Biochemistry, Genetics and Molecular Biology

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

library(robustbase) #lmrob for linear regression in the presence of heteroscedasticity

library(plyr) #for count

outlier\_threshold = 2.5
options(digits=3)
FirstCode <- 1300
LastCode <- 1315

FieldCount <- LastCode - FirstCode + 1
column.names <- c("Articles", "FirstF","FirstM","FirstP","LastF","LastM","LastP")
row.names <- c(FirstCode:LastCode)
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("US only","Gend 1st nth","Fem1 96","Fem1 18","Chg","FemN 96","FemN 18", "Chg")
BasicStats <- array(integer(8\*FieldCount),dim = c(FieldCount, 8), 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))
FieldNeedsExtraTime <- 1311

FieldCount = 0
iCode <- FirstCode
for (iCode in FirstCode:LastCode) {
 FieldCount <- FieldCount + 1
 tryCatch({
 print(""); print("")
 print("###################################")
 print(paste("Analysis of AJSC",iCode))
 print("###################################")
 SampleFile <- paste("E:\\data\\Scopus\\All fields regression\\US 1996-2018 331 fields\\cov\\United States ", iCode, "\_cov.txt", sep="")
 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$Year!=2004,] #Uncomment this to remove year with high outliers if code below reveals a problem
 AllScopusData <- AllScopusData[AllScopusData$UniqueCountries < 2,] #Comment out to include non-us authors after 1st
 AllScopusDataOlder <- AllScopusData[AllScopusData$Year<2015,] #2015 Ensure citation window of at least 3 years 2013 for 5 years
 AllScopusDataOlder$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,]
 AllYearsBothGendered2018 <-AllYearsBothGendered[AllYearsBothGendered$Year == 2018,]
 BasicStats[FieldCount,1] <- nrow(AllScopusData)
 BasicStats[FieldCount,2] <- nrow(AllYearsBothGendered) / nrow(AllScopusData) \* 100 #Percentage
 BasicStats[FieldCount,3] <- count(AllYearsBothGendered1996, vars="FirstAuthorFemale")[2,2]/ nrow(AllYearsBothGendered1996) \* 100
 BasicStats[FieldCount,4] <- count(AllYearsBothGendered2018, vars="FirstAuthorFemale")[2,2]/ nrow(AllYearsBothGendered2018) \* 100
 BasicStats[FieldCount,5] <- BasicStats[FieldCount,4] - BasicStats[FieldCount,3]
 BasicStats[FieldCount,6] <- count(AllYearsBothGendered1996, vars="LastAuthorFemale")[2,2]/ nrow(AllYearsBothGendered1996) \* 100
 BasicStats[FieldCount,7] <- count(AllYearsBothGendered2018, vars="LastAuthorFemale")[2,2]/ nrow(AllYearsBothGendered2018) \* 100
 BasicStats[FieldCount,8] <- BasicStats[FieldCount,7] - BasicStats[FieldCount,6]

 #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 becuase 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
 GenderAuthors <- split(AllScopusDataOlderFirstLastGendered$UniqueAuthors,AllScopusDataOlderFirstLastGendered$FirstAuthorFemale)
 FemaleTeamSizes <- as.numeric(GenderAuthors$'1')
 MaleTeamSizes <- as.numeric(GenderAuthors$'0')
 GenderAnalysed[FieldCount,1] <- nrow(AllScopusDataOlderFirstLastGendered)
 GenderAnalysed[FieldCount,2] <- length(FemaleTeamSizes)
 print(paste("Female first author team size geometric mean:", geometric.mean(FemaleTeamSizes)))
 print(paste("Male first author team size geometric mean:", geometric.mean(MaleTeamSizes)))
 wilc<- wilcox.test(FemaleTeamSizes, MaleTeamSizes, alternative = "two.sided")
 print(wilc)
 GenderTeamSize[FieldCount,1] <- nrow(AllScopusDataOlderFirstLastGendered)
 GenderTeamSize[FieldCount,2] <- geometric.mean(FemaleTeamSizes)
 GenderTeamSize[FieldCount,3] <- geometric.mean(MaleTeamSizes)
 GenderTeamSize[FieldCount,4] <- wilc$p.value
 #Last author gender
 GenderAuthors <- split(AllScopusDataOlderFirstLastGendered$UniqueAuthors,AllScopusDataOlderFirstLastGendered$LastAuthorFemale)
 FemaleTeamSizes <- as.numeric(GenderAuthors$'1')
 MaleTeamSizes <- as.numeric(GenderAuthors$'0')
 GenderAnalysed[FieldCount,3] <- length(FemaleTeamSizes)
 print(paste("Female last author team size geometric mean:", geometric.mean(FemaleTeamSizes)))
 print(paste("Male last author team size geometric mean:", geometric.mean(MaleTeamSizes)))
 wilc<- wilcox.test(FemaleTeamSizes, MaleTeamSizes, alternative = "two.sided")
 print(wilc)
 GenderTeamSize[FieldCount,5] <- geometric.mean(FemaleTeamSizes)
 GenderTeamSize[FieldCount,6] <- geometric.mean(MaleTeamSizes)
 GenderTeamSize[FieldCount,7] <- wilc$p.value
 }, error = function(e) return("Failed team size tests"))
 tryCatch({
 #Use 5 for 5+ authors
 for (i in 1:nrow(AllScopusDataOlderFirstLastGendered)) {
 AllScopusDataOlderFirstLastGendered$UniqueAuthors[i] <- min(5,AllScopusDataOlderFirstLastGendered$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 == FieldNeedsExtraTime) {
 FirstLastAuthorTeamLmrob <- lmrob(NLCS~FirstAuthorFemale+LastAuthorFemale+UniqueAuthors+Year, data=AllScopusDataOlderFirstLastGendered, control = lmrob.control(fast.s.large.n = Inf))
 } 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 == FieldNeedsExtraTime) {
 FirstLastAuthorLmrob <- lmrob(NLCS~FirstAuthorFemale+LastAuthorFemale+Year, data=AllScopusDataOlderFirstLastGendered, control = lmrob.control(fast.s.large.n = Inf))
 } 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 == FieldNeedsExtraTime) {
 FirstAuthorLmrob <- lmrob(NLCS~FirstAuthorFemale+Year, data=AllScopusDataOlderFirstLastGendered, control = lmrob.control(fast.s.large.n = Inf))
 } 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 == FieldNeedsExtraTime) {
 LastAuthorLmrob <- lmrob(NLCS~LastAuthorFemale+Year, data=AllScopusDataOlderFirstLastGendered, control = lmrob.control(fast.s.large.n = Inf))
 } 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 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
## 6798 4071 3206 1783 1747 1785 1760 1687 1707 1788 2002 2660 2807 3637 4370
## 2011 2012 2013 2014
## 6016 8100 9401 9375
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1814 1647 1493 1163 981 805 1178 1110 1136 1202 1353 1831 1915 2589 3092
## 2011 2012 2013 2014
## 4261 5654 6572 6637
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1553 1421 1311 1024 845 701 1033 954 978 1019 1171 1555 1611 2203 2603
## 2011 2012 2013 2014
## 3620 4808 5587 5628
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 7000, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 100, df = 1, p-value <2e-16



## [1] "Female first author team size geometric mean: 3.64482647906721"
## [1] "Male first author team size geometric mean: 3.19108418534425"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 2e+08, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.34540943118208"
## [1] "Male last author team size geometric mean: 3.34902381407545"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 1e+08, 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 1.04 1 1.02
## LastAuthorFemale 1.02 1 1.01
## UniqueAuthors 1.17 4 1.02
## Year 1.16 18 1.00



