0.0152472
##	2601	0.4896249	-0.5544091	0.1072313	-0.1174255
##	2602	0.0790948	-0.1352184	0.0073709	-0.0822052
##	2603	0.1429871	-0.1598017	0.0340833	-0.0550846
##	2604	0.0034040	0.0339331	0.0140773	0.0352762
##	2605	0.0210658	-0.1101990	-0.0133645	-0.1002578
##	2606	-0.2761039	0.3005229	-0.1544515	0.1915413
##	2607	0.0751268	-0.1199712	0.0308684	-0.0905783
##	2608	-0.1208360	0.0416796	-0.0971149	-0.0337367
##	2609	0.4395181	-0.4961554	0.0461256	-0.1258465
##	2610	0.0508704	-0.0983076	0.0150977	-0.0761098
##	2611	0.1373743	-0.0688702	0.1244895	-0.0326586
##	2612	-0.1356341	0.1799400	-0.0458346	0.0991560
##	2613	-0.0101954	-0.0025492	-0.0111866	-0.0070899
##	2614	-0.0027773	-0.0646754	-0.0248565	-0.0656534
##	2700	0.0761389	-0.0679090	0.0509319	-0.0375500
##	2701	-0.0009990	-0.0581036	-0.0180186	-0.0583926
##	2702	-0.0693787	0.0392674	-0.0589749	0.0172983
##	2703	0.0158785	0.1180610	0.0305399	0.1203406
##	2704	0.0339215	0.0940976	0.0417040	0.0982725
##	2705	0.0645239	0.0426402	0.0701990	0.0550088
##	2706	0.0053377	-0.0487207	-0.0031191	-0.0475144
##	2707	0.0814991	-0.0335699	0.0686839	0.0002664
##	2708	0.0522821	0.0156240	0.0540475	0.0212425
##	2709	NA	NA	NA	NA
##	2710	0.0030222	-0.1125336	0.0013474	-0.1123369
##	2711	0.0639039	0.0267148	0.0673058	0.0365072
##	2712	-0.0246899	-0.0648052	-0.0331626	-0.0685363
##	2713	0.0488644	0.0006446	0.0489991	0.0121422
##	2714	0.1289558	0.1553162	0.1912429	0.2080639
##	2715	0.0556110	0.0436436	0.0630221	0.0565035
##	2716	0.0149941	-0.0555894	0.0025647	-0.0519042
##	2717	0.0870431	0.0085673	0.0897144	0.0340647
##	2718	0.0602990	-0.0077209	0.0578549	0.0116214
##	2719	0.1149738	-0.0314017	0.1028895	0.0131933
##	2720	0.0524916	0.0304786	0.0563704	0.0383065
##	2721	-0.0303401	-0.0019088	-0.0303687	-0.0027510
##	2722	-0.0704125	-0.0665812	-0.0803227	-0.0811881
##	2723	-0.0082005	-0.0082130	-0.0088775	-0.0091041
##	2724	-0.0916574	-0.1362190	-0.1105397	-0.1519800
##	2725	-0.0047956	-0.0744124	-0.0083169	-0.0747238
##	2726	0.0311917	-0.0517710	0.0296288	-0.0506332
##	2727	0.0120145	0.0575861	0.0228511	0.0602664
##	2728	0.0456396	0.0647310	0.0542334	0.0721033
##	2729	0.1179530	-0.0428811	0.1065890	-0.0125283

## 2730	0.0353436	0.0274634	0.0388577	0.0327408
## 2731	0.0814563	0.0146816	0.0831138	0.0253432
## 2732	0.0479979	-0.0004565	0.0479202	0.0107120
## 2733	0.0163585	0.0835585	0.0358701	0.0887387
## 2734	0.0143317	0.0131911	0.0166399	0.0159369
## 2735	0.0670846	0.0002824	0.0671427	0.0148301
## 2736	0.0189950	-0.1399984	-0.0004919	-0.1366386
## 2737	0.0034048	-0.0316722	0.0009571	-0.0312971
## 2738	0.0318108	-0.0405315	0.0244853	-0.0346478
## 2739	0.0120788	-0.0329010	0.0014109	-0.0290174
## 2740	0.0231314	0.0222146	0.0260079	0.0265898
## 2741	0.0274218	-0.0169609	0.0249688	-0.0116633
## 2742	-0.0110507	-0.0826051	-0.0346722	-0.0855745
## 2743	0.1002015	-0.0034174	0.1001010	0.0180041
## 2744	NA	NA	NA	NA
## 2745	-0.0152360	0.0713008	-0.0055283	0.0690859
## 2746	0.0676236	-0.0023206	0.0672788	0.0109915
## 2747	0.0206873	-0.0216219	0.0172124	-0.0168944
## 2748	0.0826669	0.0327648	0.0898075	0.0559365
## 2800	0.0098864	0.0164413	0.0109570	0.0173772
## 2801	NA	NA	NA	NA
## 2802	0.0284037	0.0219931	0.0310817	0.0260055
## 2803	0.0738331	-0.0190678	0.0728518	-0.0140675
## 2804	0.0169827	-0.0072090	0.0162473	-0.0048741
## 2805	0.0203808	0.0239948	0.0252725	0.0286063
## 2806	0.0202249	0.0319480	0.0258693	0.0359434
## 2807	0.0485958	0.0757159	0.0713182	0.0892767
## 2808	0.0434017	0.0717703	0.0496039	0.0763796
## 2809	0.0189202	-0.0165293	0.0170833	-0.0141746
## 2900	-0.0840643	-0.0615448	-0.1144167	-0.0965971
## 2901	-0.1583367	-0.0244768	-0.1671873	-0.0740816
## 2902	-0.2081768	-0.2302443	-0.2923608	-0.3019329
## 2903	-0.8511793	0.1936325	-0.6890173	-0.4573897
## 2904	-0.2708035	-0.2048322	-0.3720104	-0.3908793
## 2905	0.0664726	0.1043947	0.1023939	0.1252862
## 2906	0.6506811	-0.4287581	0.5178419	-0.3938383
## 2907	0.0945020	-0.1434899	0.0088101	-0.0775607
## 2908	-0.3543593	0.0406572	-0.3330101	-0.1240203
## 2909	0.1668112	0.0632212	0.1895305	0.1185280
## 2910	0.1972008	0.0446348	0.2270844	0.1730663
## 2911	0.1877460	0.2115002	0.2940835	NA
## 2912	-0.1658753	-0.0495400	-0.1816383	-0.0875329
## 2913	1.0621799	-0.2692898	0.8383991	0.0534451
## 2914	0.5130400	-0.0782228	0.4727178	0.0708312
## 2915	NA	NA	NA	NA
## 2916	-0.0286671	-0.1058716	-0.0637090	-0.1149617
## 2917	0.5080738	0.0258712	0.5246468	0.2292551
## 2918	NA	NA	NA	NA
## 2919	0.6125618	-0.1452717	0.5238269	0.0554961
## 2920	NA	NA	NA	NA

##	2921	0.1341763	0.1678211	0.2165639	0.2320676
##	2922	0.3193348	-0.0640387	0.2727855	0.1071345
##	2923	0.0325810	0.1952969	0.0979785	0.2029931
##	3000	0.0074940	0.0612117	0.0187903	0.0625578
##	3001	NA	NA	NA	NA
##	3002	-0.0012528	0.0236931	0.0004674	0.0235217
##	3003	-0.1680642	-0.0937002	-0.1850503	-0.1306413
##	3004	0.0203941	-0.0234577	0.0191970	-0.0219633
##	3005	-0.0162495	0.0319491	-0.0131698	0.0300510
##	3100	0.0650849	0.0787544	0.0791787	0.0936906
##	3101	-0.0213762	0.0750870	-0.0092731	0.0693875
##	3102	0.0022030	-0.0558417	-0.0168264	-0.0548689
##	3103	0.1483783	-0.0644721	0.1410004	-0.0412788
##	3104	-0.0008445	0.0489349	0.0034332	0.0488210
##	3105	-0.0667790	0.0479565	-0.0649374	0.0425520
##	3106	-0.0494078	0.0220817	-0.0431384	0.0034361
##	3107	-0.0159395	-0.0492812	-0.0202769	-0.0508400
##	3108	0.0153138	0.0517590	0.0216521	0.0542209
##	3109	0.1615003	-0.0587464	0.1457580	-0.0101867
##	3110	0.0797996	0.0479508	0.0799349	0.0479323
##	3200	0.0089582	-0.0188186	0.0035551	-0.0162063
##	3201	0.0389153	-0.0598002	0.0125548	-0.0426899
##	3202	-0.0084074	-0.0181695	-0.0141172	-0.0208343
##	3203	0.0114510	-0.0137074	0.0085416	-0.0113083
##	3204	0.0529101	0.0098497	0.0556683	0.0236388
##	3205	0.0297050	0.0128062	0.0322078	0.0189191
##	3206	0.0636784	0.0404119	0.0706460	0.0519332
##	3207	-0.0050110	-0.0364120	-0.0179244	-0.0381532
##	3300	0.1407552	-0.1128515	0.0618917	-0.0121362
##	3301	0.0949580	0.0386833	0.1184258	0.0960239
##	3302	0.2338358	-0.2064209	0.1044189	-0.0470372
##	3303	0.0787137	-0.0333433	0.0551704	0.0230024
##	3304	0.0656053	0.0472854	0.0931309	0.0852115
##	3305	0.0702242	0.0342444	0.0926780	0.0815177
##	3306	0.0190627	0.0162923	0.0259141	0.0239473
##	3307	0.0313041	0.0998179	0.0537240	0.1077332
##	3308	0.0415165	-0.0376205	0.0176205	-0.0107829
##	3309	0.0782146	-0.0577245	0.0355270	0.0003380
##	3310	0.0838807	0.0063167	0.0882110	0.0632587
##	3311	-0.1284244	0.0840222	-0.0785858	0.0052948
##	3312	0.0518287	-0.0224114	0.0363401	0.0138092
##	3313	-0.0703950	0.0417584	-0.0576612	0.0182000
##	3314	0.0565410	-0.0745278	-0.0014464	-0.0303765
##	3315	0.1373614	-0.0289361	0.1172031	0.0676516
##	3316	0.0338688	0.0508775	0.0771226	0.0797140
##	3317	0.2141420	0.0659604	0.2564832	0.1964948
##	3318	0.0135531	-0.0574626	-0.0269262	-0.0485657
##	3319	-0.0353988	-0.0668120	-0.0591148	-0.0780374
##	3320	0.0545095	0.0727301	0.1172661	0.1186010
##	3321	0.0017960	0.0058347	0.0054128	0.0069074

```

## 3322 -0.1156925  0.0216673 -0.1011460 -0.0541531
## 3400  0.0458708 -0.0520564  0.0363094 -0.0428700
## 3401          NA          NA          NA          NA
## 3402 -0.2696782 -0.0458576 -0.3005939 -0.1847142
## 3403 -0.1088188 -0.0413575 -0.1150196 -0.0648265
## 3404 -0.1597666 -0.2104853 -0.2558427 -0.2871628
## 3500  0.0710584 -0.0346241  0.0636531 -0.0152919
## 3501          NA          NA          NA          NA
## 3502          NA          NA          NA          NA
## 3503          NA          NA          NA          NA
## 3504 -0.0126141  0.0881092 -0.0091437  0.0873552
## 3505  0.1301163 -0.8472116 -0.0246311 -0.7466441
## 3506          NA          NA          NA          NA
## 3600 -0.0360627  0.0178068 -0.0290005  0.0037478
## 3601  0.1810756  0.1541037  0.2643394  0.2466430
## 3602  0.0385775 -0.2373900 -0.0141761 -0.2302112
## 3603  0.0236216 -0.5004107 -0.3416157 -0.4821220
## 3604          NA          NA          NA          NA
## 3605 -0.1353539  0.0336089 -0.1252362 -0.0067782
## 3606          NA          NA          NA          NA
## 3607 -0.0914096 -0.0406591 -0.0973104 -0.0563488
## 3608          NA          NA          NA          NA
## 3609 -0.0195866 -0.0542146 -0.0299154 -0.0564914
## 3610  0.4378000 -0.3819328  0.3830059 -0.3036179
## 3611  0.0361053 -0.0940605  0.0163532 -0.0861621
## 3612  0.0544358 -0.0239430  0.0472801 -0.0068905
## 3613          NA          NA          NA          NA
## 3614 -0.0774331 -0.1465584 -0.1103722 -0.1700177
## 3615          NA          NA          NA          NA
## 3616 -0.0135131  0.0458235 -0.0017037  0.0428330