## [1] "List of 165 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 4 0030447612 4.32 1996 1300 1 3.00
## 6 0030448994 4.35 1996 1300 1 3.07
## 8 0030451819 4.40 1996 1300 1 3.27
## 9 0030460424 5.08 1996 1300 1 3.76
## 12 0030463470 4.71 1996 1300 1 3.67
## 14 0030475008 4.18 1996 1300 1 2.86
## 16 0030480322 5.32 1996 1300 1 4.04
## 57 0030582677 4.26 1996 1300 1 3.09
## 60 0030582732 4.00 1996 1300 1 2.73
## 3626 0030606239 5.32 1996 1300 1 3.99
## 3630 0030606286 3.86 1996 1300 1 2.69
## 3634 0030606320 4.48 1996 1300 1 3.16
## 3689 0030297537 4.47 1996 1300 1 3.43
## 3690 0030297538 3.97 1996 1300 1 2.98
## 3693 0030297895 4.20 1996 1300 1 2.93
## 3694 0030297912 3.84 1996 1300 1 2.85
## 3695 0030298137 4.19 1996 1300 1 3.02
## 3698 0030298375 4.69 1996 1300 1 3.61
## 3718 0029955497 4.50 1996 1300 2 3.23
## 3755 0030592544 4.37 1996 1300 1 3.09
## 3756 0030592556 3.54 1996 1300 1 2.54
## 3758 0030592564 4.05 1996 1300 1 2.73
## 3813 0029820526 4.58 1996 1300 1 3.26
## 3839 0029798819 4.73 1996 1300 2 3.41
## 3841 0029842830 5.29 1996 1300 2 3.97
## 3881 0030595342 4.36 1996 1300 1 3.08
## 3886 0029817693 4.46 1996 1300 1 3.14
## 3920 0030572695 4.06 1996 1300 1 2.74
## 3922 0030572708 3.93 1996 1300 1 2.61
## 3949 9544226448 4.17 1996 1300 1 2.85
## 3951 0029791838 4.09 1996 1300 2 3.09
## 3967 0030598829 4.93 1996 1300 1 3.89
## 3970 0030598865 3.61 1996 1300 1 2.57
## 3974 16044363014 4.45 1996 1300 1 3.13
## 4005 0030576502 4.48 1996 1300 1 3.20
## 4008 0030576518 4.44 1996 1300 1 3.40
## 4079 0030602822 4.23 1996 1300 1 2.91
## 4080 0030602823 3.78 1996 1300 1 2.79
## 4083 16044364385 4.50 1996 1300 1 3.18
## 4128 0030581152 3.81 1996 1300 1 2.68
## 4131 0030581165 4.17 1996 1300 1 2.85
## 4133 0030581174 4.38 1996 1300 1 3.10
## 4136 15844386540 4.11 1996 1300 1 2.79
## 4153 0029954860 4.28 1996 1300 1 3.29
## 4157 0030010783 4.14 1996 1300 2 3.14
## 4171 0029944241 3.65 1996 1300 1 2.86
## 4173 0029999787 4.34 1996 1300 1 3.22
## 4178 15844378825 4.21 1996 1300 1 2.89
## 4188 0029944290 4.69 1996 1300 2 3.37
## 4191 0029994529 4.29 1996 1300 2 3.17
## 4224 0030604540 3.89 1996 1300 1 2.76
## 4228 0030604722 4.69 1996 1300 1 3.37
## 4231 15844367099 4.22 1996 1300 1 2.90
## 4273 0029894165 3.85 1996 1300 1 2.58
## 4274 0029899127 3.99 1996 1300 1 2.67
## 4275 0029953780 4.08 1996 1300 1 2.76
## 4302 0001506104 4.15 1996 1300 2 2.98
## 4334 0029895156 4.20 1996 1300 1 3.20
## 4336 0029939448 4.17 1996 1300 1 3.13
## 4337 0029943141 4.03 1996 1300 1 2.75
## 4342 15844384256 4.94 1996 1300 1 3.62
## 4343 15844415946 4.83 1996 1300 1 3.51
## 4344 15844420283 4.98 1996 1300 1 3.66
## 4361 0029881125 4.62 1996 1300 1 3.30
## 4362 0029892278 3.56 1996 1300 1 2.57
## 4448 0029870085 4.65 1996 1300 1 3.48
## 4451 0029993728 3.74 1996 1300 1 2.57
## 4452 0030009544 3.96 1996 1300 1 2.97
## 4453 15844372440 3.86 1996 1300 1 2.58
## 4500 0029919935 3.65 1996 1300 1 2.62
## 4501 0029980441 4.22 1996 1300 1 3.19
## 4502 0029993450 4.96 1996 1300 1 3.64
## 4565 0029876473 4.67 1996 1300 1 3.35
## 4589 0029880651 4.20 1996 1300 2 3.03
## 4611 0029961719 3.73 1996 1300 1 2.94
## 4613 0029965130 3.96 1996 1300 1 2.89
## 4614 0029978023 3.70 1996 1300 1 2.90
## 4615 0029991047 4.20 1996 1300 1 3.03
## 4617 13344277364 4.98 1996 1300 1 3.66
## 4703 0030030905 4.20 1996 1300 1 2.93
## 4705 0030048731 4.35 1996 1300 1 3.36
## 4707 0030058657 4.09 1996 1300 1 2.97
## 4714 0030065744 3.61 1996 1300 1 2.61
## 4715 13344261952 4.26 1996 1300 1 2.94
## 4716 13344282063 4.29 1996 1300 1 3.06
## 4727 0030020590 4.57 1996 1300 2 3.44
## 4751 0030024563 4.88 1996 1300 1 3.75
## 4797 0030033699 3.98 1996 1300 1 2.66
## 4808 0030026776 5.11 1996 1300 1 3.78
## 4809 0030026934 4.45 1996 1300 1 3.17
## 4830 0030584077 4.46 1996 1300 1 3.14
## 4831 0030584078 5.02 1996 1300 1 3.70
## 4833 0030584083 3.95 1996 1300 1 2.63
## 4854 0029671310 4.40 1996 1300 1 3.40
## 4856 0030031999 4.20 1996 1300 1 3.21
## 4858 0030034731 4.29 1996 1300 1 3.16
## 4859 0030034983 4.02 1996 1300 1 3.07
## 4860 0030050396 4.34 1996 1300 1 3.30
## 4862 0030061451 3.63 1996 1300 1 2.59
## 4887 0029664992 4.11 1996 1300 1 2.79
## 4901 0029888359 4.45 1996 1300 1 3.17
## 4903 0029890229 3.83 1996 1300 1 2.84
## 4908 0029898733 3.67 1996 1300 1 2.55
## 4917 0029936764 4.43 1996 1300 1 3.25
## 4918 0029940972 4.12 1996 1300 1 3.08
## 4922 0029949784 4.50 1996 1300 1 3.18
## 4926 0029977751 3.72 1996 1300 1 2.93
## 4929 0030014157 3.96 1996 1300 1 2.84
## 4958 0030111234 3.31 1996 1300 1 2.57
## 4974 0030131182 3.67 1996 1300 1 2.88
## 4981 0030139470 3.60 1996 1300 1 2.81
## 5009 0030271388 4.00 1996 1300 1 2.96
## 5010 0030271392 4.00 1996 1300 1 2.96
## 5012 0030271890 3.84 1996 1300 1 2.67
## 5014 0030271999 4.24 1996 1300 1 2.92
## 5015 0030272047 4.28 1996 1300 1 3.10
## 5043 15844380040 4.35 1996 1300 1 3.03
## 5044 15844417385 4.49 1996 1300 1 3.17
## 5868 0030087710 3.74 1996 1100 2 2.70
## 6075 0030606018 3.88 1996 1100 2 3.08
## 7061 0031459980 4.01 1997 1300 1 2.96
## 7229 0031451777 3.98 1997 1300 2 2.80
## 7235 0031466811 3.82 1997 1300 2 2.57
## 9659 0030613551 4.25 1997 1300 1 2.92
## 9665 0030702123 4.93 1997 1300 1 3.60
## 9696 0030886099 3.75 1997 1300 1 2.70
## 9699 0030886602 3.74 1997 1300 1 2.69
## 9772 0030886852 3.38 1997 1300 1 2.62
## 9773 0030928716 3.65 1997 1300 1 2.60
## 9776 0030954870 3.67 1997 1300 1 2.67
## 9777 0030985459 3.94 1997 1300 1 2.80
## 9838 0030865245 3.67 1997 1300 2 2.53
## 9899 0030788436 3.67 1997 1300 1 2.66
## 9906 0030848970 3.66 1997 1300 1 2.65
## 9966 0030746636 4.62 1997 1300 1 3.62
## 9975 0346613495 3.65 1997 1300 1 2.85
## 9977 0642270732 4.21 1997 1300 1 3.41
## 10027 0031586174 4.07 1997 1300 1 2.78
## 10074 0031440879 3.51 1997 1300 1 2.55
## 10095 16944366965 3.60 1997 1300 1 2.60
## 10117 0030752411 3.71 1997 1300 2 2.61
## 10314 0030687987 3.58 1997 1300 1 2.53
## 10337 0030729445 3.98 1997 1300 1 2.81
## 10344 20244377493 4.04 1997 1300 1 2.70
## 10461 0030970693 4.34 1997 1300 1 3.16
## 10462 0030982264 3.77 1997 1300 1 2.59
## 10519 0030963439 3.98 1997 1300 1 2.65
## 10616 0030944985 4.74 1997 1300 1 3.40
## 10618 0030970775 3.71 1997 1300 1 2.66
## 10620 0031000884 3.82 1997 1300 1 2.64
## 10679 0030949875 3.95 1997 1300 1 2.61
## 10685 0030614893 4.10 1997 1300 1 2.91
## 10688 0031048716 3.92 1997 1300 1 2.63
## 10712 0030893115 4.03 1997 1300 1 3.02
## 10739 0345877774 3.59 1997 1300 1 2.79
## 10740 0345877775 3.87 1997 1300 1 3.06
## 10785 0031285250 3.78 1997 1300 1 2.60
## 10788 0031444245 3.33 1997 1300 1 2.53
## 10802 0031015075 3.96 1997 1300 1 2.67
## 10839 0030994017 4.01 1997 1300 1 2.83
## 10863 0001679473 3.44 1997 1300 1 2.64
## 10876 0030715563 3.88 1997 1300 1 2.59
## 10881 0030800831 3.63 1997 1300 1 2.58
## 10884 0030866897 3.70 1997 1300 1 2.56
## 11042 0031022694 3.79 1997 1300 2 2.61
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6172 -0.3533 0.0161 0.3733 4.0427
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.79105 0.03970 19.93 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.04348 0.00600 -7.25 4.3e-13 \*\*\*
## LastAuthorFemale1 -0.04352 0.00669 -6.51 7.8e-11 \*\*\*
## UniqueAuthors2 0.24840 0.01174 21.16 < 2e-16 \*\*\*
## UniqueAuthors3 0.37436 0.01414 26.47 < 2e-16 \*\*\*
## UniqueAuthors4 0.38061 0.01255 30.34 < 2e-16 \*\*\*
## UniqueAuthors5 0.52973 0.01130 46.87 < 2e-16 \*\*\*
## Year1997 0.01156 0.05206 0.22 0.82433
## Year1998 0.19611 0.05208 3.77 0.00017 \*\*\*
## Year1999 0.29644 0.04435 6.68 2.4e-11 \*\*\*
## Year2000 0.24534 0.04466 5.49 4.0e-08 \*\*\*
## Year2001 0.28760 0.04540 6.33 2.4e-10 \*\*\*
## Year2002 0.17536 0.04321 4.06 5.0e-05 \*\*\*
## Year2003 0.19142 0.04321 4.43 9.5e-06 \*\*\*
## Year2004 0.18704 0.04331 4.32 1.6e-05 \*\*\*
## Year2005 0.17653 0.04313 4.09 4.3e-05 \*\*\*
## Year2006 0.18815 0.04245 4.43 9.4e-06 \*\*\*
## Year2007 0.16631 0.04137 4.02 5.8e-05 \*\*\*
## Year2008 0.17047 0.04133 4.12 3.7e-05 \*\*\*
## Year2009 0.17123 0.04074 4.20 2.6e-05 \*\*\*
## Year2010 0.16980 0.04032 4.21 2.5e-05 \*\*\*
## Year2011 0.14400 0.04001 3.60 0.00032 \*\*\*
## Year2012 0.10155 0.03976 2.55 0.01065 \*
## Year2013 0.05566 0.03973 1.40 0.16132
## Year2014 0.05341 0.03979 1.34 0.17949
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.539
## Multiple R-squared: 0.0997, Adjusted R-squared: 0.0991
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 165 observations c(1,2,4,5,6,7,9,33,34,495,497,500,527,528,529,530,531,532,541,561,562,563,592,600,601,623,625,646,647,658,659,665,666,667,682,684,719,720,722,738,739,741,743,750,751,758,760,762,766,768,785,786,787,810,811,812,824,840,841,842,845,846,847,855,856,898,901,902,903,929,930,931,961,975,988,990,991,992,993,1035,1036,1038,1042,1043,1044,1048,1063,1084,1089,1090,1098,1099,1100,1110,1111,1112,1113,1114,1115,1126,1127,1128,1130,1132,1133,1135,1139,1140,1157,1167,1173,1190,1191,1193,1194,1195,1204,1205,1289,1442,1561,1630,1633,2112,2115,2132,2135,2171,2172,2173,2174,2206,2234,2236,2267,2270,2271,2296,2320,2328,2339,2439,2450,2452,2517,2518,2548,2596,2597,2598,2627,2632,2634,2648,2657,2658,2682,2685,2690,2710,2721,2731,2734,2736,2830)
## are outliers with |weight| = 0 ( < 2.5e-06);
## 3295 weights are ~= 1. The remaining 36165 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.860 0.950 0.896 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.52e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.02 1 1.01
## LastAuthorFemale 1.01 1 1.00
## Year 1.02 18 1.00



## [1] "List of 204 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 4 0030447612 4.32 1996 1300 1 3.29
## 6 0030448994 4.35 1996 1300 1 3.33
## 8 0030451819 4.40 1996 1300 1 3.38
## 9 0030460424 5.08 1996 1300 1 4.05
## 12 0030463470 4.71 1996 1300 1 3.68
## 14 0030475008 4.18 1996 1300 1 3.16
## 15 0030476713 3.58 1996 1300 1 2.55
## 16 0030480322 5.32 1996 1300 1 4.31
## 57 0030582677 4.26 1996 1300 1 3.23
## 60 0030582732 4.00 1996 1300 1 2.99
## 137 16144361909 3.76 1996 1300 1 2.74
## 3626 0030606239 5.32 1996 1300 1 4.29
## 3630 0030606286 3.86 1996 1300 1 2.83
## 3632 0030606315 3.68 1996 1300 1 2.65
## 3634 0030606320 4.48 1996 1300 1 3.45
## 3689 0030297537 4.47 1996 1300 1 3.45
## 3690 0030297538 3.97 1996 1300 1 2.96
## 3693 0030297895 4.20 1996 1300 1 3.19
## 3694 0030297912 3.84 1996 1300 1 2.83
## 3695 0030298137 4.19 1996 1300 1 3.16
## 3698 0030298375 4.69 1996 1300 1 3.73
## 3718 0029955497 4.50 1996 1300 2 3.49
## 3750 0001265782 3.56 1996 1300 1 2.53
## 3755 0030592544 4.37 1996 1300 1 3.36
## 3756 0030592556 3.54 1996 1300 1 2.56
## 3758 0030592564 4.05 1996 1300 1 3.02
## 3760 0030592578 3.62 1996 1300 1 2.60
## 3813 0029820526 4.58 1996 1300 1 3.55
## 3821 16044371587 3.66 1996 1300 1 2.65
## 3839 0029798819 4.73 1996 1300 2 3.70
## 3841 0029842830 5.29 1996 1300 2 4.26
## 3844 16044361810 3.81 1996 1300 2 2.78
## 3881 0030595342 4.36 1996 1300 1 3.35
## 3886 0029817693 4.46 1996 1300 1 3.43
## 3920 0030572695 4.06 1996 1300 1 3.03
## 3922 0030572708 3.93 1996 1300 1 2.90
## 3949 9544226448 4.17 1996 1300 1 3.14
## 3951 0029791838 4.09 1996 1300 2 3.11
## 3967 0030598829 4.93 1996 1300 1 3.90
## 3970 0030598865 3.61 1996 1300 1 2.58
## 3974 16044363014 4.45 1996 1300 1 3.43
## 4005 0030576502 4.48 1996 1300 1 3.46
## 4008 0030576518 4.44 1996 1300 1 3.41
## 4079 0030602822 4.23 1996 1300 1 3.21
## 4080 0030602823 3.78 1996 1300 1 2.77
## 4082 0030602838 3.72 1996 1300 1 2.70
## 4083 16044364385 4.50 1996 1300 1 3.48
## 4128 0030581152 3.81 1996 1300 1 2.80
## 4131 0030581165 4.17 1996 1300 1 3.14
## 4132 0030581169 3.53 1996 1300 1 2.50
## 4133 0030581174 4.38 1996 1300 1 3.37
## 4136 15844386540 4.11 1996 1300 1 3.09
## 4153 0029954860 4.28 1996 1300 1 3.30
## 4157 0030010783 4.14 1996 1300 2 3.13
## 4171 0029944241 3.65 1996 1300 1 2.63
## 4173 0029999787 4.34 1996 1300 1 3.32
## 4178 15844378825 4.21 1996 1300 1 3.19
## 4188 0029944290 4.69 1996 1300 2 3.66
## 4191 0029994529 4.29 1996 1300 2 3.28
## 4224 0030604540 3.89 1996 1300 1 2.87
## 4228 0030604722 4.69 1996 1300 1 3.67
## 4231 15844367099 4.22 1996 1300 1 3.19
## 4273 0029894165 3.85 1996 1300 1 2.84
## 4274 0029899127 3.99 1996 1300 1 2.96
## 4275 0029953780 4.08 1996 1300 1 3.06
## 4302 0001506104 4.15 1996 1300 2 3.12
## 4334 0029895156 4.20 1996 1300 1 3.22
## 4336 0029939448 4.17 1996 1300 1 3.14
## 4337 0029943141 4.03 1996 1300 1 3.02
## 4340 0030010590 3.81 1996 1300 1 2.79
## 4342 15844384256 4.94 1996 1300 1 3.91
## 4343 15844415946 4.83 1996 1300 1 3.81
## 4344 15844420283 4.98 1996 1300 1 3.96
## 4361 0029881125 4.62 1996 1300 1 3.60
## 4362 0029892278 3.56 1996 1300 1 2.55
## 4448 0029870085 4.65 1996 1300 1 3.62
## 4451 0029993728 3.74 1996 1300 1 2.71
## 4452 0030009544 3.96 1996 1300 1 2.95
## 4453 15844372440 3.86 1996 1300 1 2.88
## 4500 0029919935 3.65 1996 1300 1 2.63
## 4501 0029980441 4.22 1996 1300 1 3.20
## 4502 0029993450 4.96 1996 1300 1 3.93
## 4565 0029876473 4.67 1996 1300 1 3.65
## 4588 0029880254 3.67 1996 1300 2 2.64
## 4589 0029880651 4.20 1996 1300 2 3.17
## 4606 0029669982 3.53 1996 1300 1 2.51
## 4611 0029961719 3.73 1996 1300 1 2.70
## 4613 0029965130 3.96 1996 1300 1 3.00
## 4614 0029978023 3.70 1996 1300 1 2.67
## 4615 0029991047 4.20 1996 1300 1 3.17
## 4617 13344277364 4.98 1996 1300 1 3.95
## 4703 0030030905 4.20 1996 1300 1 3.23
## 4705 0030048731 4.35 1996 1300 1 3.34
## 4706 0030053650 3.62 1996 1300 1 2.59
## 4707 0030058657 4.09 1996 1300 1 3.11
## 4714 0030065744 3.61 1996 1300 1 2.59
## 4715 13344261952 4.26 1996 1300 1 3.23
## 4716 13344282063 4.29 1996 1300 1 3.33
## 4727 0030020590 4.57 1996 1300 2 3.55
## 4751 0030024563 4.88 1996 1300 1 3.87
## 4797 0030033699 3.98 1996 1300 1 2.95
## 4808 0030026776 5.11 1996 1300 1 4.08
## 4809 0030026934 4.45 1996 1300 1 3.47
## 4830 0030584077 4.46 1996 1300 1 3.44
## 4831 0030584078 5.02 1996 1300 1 4.00
## 4833 0030584083 3.95 1996 1300 1 2.92
## 4854 0029671310 4.40 1996 1300 1 3.38
## 4856 0030031999 4.20 1996 1300 1 3.19
## 4858 0030034731 4.29 1996 1300 1 3.31
## 4859 0030034983 4.02 1996 1300 1 3.06
## 4860 0030050396 4.34 1996 1300 1 3.31
## 4862 0030061451 3.63 1996 1300 1 2.61
## 4887 0029664992 4.11 1996 1300 1 3.08
## 4901 0029888359 4.45 1996 1300 1 3.43
## 4903 0029890229 3.83 1996 1300 1 2.82
## 4908 0029898733 3.67 1996 1300 1 2.66
## 4917 0029936764 4.43 1996 1300 1 3.40
## 4918 0029940972 4.12 1996 1300 1 3.10
## 4922 0029949784 4.50 1996 1300 1 3.47
## 4926 0029977751 3.72 1996 1300 1 2.69
## 4929 0030014157 3.96 1996 1300 1 2.95
## 4934 0030030057 3.70 1996 1300 1 2.67
## 4962 0030111466 3.48 1996 1300 1 2.50
## 4974 0030131182 3.67 1996 1300 1 2.64
## 4981 0030139470 3.60 1996 1300 1 2.58
## 5009 0030271388 4.00 1996 1300 1 2.97
## 5010 0030271392 4.00 1996 1300 1 2.97
## 5012 0030271890 3.84 1996 1300 1 2.81
## 5014 0030271999 4.24 1996 1300 1 3.22
## 5015 0030272047 4.28 1996 1300 1 3.25
## 5043 15844380040 4.35 1996 1300 1 3.32
## 5044 15844417385 4.49 1996 1300 1 3.47
## 5868 0030087710 3.74 1996 1100 2 2.71
## 6075 0030606018 3.88 1996 1100 2 2.85
## 7056 0031444409 3.69 1997 1300 1 2.68
## 7057 0031456065 3.65 1997 1300 1 2.64
## 7061 0031459980 4.01 1997 1300 1 2.99
## 7128 0344936739 3.68 1997 1300 1 2.65
## 7227 0031449456 3.57 1997 1300 2 2.55
## 7229 0031451777 3.98 1997 1300 2 2.96
## 7235 0031466811 3.82 1997 1300 2 2.86
## 9659 0030613551 4.25 1997 1300 1 3.23
## 9665 0030702123 4.93 1997 1300 1 3.91
## 9696 0030886099 3.75 1997 1300 1 2.73
## 9699 0030886602 3.74 1997 1300 1 2.72
## 9702 0242421666 3.60 1997 1300 1 2.59
## 9712 0030796646 3.55 1997 1300 1 2.52
## 9773 0030928716 3.65 1997 1300 1 2.62
## 9776 0030954870 3.67 1997 1300 1 2.66
## 9777 0030985459 3.94 1997 1300 1 2.93
## 9838 0030865245 3.67 1997 1300 2 2.66
## 9899 0030788436 3.67 1997 1300 1 2.69
## 9902 0030829387 3.62 1997 1300 1 2.60
## 9906 0030848970 3.66 1997 1300 1 2.68
## 9966 0030746636 4.62 1997 1300 1 3.61
## 9975 0346613495 3.65 1997 1300 1 2.63
## 9977 0642270732 4.21 1997 1300 1 3.19
## 10022 0030755579 3.56 1997 1300 1 2.54
## 10026 0030877659 3.57 1997 1300 1 2.56
## 10027 0031586174 4.07 1997 1300 1 3.05
## 10074 0031440879 3.51 1997 1300 1 2.55
## 10075 0031444148 3.60 1997 1300 1 2.58
## 10083 0031472234 3.71 1997 1300 1 2.70
## 10095 16944366965 3.60 1997 1300 1 2.59
## 10096 0030970602 3.62 1997 1300 2 2.59
## 10117 0030752411 3.71 1997 1300 2 2.75
## 10314 0030687987 3.58 1997 1300 1 2.56
## 10317 0030712145 3.60 1997 1300 1 2.62
## 10337 0030729445 3.98 1997 1300 1 2.96
## 10343 18844476167 3.64 1997 1300 1 2.63
## 10344 20244377493 4.04 1997 1300 1 3.01
## 10376 0031007189 3.76 1997 1300 1 2.73
## 10461 0030970693 4.34 1997 1300 1 3.32
## 10462 0030982264 3.77 1997 1300 1 2.75
## 10515 0030890721 3.60 1997 1300 1 2.58
## 10519 0030963439 3.98 1997 1300 1 2.95
## 10585 0030933978 3.70 1997 1300 1 2.67
## 10616 0030944985 4.74 1997 1300 1 3.71
## 10618 0030970775 3.71 1997 1300 1 2.69
## 10620 0031000884 3.82 1997 1300 1 2.79
## 10678 0030909050 3.53 1997 1300 1 2.51
## 10679 0030949875 3.95 1997 1300 1 2.92
## 10685 0030614893 4.10 1997 1300 1 3.07
## 10688 0031048716 3.92 1997 1300 1 2.91
## 10712 0030893115 4.03 1997 1300 1 3.05
## 10739 0345877774 3.59 1997 1300 1 2.57
## 10740 0345877775 3.87 1997 1300 1 2.84
## 10785 0031285250 3.78 1997 1300 1 2.76
## 10802 0031015075 3.96 1997 1300 1 2.95
## 10839 0030994017 4.01 1997 1300 1 2.99
## 10876 0030715563 3.88 1997 1300 1 2.87
## 10881 0030800831 3.63 1997 1300 1 2.60
## 10884 0030866897 3.70 1997 1300 1 2.73
## 10958 0031587821 3.70 1997 1300 1 2.74
## 10959 0031587826 3.61 1997 1300 1 2.59
## 10961 0031587830 3.65 1997 1300 1 2.63
## 10974 16944364768 3.66 1997 1300 1 2.64
## 11042 0031022694 3.79 1997 1300 2 2.77
## 11356 0032446607 3.74 1998 1300 1 2.54
## 12409 0031720906 3.75 1998 1300 2 2.56
## 13176 0032540267 3.75 1998 1300 1 2.55
## 13338 0032549708 3.70 1998 1300 1 2.50
## 15773 0032953920 3.85 1999 1300 2 2.52
## 15917 0032976690 3.86 1999 1300 2 2.52
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3745 -0.3663 0.0259 0.3819 4.3056
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02656 0.03839 26.74 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.01217 0.00617 -1.97 0.04867 \*
## LastAuthorFemale1 -0.04919 0.00699 -7.04 2.0e-12 \*\*\*
## Year1997 -0.00330 0.05288 -0.06 0.95019
## Year1998 0.17370 0.05270 3.30 0.00098 \*\*\*
## Year1999 0.30995 0.04427 7.00 2.6e-12 \*\*\*
## Year2000 0.29104 0.04479 6.50 8.3e-11 \*\*\*
## Year2001 0.34789 0.04533 7.67 1.7e-14 \*\*\*
## Year2002 0.23309 0.04290 5.43 5.6e-08 \*\*\*
## Year2003 0.26073 0.04287 6.08 1.2e-09 \*\*\*
## Year2004 0.26109 0.04297 6.08 1.2e-09 \*\*\*
## Year2005 0.25429 0.04292 5.92 3.2e-09 \*\*\*
## Year2006 0.27977 0.04214 6.64 3.2e-11 \*\*\*
## Year2007 0.25885 0.04095 6.32 2.6e-10 \*\*\*
## Year2008 0.27357 0.04106 6.66 2.7e-11 \*\*\*
## Year2009 0.28167 0.04029 6.99 2.8e-12 \*\*\*
## Year2010 0.29567 0.03977 7.43 1.1e-13 \*\*\*
## Year2011 0.27513 0.03939 6.98 2.9e-12 \*\*\*
## Year2012 0.24349 0.03907 6.23 4.6e-10 \*\*\*
## Year2013 0.20686 0.03903 5.30 1.2e-07 \*\*\*
## Year2014 0.20337 0.03908 5.20 2.0e-07 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared: 0.0159, Adjusted R-squared: 0.0154
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 180 observations c(1,2,4,5,6,7,9,33,34,67,495,497,499,500,527,528,529,530,531,532,541,561,563,564,592,595,600,601,602,623,625,646,647,658,659,665,666,667,682,684,719,720,721,722,738,739,741,743,750,751,758,760,762,766,768,785,786,787,810,811,812,824,840,841,842,843,845,846,847,855,898,901,902,903,929,930,931,961,974,975,988,990,991,992,993,1035,1036,1037,1038,1042,1043,1044,1048,1063,1084,1089,1090,1098,1099,1100,1110,1111,1112,1113,1114,1115,1126,1127,1128,1130,1132,1133,1135,1139,1140,1144,1167,1190,1191,1193,1194,1195,1204,1205,1289,1442,1557,1558,1561,1589,1630,1633,2112,2115,2132,2135,2137,2172,2173,2174,2206,2234,2235,2236,2267,2270,2271,2296,2324,2328,2329,2339,2440,2450,2451,2452,2471,2517,2518,2548,2578,2596,2597,2598,2627,2632,2634,2648,2658,2682,2690,2710,2731,2734,2736,2782,2783,2784,2788,2830)
## are outliers with |weight| <= 1.1e-06 ( < 2.5e-06);
## 3322 weights are ~= 1. The remaining 36123 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.858 0.950 0.892 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.52e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.02 18 1.00