```

```
print(RegP)
```

```

##          FFA1p      FLA1p          2p          3p          4p          5+p
## 1000 3.018e-01 3.163e-01 5.500e-17 1.643e-16 2.808e-23 9.287e-27
## 1100 2.470e-01 8.573e-01 3.070e-06 1.304e-04 4.478e-08 4.015e-12
## 1101 4.383e-01 3.073e-01 5.843e-06 4.926e-03 2.985e-05 1.593e-03
## 1102 8.525e-04 9.461e-01 9.009e-04 6.336e-06 6.964e-09 1.016e-11
## 1103 6.626e-01 5.034e-01 1.998e-07 2.738e-07 8.090e-14 4.069e-10
## 1104 5.145e-01 6.740e-01 8.083e-08 1.395e-06 3.216e-12 4.493e-18
## 1105 4.966e-02 1.852e-02 4.704e-18 9.256e-20 3.775e-25 8.804e-37
## 1106 6.096e-02 9.163e-01 3.535e-14 5.774e-16 1.994e-15 1.001e-16
## 1107 7.631e-02 7.846e-01 4.444e-09 1.550e-12 1.715e-14 5.912e-20
## 1108 2.104e-02 4.865e-01 6.921e-01 6.745e-01 8.352e-01 5.059e-01
## 1109 2.925e-06 5.679e-01 3.470e-16 8.173e-15 9.311e-20 1.475e-29
## 1110 1.227e-04 1.847e-01 5.895e-11 1.295e-12 5.694e-16 5.315e-27
## 1111 7.919e-01 3.608e-01 6.569e-03 2.364e-02 5.102e-04 1.989e-06
## 1200 3.295e-01 6.396e-01 1.166e-01 5.785e-01 5.361e-06 1.038e-01
## 1201 4.465e-01 4.808e-01 2.519e-23 3.897e-25 7.508e-11 1.650e-22
## 1202 7.131e-02 6.096e-01 1.408e-04 1.729e-10 7.025e-04 2.884e-17

```

##	1203	6.057e-01	6.138e-02	5.466e-45	6.771e-30	1.329e-21	8.391e-10
##	1204	1.081e-02	2.336e-01	1.413e-01	4.936e-07	6.569e-06	9.402e-04
##	1205	7.300e-01	8.642e-01	2.535e-01	4.400e-01	1.572e-01	3.345e-01
##	1206	2.359e-01	1.114e-01	1.622e-01	7.224e-06	3.997e-01	3.476e-02
##	1207	2.242e-03	4.166e-02	2.017e-09	6.892e-19	2.111e-13	1.198e-20
##	1208	8.791e-01	3.581e-01	1.864e-01	3.467e-07	7.775e-02	1.000e-05
##	1209	3.803e-01	6.532e-01	7.981e-03	4.084e-02	1.823e-01	9.357e-03
##	1210	3.456e-01	8.682e-01	1.805e-05	5.924e-10	4.584e-01	4.207e-03
##	1211	2.423e-01	1.140e-01	6.858e-04	2.264e-01	3.697e-04	1.150e-01
##	1212	9.276e-01	5.929e-01	2.243e-01	2.372e-01	3.207e-02	6.469e-01
##	1213	1.764e-02	1.890e-01	2.464e-01	3.401e-04	5.269e-03	9.921e-01
##	1300	1.307e-02	6.251e-01	7.352e-02	8.109e-01	4.045e-03	2.970e-07
##	1301	NA	NA	NA	NA	NA	NA
##	1302	2.878e-01	2.918e-01	6.599e-08	3.113e-14	1.529e-19	2.316e-23
##	1303	3.169e-02	2.249e-01	3.056e-11	1.729e-13	1.686e-16	1.027e-22
##	1304	9.534e-01	2.408e-01	2.863e-03	1.301e-04	3.009e-04	4.658e-05
##	1305	7.253e-01	1.003e-02	1.827e-09	2.073e-10	8.840e-13	5.065e-18
##	1306	1.032e-01	8.092e-01	1.119e-08	7.203e-10	7.423e-13	1.170e-20
##	1307	1.263e-01	4.772e-01	2.418e-08	1.489e-10	1.108e-12	1.512e-20
##	1308	6.703e-01	1.895e-02	1.383e-03	1.452e-03	7.115e-04	1.343e-05
##	1309	4.973e-01	1.720e-01	4.590e-02	6.444e-02	2.926e-02	1.410e-04
##	1310	3.605e-01	1.861e-01	1.733e-04	1.109e-04	8.261e-05	5.706e-07
##	1311	2.166e-01	2.214e-01	2.517e-03	5.540e-05	2.709e-04	3.063e-10
##	1312	9.117e-01	9.963e-01	3.025e-11	5.328e-15	7.866e-17	4.102e-30
##	1313	4.557e-01	6.446e-02	1.233e-04	1.760e-05	1.058e-06	1.997e-08
##	1314	8.132e-01	8.793e-02	9.733e-06	1.381e-07	1.919e-07	6.995e-13
##	1315	4.113e-02	7.153e-01	1.669e-02	2.618e-02	1.612e-03	5.262e-09
##	1400	4.533e-01	6.899e-01	6.925e-05	1.669e-08	1.649e-02	2.174e-01
##	1401	9.582e-01	6.077e-01	6.721e-03	2.884e-03	2.240e-02	2.080e-01
##	1402	7.977e-01	3.910e-01	1.215e-03	1.953e-04	6.121e-01	7.554e-01
##	1403	4.551e-01	3.357e-02	4.949e-10	1.764e-05	5.562e-02	1.585e-03
##	1404	1.225e-02	7.850e-01	3.312e-03	2.916e-03	1.985e-02	3.616e-01
##	1405	9.548e-01	8.781e-01	2.707e-01	6.688e-01	2.035e-01	1.922e-01
##	1406	4.005e-01	1.683e-01	6.719e-03	5.428e-05	4.533e-02	8.461e-01
##	1407	2.474e-01	3.440e-01	1.052e-02	2.527e-02	7.405e-03	6.920e-01
##	1408	3.209e-01	2.359e-01	1.363e-06	1.428e-04	1.421e-02	1.118e-05
##	1409	7.364e-01	6.022e-01	2.181e-01	1.476e-02	4.514e-04	1.473e-02
##	1410	6.894e-01	2.028e-01	1.218e-05	2.457e-03	5.723e-02	9.985e-02
##	1500	3.168e-02	5.117e-01	6.130e-11	2.973e-13	3.867e-16	4.216e-14
##	1501	2.617e-01	3.714e-01	1.932e-02	2.892e-03	4.271e-02	6.630e-01
##	1502	3.164e-01	4.149e-02	2.524e-05	2.941e-05	6.065e-08	1.463e-07
##	1503	7.487e-03	6.697e-01	9.596e-03	2.658e-03	1.399e-04	1.493e-04
##	1504	NA	NA	NA	NA	NA	NA
##	1505	7.701e-03	9.000e-01	1.874e-03	1.249e-02	1.747e-04	2.252e-04
##	1506	1.184e-01	2.634e-01	5.072e-01	5.981e-01	8.315e-01	8.305e-01
##	1507	4.617e-01	4.557e-01	1.068e-03	9.113e-04	1.446e-05	4.196e-03
##	1508	1.619e-01	6.795e-01	6.482e-04	1.276e-04	8.188e-05	9.758e-05
##	1600	3.992e-01	3.780e-01	1.208e-12	1.412e-12	2.036e-17	4.120e-23
##	1601	2.635e-01	7.190e-01	2.593e-01	3.822e-01	5.588e-02	1.183e-02
##	1602	3.817e-01	1.570e-01	1.718e-08	5.243e-10	9.814e-10	2.453e-11

##	1603	3.393e-01	6.279e-01	1.320e-03	4.021e-04	6.251e-05	3.142e-06
##	1604	7.202e-01	8.664e-01	5.572e-01	2.755e-01	1.661e-01	4.778e-02
##	1605	2.590e-04	5.236e-01	1.722e-01	3.390e-02	4.961e-03	6.882e-03
##	1606	4.795e-01	5.600e-01	1.462e-04	2.006e-06	2.144e-08	6.779e-10
##	1607	3.209e-01	7.137e-01	8.968e-01	5.305e-01	4.473e-01	2.587e-01
##	1700	2.246e-01	7.759e-01	1.568e-01	3.869e-08	1.098e-03	9.829e-03
##	1701	5.923e-01	1.791e-02	5.703e-01	5.097e-01	8.747e-01	3.176e-01
##	1702	5.433e-01	4.869e-02	7.961e-03	2.528e-04	4.812e-03	9.991e-04
##	1703	4.627e-01	1.100e-01	3.467e-05	1.014e-06	5.240e-05	5.283e-07
##	1704	7.234e-01	9.362e-01	3.386e-02	5.092e-03	1.147e-02	4.278e-01
##	1705	3.987e-01	2.035e-01	4.377e-04	9.574e-07	1.136e-01	2.369e-05
##	1706	5.961e-01	2.745e-01	1.231e-09	1.595e-14	4.102e-06	1.704e-13
##	1707	1.260e-01	3.044e-01	9.947e-02	5.674e-01	1.335e-02	1.141e-01
##	1708	2.378e-01	4.444e-01	8.966e-02	5.305e-04	1.536e-01	1.282e-02
##	1709	2.737e-01	6.311e-01	6.207e-02	8.483e-07	2.415e-02	3.414e-03
##	1710	9.815e-02	6.118e-01	8.980e-07	6.615e-06	2.983e-03	2.968e-04
##	1711	9.739e-01	4.084e-01	1.785e-02	1.469e-01	2.031e-03	1.922e-01
##	1712	8.880e-02	3.382e-01	4.757e-05	4.398e-09	2.669e-03	2.853e-06
##	1800	6.952e-01	4.200e-02	2.389e-01	1.149e-02	1.481e-02	4.081e-02
##	1801	NA	NA	NA	NA	NA	NA
##	1802	1.778e-01	9.850e-01	2.683e-05	5.839e-07	1.194e-02	1.261e-01
##	1803	3.462e-01	4.382e-01	4.407e-05	7.880e-12	3.725e-02	2.340e-04
##	1804	3.559e-03	1.620e-01	1.569e-04	4.786e-03	5.220e-01	8.226e-01
##	1900	6.143e-01	7.186e-01	3.331e-10	1.227e-12	1.533e-11	1.492e-15
##	1901	6.228e-02	1.196e-04	5.950e-07	2.337e-05	1.728e-05	8.412e-09
##	1902	1.147e-01	3.237e-01	4.791e-01	9.566e-01	1.497e-02	2.649e-05
##	1903	9.739e-01	3.626e-01	2.183e-01	1.927e-02	1.367e-02	9.588e-03
##	1904	2.397e-01	7.319e-04	3.945e-10	1.730e-06	7.175e-05	5.969e-09
##	1905	8.894e-01	9.272e-01	3.022e-01	4.994e-01	8.838e-01	8.254e-01
##	1906	6.848e-01	4.917e-01	2.303e-01	7.650e-01	4.358e-02	1.288e-01
##	1907	4.585e-01	3.936e-01	1.815e-02	5.475e-04	2.152e-06	3.543e-06
##	1908	4.139e-01	2.387e-01	2.558e-01	3.089e-01	2.565e-04	1.718e-03
##	1909	4.362e-01	3.029e-01	2.697e-06	1.462e-04	1.083e-02	6.233e-02
##	1910	7.257e-01	6.349e-01	1.420e-04	3.248e-05	3.740e-09	2.769e-09
##	1911	9.481e-01	8.883e-02	1.834e-02	2.797e-03	6.730e-05	7.019e-07
##	1912	3.130e-03	3.560e-02	6.707e-02	2.418e-02	1.130e-02	1.825e-05
##	1913	7.192e-01	2.756e-02	9.921e-01	1.149e-01	2.354e-02	1.072e-02
##	2000	3.192e-01	8.467e-01	8.546e-02	9.989e-02	4.023e-06	6.932e-01
##	2001	3.977e-02	4.820e-01	1.769e-02	1.686e-04	9.188e-07	4.975e-01
##	2002	3.382e-02	6.764e-01	2.762e-12	4.651e-08	1.091e-02	3.038e-01
##	2003	1.437e-03	1.581e-02	4.101e-07	6.953e-05	5.487e-01	2.418e-01
##	2100	NA	NA	NA	NA	NA	NA
##	2101	1.131e-01	7.380e-02	3.094e-03	2.140e-02	5.735e-04	5.998e-07
##	2102	3.781e-01	8.482e-01	7.984e-11	1.707e-16	1.120e-09	4.391e-12
##	2103	3.742e-02	4.240e-01	2.211e-04	3.949e-09	3.351e-05	4.552e-04
##	2104	4.870e-01	8.962e-01	9.602e-01	7.027e-01	7.450e-01	3.823e-01
##	2105	2.861e-02	6.867e-02	1.090e-03	5.720e-07	1.526e-06	5.322e-05
##	2200	4.375e-01	6.207e-02	5.581e-04	1.634e-07	1.689e-02	1.978e-03
##	2201	1.077e-01	1.621e-01	2.730e-05	1.851e-06	3.148e-07	6.232e-07
##	2202	6.083e-01	1.949e-01	8.695e-06	1.123e-06	2.516e-02	5.745e-06