## [1] "List of 204 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 4 0030447612 4.32 1996 1300 1 3.29
## 6 0030448994 4.35 1996 1300 1 3.33
## 8 0030451819 4.40 1996 1300 1 3.38
## 9 0030460424 5.08 1996 1300 1 4.05
## 12 0030463470 4.71 1996 1300 1 3.68
## 14 0030475008 4.18 1996 1300 1 3.16
## 15 0030476713 3.58 1996 1300 1 2.55
## 16 0030480322 5.32 1996 1300 1 4.31
## 57 0030582677 4.26 1996 1300 1 3.23
## 60 0030582732 4.00 1996 1300 1 2.99
## 137 16144361909 3.76 1996 1300 1 2.74
## 3626 0030606239 5.32 1996 1300 1 4.29
## 3630 0030606286 3.86 1996 1300 1 2.83
## 3632 0030606315 3.68 1996 1300 1 2.65
## 3634 0030606320 4.48 1996 1300 1 3.45
## 3689 0030297537 4.47 1996 1300 1 3.45
## 3690 0030297538 3.97 1996 1300 1 2.96
## 3693 0030297895 4.20 1996 1300 1 3.19
## 3694 0030297912 3.84 1996 1300 1 2.83
## 3695 0030298137 4.19 1996 1300 1 3.16
## 3698 0030298375 4.69 1996 1300 1 3.73
## 3718 0029955497 4.50 1996 1300 2 3.49
## 3750 0001265782 3.56 1996 1300 1 2.53
## 3755 0030592544 4.37 1996 1300 1 3.36
## 3756 0030592556 3.54 1996 1300 1 2.56
## 3758 0030592564 4.05 1996 1300 1 3.02
## 3760 0030592578 3.62 1996 1300 1 2.60
## 3813 0029820526 4.58 1996 1300 1 3.55
## 3821 16044371587 3.66 1996 1300 1 2.65
## 3839 0029798819 4.73 1996 1300 2 3.70
## 3841 0029842830 5.29 1996 1300 2 4.26
## 3844 16044361810 3.81 1996 1300 2 2.78
## 3881 0030595342 4.36 1996 1300 1 3.35
## 3886 0029817693 4.46 1996 1300 1 3.43
## 3920 0030572695 4.06 1996 1300 1 3.03
## 3922 0030572708 3.93 1996 1300 1 2.90
## 3949 9544226448 4.17 1996 1300 1 3.14
## 3951 0029791838 4.09 1996 1300 2 3.11
## 3967 0030598829 4.93 1996 1300 1 3.90
## 3970 0030598865 3.61 1996 1300 1 2.58
## 3974 16044363014 4.45 1996 1300 1 3.43
## 4005 0030576502 4.48 1996 1300 1 3.46
## 4008 0030576518 4.44 1996 1300 1 3.41
## 4079 0030602822 4.23 1996 1300 1 3.21
## 4080 0030602823 3.78 1996 1300 1 2.77
## 4082 0030602838 3.72 1996 1300 1 2.70
## 4083 16044364385 4.50 1996 1300 1 3.48
## 4128 0030581152 3.81 1996 1300 1 2.80
## 4131 0030581165 4.17 1996 1300 1 3.14
## 4132 0030581169 3.53 1996 1300 1 2.50
## 4133 0030581174 4.38 1996 1300 1 3.37
## 4136 15844386540 4.11 1996 1300 1 3.09
## 4153 0029954860 4.28 1996 1300 1 3.30
## 4157 0030010783 4.14 1996 1300 2 3.13
## 4171 0029944241 3.65 1996 1300 1 2.63
## 4173 0029999787 4.34 1996 1300 1 3.32
## 4178 15844378825 4.21 1996 1300 1 3.19
## 4188 0029944290 4.69 1996 1300 2 3.66
## 4191 0029994529 4.29 1996 1300 2 3.28
## 4224 0030604540 3.89 1996 1300 1 2.87
## 4228 0030604722 4.69 1996 1300 1 3.67
## 4231 15844367099 4.22 1996 1300 1 3.19
## 4273 0029894165 3.85 1996 1300 1 2.84
## 4274 0029899127 3.99 1996 1300 1 2.96
## 4275 0029953780 4.08 1996 1300 1 3.06
## 4302 0001506104 4.15 1996 1300 2 3.12
## 4334 0029895156 4.20 1996 1300 1 3.22
## 4336 0029939448 4.17 1996 1300 1 3.14
## 4337 0029943141 4.03 1996 1300 1 3.02
## 4340 0030010590 3.81 1996 1300 1 2.79
## 4342 15844384256 4.94 1996 1300 1 3.91
## 4343 15844415946 4.83 1996 1300 1 3.81
## 4344 15844420283 4.98 1996 1300 1 3.96
## 4361 0029881125 4.62 1996 1300 1 3.60
## 4362 0029892278 3.56 1996 1300 1 2.55
## 4448 0029870085 4.65 1996 1300 1 3.62
## 4451 0029993728 3.74 1996 1300 1 2.71
## 4452 0030009544 3.96 1996 1300 1 2.95
## 4453 15844372440 3.86 1996 1300 1 2.88
## 4500 0029919935 3.65 1996 1300 1 2.63
## 4501 0029980441 4.22 1996 1300 1 3.20
## 4502 0029993450 4.96 1996 1300 1 3.93
## 4565 0029876473 4.67 1996 1300 1 3.65
## 4588 0029880254 3.67 1996 1300 2 2.64
## 4589 0029880651 4.20 1996 1300 2 3.17
## 4606 0029669982 3.53 1996 1300 1 2.51
## 4611 0029961719 3.73 1996 1300 1 2.70
## 4613 0029965130 3.96 1996 1300 1 3.00
## 4614 0029978023 3.70 1996 1300 1 2.67
## 4615 0029991047 4.20 1996 1300 1 3.17
## 4617 13344277364 4.98 1996 1300 1 3.95
## 4703 0030030905 4.20 1996 1300 1 3.23
## 4705 0030048731 4.35 1996 1300 1 3.34
## 4706 0030053650 3.62 1996 1300 1 2.59
## 4707 0030058657 4.09 1996 1300 1 3.11
## 4714 0030065744 3.61 1996 1300 1 2.59
## 4715 13344261952 4.26 1996 1300 1 3.23
## 4716 13344282063 4.29 1996 1300 1 3.33
## 4727 0030020590 4.57 1996 1300 2 3.55
## 4751 0030024563 4.88 1996 1300 1 3.87
## 4797 0030033699 3.98 1996 1300 1 2.95
## 4808 0030026776 5.11 1996 1300 1 4.08
## 4809 0030026934 4.45 1996 1300 1 3.47
## 4830 0030584077 4.46 1996 1300 1 3.44
## 4831 0030584078 5.02 1996 1300 1 4.00
## 4833 0030584083 3.95 1996 1300 1 2.92
## 4854 0029671310 4.40 1996 1300 1 3.38
## 4856 0030031999 4.20 1996 1300 1 3.19
## 4858 0030034731 4.29 1996 1300 1 3.31
## 4859 0030034983 4.02 1996 1300 1 3.06
## 4860 0030050396 4.34 1996 1300 1 3.31
## 4862 0030061451 3.63 1996 1300 1 2.61
## 4887 0029664992 4.11 1996 1300 1 3.08
## 4901 0029888359 4.45 1996 1300 1 3.43
## 4903 0029890229 3.83 1996 1300 1 2.82
## 4908 0029898733 3.67 1996 1300 1 2.66
## 4917 0029936764 4.43 1996 1300 1 3.40
## 4918 0029940972 4.12 1996 1300 1 3.10
## 4922 0029949784 4.50 1996 1300 1 3.47
## 4926 0029977751 3.72 1996 1300 1 2.69
## 4929 0030014157 3.96 1996 1300 1 2.95
## 4934 0030030057 3.70 1996 1300 1 2.67
## 4962 0030111466 3.48 1996 1300 1 2.50
## 4974 0030131182 3.67 1996 1300 1 2.64
## 4981 0030139470 3.60 1996 1300 1 2.58
## 5009 0030271388 4.00 1996 1300 1 2.97
## 5010 0030271392 4.00 1996 1300 1 2.97
## 5012 0030271890 3.84 1996 1300 1 2.81
## 5014 0030271999 4.24 1996 1300 1 3.22
## 5015 0030272047 4.28 1996 1300 1 3.25
## 5043 15844380040 4.35 1996 1300 1 3.32
## 5044 15844417385 4.49 1996 1300 1 3.47
## 5868 0030087710 3.74 1996 1100 2 2.71
## 6075 0030606018 3.88 1996 1100 2 2.85
## 7056 0031444409 3.69 1997 1300 1 2.68
## 7057 0031456065 3.65 1997 1300 1 2.64
## 7061 0031459980 4.01 1997 1300 1 2.99
## 7128 0344936739 3.68 1997 1300 1 2.65
## 7227 0031449456 3.57 1997 1300 2 2.55
## 7229 0031451777 3.98 1997 1300 2 2.96
## 7235 0031466811 3.82 1997 1300 2 2.86
## 9659 0030613551 4.25 1997 1300 1 3.23
## 9665 0030702123 4.93 1997 1300 1 3.91
## 9696 0030886099 3.75 1997 1300 1 2.73
## 9699 0030886602 3.74 1997 1300 1 2.72
## 9702 0242421666 3.60 1997 1300 1 2.59
## 9712 0030796646 3.55 1997 1300 1 2.52
## 9773 0030928716 3.65 1997 1300 1 2.62
## 9776 0030954870 3.67 1997 1300 1 2.66
## 9777 0030985459 3.94 1997 1300 1 2.93
## 9838 0030865245 3.67 1997 1300 2 2.66
## 9899 0030788436 3.67 1997 1300 1 2.69
## 9902 0030829387 3.62 1997 1300 1 2.60
## 9906 0030848970 3.66 1997 1300 1 2.68
## 9966 0030746636 4.62 1997 1300 1 3.61
## 9975 0346613495 3.65 1997 1300 1 2.63
## 9977 0642270732 4.21 1997 1300 1 3.19
## 10022 0030755579 3.56 1997 1300 1 2.54
## 10026 0030877659 3.57 1997 1300 1 2.56
## 10027 0031586174 4.07 1997 1300 1 3.05
## 10074 0031440879 3.51 1997 1300 1 2.55
## 10075 0031444148 3.60 1997 1300 1 2.58
## 10083 0031472234 3.71 1997 1300 1 2.70
## 10095 16944366965 3.60 1997 1300 1 2.59
## 10096 0030970602 3.62 1997 1300 2 2.59
## 10117 0030752411 3.71 1997 1300 2 2.75
## 10314 0030687987 3.58 1997 1300 1 2.56
## 10317 0030712145 3.60 1997 1300 1 2.62
## 10337 0030729445 3.98 1997 1300 1 2.96
## 10343 18844476167 3.64 1997 1300 1 2.63
## 10344 20244377493 4.04 1997 1300 1 3.01
## 10376 0031007189 3.76 1997 1300 1 2.73
## 10461 0030970693 4.34 1997 1300 1 3.32
## 10462 0030982264 3.77 1997 1300 1 2.75
## 10515 0030890721 3.60 1997 1300 1 2.58
## 10519 0030963439 3.98 1997 1300 1 2.95
## 10585 0030933978 3.70 1997 1300 1 2.67
## 10616 0030944985 4.74 1997 1300 1 3.71
## 10618 0030970775 3.71 1997 1300 1 2.69
## 10620 0031000884 3.82 1997 1300 1 2.79
## 10678 0030909050 3.53 1997 1300 1 2.51
## 10679 0030949875 3.95 1997 1300 1 2.92
## 10685 0030614893 4.10 1997 1300 1 3.07
## 10688 0031048716 3.92 1997 1300 1 2.91
## 10712 0030893115 4.03 1997 1300 1 3.05
## 10739 0345877774 3.59 1997 1300 1 2.57
## 10740 0345877775 3.87 1997 1300 1 2.84
## 10785 0031285250 3.78 1997 1300 1 2.76
## 10802 0031015075 3.96 1997 1300 1 2.95
## 10839 0030994017 4.01 1997 1300 1 2.99
## 10876 0030715563 3.88 1997 1300 1 2.87
## 10881 0030800831 3.63 1997 1300 1 2.60
## 10884 0030866897 3.70 1997 1300 1 2.73
## 10958 0031587821 3.70 1997 1300 1 2.74
## 10959 0031587826 3.61 1997 1300 1 2.59
## 10961 0031587830 3.65 1997 1300 1 2.63
## 10974 16944364768 3.66 1997 1300 1 2.64
## 11042 0031022694 3.79 1997 1300 2 2.77
## 11356 0032446607 3.74 1998 1300 1 2.54
## 12409 0031720906 3.75 1998 1300 2 2.56
## 13176 0032540267 3.75 1998 1300 1 2.55
## 13338 0032549708 3.70 1998 1300 1 2.50
## 15773 0032953920 3.85 1999 1300 2 2.52
## 15917 0032976690 3.86 1999 1300 2 2.52
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3654 -0.3630 0.0272 0.3826 4.3183
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02014 0.03844 26.54 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.01847 0.00619 -2.98 0.0028 \*\*
## Year1997 -0.00573 0.05296 -0.11 0.9139
## Year1998 0.17202 0.05280 3.26 0.0011 \*\*
## Year1999 0.30953 0.04432 6.98 2.9e-12 \*\*\*
## Year2000 0.28896 0.04485 6.44 1.2e-10 \*\*\*
## Year2001 0.34531 0.04539 7.61 2.8e-14 \*\*\*
## Year2002 0.23231 0.04297 5.41 6.5e-08 \*\*\*
## Year2003 0.25837 0.04291 6.02 1.7e-09 \*\*\*
## Year2004 0.25873 0.04299 6.02 1.8e-09 \*\*\*
## Year2005 0.25179 0.04298 5.86 4.7e-09 \*\*\*
## Year2006 0.27660 0.04218 6.56 5.6e-11 \*\*\*
## Year2007 0.25648 0.04100 6.26 4.0e-10 \*\*\*
## Year2008 0.27097 0.04113 6.59 4.5e-11 \*\*\*
## Year2009 0.27910 0.04036 6.91 4.7e-12 \*\*\*
## Year2010 0.29307 0.03983 7.36 1.9e-13 \*\*\*
## Year2011 0.27189 0.03945 6.89 5.6e-12 \*\*\*
## Year2012 0.24002 0.03912 6.14 8.6e-10 \*\*\*
## Year2013 0.20264 0.03908 5.19 2.2e-07 \*\*\*
## Year2014 0.19925 0.03913 5.09 3.6e-07 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared: 0.0146, Adjusted R-squared: 0.0142
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 183 observations c(1,2,4,5,6,7,9,33,34,67,495,497,499,500,527,528,529,530,531,532,541,561,563,564,592,595,600,601,602,623,625,646,647,658,659,665,666,667,682,684,719,720,721,722,738,739,741,743,750,751,758,760,762,766,768,785,786,787,810,811,812,824,840,841,842,843,845,846,847,855,898,901,902,903,929,930,931,961,974,975,988,990,991,992,993,1035,1036,1037,1038,1042,1043,1044,1048,1063,1084,1089,1090,1098,1099,1100,1110,1111,1112,1113,1114,1115,1126,1127,1128,1130,1132,1133,1135,1139,1140,1144,1167,1173,1190,1191,1193,1194,1195,1204,1205,1289,1442,1557,1558,1561,1589,1630,1633,2112,2115,2132,2135,2137,2172,2173,2174,2206,2234,2235,2236,2267,2270,2271,2296,2321,2324,2328,2329,2339,2440,2450,2451,2452,2471,2517,2518,2546,2548,2578,2596,2597,2598,2627,2632,2634,2648,2658,2682,2690,2710,2731,2734,2736,2782,2783,2784,2788,2830)
## are outliers with |weight| <= 1.1e-06 ( < 2.5e-06);
## 3345 weights are ~= 1. The remaining 36097 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.859 0.950 0.892 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.52e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1.01 18 1