##	2203	7.554e-01	9.914e-01	3.667e-05	1.370e-05	3.969e-04	1.225e-02
##	2204	4.070e-01	4.314e-01	9.172e-09	9.547e-08	6.814e-10	1.647e-09
##	2205	8.997e-01	4.168e-01	1.834e-02	9.073e-03	3.609e-01	6.719e-02
##	2206	7.064e-01	4.094e-01	2.594e-01	1.547e-01	3.937e-01	1.354e-01
##	2207	5.407e-01	5.791e-01	1.705e-02	8.621e-06	4.941e-04	3.058e-03
##	2208	9.555e-01	9.860e-01	4.388e-09	2.556e-12	2.119e-05	3.189e-08
##	2209	5.882e-01	5.452e-01	1.620e-07	6.830e-08	1.309e-10	1.357e-06
##	2210	6.937e-01	2.266e-01	1.900e-14	1.651e-15	6.584e-11	2.656e-21
##	2211	1.006e-01	7.830e-02	3.487e-06	2.933e-06	1.184e-05	5.193e-12
##	2212	5.548e-01	1.961e-01	8.168e-04	8.429e-06	1.667e-03	4.298e-03
##	2213	7.140e-01	6.998e-02	1.988e-06	1.185e-09	6.973e-07	1.226e-11
##	2214	1.582e-01	2.828e-01	5.546e-04	2.770e-04	5.850e-01	1.734e-02
##	2215	3.824e-01	9.903e-01	3.631e-04	1.131e-06	4.145e-04	1.671e-03
##	2216	9.964e-01	6.969e-01	3.999e-01	2.481e-02	3.411e-04	4.173e-06
##	2300	3.176e-02	3.086e-01	1.336e-04	5.018e-04	1.154e-05	8.940e-11
##	2301	2.837e-03	8.191e-01	5.449e-02	8.367e-01	3.681e-01	9.184e-01
##	2302	7.494e-01	2.661e-01	6.751e-03	6.561e-04	6.954e-03	9.509e-05
##	2303	3.621e-01	4.532e-01	7.315e-42	1.563e-42	7.663e-44	2.079e-66
##	2304	5.969e-01	5.735e-01	3.250e-08	3.748e-09	7.212e-08	1.037e-15
##	2305	1.316e-01	5.984e-04	2.284e-06	1.829e-05	5.006e-05	3.270e-08
##	2306	5.852e-02	1.947e-02	3.195e-06	1.549e-07	9.994e-07	1.792e-09
##	2307	8.672e-01	5.822e-01	2.499e-04	3.285e-04	1.455e-06	2.148e-08
##	2308	6.704e-01	9.980e-01	6.484e-09	1.216e-12	7.526e-07	2.478e-15
##	2309	4.787e-01	2.611e-02	1.678e-07	7.680e-08	3.225e-08	8.099e-06
##	2310	5.323e-01	1.600e-01	5.297e-03	2.133e-04	9.032e-04	4.158e-06
##	2311	6.069e-01	6.824e-02	1.324e-04	3.897e-05	7.413e-04	2.300e-06
##	2312	1.015e-01	9.573e-01	4.519e-13	6.106e-14	4.074e-12	1.608e-16
##	2400	3.487e-01	4.439e-01	1.324e-01	9.331e-01	1.073e-01	8.572e-02
##	2401	NA	NA	NA	NA	NA	NA
##	2402	5.086e-02	2.297e-01	4.008e-04	1.047e-04	6.125e-05	9.559e-07
##	2403	4.791e-01	9.853e-01	6.217e-03	6.509e-04	5.968e-06	7.022e-10
##	2404	2.491e-01	5.179e-01	6.017e-03	4.694e-04	5.242e-04	5.856e-05
##	2405	3.550e-02	7.593e-01	1.942e-01	2.695e-02	4.157e-02	5.067e-05
##	2406	2.047e-01	9.839e-01	4.554e-01	5.243e-01	4.624e-01	1.992e-02
##	2500	2.493e-01	9.576e-01	1.794e-04	5.246e-02	6.317e-06	3.228e-13
##	2501	7.217e-01	5.386e-01	8.857e-02	2.425e-01	1.026e-01	1.896e-02
##	2502	9.086e-01	2.124e-03	2.513e-01	2.107e-01	5.639e-02	9.872e-02
##	2503	5.895e-01	1.615e-01	1.400e-02	7.667e-03	5.097e-03	5.307e-03
##	2504	5.966e-01	6.853e-01	1.822e-01	1.611e-02	8.087e-03	8.210e-05
##	2505	8.048e-01	4.933e-01	1.356e-03	2.486e-03	9.736e-05	1.771e-06
##	2506	3.977e-01	1.407e-01	3.237e-11	2.952e-11	1.539e-11	2.655e-15
##	2507	3.018e-01	9.718e-01	7.717e-04	2.008e-04	8.062e-05	2.085e-06
##	2508	3.628e-01	7.624e-01	4.353e-03	2.876e-03	1.114e-04	3.187e-05
##	2600	4.796e-01	8.012e-01	6.561e-04	5.575e-06	9.051e-04	2.948e-04
##	2601	8.912e-02	8.077e-02	7.777e-02	1.548e-03	7.284e-01	7.884e-01
##	2602	7.293e-01	2.709e-01	2.803e-04	1.861e-01	2.735e-01	6.770e-30
##	2603	1.819e-01	1.175e-01	3.363e-01	2.002e-01	7.216e-01	5.573e-02
##	2604	4.819e-01	4.178e-01	1.403e-09	6.801e-12	6.662e-03	2.612e-05
##	2605	8.917e-01	3.169e-01	2.525e-01	6.313e-02	8.005e-02	7.107e-03
##	2606	7.121e-03	4.865e-02	1.407e-02	8.932e-04	2.939e-02	4.736e-03

##	2607	2.636e-01	1.046e-01	7.534e-01	5.380e-01	7.471e-01	2.501e-01
##	2608	9.147e-02	6.509e-01	4.048e-03	1.600e-02	1.060e-01	1.608e-02
##	2609	1.430e-01	6.657e-02	1.748e-03	1.020e-01	2.564e-22	7.183e-01
##	2610	6.580e-01	2.120e-01	3.091e-02	1.746e-01	2.693e-02	2.035e-01
##	2611	1.748e-02	2.090e-01	7.424e-04	5.894e-08	1.296e-03	6.276e-03
##	2612	3.614e-01	2.418e-01	2.196e-02	3.452e-02	4.809e-01	5.468e-01
##	2613	4.305e-01	9.640e-01	2.079e-05	2.295e-06	2.233e-05	2.726e-06
##	2614	8.463e-01	2.631e-01	4.142e-01	9.252e-02	1.212e-01	3.094e-01
##	2700	6.858e-01	3.780e-01	5.603e-83	6.527e-133	7.795e-166	0.000e+00
##	2701	5.977e-01	2.723e-02	7.112e-04	7.033e-04	3.503e-05	4.796e-10
##	2702	2.282e-01	2.640e-01	3.246e-01	4.951e-02	1.681e-02	4.581e-03
##	2703	6.224e-01	7.160e-05	2.574e-08	8.921e-15	3.800e-20	8.859e-27
##	2704	8.214e-01	8.622e-02	4.687e-02	2.975e-03	1.442e-03	2.959e-05
##	2705	8.784e-04	9.866e-04	4.956e-10	1.831e-21	1.290e-25	7.226e-77
##	2706	6.751e-01	7.913e-01	2.126e-07	1.156e-16	8.566e-24	1.433e-33
##	2707	8.946e-01	7.185e-01	5.027e-01	4.343e-03	4.049e-02	1.818e-04
##	2708	6.626e-01	9.557e-01	1.404e-04	4.179e-03	6.625e-04	4.601e-07
##	2709	NA	NA	NA	NA	NA	NA
##	2710	6.453e-01	6.166e-02	1.722e-01	1.448e-01	1.359e-01	1.255e-02
##	2711	8.586e-01	6.267e-01	6.656e-12	3.420e-21	7.698e-31	1.829e-49
##	2712	3.156e-02	1.474e-02	4.265e-06	7.775e-07	6.472e-08	3.436e-13
##	2713	1.687e-01	9.230e-01	8.536e-03	2.681e-04	1.478e-04	4.132e-06
##	2714	3.578e-01	1.609e-02	7.264e-03	5.114e-03	4.343e-05	2.312e-13
##	2715	5.237e-01	3.313e-01	2.114e-02	2.201e-06	4.228e-07	1.012e-18
##	2716	9.296e-01	2.731e-02	7.069e-03	7.373e-03	9.951e-03	2.345e-05
##	2717	2.462e-02	1.675e-01	4.613e-08	6.293e-17	7.685e-19	3.391e-28
##	2718	3.957e-01	7.171e-01	1.016e-02	1.823e-02	3.321e-05	9.861e-08
##	2719	5.303e-02	4.204e-01	5.381e-05	8.275e-11	5.983e-19	4.557e-28
##	2720	1.630e-01	3.341e-02	5.559e-02	1.208e-05	6.344e-06	7.965e-14
##	2721	5.012e-01	7.810e-01	5.838e-01	2.912e-01	5.213e-01	2.547e-02
##	2722	9.238e-02	3.339e-01	2.294e-02	2.505e-02	5.819e-02	1.550e-02
##	2723	1.108e-01	8.041e-01	4.840e-04	1.591e-04	9.219e-09	5.832e-15
##	2724	1.196e-04	2.014e-02	7.520e-08	3.814e-11	1.300e-13	2.296e-25
##	2725	1.365e-01	5.203e-04	2.793e-07	7.363e-10	8.409e-13	5.309e-18
##	2726	6.458e-01	1.296e-02	6.562e-06	6.672e-09	2.443e-14	1.792e-22
##	2727	8.091e-01	5.000e-02	1.622e-02	4.870e-04	1.310e-04	3.371e-12
##	2728	1.586e-02	7.445e-07	1.077e-08	2.243e-11	3.529e-16	3.976e-26
##	2729	1.246e-03	5.531e-01	5.754e-10	4.199e-12	1.007e-13	1.800e-28
##	2730	1.102e-01	1.162e-02	8.412e-17	7.338e-26	4.880e-38	3.350e-65
##	2731	1.100e-02	8.282e-01	3.176e-04	1.999e-03	5.575e-06	5.333e-10
##	2732	1.091e-02	2.090e-01	3.866e-10	1.038e-15	1.920e-18	3.932e-24
##	2733	2.409e-01	6.050e-03	3.325e-02	9.161e-03	4.876e-04	8.893e-07
##	2734	9.053e-01	3.707e-01	6.811e-05	1.836e-05	3.706e-05	1.727e-12
##	2735	4.357e-03	1.846e-01	6.233e-13	9.800e-17	3.991e-28	5.560e-50
##	2736	8.464e-01	9.764e-06	3.382e-09	3.291e-12	8.412e-17	3.537e-28
##	2737	7.990e-01	2.243e-01	6.104e-03	1.305e-03	1.737e-03	3.106e-08
##	2738	2.535e-01	1.888e-03	2.072e-33	5.900e-47	1.397e-52	2.629e-80
##	2739	8.810e-01	2.105e-01	3.622e-09	1.033e-17	5.486e-21	4.125e-38
##	2740	8.848e-01	7.722e-02	1.507e-14	1.257e-20	3.788e-32	1.864e-60
##	2741	9.554e-01	8.750e-01	1.509e-02	2.082e-05	1.203e-09	1.234e-27

##	2742	4.661e-01	8.910e-02	6.804e-10	3.096e-16	1.627e-19	4.894e-35
##	2743	3.033e-02	6.164e-01	3.652e-02	1.320e-02	1.768e-02	3.291e-03
##	2744	NA	NA	NA	NA	NA	NA
##	2745	2.538e-01	4.271e-03	4.910e-01	8.734e-02	2.922e-03	4.915e-05
##	2746	1.118e-03	6.980e-01	4.602e-10	7.082e-14	2.141e-24	8.262e-49
##	2747	7.074e-01	6.519e-01	4.638e-02	1.340e-01	5.795e-02	2.468e-04
##	2748	3.633e-01	4.063e-01	1.072e-06	1.079e-12	1.048e-17	1.110e-40
##	2800	6.234e-01	3.412e-01	1.420e-07	2.347e-08	1.235e-09	2.298e-16
##	2801	NA	NA	NA	NA	NA	NA
##	2802	1.514e-01	1.822e-01	9.484e-02	7.049e-02	9.153e-02	1.948e-03
##	2803	1.552e-02	4.924e-01	1.436e-02	4.157e-03	8.538e-04	6.712e-04
##	2804	3.047e-01	8.306e-01	9.723e-01	8.663e-01	5.250e-01	1.494e-01
##	2805	3.556e-01	2.386e-01	1.375e-01	2.822e-01	4.757e-01	2.017e-02
##	2806	5.281e-01	2.987e-01	8.205e-03	4.089e-02	2.382e-02	2.034e-03
##	2807	2.060e-01	5.794e-02	8.114e-01	9.655e-01	5.328e-01	2.526e-01
##	2808	1.265e-01	4.213e-05	1.913e-09	3.388e-13	3.457e-12	3.003e-20
##	2809	6.046e-01	6.486e-01	2.357e-04	8.650e-05	1.855e-05	2.438e-07
##	2900	5.079e-03	5.650e-01	1.975e-23	1.307e-24	9.080e-32	1.407e-48
##	2901	9.205e-02	9.342e-01	8.049e-01	1.308e-03	5.847e-01	9.313e-01
##	2902	4.525e-03	6.887e-01	4.920e-04	1.384e-09	1.092e-14	1.371e-29
##	2903	NA	NA	NA	NA	NA	NA
##	2904	7.026e-04	6.729e-01	7.302e-01	1.898e-02	3.154e-06	4.665e-02
##	2905	3.812e-01	3.362e-02	3.273e-01	1.301e-03	4.962e-01	4.148e-01
##	2906	2.110e-01	5.266e-02	2.202e-02	9.285e-04	1.225e-04	1.271e-05
##	2907	5.001e-01	6.927e-01	7.329e-02	4.017e-05	4.521e-05	4.715e-09
##	2908	7.283e-01	5.682e-01	3.414e-03	2.355e-03	5.110e-03	3.488e-08
##	2909	1.018e-02	2.460e-02	8.347e-03	1.321e-10	3.834e-09	4.187e-10
##	2910	1.732e-01	4.854e-01	3.938e-04	4.910e-03	2.161e-04	1.095e-05
##	2911	3.677e-01	6.549e-02	9.399e-02	1.089e-03	1.388e-06	3.397e-05
##	2912	3.330e-01	5.836e-01	9.377e-03	2.376e-04	5.589e-06	6.052e-02
##	2913	6.085e-02	3.442e-01	9.513e-01	6.259e-01	2.360e-01	1.996e-02
##	2914	2.210e-01	4.073e-01	2.515e-01	9.780e-03	1.659e-02	3.018e-01
##	2915	NA	NA	NA	NA	NA	NA
##	2916	1.426e-01	2.312e-04	4.522e-06	1.178e-06	8.680e-07	5.103e-09
##	2917	1.249e-03	4.774e-01	2.200e-02	1.878e-03	8.149e-03	2.792e-03
##	2918	NA	NA	NA	NA	NA	NA
##	2919	1.660e-02	5.169e-01	8.061e-01	2.409e-02	1.978e-02	2.237e-06
##	2920	NA	NA	NA	NA	NA	NA
##	2921	1.198e-01	2.717e-02	1.922e-02	2.985e-04	3.762e-08	1.242e-05
##	2922	9.375e-02	9.942e-01	3.944e-01	2.205e-03	8.096e-01	2.169e-02
##	2923	9.929e-01	3.528e-01	9.568e-01	1.649e-01	5.266e-01	4.892e-01
##	3000	9.422e-01	3.457e-01	6.184e-01	1.337e-01	9.579e-02	9.729e-02
##	3001	NA	NA	NA	NA	NA	NA
##	3002	7.964e-01	5.545e-01	1.982e-02	1.980e-02	4.712e-03	2.134e-04
##	3003	5.869e-05	7.005e-01	2.889e-25	2.377e-31	9.280e-35	4.254e-58
##	3004	2.820e-01	4.210e-01	8.823e-06	1.331e-07	2.558e-09	1.843e-13
##	3005	2.979e-01	6.507e-02	1.397e-01	7.782e-02	1.095e-02	4.484e-03
##	3100	3.987e-01	1.144e-01	2.248e-16	1.550e-13	4.320e-17	2.529e-12
##	3101	1.490e-01	5.471e-01	3.804e-04	2.711e-05	7.768e-03	3.213e-06
##	3102	4.144e-01	4.233e-01	2.878e-08	7.613e-08	1.000e-04	4.134e-06

```

## 3103 6.529e-02 6.062e-02 7.878e-05 1.783e-06 2.540e-04 6.915e-04
## 3104 5.134e-01 1.427e-01 3.414e-03 7.611e-02 8.945e-03 3.097e-07
## 3105 2.966e-02 6.471e-01 9.516e-06 2.430e-04 7.628e-05 8.377e-05
## 3106 1.533e-01 7.997e-01 2.965e-03 1.783e-04 3.078e-06 2.438e-01
## 3107 4.535e-01 1.067e-01 1.560e-02 2.342e-03 4.784e-08 1.782e-12
## 3108 3.701e-01 6.327e-01 2.528e-02 1.724e-04 2.018e-03 4.260e-08
## 3109 3.551e-02 3.364e-01 2.645e-02 3.364e-02 1.120e-02 1.736e-03
## 3110 8.280e-02 3.180e-01 1.752e-01 4.890e-01 3.112e-01 7.425e-02
## 3200 3.368e-01 8.245e-02 3.154e-18 5.687e-23 4.254e-19 5.246e-23
## 3201 5.816e-01 3.037e-01 5.315e-01 9.079e-02 7.490e-01 6.380e-01
## 3202 2.889e-01 3.324e-01 1.589e-09 4.143e-13 9.246e-08 2.102e-10
## 3203 2.929e-01 2.217e-01 1.036e-25 1.580e-24 2.408e-31 1.807e-37
## 3204 2.909e-02 3.730e-01 2.045e-15 6.499e-21 2.410e-14 4.012e-24
## 3205 7.542e-02 3.629e-01 1.378e-01 1.668e-02 2.407e-02 3.960e-05
## 3206 9.301e-02 1.677e-01 3.231e-04 1.642e-04 5.093e-02 2.441e-05
## 3207 1.824e-01 1.103e-01 8.001e-19 1.337e-17 1.550e-14 2.996e-11
## 3300 2.634e-02 3.519e-02 2.337e-04 1.725e-07 3.400e-02 3.853e-02
## 3301 4.924e-02 2.093e-01 2.235e-13 1.025e-15 1.927e-13 2.384e-14
## 3302 4.944e-02 5.497e-02 6.535e-02 2.499e-07 9.759e-06 1.528e-03
## 3303 1.303e-01 6.416e-01 1.952e-03 4.204e-04 8.866e-01 1.982e-01
## 3304 4.554e-02 7.619e-03 1.335e-33 1.338e-44 7.997e-42 1.059e-44
## 3305 6.967e-02 7.797e-02 1.183e-07 6.754e-07 2.324e-06 1.781e-13
## 3306 8.053e-01 7.387e-02 1.106e-04 1.287e-16 2.668e-12 1.245e-18
## 3307 6.368e-01 2.916e-02 2.008e-02 4.735e-03 7.028e-02 1.398e-03
## 3308 4.856e-01 3.217e-01 2.808e-20 1.925e-26 1.393e-14 1.554e-19
## 3309 1.048e-01 3.272e-01 1.952e-16 6.357e-09 8.184e-06 1.766e-02
## 3310 7.860e-01 2.998e-01 1.196e-48 2.359e-28 1.802e-18 1.099e-06
## 3311 1.297e-01 3.479e-01 1.015e-03 2.296e-03 9.726e-01 7.597e-03
## 3312 1.901e-01 4.346e-01 3.266e-23 1.681e-22 1.243e-10 1.726e-09
## 3313 2.655e-01 4.199e-01 4.083e-06 7.011e-07 1.577e-01 6.465e-05
## 3314 8.259e-01 3.779e-01 8.182e-10 1.565e-03 1.765e-02 5.638e-03
## 3315 2.650e-02 4.875e-01 6.809e-07 1.434e-06 1.223e-04 1.528e-03
## 3316 8.902e-01 2.559e-01 1.032e-04 1.281e-04 6.522e-01 3.336e-01
## 3317 1.469e-02 8.439e-02 1.332e-02 1.322e-01 3.357e-03 8.395e-03
## 3318 7.497e-01 7.589e-01 1.886e-07 3.379e-05 1.554e-02 6.686e-07
## 3319 3.992e-01 2.063e-01 4.225e-04 1.111e-03 5.846e-03 2.689e-03
## 3320 2.668e-01 5.042e-01 5.395e-02 1.010e-03 2.028e-01 2.702e-04
## 3321 9.830e-01 9.243e-01 9.751e-02 8.765e-01 6.142e-01 2.366e-01
## 3322 5.490e-02 6.627e-01 2.834e-01 3.275e-01 2.081e-01 8.956e-01
## 3400 4.442e-02 1.500e-01 2.026e-10 3.782e-20 5.181e-35 1.879e-52
## 3401 NA NA NA NA NA NA NA
## 3402 NA NA NA NA NA NA NA
## 3403 1.024e-01 7.210e-01 1.235e-01 3.526e-01 4.833e-02 2.087e-02
## 3404 5.061e-01 3.452e-01 2.779e-01 1.622e-01 1.606e-01 5.568e-01
## 3500 9.210e-01 5.228e-01 2.528e-09 1.886e-15 1.152e-11 6.231e-17
## 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 8.828e-01 7.963e-01 9.449e-02 1.479e-02 6.377e-03 1.371e-04
## 3505 4.395e-01 5.648e-01 7.270e-01 8.475e-01 7.667e-01 6.224e-01