## [1] "List of 204 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 4 0030447612 4.32 1996 1300 1 3.29
## 6 0030448994 4.35 1996 1300 1 3.33
## 8 0030451819 4.40 1996 1300 1 3.38
## 9 0030460424 5.08 1996 1300 1 4.05
## 12 0030463470 4.71 1996 1300 1 3.68
## 14 0030475008 4.18 1996 1300 1 3.16
## 15 0030476713 3.58 1996 1300 1 2.55
## 16 0030480322 5.32 1996 1300 1 4.31
## 57 0030582677 4.26 1996 1300 1 3.23
## 60 0030582732 4.00 1996 1300 1 2.99
## 137 16144361909 3.76 1996 1300 1 2.74
## 3626 0030606239 5.32 1996 1300 1 4.29
## 3630 0030606286 3.86 1996 1300 1 2.83
## 3632 0030606315 3.68 1996 1300 1 2.65
## 3634 0030606320 4.48 1996 1300 1 3.45
## 3689 0030297537 4.47 1996 1300 1 3.45
## 3690 0030297538 3.97 1996 1300 1 2.96
## 3693 0030297895 4.20 1996 1300 1 3.19
## 3694 0030297912 3.84 1996 1300 1 2.83
## 3695 0030298137 4.19 1996 1300 1 3.16
## 3698 0030298375 4.69 1996 1300 1 3.73
## 3718 0029955497 4.50 1996 1300 2 3.49
## 3750 0001265782 3.56 1996 1300 1 2.53
## 3755 0030592544 4.37 1996 1300 1 3.36
## 3756 0030592556 3.54 1996 1300 1 2.56
## 3758 0030592564 4.05 1996 1300 1 3.02
## 3760 0030592578 3.62 1996 1300 1 2.60
## 3813 0029820526 4.58 1996 1300 1 3.55
## 3821 16044371587 3.66 1996 1300 1 2.65
## 3839 0029798819 4.73 1996 1300 2 3.70
## 3841 0029842830 5.29 1996 1300 2 4.26
## 3844 16044361810 3.81 1996 1300 2 2.78
## 3881 0030595342 4.36 1996 1300 1 3.35
## 3886 0029817693 4.46 1996 1300 1 3.43
## 3920 0030572695 4.06 1996 1300 1 3.03
## 3922 0030572708 3.93 1996 1300 1 2.90
## 3949 9544226448 4.17 1996 1300 1 3.14
## 3951 0029791838 4.09 1996 1300 2 3.11
## 3967 0030598829 4.93 1996 1300 1 3.90
## 3970 0030598865 3.61 1996 1300 1 2.58
## 3974 16044363014 4.45 1996 1300 1 3.43
## 4005 0030576502 4.48 1996 1300 1 3.46
## 4008 0030576518 4.44 1996 1300 1 3.41
## 4079 0030602822 4.23 1996 1300 1 3.21
## 4080 0030602823 3.78 1996 1300 1 2.77
## 4082 0030602838 3.72 1996 1300 1 2.70
## 4083 16044364385 4.50 1996 1300 1 3.48
## 4128 0030581152 3.81 1996 1300 1 2.80
## 4131 0030581165 4.17 1996 1300 1 3.14
## 4132 0030581169 3.53 1996 1300 1 2.50
## 4133 0030581174 4.38 1996 1300 1 3.37
## 4136 15844386540 4.11 1996 1300 1 3.09
## 4153 0029954860 4.28 1996 1300 1 3.30
## 4157 0030010783 4.14 1996 1300 2 3.13
## 4171 0029944241 3.65 1996 1300 1 2.63
## 4173 0029999787 4.34 1996 1300 1 3.32
## 4178 15844378825 4.21 1996 1300 1 3.19
## 4188 0029944290 4.69 1996 1300 2 3.66
## 4191 0029994529 4.29 1996 1300 2 3.28
## 4224 0030604540 3.89 1996 1300 1 2.87
## 4228 0030604722 4.69 1996 1300 1 3.67
## 4231 15844367099 4.22 1996 1300 1 3.19
## 4273 0029894165 3.85 1996 1300 1 2.84
## 4274 0029899127 3.99 1996 1300 1 2.96
## 4275 0029953780 4.08 1996 1300 1 3.06
## 4302 0001506104 4.15 1996 1300 2 3.12
## 4334 0029895156 4.20 1996 1300 1 3.22
## 4336 0029939448 4.17 1996 1300 1 3.14
## 4337 0029943141 4.03 1996 1300 1 3.02
## 4340 0030010590 3.81 1996 1300 1 2.79
## 4342 15844384256 4.94 1996 1300 1 3.91
## 4343 15844415946 4.83 1996 1300 1 3.81
## 4344 15844420283 4.98 1996 1300 1 3.96
## 4361 0029881125 4.62 1996 1300 1 3.60
## 4362 0029892278 3.56 1996 1300 1 2.55
## 4448 0029870085 4.65 1996 1300 1 3.62
## 4451 0029993728 3.74 1996 1300 1 2.71
## 4452 0030009544 3.96 1996 1300 1 2.95
## 4453 15844372440 3.86 1996 1300 1 2.88
## 4500 0029919935 3.65 1996 1300 1 2.63
## 4501 0029980441 4.22 1996 1300 1 3.20
## 4502 0029993450 4.96 1996 1300 1 3.93
## 4565 0029876473 4.67 1996 1300 1 3.65
## 4588 0029880254 3.67 1996 1300 2 2.64
## 4589 0029880651 4.20 1996 1300 2 3.17
## 4606 0029669982 3.53 1996 1300 1 2.51
## 4611 0029961719 3.73 1996 1300 1 2.70
## 4613 0029965130 3.96 1996 1300 1 3.00
## 4614 0029978023 3.70 1996 1300 1 2.67
## 4615 0029991047 4.20 1996 1300 1 3.17
## 4617 13344277364 4.98 1996 1300 1 3.95
## 4703 0030030905 4.20 1996 1300 1 3.23
## 4705 0030048731 4.35 1996 1300 1 3.34
## 4706 0030053650 3.62 1996 1300 1 2.59
## 4707 0030058657 4.09 1996 1300 1 3.11
## 4714 0030065744 3.61 1996 1300 1 2.59
## 4715 13344261952 4.26 1996 1300 1 3.23
## 4716 13344282063 4.29 1996 1300 1 3.33
## 4727 0030020590 4.57 1996 1300 2 3.55
## 4751 0030024563 4.88 1996 1300 1 3.87
## 4797 0030033699 3.98 1996 1300 1 2.95
## 4808 0030026776 5.11 1996 1300 1 4.08
## 4809 0030026934 4.45 1996 1300 1 3.47
## 4830 0030584077 4.46 1996 1300 1 3.44
## 4831 0030584078 5.02 1996 1300 1 4.00
## 4833 0030584083 3.95 1996 1300 1 2.92
## 4854 0029671310 4.40 1996 1300 1 3.38
## 4856 0030031999 4.20 1996 1300 1 3.19
## 4858 0030034731 4.29 1996 1300 1 3.31
## 4859 0030034983 4.02 1996 1300 1 3.06
## 4860 0030050396 4.34 1996 1300 1 3.31
## 4862 0030061451 3.63 1996 1300 1 2.61
## 4887 0029664992 4.11 1996 1300 1 3.08
## 4901 0029888359 4.45 1996 1300 1 3.43
## 4903 0029890229 3.83 1996 1300 1 2.82
## 4908 0029898733 3.67 1996 1300 1 2.66
## 4917 0029936764 4.43 1996 1300 1 3.40
## 4918 0029940972 4.12 1996 1300 1 3.10
## 4922 0029949784 4.50 1996 1300 1 3.47
## 4926 0029977751 3.72 1996 1300 1 2.69
## 4929 0030014157 3.96 1996 1300 1 2.95
## 4934 0030030057 3.70 1996 1300 1 2.67
## 4962 0030111466 3.48 1996 1300 1 2.50
## 4974 0030131182 3.67 1996 1300 1 2.64
## 4981 0030139470 3.60 1996 1300 1 2.58
## 5009 0030271388 4.00 1996 1300 1 2.97
## 5010 0030271392 4.00 1996 1300 1 2.97
## 5012 0030271890 3.84 1996 1300 1 2.81
## 5014 0030271999 4.24 1996 1300 1 3.22
## 5015 0030272047 4.28 1996 1300 1 3.25
## 5043 15844380040 4.35 1996 1300 1 3.32
## 5044 15844417385 4.49 1996 1300 1 3.47
## 5868 0030087710 3.74 1996 1100 2 2.71
## 6075 0030606018 3.88 1996 1100 2 2.85
## 7056 0031444409 3.69 1997 1300 1 2.68
## 7057 0031456065 3.65 1997 1300 