```

## 3506	NA	NA	NA	NA	NA	NA
## 3600	4.267e-01	6.057e-01	3.015e-01	1.963e-02	9.912e-02	1.154e-05
## 3601	3.391e-01	1.959e-01	9.000e-04	1.938e-03	5.582e-04	5.428e-03
## 3602	5.794e-01	7.401e-02	1.428e-01	7.621e-03	3.021e-04	5.047e-03
## 3603	9.501e-01	1.434e-01	3.781e-01	8.594e-01	9.090e-01	8.027e-01
## 3604	NA	NA	NA	NA	NA	NA
## 3605	1.803e-01	9.739e-01	2.769e-01	3.937e-02	3.394e-03	1.597e-02
## 3606	NA	NA	NA	NA	NA	NA
## 3607	3.299e-01	5.266e-01	3.614e-02	1.530e-01	4.774e-02	1.063e-02
## 3608	NA	NA	NA	NA	NA	NA
## 3609	9.057e-01	8.630e-01	9.305e-02	2.758e-01	5.965e-03	4.215e-05
## 3610	7.316e-04	8.436e-04	2.108e-01	1.313e-02	1.706e-03	4.320e-03
## 3611	4.884e-01	1.087e-01	1.059e-02	2.443e-02	3.091e-02	1.025e-05
## 3612	1.343e-01	8.370e-01	1.166e-10	1.015e-18	2.311e-23	1.252e-35
## 3613	NA	NA	NA	NA	NA	NA
## 3614	1.300e-02	2.569e-02	7.049e-04	3.069e-07	5.669e-07	2.007e-13
## 3615	NA	NA	NA	NA	NA	NA
## 3616	5.919e-01	3.533e-01	1.826e-02	1.063e-01	1.068e-01	1.284e-01
##	FFA2p	FLA2p	FFA3p	FLA4p		
## 1000	9.151e-01	4.381e-01	9.815e-01	4.514e-01		
## 1100	8.075e-01	6.927e-01	8.460e-01	7.141e-01		
## 1101	9.484e-02	1.714e-01	1.425e-01	3.170e-01		
## 1102	2.509e-04	8.116e-01	2.643e-04	7.828e-01		
## 1103	3.296e-01	3.321e-01	4.228e-01	4.224e-01		
## 1104	8.240e-01	4.135e-01	9.040e-01	4.283e-01		
## 1105	1.883e-03	6.855e-03	7.516e-03	2.698e-02		
## 1106	6.195e-03	4.943e-01	9.018e-03	7.656e-01		
## 1107	2.965e-01	9.215e-01	2.852e-01	7.770e-01		
## 1108	1.219e-02	4.433e-01	1.092e-02	3.749e-01		
## 1109	1.311e-08	6.635e-01	1.705e-08	6.273e-01		
## 1110	2.528e-06	3.462e-01	1.016e-06	1.007e-01		
## 1111	5.449e-01	4.914e-01	5.827e-01	5.199e-01		
## 1200	9.349e-01	1.362e-01	5.865e-02	2.371e-02		
## 1201	3.206e-02	5.487e-01	4.910e-02	6.591e-01		
## 1202	4.296e-02	4.847e-01	2.614e-03	1.618e-02		
## 1203	6.590e-02	7.365e-01	4.205e-03	2.480e-02		
## 1204	8.054e-04	3.698e-02	2.519e-02	8.312e-01		
## 1205	4.887e-01	6.515e-01	3.584e-01	4.419e-01		
## 1206	5.234e-01	3.220e-01	8.361e-01	5.003e-01		
## 1207	3.981e-07	7.794e-02	2.757e-16	1.232e-11		
## 1208	7.324e-01	6.389e-01	2.194e-03	2.167e-03		
## 1209	8.649e-01	2.519e-01	2.154e-01	1.002e-01		
## 1210	2.402e-01	4.274e-01	2.184e-01	5.533e-01		
## 1211	2.678e-01	1.033e-01	2.932e-01	1.088e-01		
## 1212	8.199e-01	7.651e-01	1.417e-01	1.361e-01		
## 1213	1.878e-02	2.045e-01	8.382e-03	5.313e-02		
## 1300	2.635e-02	5.939e-01	2.079e-02	4.069e-01		
## 1301	NA	NA	NA	NA		
## 1302	2.206e-02	3.940e-01	3.859e-02	9.365e-01		
## 1303	3.041e-01	4.664e-01	2.790e-01	4.199e-01		

```
## 1304 5.639e-01 1.627e-01 4.743e-01 1.454e-01
## 1305 1.982e-01 2.013e-02 1.297e-01 1.449e-02
## 1306 2.146e-01 9.730e-01 2.149e-01 8.905e-01
## 1307 3.879e-01 6.191e-01 3.693e-01 5.779e-01
## 1308 5.365e-01 2.731e-02 4.212e-01 2.326e-02
## 1309 6.140e-01 1.475e-01 7.331e-01 1.624e-01
## 1310 6.806e-01 1.017e-01 5.192e-01 8.946e-02
## 1311 5.103e-01 2.215e-01 4.044e-01 1.863e-01
## 1312 2.570e-01 6.269e-01 2.416e-01 5.667e-01
## 1313 6.294e-01 1.380e-01 7.476e-01 1.518e-01
## 1314 5.171e-01 4.956e-02 6.184e-01 5.552e-02
## 1315 4.460e-03 5.699e-01 5.309e-03 8.657e-01
## 1400 4.067e-01 6.583e-01 4.691e-01 9.548e-01
## 1401 7.199e-01 5.382e-01 9.415e-01 6.179e-01
## 1402 9.838e-01 5.886e-01 7.301e-01 5.263e-01
## 1403 8.010e-01 5.483e-02 4.250e-01 3.960e-02
## 1404 2.664e-02 5.896e-01 1.294e-02 4.058e-01
## 1405 9.317e-01 7.726e-01 9.413e-01 7.677e-01
## 1406 4.683e-01 1.705e-01 8.927e-01 2.586e-01
## 1407 2.316e-01 3.555e-01 5.659e-02 8.039e-02
## 1408 4.518e-01 2.580e-01 8.248e-01 3.842e-01
## 1409 9.619e-01 6.691e-01 8.130e-01 6.435e-01
## 1410 2.434e-01 1.082e-01 8.365e-01 2.515e-01
## 1500 5.600e-01 8.343e-01 5.501e-01 8.033e-01
## 1501 3.014e-01 4.825e-01 3.059e-01 3.482e-01
## 1502 6.705e-01 4.454e-02 9.490e-01 5.123e-02
## 1503 1.039e-02 6.804e-01 9.419e-03 5.604e-01
## 1504 1.453e-01 4.760e-01 1.730e-01 8.156e-01
## 1505 6.432e-03 8.874e-01 5.038e-03 5.787e-01
## 1506 1.621e-01 1.714e-01 2.019e-01 2.160e-01
## 1507 5.500e-01 3.302e-01 7.877e-01 3.741e-01
## 1508 4.232e-02 8.036e-01 3.359e-02 6.111e-01
## 1600 7.949e-01 2.978e-01 7.368e-01 2.894e-01
## 1601 3.313e-01 4.811e-01 2.729e-01 4.276e-01
## 1602 9.671e-01 2.105e-01 9.916e-01 2.121e-01
## 1603 9.267e-01 2.988e-01 8.947e-01 2.953e-01
## 1604 6.068e-01 8.233e-01 5.980e-01 8.038e-01
## 1605 5.634e-04 6.346e-01 5.062e-04 4.886e-01
## 1606 1.212e-01 7.434e-01 1.259e-01 8.089e-01
## 1607 4.223e-01 7.408e-01 4.155e-01 7.189e-01
## 1700 4.721e-01 8.170e-01 4.154e-01 6.600e-01
## 1701 7.202e-01 3.328e-03 2.930e-01 4.518e-03
## 1702 5.670e-01 4.041e-02 7.278e-01 5.824e-02
## 1703 7.461e-01 1.573e-01 3.622e-01 1.090e-01
## 1704 8.338e-01 9.673e-01 8.418e-01 9.959e-01
## 1705 4.524e-01 1.290e-01 7.607e-01 1.764e-01
## 1706 5.582e-01 1.572e-01 9.744e-01 2.041e-01
## 1707 1.055e-01 3.364e-01 1.489e-01 5.259e-01
## 1708 2.240e-01 4.312e-01 3.024e-01 6.490e-01
## 1709 1.978e-01 7.163e-01 9.855e-02 2.926e-01
```

```
## 1710 1.241e-01 8.121e-01 5.327e-02 2.296e-01
## 1711 9.710e-01 3.368e-01 8.843e-01 3.335e-01
## 1712 1.149e-01 4.264e-01 1.693e-01 7.620e-01
## 1800 4.935e-01 9.295e-02 1.187e-01 3.774e-02
## 1801      NA      NA      NA      NA
## 1802 7.438e-02 5.610e-01 9.725e-02 9.665e-01
## 1803 6.644e-01 3.839e-01 8.485e-01 4.428e-01
## 1804 1.616e-02 1.720e-01 4.538e-02 7.944e-01
## 1900 9.103e-01 9.940e-01 9.076e-01 9.709e-01
## 1901 1.768e-01 1.272e-04 3.200e-02 9.852e-05
## 1902 1.128e-01 1.209e-01 6.293e-02 6.861e-02
## 1903 7.077e-01 2.372e-01 6.916e-01 2.354e-01
## 1904 6.057e-02 6.106e-05 5.618e-01 6.217e-04
## 1905 6.418e-01 7.751e-01 6.683e-01 8.022e-01
## 1906 5.939e-01 3.838e-01 7.015e-01 4.307e-01
## 1907 6.616e-01 6.343e-01 5.822e-01 5.634e-01
## 1908 3.573e-01 2.582e-01 4.806e-01 3.435e-01
## 1909 1.685e-01 3.803e-01 2.371e-01 5.184e-01
## 1910 7.737e-02 6.569e-01 6.330e-02 4.458e-01
## 1911 3.365e-01 2.729e-02 7.224e-01 5.193e-02
## 1912 1.141e-04 6.918e-02 3.638e-04 1.867e-01
## 1913 8.101e-01 4.285e-02 9.739e-01 5.541e-02
## 2000 2.094e-01 9.795e-01 1.042e-01 2.473e-01
## 2001 1.307e-01 8.465e-01 1.413e-01 5.644e-01
## 2002 1.752e-02 6.785e-01 1.015e-03 1.856e-02
## 2003 9.629e-04 2.316e-02 1.752e-02 5.297e-01
## 2100      NA      NA      NA      NA
## 2101 6.514e-02 1.845e-01 1.121e-01 6.034e-01
## 2102 9.801e-01 9.824e-01 9.758e-01 9.787e-01
## 2103 2.690e-01 3.114e-01 3.590e-01 4.092e-01
## 2104 4.254e-01 7.452e-01 3.734e-01 6.587e-01
## 2105 5.127e-02 1.320e-01 1.058e-01 2.642e-01
## 2200 4.323e-01 6.207e-02 1.096e-01 2.057e-02
## 2201 2.892e-02 1.432e-02 2.230e-01 7.625e-02
## 2202 7.798e-01 8.866e-01 8.065e-01 9.660e-01
## 2203 8.986e-01 6.191e-01 9.670e-01 6.324e-01
## 2204 5.936e-01 6.446e-01 6.456e-01 7.137e-01
## 2205 8.210e-01 3.839e-01 9.038e-01 3.975e-01
## 2206 6.964e-01 4.433e-01 8.477e-01 4.552e-01
## 2207 8.189e-01 7.117e-01 9.050e-01 7.548e-01
## 2208 6.981e-01 9.396e-01 7.081e-01 9.840e-01
## 2209 1.305e-01 6.030e-01 1.626e-01 7.744e-01
## 2210 1.318e-01 3.224e-01 1.706e-01 4.125e-01
## 2211 4.917e-03 9.275e-03 1.587e-03 2.378e-03
## 2212 2.669e-01 2.925e-01 4.183e-01 3.734e-01
## 2213 2.946e-01 1.320e-01 1.041e-01 5.068e-02
## 2214 6.601e-02 3.902e-01 3.012e-02 1.288e-01
## 2215 7.486e-01 9.182e-01 7.575e-01 9.446e-01
## 2216 6.685e-01 7.497e-01 7.709e-01 9.251e-01
## 2300 7.745e-03 5.205e-01 3.936e-03 1.881e-01
```

```
## 2301 2.837e-03 7.702e-01 1.378e-04 6.950e-03
## 2302 4.557e-01 2.823e-01 3.953e-01 2.474e-01
## 2303 3.914e-03 1.883e-01 8.506e-03 4.169e-01
## 2304 8.272e-01 8.752e-01 8.152e-01 8.588e-01
## 2305 2.796e-01 2.559e-04 1.597e-01 1.844e-04
## 2306 2.238e-01 2.448e-02 9.754e-02 1.431e-02
## 2307 6.309e-01 7.008e-01 6.100e-01 6.726e-01
## 2308 7.504e-01 7.295e-01 8.489e-01 8.173e-01
## 2309 8.298e-01 3.119e-02 5.174e-01 2.698e-02
## 2310 9.033e-01 3.461e-01 9.617e-01 3.498e-01
## 2311 8.125e-01 5.355e-02 6.469e-01 5.261e-02
## 2312 5.974e-01 7.909e-01 5.731e-01 7.347e-01
## 2400 4.208e-01 5.281e-01 4.499e-01 5.703e-01
## 2401      NA      NA      NA      NA
## 2402 1.261e-02 2.883e-01 8.301e-03 1.777e-01
## 2403 8.857e-01 7.248e-01 8.658e-01 7.173e-01
## 2404 1.245e-01 4.247e-01 1.281e-01 4.399e-01
## 2405 6.666e-04 8.040e-01 7.062e-04 9.758e-01
## 2406 1.044e-01 7.581e-01 1.035e-01 7.432e-01
## 2500 4.997e-02 6.253e-01 4.537e-02 5.054e-01
## 2501 9.925e-01 3.730e-01 8.022e-01 3.209e-01
## 2502 9.815e-01 1.148e-03 6.680e-01 1.087e-03
## 2503 9.175e-01 9.382e-02 8.626e-01 9.146e-02
## 2504 3.881e-01 7.609e-01 3.685e-01 7.040e-01
## 2505 4.036e-01 2.925e-01 3.680e-01 2.698e-01
## 2506 7.694e-02 1.217e-01 1.286e-01 1.791e-01
## 2507 5.829e-01 6.652e-01 6.231e-01 7.181e-01
## 2508 1.784e-01 5.714e-01 1.672e-01 5.155e-01
## 2600 6.305e-01 9.590e-01 5.874e-01 7.501e-01
## 2601 4.042e-02 4.436e-02 4.555e-01 4.178e-01
## 2602 3.343e-01 1.626e-01 9.178e-01 3.085e-01
## 2603 1.656e-01 1.140e-01 6.665e-01 4.739e-01
## 2604 9.205e-01 4.034e-01 6.647e-01 3.620e-01
## 2605 7.513e-01 2.237e-01 8.283e-01 2.248e-01
## 2606 1.538e-02 7.375e-03 1.845e-01 1.093e-01
## 2607 2.808e-01 1.179e-01 6.363e-01 2.127e-01
## 2608 2.191e-01 7.067e-01 2.526e-01 7.292e-01
## 2609 3.921e-02 2.777e-02 7.404e-01 2.825e-01
## 2610 5.409e-01 2.423e-01 8.474e-01 3.187e-01
## 2611 3.041e-03 2.529e-01 6.457e-03 5.789e-01
## 2612 4.785e-01 3.189e-01 7.554e-01 4.987e-01
## 2613 7.951e-01 9.506e-01 7.567e-01 8.516e-01
## 2614 9.613e-01 2.956e-01 6.495e-01 2.636e-01
## 2700 4.676e-04 2.822e-03 1.355e-02 7.996e-02
## 2701 9.641e-01 7.751e-03 4.042e-01 5.951e-03
## 2702 1.921e-01 4.939e-01 2.320e-01 7.473e-01
## 2703 5.709e-01 6.591e-05 2.738e-01 4.504e-05
## 2704 4.986e-01 9.447e-02 4.105e-01 7.998e-02
## 2705 5.856e-05 2.920e-02 1.040e-05 4.421e-03
## 2706 8.796e-01 2.524e-01 9.285e-01 2.542e-01
```



```
## 2707 3.170e-01 6.920e-01 3.770e-01 9.974e-01
## 2708 1.862e-01 7.124e-01 1.691e-01 6.139e-01
## 2709      NA      NA      NA      NA
## 2710 9.544e-01 2.643e-02 9.803e-01 2.858e-02
## 2711 1.918e-01 6.220e-01 1.693e-01 4.971e-01
## 2712 1.806e-01 1.085e-03 7.062e-02 5.130e-04
## 2713 3.770e-02 9.788e-01 3.430e-02 6.112e-01
## 2714 5.341e-02 1.900e-02 2.008e-03 7.861e-04
## 2715 6.718e-02 2.373e-01 3.488e-02 1.190e-01
## 2716 5.540e-01 3.337e-02 9.179e-01 4.315e-02
## 2717 3.133e-03 7.544e-01 1.922e-03 2.085e-01
## 2718 1.994e-01 8.685e-01 2.135e-01 8.014e-01
## 2719 2.607e-05 2.496e-01 7.643e-05 6.119e-01
## 2720 2.849e-02 2.276e-01 1.970e-02 1.333e-01
## 2721 4.838e-01 9.730e-01 4.825e-01 9.609e-01
## 2722 1.164e-01 2.165e-01 7.262e-02 1.273e-01
## 2723 6.456e-01 6.978e-01 6.164e-01 6.650e-01
## 2724 8.514e-03 5.798e-04 1.670e-03 1.481e-04
## 2725 7.913e-01 8.024e-04 6.475e-01 7.678e-04
## 2726 2.569e-01 8.209e-02 2.832e-01 8.963e-02
## 2727 7.248e-01 1.136e-01 4.947e-01 9.118e-02
## 2728 5.167e-04 6.625e-06 3.544e-05 4.813e-07
## 2729 2.827e-05 1.106e-01 1.348e-04 6.394e-01
## 2730 1.018e-02 6.025e-02 5.069e-03 2.617e-02
## 2731 2.839e-03 6.179e-01 2.145e-03 3.870e-01
## 2732 5.535e-03 9.819e-01 5.334e-03 5.926e-01
## 2733 6.328e-01 3.778e-02 2.784e-01 2.212e-02
## 2734 5.702e-01 6.111e-01 5.061e-01 5.354e-01
## 2735 1.654e-04 9.876e-01 1.189e-04 4.029e-01
## 2736 3.841e-01 1.232e-07 9.822e-01 2.323e-07
## 2737 8.176e-01 6.866e-02 9.483e-01 7.172e-02
## 2738 7.807e-03 7.184e-04 4.126e-02 3.938e-03
## 2739 3.679e-01 1.363e-02 9.127e-01 2.334e-02
## 2740 3.170e-01 4.356e-01 2.587e-01 3.470e-01
## 2741 1.676e-01 4.609e-01 2.065e-01 6.101e-01
## 2742 6.879e-01 2.399e-03 1.961e-01 1.277e-03
## 2743 2.108e-02 9.405e-01 2.154e-02 6.919e-01
## 2744      NA      NA      NA      NA
## 2745 5.641e-01 8.536e-03 8.353e-01 1.068e-02
## 2746 1.121e-04 9.092e-01 1.090e-04 5.848e-01
## 2747 5.163e-01 5.743e-01 5.855e-01 6.556e-01
## 2748 3.174e-02 4.340e-01 1.770e-02 1.778e-01
## 2800 3.862e-01 2.415e-01 3.370e-01 2.158e-01
## 2801      NA      NA      NA      NA
## 2802 9.662e-02 2.155e-01 6.785e-02 1.419e-01
## 2803 4.754e-03 5.199e-01 5.693e-03 6.411e-01
## 2804 3.450e-01 7.320e-01 3.630e-01 8.156e-01
## 2805 3.739e-01 3.110e-01 2.669e-01 2.239e-01
## 2806 5.341e-01 3.565e-01 4.237e-01 2.948e-01
## 2807 2.985e-01 9.039e-02 1.420e-01 4.969e-02
```

```
## 2808 1.677e-02 2.317e-04 6.419e-03 9.415e-05
## 2809 5.054e-01 5.851e-01 5.462e-01 6.388e-01
## 2900 2.979e-02 8.136e-02 1.356e-03 3.182e-03
## 2901 2.060e-01 8.585e-01 1.597e-01 5.783e-01
## 2902 9.434e-05 1.491e-05 2.794e-08 1.309e-08
## 2903 8.522e-02 6.527e-01 2.676e-02 1.743e-01
## 2904 5.590e-01 7.326e-01 3.231e-01 4.776e-01
## 2905 2.446e-01 6.881e-02 6.389e-02 2.325e-02
## 2906 1.450e-01 4.957e-03 1.733e-01 7.422e-02
## 2907 6.656e-01 5.303e-01 9.619e-01 6.808e-01
## 2908 2.737e-01 8.833e-01 2.805e-01 6.227e-01
## 2909 4.858e-04 1.619e-01 3.651e-05 6.691e-03
## 2910 1.812e-02 5.848e-01 4.132e-04 6.453e-03
## 2911 1.143e-01 6.353e-02 4.456e-03 NA
## 2912 1.972e-01 6.652e-01 1.931e-01 4.679e-01
## 2913 4.462e-03 1.736e-01 4.342e-02 8.146e-01
## 2914 9.620e-02 6.999e-01 7.707e-02 7.055e-01
## 2915 NA NA NA NA
## 2916 2.671e-01 3.310e-05 1.088e-02 2.819e-06
## 2917 1.546e-04 7.834e-01 8.773e-06 1.576e-02
## 2918 NA NA NA NA
## 2919 1.259e-03 3.016e-01 9.680e-04 6.641e-01
## 2920 NA NA NA NA
## 2921 1.058e-01 3.638e-02 3.300e-03 1.015e-03
## 2922 3.557e-02 6.466e-01 2.508e-02 3.599e-01
## 2923 9.086e-01 4.521e-01 7.293e-01 4.036e-01
## 3000 9.030e-01 4.227e-01 7.656e-01 4.220e-01
## 3001 NA NA NA NA
## 3002 9.501e-01 4.097e-01 9.815e-01 4.142e-01
## 3003 1.039e-06 1.947e-02 1.667e-07 2.132e-03
## 3004 1.504e-01 1.781e-01 1.773e-01 2.075e-01
## 3005 4.169e-01 1.230e-01 5.079e-01 1.446e-01
## 3100 5.123e-02 4.181e-02 1.631e-02 1.400e-02
## 3101 7.236e-01 3.932e-01 8.784e-01 4.279e-01
## 3102 9.694e-01 4.047e-01 7.490e-01 3.699e-01
## 3103 1.577e-03 2.940e-01 3.454e-03 5.085e-01
## 3104 9.708e-01 9.864e-02 8.815e-01 9.834e-02
## 3105 1.543e-01 4.988e-01 1.689e-01 5.496e-01
## 3106 3.629e-01 7.300e-01 4.230e-01 9.566e-01
## 3107 6.795e-01 2.328e-01 5.988e-01 2.172e-01
## 3108 6.772e-01 2.049e-01 5.599e-01 1.879e-01
## 3109 1.147e-02 3.918e-01 1.739e-02 8.747e-01
## 3110 4.437e-02 2.588e-01 4.465e-02 2.651e-01
## 3200 6.375e-01 3.208e-01 8.493e-01 3.848e-01
## 3201 4.580e-01 2.409e-01 7.853e-01 3.418e-01
## 3202 7.156e-01 4.408e-01 5.279e-01 3.617e-01
## 3203 5.271e-01 4.392e-01 6.385e-01 5.255e-01
## 3204 1.496e-03 5.358e-01 7.142e-04 1.329e-01
## 3205 5.639e-02 4.257e-01 3.732e-02 2.366e-01
## 3206 2.379e-02 1.595e-01 1.222e-02 7.060e-02
```