1 2.64
## 7061 0031459980 4.01 1997 1300 1 2.99
## 7128 0344936739 3.68 1997 1300 1 2.65
## 7227 0031449456 3.57 1997 1300 2 2.55
## 7229 0031451777 3.98 1997 1300 2 2.96
## 7235 0031466811 3.82 1997 1300 2 2.86
## 9659 0030613551 4.25 1997 1300 1 3.23
## 9665 0030702123 4.93 1997 1300 1 3.91
## 9696 0030886099 3.75 1997 1300 1 2.73
## 9699 0030886602 3.74 1997 1300 1 2.72
## 9702 0242421666 3.60 1997 1300 1 2.59
## 9712 0030796646 3.55 1997 1300 1 2.52
## 9773 0030928716 3.65 1997 1300 1 2.62
## 9776 0030954870 3.67 1997 1300 1 2.66
## 9777 0030985459 3.94 1997 1300 1 2.93
## 9838 0030865245 3.67 1997 1300 2 2.66
## 9899 0030788436 3.67 1997 1300 1 2.69
## 9902 0030829387 3.62 1997 1300 1 2.60
## 9906 0030848970 3.66 1997 1300 1 2.68
## 9966 0030746636 4.62 1997 1300 1 3.61
## 9975 0346613495 3.65 1997 1300 1 2.63
## 9977 0642270732 4.21 1997 1300 1 3.19
## 10022 0030755579 3.56 1997 1300 1 2.54
## 10026 0030877659 3.57 1997 1300 1 2.56
## 10027 0031586174 4.07 1997 1300 1 3.05
## 10074 0031440879 3.51 1997 1300 1 2.55
## 10075 0031444148 3.60 1997 1300 1 2.58
## 10083 0031472234 3.71 1997 1300 1 2.70
## 10095 16944366965 3.60 1997 1300 1 2.59
## 10096 0030970602 3.62 1997 1300 2 2.59
## 10117 0030752411 3.71 1997 1300 2 2.75
## 10314 0030687987 3.58 1997 1300 1 2.56
## 10317 0030712145 3.60 1997 1300 1 2.62
## 10337 0030729445 3.98 1997 1300 1 2.96
## 10343 18844476167 3.64 1997 1300 1 2.63
## 10344 20244377493 4.04 1997 1300 1 3.01
## 10376 0031007189 3.76 1997 1300 1 2.73
## 10461 0030970693 4.34 1997 1300 1 3.32
## 10462 0030982264 3.77 1997 1300 1 2.75
## 10515 0030890721 3.60 1997 1300 1 2.58
## 10519 0030963439 3.98 1997 1300 1 2.95
## 10585 0030933978 3.70 1997 1300 1 2.67
## 10616 0030944985 4.74 1997 1300 1 3.71
## 10618 0030970775 3.71 1997 1300 1 2.69
## 10620 0031000884 3.82 1997 1300 1 2.79
## 10678 0030909050 3.53 1997 1300 1 2.51
## 10679 0030949875 3.95 1997 1300 1 2.92
## 10685 0030614893 4.10 1997 1300 1 3.07
## 10688 0031048716 3.92 1997 1300 1 2.91
## 10712 0030893115 4.03 1997 1300 1 3.05
## 10739 0345877774 3.59 1997 1300 1 2.57
## 10740 0345877775 3.87 1997 1300 1 2.84
## 10785 0031285250 3.78 1997 1300 1 2.76
## 10802 0031015075 3.96 1997 1300 1 2.95
## 10839 0030994017 4.01 1997 1300 1 2.99
## 10876 0030715563 3.88 1997 1300 1 2.87
## 10881 0030800831 3.63 1997 1300 1 2.60
## 10884 0030866897 3.70 1997 1300 1 2.73
## 10958 0031587821 3.70 1997 1300 1 2.74
## 10959 0031587826 3.61 1997 1300 1 2.59
## 10961 0031587830 3.65 1997 1300 1 2.63
## 10974 16944364768 3.66 1997 1300 1 2.64
## 11042 0031022694 3.79 1997 1300 2 2.77
## 11356 0032446607 3.74 1998 1300 1 2.54
## 12409 0031720906 3.75 1998 1300 2 2.56
## 13176 0032540267 3.75 1998 1300 1 2.55
## 13338 0032549708 3.70 1998 1300 1 2.50
## 15773 0032953920 3.85 1999 1300 2 2.52
## 15917 0032976690 3.86 1999 1300 2 2.52
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3715 -0.3643 0.0261 0.3813 4.2961
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02389 0.03835 26.70 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.05120 0.00701 -7.31 2.8e-13 \*\*\*
## Year1997 -0.00280 0.05281 -0.05 0.95772
## Year1998 0.17359 0.05266 3.30 0.00098 \*\*\*
## Year1999 0.30967 0.04426 7.00 2.7e-12 \*\*\*
## Year2000 0.29070 0.04477 6.49 8.5e-11 \*\*\*
## Year2001 0.34757 0.04531 7.67 1.8e-14 \*\*\*
## Year2002 0.23230 0.04288 5.42 6.1e-08 \*\*\*
## Year2003 0.26002 0.04285 6.07 1.3e-09 \*\*\*
## Year2004 0.26034 0.04295 6.06 1.4e-09 \*\*\*
## Year2005 0.25310 0.04290 5.90 3.7e-09 \*\*\*
## Year2006 0.27883 0.04211 6.62 3.6e-11 \*\*\*
## Year2007 0.25759 0.04092 6.30 3.1e-10 \*\*\*
## Year2008 0.27247 0.04104 6.64 3.2e-11 \*\*\*
## Year2009 0.28061 0.04027 6.97 3.2e-12 \*\*\*
## Year2010 0.29402 0.03974 7.40 1.4e-13 \*\*\*
## Year2011 0.27342 0.03936 6.95 3.8e-12 \*\*\*
## Year2012 0.24168 0.03903 6.19 6.0e-10 \*\*\*
## Year2013 0.20524 0.03899 5.26 1.4e-07 \*\*\*
## Year2014 0.20142 0.03904 5.16 2.5e-07 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared: 0.0158, Adjusted R-squared: 0.0153
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 179 observations c(1,2,4,5,6,7,9,33,34,67,495,497,499,500,527,528,529,530,531,532,541,561,563,564,592,595,600,601,602,623,625,646,647,658,659,665,666,667,682,684,719,720,721,722,738,739,741,743,750,751,758,760,762,766,768,785,786,787,810,811,812,824,840,841,842,843,845,846,847,855,898,901,902,903,929,930,931,961,974,975,988,990,991,992,993,1035,1036,1037,1038,1042,1043,1044,1048,1063,1084,1089,1090,1098,1099,1100,1110,1111,1112,1113,1114,1115,1126,1127,1128,1130,1132,1133,1135,1139,1140,1144,1167,1190,1191,1193,1194,1195,1204,1205,1289,1442,1557,1558,1561,1589,1630,1633,2112,2115,2132,2135,2137,2172,2173,2174,2206,2234,2235,2236,2267,2270,2271,2296,2324,2329,2339,2440,2450,2451,2452,2471,2517,2518,2548,2578,2596,2597,2598,2627,2632,2634,2648,2658,2682,2690,2710,2731,2734,2736,2782,2783,2784,2788,2830)
## are outliers with |weight| <= 1.4e-06 ( < 2.5e-06);
## 3244 weights are ~= 1. The remaining 36202 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.859 0.950 0.893 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.52e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 39625"
## [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]"
##
## 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014
## 14 12 4 13 31 37 51 61 77 125 114
##
## 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014
## 10 5 3 5 19 29 32 46 49 84 81
##
## 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014
## 7 5 3 4 16 24 27 38 36 72 70
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 30, df = 10, p-value = 2e-04