##	3207	8.288e-01	1.110e-01	4.252e-01	8.552e-02
##	3300	5.662e-03	3.151e-02	9.761e-02	7.516e-01
##	3301	1.794e-03	1.988e-01	1.949e-06	1.029e-04
##	3302	2.708e-03	1.316e-02	1.212e-01	5.056e-01
##	3303	6.762e-02	4.445e-01	1.006e-01	4.983e-01
##	3304	4.663e-04	1.124e-02	2.378e-09	4.097e-08
##	3305	8.708e-03	2.098e-01	7.421e-06	1.102e-04
##	3306	4.242e-01	4.780e-01	2.481e-01	2.676e-01
##	3307	5.126e-01	5.366e-02	2.577e-01	3.534e-02
##	3308	2.210e-01	2.694e-01	5.028e-01	6.822e-01
##	3309	1.463e-01	2.834e-01	3.418e-01	9.928e-01
##	3310	1.614e-02	8.553e-01	8.107e-04	1.571e-02
##	3311	1.894e-01	4.067e-01	2.828e-01	9.446e-01
##	3312	1.028e-02	2.676e-01	1.571e-02	3.590e-01
##	3313	4.087e-01	6.521e-01	4.796e-01	8.375e-01
##	3314	2.483e-01	1.303e-01	9.645e-01	3.522e-01
##	3315	2.140e-02	6.268e-01	1.067e-02	1.405e-01
##	3316	5.671e-01	3.888e-01	1.555e-02	1.219e-02
##	3317	1.282e-03	3.034e-01	7.823e-07	1.067e-04
##	3318	8.084e-01	2.817e-01	5.547e-01	2.657e-01
##	3319	4.999e-01	1.645e-01	2.404e-01	8.939e-02
##	3320	3.996e-01	2.528e-01	6.856e-04	4.798e-04
##	3321	9.677e-01	8.942e-01	8.772e-01	8.418e-01
##	3322	4.884e-02	7.076e-01	2.974e-02	2.388e-01
##	3400	5.418e-02	3.741e-02	1.211e-01	7.998e-02
##	3401	NA	NA	NA	NA
##	3402	5.402e-02	6.934e-01	2.396e-02	1.704e-01
##	3403	7.327e-02	5.397e-01	5.729e-02	3.398e-01
##	3404	3.719e-01	2.631e-01	8.812e-02	7.622e-02
##	3500	1.695e-01	5.793e-01	2.341e-01	8.085e-01
##	3501	NA	NA	NA	NA
##	3502	NA	NA	NA	NA
##	3503	NA	NA	NA	NA
##	3504	8.671e-01	3.420e-01	9.045e-01	3.509e-01
##	3505	4.260e-01	3.033e-01	9.444e-01	3.487e-01
##	3506	NA	NA	NA	NA
##	3600	5.419e-01	7.561e-01	5.851e-01	9.421e-01
##	3601	1.444e-01	2.491e-01	4.778e-03	1.737e-02
##	3602	8.250e-01	1.497e-01	9.369e-01	1.608e-01
##	3603	9.541e-01	1.464e-01	3.884e-02	5.797e-03
##	3604	NA	NA	NA	NA
##	3605	2.894e-01	7.944e-01	3.328e-01	9.592e-01
##	3606	NA	NA	NA	NA
##	3607	2.889e-01	6.481e-01	2.581e-01	5.341e-01
##	3608	NA	NA	NA	NA
##	3609	8.521e-01	4.249e-01	7.787e-01	4.126e-01
##	3610	1.006e-04	1.283e-03	2.923e-03	2.160e-02
##	3611	5.805e-01	1.684e-01	7.967e-01	1.927e-01
##	3612	2.135e-02	3.339e-01	4.073e-02	7.757e-01
##	3613	NA	NA	NA	NA

```
## 3614 5.817e-02 3.332e-03 6.592e-03 5.428e-04
## 3615          NA          NA          NA          NA
## 3616 7.915e-01 3.224e-01 9.736e-01 3.581e-01
```

```
print(RegStar)
```

```
##      FFA1p FLA1p 2p    3p    4p    5+p    FFA2p FLA2p FFA3p FLA4p
## 1000 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1100 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1101 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1102 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1103 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1104 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1105 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1106 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1107 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1108 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1109 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1110 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1111 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1200 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1201 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1202 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1203 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1204 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1205 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1206 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1207 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1208 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1209 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1210 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1211 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1212 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1213 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1300 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1301 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1302 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1303 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1304 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1305 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1306 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1307 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1308 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1309 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1310 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1311 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1312 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1313 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1314 ""      ""    ""    ""    ""    ""    ""    ""    ""    ""
## 1315 ""    ""    ""    ""    ""    ""    ""    ""    ""    ""
```

##	1400	"	"	"**"	"**"	"	"	"	"	"	"
##	1401	"	"	"**"	"**"	"	"	"	"	"	"
##	1402	"	"	"**"	"**"	"	"	"	"	"	"
##	1403	"	"*	"**"	"**"	"	"**"	"	"	"	"*
##	1404	"*	"	"**"	"**"	"	"	"**"	"	"**"	"
##	1405	"	"	"	"	"	"	"	"	"	"
##	1406	"	"	"**"	"**"	"	"	"	"	"	"
##	1407	"	"	"*	"*	"**"	"	"	"	"	"
##	1408	"	"	"**"	"**"	"	"**"	"	"	"	"
##	1409	"	"	"	"*	"**"	"	"	"	"	"
##	1410	"	"	"**"	"**"	"	"	"	"	"	"
##	1500	"*	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1501	"	"	"*	"**"	"	"	"	"	"	"
##	1502	"	"*	"**"	"**"	"**"	"**"	"	"**"	"	"
##	1503	"**"	"	"**"	"**"	"**"	"**"	"**"	"	"**"	"
##	1504	"	"	"	"	"	"	"	"	"	"
##	1505	"**"	"	"**"	"*	"**"	"**"	"**"	"	"**"	"
##	1506	"	"	"	"	"	"	"	"	"	"
##	1507	"	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1508	"	"	"**"	"**"	"**"	"**"	"**"	"	"**"	"
##	1600	"	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1601	"	"	"	"	"	"*	"	"	"	"
##	1602	"	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1603	"	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1604	"	"	"	"	"	"*	"	"	"	"
##	1605	"**"	"	"	"*	"**"	"**"	"**"	"	"**"	"
##	1606	"	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1607	"	"	"	"	"	"	"	"	"	"
##	1700	"	"	"	"**"	"**"	"**"	"	"	"	"
##	1701	"	"*	"	"	"	"	"	"**"	"	"**"
##	1702	"	"*	"**"	"**"	"**"	"**"	"	"**"	"	"
##	1703	"	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1704	"	"	"*	"**"	"*	"	"	"	"	"
##	1705	"	"	"**"	"**"	"	"**"	"	"	"	"
##	1706	"	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1707	"	"	"	"	"*	"	"	"	"	"
##	1708	"	"	"	"**"	"	"**"	"	"	"	"
##	1709	"	"	"	"**"	"**"	"**"	"	"	"	"
##	1710	"	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1711	"	"	"*	"	"**"	"	"	"	"	"
##	1712	"	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1800	"	"*	"	"*	"**"	"**"	"	"	"	"*
##	1801	"	"	"	"	"	"	"	"	"	"
##	1802	"	"	"**"	"**"	"**"	"	"	"	"	"
##	1803	"	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1804	"**"	"	"**"	"**"	"	"	"**"	"	"**"	"
##	1900	"	"	"**"	"**"	"**"	"**"	"	"	"	"
##	1901	"	"**"	"**"	"**"	"**"	"**"	"	"**"	"**"	"**"
##	1902	"	"	"	"	"*	"**"	"	"	"	"
##	1903	"	"	"	"*	"**"	"**"	"	"	"	"