##
## 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 geometric mean: 2.65554386212592"
## [1] "Male first author team size geometric mean: 2.79120672828282"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 9000, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 2.6713254923574"
## [1] "Male last author team size geometric mean: 2.77195313462523"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 7000, 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 1.57 1 1.25
## LastAuthorFemale 1.49 1 1.22
## UniqueAuthors 2.90 4 1.14
## Year 3.68 10 1.07



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.332 -0.376 0.037 0.385 1.685
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.72564 0.16353 4.44 1.3e-05 \*\*\*
## FirstAuthorFemale1 0.15062 0.08404 1.79 0.0742 .
## LastAuthorFemale1 -0.21860 0.09703 -2.25 0.0250 \*
## UniqueAuthors2 0.38327 0.09576 4.00 8.0e-05 \*\*\*
## UniqueAuthors3 0.53485 0.20387 2.62 0.0092 \*\*
## UniqueAuthors4 0.30344 0.16271 1.86 0.0632 .
## UniqueAuthors5 0.60911 0.09543 6.38 7.0e-10 \*\*\*
## Year2005 -0.03972 0.20366 -0.20 0.8455
## Year2006 -0.12622 0.19049 -0.66 0.5081
## Year2007 0.64184 0.23250 2.76 0.0061 \*\*
## Year2008 0.21802 0.40589 0.54 0.5916
## Year2009 0.19990 0.23304 0.86 0.3917
## Year2010 0.00875 0.18804 0.05 0.9629
## Year2011 -0.09066 0.17668 -0.51 0.6083
## Year2012 0.08979 0.18502 0.49 0.6279
## Year2013 -0.00303 0.16446 -0.02 0.9853
## Year2014 0.13673 0.16580 0.82 0.4103
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.572
## Multiple R-squared: 0.171, Adjusted R-squared: 0.125
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 279 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.365 0.872 0.954 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 3.31e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.40 1 1.18
## LastAuthorFemale 1.31 1 1.15
## Year 1.70 10 1.03



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2700 -0.4007 0.0449 0.4156 2.0377
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0551 0.1208 8.73 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.1614 0.0820 1.97 0.04995 \*
## LastAuthorFemale1 -0.2667 0.1029 -2.59 0.01001 \*
## Year2005 0.1734 0.2066 0.84 0.40202
## Year2006 -0.1505 0.1397 -1.08 0.28251
## Year2007 0.7601 0.1967 3.87 0.00014 \*\*\*
## Year2008 0.1452 0.4014 0.36 0.71780
## Year2009 0.1734 0.2103 0.82 0.41017
## Year2010 0.0394 0.1595 0.25 0.80479
## Year2011 -0.0350 0.1503 -0.23 0.81602
## Year2012 0.1593 0.1675 0.95 0.34239
## Year2013 0.0551 0.1439 0.38 0.70188
## Year2014 0.2149 0.1434 1.50 0.13505
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared: 0.0601, Adjusted R-squared: 0.0211
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 277 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.237 0.865 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 3.31e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.29 1 1.13
## Year 1.29 10 1.01



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3475 -0.4101 0.0319 0.4132 2.1210
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06331 0.11791 9.02 <2e-16 \*\*\*
## FirstAuthorFemale1 0.10534 0.08329 1.26 0.207
## Year2005 0.12950 0.21734 0.60 0.552
## Year2006 -0.15868 0.13722 -1.16 0.248
## Year2007 0.61809 0.25125 2.46 0.014 \*
## Year2008 0.05370 0.42435 0.13 0.899
## Year2009 0.14924 0.20679 0.72 0.471
## Year2010 0.02653 0.15475 0.17 0.864
## Year2011 -0.07605 0.14986 -0.51 0.612
## Year2012 0.09748 0.16338 0.60 0.551
## Year2013 0.00299 0.14074 0.02 0.983
## Year2014 0.17880 0.14130 1.27 0.207
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.613
## Multiple R-squared: 0.0357, Adjusted R-squared: -0.000905
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 280 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.207 0.868 0.950 0.908 0.987 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 3.31e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## LastAuthorFemale 1.27 1 1.12
## Year 1.27 10 1.01



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30843 -0.40618 0.00876 0.42673 2.02118
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0789 0.1129 9.55 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.2194 0.1000 -2.19 0.02911 \*
## Year2005 0.2040 0.2194 0.93 0.35331
## Year2006 -0.1742 0.1330 -1.31 0.19111
## Year2007 0.7126 0.2007 3.55 0.00045 \*\*\*
## Year2008 0.1380 0.4508 0.31 0.75979
## Year2009 0.2183 0.2093 1.04 0.29796
## Year2010 0.0600 0.1516 0.40 0.69280
## Year2011 -0.0282 0.1464 -0.19 0.84754
## Year2012 0.1674 0.1593 1.05 0.29409
## Year2013 0.0772 0.1371 0.56 0.57378
## Year2014 0.2296 0.1361 1.69 0.09269 .
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.606
## Multiple R-squared: 0.049, Adjusted R-squared: 0.0129
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 34 weights are ~= 1. The remaining 268 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.244 0.872 0.945 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 3.31e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 302"
## [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
## 333 283 305 278 303 338 340 272 308 351 388 433 446 420 460
## 2011 2012 2013 2014
## 467 566 521 526
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 239 169 220 200 159 140 260 199 234 274 295 327 332 330 335
## 2011 2012 2013 2014
## 340 413 387 390
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 221 156 207 177 143 122 228 174 204 244 264 291 287 291 291
## 2011 2012 2013 2014
## 310 368 340 341
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 70, df = 20, p-value = 8e-08



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2, df = 1, p-value = 0.2



## [1] "Female first author team size geometric mean: 3.35947472918858"
## [1] "Male first author team size geometric mean: 3.26612203035433"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 3e+06, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.08521365727462"
## [1] "Male last author team size geometric mean: 3.42029352231071"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 2e+06, p-value = 8e-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 1.04 1 1.02
## LastAuthorFemale 1.04 1 1.02
## UniqueAuthors 1.14 4 1.02
## Year 1.17 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4106 -0.2597 0.0119 0.2513 1.5993
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11079 0.03702 30.00 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.00061 0.01200 0.05 0.95948
## LastAuthorFemale1 0.02537 0.01282 1.98 0.04784 \*
## UniqueAuthors2 0.09882 0.02362 4.18 2.9e-05 \*\*\*
## UniqueAuthors3 0.18960 0.02855 6.64 3.5e-11 \*\*\*
## UniqueAuthors4 0.21077 0.02472 8.53 < 2e-16 \*\*\*
## UniqueAuthors5 0.28576 0.02176 13.13 < 2e-16 \*\*\*
## Year1997 -0.05065 0.04682 -1.08 0.27936
## Year1998 -0.10491 0.04702 -2.23 0.02570 \*
## Year1999 -0.10067 0.04476 -2.25 0.02454 \*
## Year2000 -0.01192 0.04799 -0.25 0.80378
## Year2001 -0.07042 0.05846 -1.20 0.22838
## Year2002 -0.02946 0.04122 -0.71 0.47488
## Year2003 -0.05428 0.04500 -1.21 0.22779
## Year2004 -0.01987 0.04097 -0.48 0.62773
## Year2005 -0.06971 0.04038 -1.73 0.08436 .
## Year2006 -0.07109 0.03917 -1.81 0.06961 .
## Year2007 -0.10616 0.03913 -2.71 0.00669 \*\*
## Year2008 -0.14606 0.03949 -3.70 0.00022 \*\*\*
## Year2009 -0.06787 0.04037 -1.68 0.09277 .
## Year2010 -0.08005 0.03774 -2.12 0.03394 \*
## Year2011 -0.11381 0.03911 -2.91 0.00363 \*\*
## Year2012 -0.13041 0.03743 -3.48 0.00050 \*\*\*
## Year2013 -0.08817 0.03834 -2.30 0.02153 \*
## Year2014 -0.17259 0.03921 -4.40 1.1e-05 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared: 0.0696, Adjusted R-squared: 0.0648
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 380 weights are ~= 1. The remaining 4279 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.037 0.867 0.950 0.894 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.15e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.03 1 1.01
## LastAuthorFemale 1.03 1 1.02
## Year 1.04 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.290 -0.265 0.010 0.261 1.683
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26010 0.03224 39.09 <2e-16 \*\*\*
## FirstAuthorFemale1 0.00552 0.01228 0.45 0.653
## LastAuthorFemale1 0.01209 0.01309 0.92 0.356
## Year1997 -0.05968 0.04808 -1.24 0.215
## Year1998 -0.10404 0.04679 -2.22 0.026 \*
## Year1999 -0.09397 0.04543 -2.07 0.039 \*
## Year2000 0.01236 0.04807 0.26 0.797
## Year2001 -0.05469 0.05810 -0.94 0.347
## Year2002 -0.01796 0.04217 -0.43 0.670
## Year2003 -0.02212 0.04565 -0.48 0.628
## Year2004 0.00512 0.04096 0.13 0.900
## Year2005 -0.02063 0.04071 -0.51 0.612
## Year2006 -0.03510 0.03986 -0.88 0.379
## Year2007 -0.07119 0.04005 -1.78 0.076 .
## Year2008 -0.09984 0.03954 -2.52 0.012 \*
## Year2009 -0.02661 0.04062 -0.66 0.512
## Year2010 -0.02865 0.03768 -0.76 0.447
## Year2011 -0.06078 0.03864 -1.57 0.116
## Year2012 -0.06301 0.03730 -1.69 0.091 .
## Year2013 -0.01491 0.03841 -0.39 0.698
## Year2014 -0.09932 0.03948 -2.52 0.012 \*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared: 0.00786, Adjusted R-squared: 0.00358
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 384 weights are ~= 1. The remaining 4275 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.022 0.868 0.950 0.893 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.15e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.02 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2821 -0.2657 0.0103 0.2603 1.6801
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26323 0.03211 39.35 <2e-16 \*\*\*
## FirstAuthorFemale1 0.00751 0.01225 0.61 0.540
## Year1997 -0.05933 0.04810 -1.23 0.217
## Year1998 -0.10351 0.04678 -2.21 0.027 \*
## Year1999 -0.09372 0.04540 -2.06 0.039 \*
## Year2000 0.01134 0.04805 0.24 0.813
## Year2001 -0.05507 0.05812 -0.95 0.343
## Year2002 -0.01748 0.04211 -0.41 0.678
## Year2003 -0.02147 0.04562 -0.47 0.638
## Year2004 0.00542 0.04093 0.13 0.895
## Year2005 -0.02029 0.04071 -0.50 0.618
## Year2006 -0.03552 0.03983 -0.89 0.373
## Year2007 -0.07103 0.04002 -1.77 0.076 .
## Year2008 -0.10005 0.03951 -2.53 0.011 \*
## Year2009 -0.02655 0.04060 -0.65 0.513
## Year2010 -0.02935 0.03764 -0.78 0.436
## Year2011 -0.06133 0.03862 -1.59 0.112
## Year2012 -0.06309 0.03726 -1.69 0.090 .
## Year2013 -0.01537 0.03836 -0.40 0.689
## Year2014 -0.09956 0.03946 -2.52 0.012 \*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared: 0.00764, Adjusted R-squared: 0.00358
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 386 weights are ~= 1. The remaining 4273 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.023 0.867 0.951 0.893 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.15e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.02 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2880 -0.2641 0.0109 0.2620 1.6806
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26226 0.03179 39.71 <2e-16 \*\*\*
## LastAuthorFemale1 0.01311 0.01305 1.01 0.315
## Year1997 -0.06006 0.04808 -1.25 0.212
## Year1998 -0.10440 0.04678 -2.23 0.026 \*
## Year1999 -0.09434 0.04540 -2.08 0.038 \*
## Year2000 0.01267 0.04805 0.26 0.792
## Year2001 -0.05468 0.05811 -0.94 0.347
## Year2002 -0.01827 0.04218 -0.43 0.665
## Year2003 -0.02187 0.04563 -0.48 0.632
## Year2004 0.00509 0.04097 0.12 0.901
## Year2005 -0.02059 0.04072 -0.51 0.613
## Year2006 -0.03536 0.03986 -0.89 0.375
## Year2007 -0.07113 0.04006 -1.78 0.076 .
## Year2008 -0.09948 0.03954 -2.52 0.012 \*
## Year2009 -0.02635 0.04064 -0.65 0.517
## Year2010 -0.02848 0.03768 -0.76 0.450
## Year2011 -0.06090 0.03865 -1.58 0.115
## Year2012 -0.06269 0.03730 -1.68 0.093 .
## Year2013 -0.01492 0.03842 -0.39 0.698
## Year2014 -0.09950 0.03948 -2.52 0.012 \*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared: 0.00782, Adjusted R-squared: 0.00375
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 387 weights are ~= 1. The remaining 4272 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.023 0.868 0.950 0.893 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.15e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4659"
## [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
## 18463 15727 13926 11208 12065 10650 12430 10983 11797 11520 11649 11447
## 2008 2009 2010 2011 2012 2013 2014
## 11686 12238 12150 12467 11532 10956 10245
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8787 8151 7598 7200 6632 5209 8009 6876 7295 7140 7165 7173 7412 7670 7561
## 2011 2012 2013 2014
## 7831 7353 7028 6641
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7376 6870 6436 6014 5588 4357 6690 5784 6081 5895 5901 5897 6131 6338 6195
## 2011 2012 2013 2014
## 6422 6064 5786 5461
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 8000, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 100, df = 1, p-value <2e-16