##	1904	" "	"**"	"**"	"**"	"**"	"**"	" "	"**"	" "	"**"
##	1905	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	1906	" "	" "	" "	" "	"*	" "	" "	" "	" "	" "
##	1907	" "	" "	"*	"**"	"**"	"**"	" "	" "	" "	" "
##	1908	" "	" "	" "	" "	"**"	"**"	" "	" "	" "	" "
##	1909	" "	" "	"**"	"**"	"*	" "	" "	" "	" "	" "
##	1910	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	1911	" "	" "	"*	"**"	"**"	"**"	" "	"*	" "	" "
##	1912	"**"	"**"	" "	"**"	"**"	"**"	"**"	" "	"**"	" "
##	1913	" "	"**"	" "	" "	"**"	"**"	" "	"**"	" "	" "
##	2000	" "	" "	" "	" "	"**"	" "	" "	" "	" "	" "
##	2001	"**"	" "	"**"	"**"	"**"	" "	" "	" "	" "	" "
##	2002	"**"	" "	"**"	"**"	"**"	" "	"**"	" "	"**"	"**"
##	2003	"**"	"**"	"**"	"**"	" "	" "	"**"	"**"	"**"	" "
##	2100	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	2101	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2102	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2103	"**"	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2104	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	2105	"**"	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2200	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	"**"
##	2201	" "	" "	"**"	"**"	"**"	"**"	"**"	"**"	" "	" "
##	2202	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2203	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2204	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2205	" "	" "	"**"	"**"	" "	" "	" "	" "	" "	" "
##	2206	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	2207	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2208	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2209	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2210	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2211	" "	" "	"**"	"**"	"**"	"**"	"**"	"**"	"**"	"**"
##	2212	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2213	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2214	" "	" "	"**"	"**"	" "	"**"	" "	" "	"**"	" "
##	2215	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2216	" "	" "	" "	"**"	"**"	"**"	" "	" "	" "	" "
##	2300	"**"	" "	"**"	"**"	"**"	"**"	"**"	" "	"**"	" "
##	2301	"**"	" "	" "	" "	" "	" "	"**"	" "	"**"	"**"
##	2302	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2303	" "	" "	"**"	"**"	"**"	"**"	"**"	" "	"**"	" "
##	2304	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2305	" "	"**"	"**"	"**"	"**"	"**"	" "	"**"	" "	"**"
##	2306	" "	"**"	"**"	"**"	"**"	"**"	" "	"**"	" "	"**"
##	2307	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2308	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2309	" "	"**"	"**"	"**"	"**"	"**"	" "	"**"	" "	"**"
##	2310	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2311	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "
##	2312	" "	" "	"**"	"**"	"**"	"**"	" "	" "	" "	" "

## 2400	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 2401	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 2402	" "	" "	"***"	"***"	"***"	"***"	"*"	" "	"**"
## 2403	" "	" "	"**"	"***"	"***"	"***"	" "	" "	" "
## 2404	" "	" "	"**"	"***"	"***"	"***"	" "	" "	" "
## 2405	"*"	" "	" "	"*"	"*"	"***"	"***"	" "	"***"
## 2406	" "	" "	" "	" "	" "	"*"	" "	" "	" "
## 2500	" "	" "	"***"	" "	"***"	"***"	"*"	" "	"*"
## 2501	" "	" "	" "	" "	" "	"*"	" "	" "	" "
## 2502	" "	"**"	" "	" "	" "	" "	" "	"**"	"**"
## 2503	" "	" "	"*"	"**"	"**"	"**"	" "	" "	" "
## 2504	" "	" "	" "	"*"	"**"	"***"	" "	" "	" "
## 2505	" "	" "	"**"	"**"	"***"	"***"	" "	" "	" "
## 2506	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "
## 2507	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "
## 2508	" "	" "	"**"	"**"	"***"	"***"	" "	" "	" "
## 2600	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "
## 2601	" "	" "	" "	"**"	" "	" "	"*"	"*"	" "
## 2602	" "	" "	"***"	" "	" "	"***"	" "	" "	" "
## 2603	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 2604	" "	" "	"***"	"***"	"**"	"***"	" "	" "	" "
## 2605	" "	" "	" "	" "	" "	"**"	" "	" "	" "
## 2606	"**"	"*"	"*"	"***"	"*"	"***"	"*"	"**"	" "
## 2607	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 2608	" "	" "	"**"	"*"	" "	"*"	" "	" "	" "
## 2609	" "	" "	"**"	" "	"***"	" "	"*"	"*"	" "
## 2610	" "	" "	"*"	" "	"*"	" "	" "	" "	" "
## 2611	"*"	" "	"***"	"***"	"**"	"**"	"***"	" "	"**"
## 2612	" "	" "	"*"	"*"	" "	" "	" "	" "	" "
## 2613	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "
## 2614	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 2700	" "	" "	"***"	"***"	"***"	"***"	"***"	"**"	"*"
## 2701	" "	"*"	"***"	"***"	"***"	"***"	" "	"**"	"**"
## 2702	" "	" "	" "	"*"	"*"	"**"	" "	" "	" "
## 2703	" "	"***"	"***"	"***"	"***"	"***"	" "	"***"	"***"
## 2704	" "	" "	"*"	"**"	"**"	"***"	" "	" "	" "
## 2705	"***"	"***"	"***"	"***"	"***"	"***"	"***"	"*"	"***"
## 2706	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "
## 2707	" "	" "	" "	"**"	"*"	"***"	" "	" "	" "
## 2708	" "	" "	"***"	"**"	"***"	"***"	" "	" "	" "
## 2709	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 2710	" "	" "	" "	" "	" "	"*"	" "	"*"	"*"
## 2711	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "
## 2712	"*"	"*"	"***"	"***"	"***"	"***"	" "	"**"	"***"
## 2713	" "	" "	"**"	"***"	"***"	"***"	"*"	" "	" "
## 2714	" "	"*"	"**"	"**"	"***"	"***"	" "	"*"	"***"
## 2715	" "	" "	"*"	"***"	"***"	"***"	" "	" "	" "
## 2716	" "	"*"	"**"	"**"	"**"	"***"	" "	"*"	"*"
## 2717	"*"	" "	"***"	"***"	"***"	"***"	"**"	" "	"**"
## 2718	" "	" "	"*"	"*"	"***"	"***"	" "	" "	" "

##	2719	" "	" "	" * * *	" * * *	" * * *	" * * *	" * * *	" "	" * * *	" "
##	2720	" "	" * "	" "	" * * * *	" * * * *	" * * * *	" * "	" "	" * "	" "
##	2721	" "	" "	" "	" "	" "	" * "	" "	" "	" "	" "
##	2722	" "	" "	" * "	" * "	" "	" * "	" "	" "	" "	" "
##	2723	" "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" "	" "	" "	" "
##	2724	" * * * "	" * "	" * * * *	" * * * *	" * * * *	" * * * *	" * * "	" * * * *	" * * "	" * * * *
##	2725	" "	" * * * "	" * * * *	" * * * *	" * * * *	" * * * *	" "	" * * * *	" "	" * * * *
##	2726	" "	" * "	" * * * *	" * * * *	" * * * *	" * * * *	" "	" "	" "	" "
##	2727	" "	" "	" * "	" * * * *	" * * * *	" * * * *	" "	" "	" "	" "
##	2728	" * "	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *
##	2729	" * * "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *	" "	" * * * *	" "
##	2730	" "	" * "	" * * * *	" * * * *	" * * * *	" * * * *	" * "	" "	" * * "	" * "
##	2731	" * "	" "	" * * * *	" * * "	" * * * *	" * * * *	" * * "	" "	" * * "	" "
##	2732	" * "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" * * "	" "	" * * "	" "
##	2733	" "	" * * "	" * "	" * * "	" * * * *	" * * * *	" "	" * "	" "	" * "
##	2734	" "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" "	" "	" "	" "
##	2735	" * * "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *	" "	" * * * *	" "
##	2736	" "	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *	" "	" * * * *	" "	" * * * *
##	2737	" "	" "	" * * "	" * * "	" * * "	" * * * *	" "	" "	" "	" "
##	2738	" "	" * * "	" * * * *	" * * * *	" * * * *	" * * * *	" * * "	" * * * *	" * "	" * * "
##	2739	" "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" "	" * "	" "	" * "
##	2740	" "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" "	" "	" "	" "
##	2741	" "	" "	" * "	" * * * *	" * * * *	" * * * *	" "	" "	" "	" "
##	2742	" "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" "	" * * "	" "	" * * "
##	2743	" * "	" "	" * "	" * "	" * "	" * * "	" * "	" "	" * "	" "
##	2744	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	2745	" "	" * * "	" "	" "	" * * "	" * * * *	" "	" * * "	" "	" * "
##	2746	" * * "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *	" "	" * * * *	" "
##	2747	" "	" "	" * "	" "	" "	" * * * *	" "	" "	" "	" "
##	2748	" "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" * "	" "	" * "	" "
##	2800	" "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" "	" "	" "	" "
##	2801	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	2802	" "	" "	" "	" "	" "	" * * "	" "	" "	" "	" "
##	2803	" * "	" "	" * "	" * * "	" * * * *	" * * * *	" * * "	" "	" * * "	" "
##	2804	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	2805	" "	" "	" "	" "	" "	" * "	" "	" "	" "	" "
##	2806	" "	" "	" * * "	" * "	" * "	" * * "	" "	" "	" "	" "
##	2807	" "	" "	" "	" "	" "	" "	" "	" "	" "	" * "
##	2808	" "	" * * * "	" * * * *	" * * * *	" * * * *	" * * * *	" * "	" * * * *	" * * "	" * * * *
##	2809	" "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" "	" "	" "	" "
##	2900	" * * "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" * "	" "	" * * "	" * * "
##	2901	" "	" "	" "	" * * "	" "	" "	" "	" "	" "	" "
##	2902	" * * "	" "	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *	" * * * *
##	2903	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	2904	" * * * "	" "	" "	" * "	" * * * *	" * "	" "	" "	" "	" "
##	2905	" "	" * "	" "	" * * "	" "	" "	" "	" "	" "	" * "
##	2906	" "	" "	" * "	" * * * *	" * * * *	" * * * *	" "	" * * "	" "	" "
##	2907	" "	" "	" "	" * * * *	" * * * *	" * * * *	" "	" "	" "	" "
##	2908	" "	" "	" * * "	" * * "	" * * "	" * * * *	" "	" "	" "	" "
##	2909	" * "	" * "	" * * "	" * * * *	" * * * *	" * * * *	" * * * *	" "	" * * * *	" * * "

##	2910	" "	" "	"***"	"**"	"***"	"***"	"**"	" "	"***"	"**"
##	2911	" "	" "	" "	"**"	"***"	"***"	" "	" "	"**"	" "
##	2912	" "	" "	"**"	"***"	"***"	" "	" "	" "	" "	" "
##	2913	" "	" "	" "	" "	" "	"**"	"***"	" "	"**"	" "
##	2914	" "	" "	" "	"**"	"**"	" "	" "	" "	" "	" "
##	2915	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	2916	" "	"***"	"***"	"***"	"***"	"***"	" "	"***"	"**"	"***"
##	2917	"**"	" "	"**"	"**"	"**"	"**"	"***"	" "	"***"	"**"
##	2918	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	2919	"**"	" "	" "	"**"	"**"	"***"	"**"	" "	"***"	" "
##	2920	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	2921	" "	"**"	"**"	"***"	"***"	"***"	" "	"**"	"**"	"**"
##	2922	" "	" "	" "	"**"	" "	"**"	"**"	" "	"**"	" "
##	2923	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	3000	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	3001	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	3002	" "	" "	"**"	"**"	"***"	"***"	" "	" "	" "	" "
##	3003	"***"	" "	"***"	"***"	"***"	"***"	"***"	"**"	"***"	"**"
##	3004	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	3005	" "	" "	" "	" "	"**"	"**"	" "	" "	" "	" "
##	3100	" "	" "	"***"	"***"	"***"	"***"	" "	"**"	"**"	"**"
##	3101	" "	" "	"***"	"***"	"**"	"***"	" "	" "	" "	" "
##	3102	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	3103	" "	" "	"***"	"***"	"***"	"***"	"**"	" "	"**"	" "
##	3104	" "	" "	"**"	" "	"**"	"***"	" "	" "	" "	" "
##	3105	"**"	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	3106	" "	" "	"**"	"***"	"***"	" "	" "	" "	" "	" "
##	3107	" "	" "	"**"	"**"	"***"	"***"	" "	" "	" "	" "
##	3108	" "	" "	"**"	"***"	"**"	"***"	" "	" "	" "	" "
##	3109	"**"	" "	"**"	"**"	"**"	"**"	"**"	" "	"**"	" "
##	3110	" "	" "	" "	" "	" "	" "	"**"	" "	"**"	" "
##	3200	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	3201	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	3202	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	3203	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	3204	"**"	" "	"***"	"***"	"***"	"***"	"**"	" "	"***"	" "
##	3205	" "	" "	" "	"**"	"**"	"***"	" "	" "	"**"	" "
##	3206	" "	" "	"***"	"***"	" "	"***"	"**"	" "	"**"	" "
##	3207	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	3300	"**"	"**"	"***"	"***"	"**"	"**"	"**"	"**"	" "	" "
##	3301	"**"	" "	"***"	"***"	"***"	"***"	"**"	" "	"***"	"***"
##	3302	"**"	" "	" "	"***"	"***"	"**"	"**"	"**"	" "	" "
##	3303	" "	" "	"**"	"***"	" "	" "	" "	" "	" "	" "
##	3304	"**"	"**"	"***"	"***"	"***"	"***"	"***"	"**"	"***"	"***"
##	3305	" "	" "	"***"	"***"	"***"	"***"	"**"	" "	"***"	"***"
##	3306	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	3307	" "	"**"	"**"	"**"	" "	"**"	" "	" "	" "	"**"
##	3308	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	3309	" "	" "	"***"	"***"	"***"	"**"	" "	" "	" "	" "
##	3310	" "	" "	"***"	"***"	"***"	"***"	"**"	" "	"***"	"**"

```

## 3311 "" "" "" "" "" "" "" "" "" ""
## 3312 "" "" "" "" "" "" "" "" "" ""
## 3313 "" "" "" "" "" "" "" "" "" ""
## 3314 "" "" "" "" "" "" "" "" "" ""
## 3315 "" "" "" "" "" "" "" "" "" ""
## 3316 "" "" "" "" "" "" "" "" "" ""
## 3317 "" "" "" "" "" "" "" "" "" ""
## 3318 "" "" "" "" "" "" "" "" "" ""
## 3319 "" "" "" "" "" "" "" "" "" ""
## 3320 "" "" "" "" "" "" "" "" "" ""
## 3321 "" "" "" "" "" "" "" "" "" ""
## 3322 "" "" "" "" "" "" "" "" "" ""
## 3400 "" "" "" "" "" "" "" "" "" ""
## 3401 "" "" "" "" "" "" "" "" "" ""
## 3402 "" "" "" "" "" "" "" "" "" ""
## 3403 "" "" "" "" "" "" "" "" "" ""
## 3404 "" "" "" "" "" "" "" "" "" ""
## 3500 "" "" "" "" "" "" "" "" "" ""
## 3501 "" "" "" "" "" "" "" "" "" ""
## 3502 "" "" "" "" "" "" "" "" "" ""
## 3503 "" "" "" "" "" "" "" "" "" ""
## 3504 "" "" "" "" "" "" "" "" "" ""
## 3505 "" "" "" "" "" "" "" "" "" ""
## 3506 "" "" "" "" "" "" "" "" "" ""
## 3600 "" "" "" "" "" "" "" "" "" ""
## 3601 "" "" "" "" "" "" "" "" "" ""
## 3602 "" "" "" "" "" "" "" "" "" ""
## 3603 "" "" "" "" "" "" "" "" "" ""
## 3604 "" "" "" "" "" "" "" "" "" ""
## 3605 "" "" "" "" "" "" "" "" "" ""
## 3606 "" "" "" "" "" "" "" "" "" ""
## 3607 "" "" "" "" "" "" "" "" "" ""
## 3608 "" "" "" "" "" "" "" "" "" ""
## 3609 "" "" "" "" "" "" "" "" "" ""
## 3610 "" "" "" "" "" "" "" "" "" ""
## 3611 "" "" "" "" "" "" "" "" "" ""
## 3612 "" "" "" "" "" "" "" "" "" ""
## 3613 "" "" "" "" "" "" "" "" "" ""
## 3614 "" "" "" "" "" "" "" "" "" ""
## 3615 "" "" "" "" "" "" "" "" "" ""
## 3616 "" "" "" "" "" "" "" "" "" ""

```

```
print("Gender-based average team sizes in 2018")
```

```
## [1] "Gender-based average team sizes in 2018"
```

```
print(GenderAnalysed)
```

```

##      CitationSet Female1st FemaleLast
## 1000          260          96         63
## 1100          431         188        131

```

## 1101	25	12	9
## 1102	149	47	39
## 1103	310	172	93
## 1104	315	153	85
## 1105	747	333	184
## 1106	198	117	88
## 1107	160	51	39
## 1108	25	4	5
## 1109	106	50	31
## 1110	181	69	42
## 1111	82	27	19
## 1200	27	13	14
## 1201	330	167	157
## 1202	296	132	128
## 1203	184	98	104
## 1204	64	23	17
## 1205	10	1	1
## 1206	11	7	7
## 1207	114	43	52
## 1208	159	83	81
## 1209	9	6	6
## 1210	35	19	20
## 1211	278	97	94
## 1212	145	42	40
## 1213	139	83	84
## 1300	524	224	154
## 1301	9	2	0
## 1302	30	20	19
## 1303	431	145	74
## 1304	216	91	49
## 1305	177	66	28
## 1306	212	98	64
## 1307	303	135	60
## 1308	99	52	22
## 1309	116	60	33
## 1310	167	99	53
## 1311	412	184	104
## 1312	548	229	122
## 1313	166	63	41
## 1314	378	181	93
## 1315	49	20	13
## 1400	84	33	22
## 1401	35	15	18
## 1402	55	16	11
## 1403	140	55	47
## 1404	14	3	3
## 1405	92	33	32
## 1406	73	32	25
## 1407	82	49	48
## 1408	190	59	69

## 1409	58	27	25
## 1410	19	9	9
## 1500	238	79	42
## 1501	11	4	2
## 1502	114	37	23
## 1503	161	51	31
## 1504	4	3	2
## 1505	41	11	7
## 1506	12	5	0
## 1507	66	7	5
## 1508	39	10	5
## 1600	492	166	79
## 1601	17	9	5
## 1602	135	46	38
## 1603	38	12	11
## 1604	115	34	27
## 1605	239	75	43
## 1606	299	73	56
## 1607	112	47	40
## 1700	112	21	16
## 1701	8	2	1
## 1702	14	2	2
## 1703	64	13	6
## 1704	40	4	2
## 1705	47	13	4
## 1706	186	47	40
## 1707	32	9	5
## 1708	48	6	5
## 1709	62	21	23
## 1710	105	29	25
## 1711	68	11	6
## 1712	130	21	12
## 1800	27	12	10
## 1801	2	2	0
## 1802	39	13	10
## 1803	101	31	23
## 1804	51	10	8
## 1900	136	32	24
## 1901	54	17	8
## 1902	127	47	30
## 1903	41	10	9
## 1904	146	56	38
## 1905	23	7	4
## 1906	97	35	17
## 1907	136	36	26
## 1908	77	25	17
## 1909	121	29	16
## 1910	81	37	24
## 1911	42	12	13
## 1912	58	17	11

## 1913	15	3	3
## 2000	40	10	8
## 2001	33	8	3
## 2002	315	73	67
## 2003	121	25	21
## 2100	76	14	11
## 2101	16	0	0
## 2102	164	30	19
## 2103	124	26	17
## 2104	46	9	7
## 2105	227	54	32
## 2200	55	13	6
## 2201	40	10	9
## 2202	60	1	2
## 2203	44	7	4
## 2204	228	69	48
## 2205	301	56	33
## 2206	24	1	1
## 2207	58	8	10
## 2208	184	21	19
## 2209	227	43	34
## 2210	373	44	38
## 2211	235	32	29
## 2212	43	8	6
## 2213	111	31	20
## 2214	24	8	8
## 2215	165	29	15
## 2216	36	8	6
## 2300	229	77	65
## 2301	78	27	19
## 2302	57	26	10
## 2303	422	178	102
## 2304	287	130	85
## 2305	165	56	28
## 2306	122	47	31
## 2307	193	99	68
## 2308	292	116	83
## 2309	159	69	49
## 2310	260	110	65
## 2311	153	57	33
## 2312	234	86	46
## 2400	95	34	35
## 2401	6	3	4
## 2402	52	24	10
## 2403	240	130	84
## 2404	156	73	41
## 2405	35	13	7
## 2406	79	40	19
## 2500	382	69	48
## 2501	13	2	2

## 2502	78	21	21
## 2503	72	14	9
## 2504	133	23	16
## 2505	95	31	15
## 2506	83	20	10
## 2507	100	22	14
## 2508	118	25	23
## 2600	139	26	22
## 2601	13	4	3
## 2602	56	7	7
## 2603	46	8	6
## 2604	237	38	31
## 2605	67	7	2
## 2606	52	3	6
## 2607	47	8	12
## 2608	39	7	8
## 2609	8	0	0
## 2610	39	2	1
## 2611	136	24	19
## 2612	25	4	3
## 2613	134	21	16
## 2614	71	15	12
## 2700	402	200	170
## 2701	257	151	108
## 2702	63	35	24
## 2703	133	62	52
## 2704	22	14	8
## 2705	402	151	116
## 2706	114	54	39
## 2707	30	15	12
## 2708	107	61	41
## 2709	0	0	0
## 2710	10	6	3
## 2711	162	63	59
## 2712	216	131	70
## 2713	184	117	94
## 2714	83	56	49
## 2715	74	31	26
## 2716	109	60	32
## 2717	179	122	99
## 2718	146	87	78
## 2719	418	258	221
## 2720	133	71	43
## 2721	28	13	9
## 2722	53	21	11
## 2723	189	109	80
## 2724	102	59	35
## 2725	221	105	72
## 2726	108	48	33
## 2727	68	34	25

## 2728	547	295	180
## 2729	251	165	124
## 2730	409	203	142
## 2731	129	58	46
## 2732	384	155	103
## 2733	103	38	28
## 2734	159	76	72
## 2735	482	323	234
## 2736	191	102	75
## 2737	236	106	57
## 2738	686	446	329
## 2739	877	577	435
## 2740	204	102	70
## 2741	368	140	107
## 2742	252	163	136
## 2743	64	36	25
## 2744	0	0	0
## 2745	68	35	35
## 2746	536	207	146
## 2747	61	31	18
## 2748	154	50	31
## 2800	330	177	103
## 2801	51	22	11
## 2802	143	83	65
## 2803	92	62	38
## 2804	143	70	38
## 2805	181	98	77
## 2806	62	41	27
## 2807	24	17	13
## 2808	321	178	99
## 2809	79	40	22
## 2900	145	117	106
## 2901	27	20	17
## 2902	72	53	47
## 2903	11	5	5
## 2904	11	8	7
## 2905	46	40	33
## 2906	9	3	6
## 2907	20	6	6
## 2908	15	10	8
## 2909	82	61	49
## 2910	52	31	41
## 2911	25	19	20
## 2912	18	12	10
## 2913	15	14	13
## 2914	7	7	0
## 2915	1	0	0
## 2916	184	127	88
## 2917	20	18	9
## 2918	0	0	0

## 2919	20	16	14
## 2920	2	2	0
## 2921	40	31	27
## 2922	13	8	10
## 2923	7	6	6
## 3000	20	9	10
## 3001	14	8	4
## 3002	95	39	17
## 3003	103	50	52
## 3004	214	97	69
## 3005	125	64	40
## 3100	222	36	29
## 3101	64	6	4
## 3102	45	10	7
## 3103	41	10	9
## 3104	344	53	45
## 3105	150	25	23
## 3106	57	4	5
## 3107	208	30	25
## 3108	64	20	14
## 3109	46	3	2
## 3110	47	17	9
## 3200	316	171	143
## 3201	35	20	20
## 3202	192	132	96
## 3203	281	178	134
## 3204	385	286	212
## 3205	253	138	112
## 3206	107	70	48
## 3207	275	170	144
## 3300	163	83	81
## 3301	228	144	134
## 3302	61	25	23
## 3303	131	60	54
## 3304	762	491	456
## 3305	449	192	172
## 3306	364	240	232
## 3307	51	18	19
## 3308	323	151	154
## 3309	110	65	63
## 3310	214	123	129
## 3311	44	21	14
## 3312	760	379	360
## 3313	79	27	15
## 3314	203	123	114
## 3315	135	70	63
## 3316	355	187	188
## 3317	107	67	58
## 3318	116	92	92
## 3319	69	48	38

## 3320	222	80	77
## 3321	116	56	55
## 3322	75	30	22
## 3400	166	104	76
## 3401	3	2	2
## 3402	4	3	0
## 3403	33	25	15
## 3404	9	6	3
## 3500	25	15	12
## 3501	1	0	0
## 3502	0	0	0
## 3503	0	0	0
## 3504	11	6	3
## 3505	6	2	3
## 3506	3	1	1
## 3600	10	6	7
## 3601	18	12	12
## 3602	13	5	3
## 3603	3	3	0
## 3604	7	2	4
## 3605	21	5	6
## 3606	0	0	0
## 3607	13	7	5
## 3608	0	0	0
## 3609	24	24	0
## 3610	10	7	4
## 3611	55	38	35
## 3612	298	185	146
## 3613	1	0	0
## 3614	78	35	32
## 3615	0	0	0
## 3616	59	44	38

```

write.table(BasicStats, file=paste("D:\\Downloads\\basicStats", Country,
MaxCountriesAllowed, " ", FirstYearForCitationAnalysis, "-",
LastYearForCitationAnalysis, "var3.txt", sep=""))
write.table(BasicStats2, file=paste("D:\\Downloads\\basicStats2", Country,
MaxCountriesAllowed, " ", FirstYearForCitationAnalysis, "-",
LastYearForCitationAnalysis, "var3.txt", sep=""))
write.table(GenderTeamSize, file=paste("D:\\Downloads\\GenderTeamSize",
Country, MaxCountriesAllowed, " ", FirstYearForCitationAnalysis, "-",
LastYearForCitationAnalysis, "var3.txt", sep=""))
write.table(RegCoef, file=paste("D:\\Downloads\\RegCoef", Country,
MaxCountriesAllowed, " ", FirstYearForCitationAnalysis, "-",
LastYearForCitationAnalysis, "var3.txt", sep=""))
write.table(RegP, file=paste("D:\\Downloads\\RegP", Country,
MaxCountriesAllowed, " ", FirstYearForCitationAnalysis, "-",
LastYearForCitationAnalysis, "var3.txt", sep=""))
write.table(RegStar, file=paste("D:\\Downloads\\RegStar", Country,

```

```
MaxCountriesAllowed," ", FirstYearForCitationAnalysis, "-",  
LastYearForCitationAnalysis, "var3.txt", sep=""))
```