## [1] "Female first author team size geometric mean: 3.75521546448353"
## [1] "Male first author team size geometric mean: 3.47550747279925"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 2e+09, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.60897209160943"
## [1] "Male last author team size geometric mean: 3.55310623873974"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 1e+09, p-value = 3e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.01 1 1.01
## LastAuthorFemale 1.01 1 1.01
## UniqueAuthors 1.04 4 1.00
## Year 1.04 18 1.00



## [1] "List of 3 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 52286 0031773680 3.96 1998 1303 6 2.78
## 192756 65449136284 3.76 2009 1303 7 2.67
## 253585 84871809302 4.00 2013 1303 6 2.79
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.37364 -0.23692 0.00292 0.23094 2.78637
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08899 0.00903 120.60 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.01197 0.00232 -5.15 2.6e-07 \*\*\*
## LastAuthorFemale1 -0.01496 0.00286 -5.24 1.6e-07 \*\*\*
## UniqueAuthors2 0.16666 0.00711 23.44 < 2e-16 \*\*\*
## UniqueAuthors3 0.22243 0.00775 28.72 < 2e-16 \*\*\*
## UniqueAuthors4 0.21438 0.00726 29.55 < 2e-16 \*\*\*
## UniqueAuthors5 0.28464 0.00701 40.60 < 2e-16 \*\*\*
## Year1997 -0.01871 0.00879 -2.13 0.033 \*
## Year1998 -0.07654 0.00820 -9.34 < 2e-16 \*\*\*
## Year1999 -0.13131 0.00755 -17.40 < 2e-16 \*\*\*
## Year2000 -0.13831 0.00798 -17.34 < 2e-16 \*\*\*
## Year2001 -0.15394 0.00792 -19.44 < 2e-16 \*\*\*
## Year2002 -0.16213 0.00730 -22.21 < 2e-16 \*\*\*
## Year2003 -0.18378 0.00741 -24.79 < 2e-16 \*\*\*
## Year2004 -0.17855 0.00727 -24.55 < 2e-16 \*\*\*
## Year2005 -0.18435 0.00738 -24.98 < 2e-16 \*\*\*
## Year2006 -0.19274 0.00742 -25.99 < 2e-16 \*\*\*
## Year2007 -0.16948 0.00753 -22.50 < 2e-16 \*\*\*
## Year2008 -0.16723 0.00746 -22.40 < 2e-16 \*\*\*
## Year2009 -0.16414 0.00753 -21.80 < 2e-16 \*\*\*
## Year2010 -0.16595 0.00756 -21.96 < 2e-16 \*\*\*
## Year2011 -0.16584 0.00756 -21.93 < 2e-16 \*\*\*
## Year2012 -0.17284 0.00781 -22.13 < 2e-16 \*\*\*
## Year2013 -0.16201 0.00804 -20.15 < 2e-16 \*\*\*
## Year2014 -0.15284 0.00845 -18.08 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared: 0.0534, Adjusted R-squared: 0.0532
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 44 observations c(1531,5850,6115,7919,11149,19539,20586,24556,36589,36590,43261,49737,51165,53965,58649,60019,68414,69355,72907,75615,82610,82858,88052,89379,90517,92684,94109,95291,96853,96855,97334,101548,101798,102174,102831,104323,108419,109806,111811,111818,112703,114417,114713,115251)
## are outliers with |weight| = 0 ( < 8.7e-07);
## 9642 weights are ~= 1. The remaining 105600 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.863 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 8.67e-07 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.01 1 1
## LastAuthorFemale 1.01 1 1
## Year 1.01 18 1



## [1] "List of 3 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 52286 0031773680 3.96 1998 1303 6 2.75
## 192756 65449136284 3.76 2009 1303 7 2.61
## 253585 84871809302 4.00 2013 1303 6 2.84
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28730 -0.24125 0.00424 0.23127 2.84267
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28730 0.00618 208.27 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.00285 0.00235 -1.21 0.23
## LastAuthorFemale1 -0.01465 0.00289 -5.07 3.9e-07 \*\*\*
## Year1997 -0.01369 0.00888 -1.54 0.12
## Year1998 -0.07402 0.00829 -8.93 < 2e-16 \*\*\*
## Year1999 -0.12776 0.00763 -16.75 < 2e-16 \*\*\*
## Year2000 -0.13193 0.00807 -16.35 < 2e-16 \*\*\*
## Year2001 -0.14108 0.00804 -17.56 < 2e-16 \*\*\*
## Year2002 -0.14856 0.00738 -20.13 < 2e-16 \*\*\*
## Year2003 -0.16670 0.00749 -22.24 < 2e-16 \*\*\*
## Year2004 -0.16102 0.00736 -21.88 < 2e-16 \*\*\*
## Year2005 -0.16619 0.00747 -22.23 < 2e-16 \*\*\*
## Year2006 -0.17161 0.00749 -22.91 < 2e-16 \*\*\*
## Year2007 -0.14743 0.00763 -19.33 < 2e-16 \*\*\*
## Year2008 -0.14320 0.00756 -18.95 < 2e-16 \*\*\*
## Year2009 -0.13944 0.00762 -18.31 < 2e-16 \*\*\*
## Year2010 -0.13907 0.00764 -18.20 < 2e-16 \*\*\*
## Year2011 -0.13930 0.00764 -18.23 < 2e-16 \*\*\*
## Year2012 -0.14479 0.00791 -18.31 < 2e-16 \*\*\*
## Year2013 -0.13197 0.00812 -16.25 < 2e-16 \*\*\*
## Year2014 -0.12378 0.00855 -14.47 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.353
## Multiple R-squared: 0.0169, Adjusted R-squared: 0.0167
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 34 observations c(1531,5850,6115,7919,19539,20586,36590,44926,51165,60019,68414,69355,72907,82858,88052,89379,90517,92684,94109,95291,96853,97334,101548,101798,102174,102831,104323,108419,109806,111811,111818,112703,114713,115251)
## are outliers with |weight| = 0 ( < 8.7e-07);
## 9921 weights are ~= 1. The remaining 105331 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.862 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 8.67e-07 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1 18 1



## [1] "List of 3 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 52286 0031773680 3.96 1998 1303 6 2.75
## 192756 65449136284 3.76 2009 1303 7 2.61
## 253585 84871809302 4.00 2013 1303 6 2.84
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28555 -0.24080 0.00441 0.23159 2.84541
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28555 0.00618 208.10 <2e-16 \*\*\*
## FirstAuthorFemale1 -0.00399 0.00235 -1.70 0.089 .
## Year1997 -0.01375 0.00888 -1.55 0.122
## Year1998 -0.07418 0.00829 -8.95 <2e-16 \*\*\*
## Year1999 -0.12814 0.00763 -16.80 <2e-16 \*\*\*
## Year2000 -0.13208 0.00807 -16.36 <2e-16 \*\*\*
## Year2001 -0.14170 0.00804 -17.63 <2e-16 \*\*\*
## Year2002 -0.14899 0.00738 -20.18 <2e-16 \*\*\*
## Year2003 -0.16709 0.00750 -22.29 <2e-16 \*\*\*
## Year2004 -0.16153 0.00736 -21.94 <2e-16 \*\*\*
## Year2005 -0.16656 0.00748 -22.28 <2e-16 \*\*\*
## Year2006 -0.17228 0.00749 -22.99 <2e-16 \*\*\*
## Year2007 -0.14812 0.00763 -19.42 <2e-16 \*\*\*
## Year2008 -0.14406 0.00755 -19.07 <2e-16 \*\*\*
## Year2009 -0.14009 0.00762 -18.39 <2e-16 \*\*\*
## Year2010 -0.13971 0.00764 -18.28 <2e-16 \*\*\*
## Year2011 -0.14029 0.00764 -18.36 <2e-16 \*\*\*
## Year2012 -0.14554 0.00791 -18.40 <2e-16 \*\*\*
## Year2013 -0.13296 0.00812 -16.37 <2e-16 \*\*\*
## Year2014 -0.12479 0.00855 -14.60 <2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.353
## Multiple R-squared: 0.0167, Adjusted R-squared: 0.0165
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 34 observations c(1531,5850,6115,7919,19539,20586,36590,43261,44926,51165,60019,68414,69355,72907,82858,88052,89379,90517,92684,94109,95291,96853,97334,101548,101798,102174,102831,104323,108419,109806,111811,112703,114713,115251)
## are outliers with |weight| = 0 ( < 8.7e-07);
## 9917 weights are ~= 1. The remaining 105335 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.862 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 8.67e-07 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1.01 18 1



## [1] "List of 3 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 52286 0031773680 3.96 1998 1303 6 2.75
## 192756 65449136284 3.76 2009 1303 7 2.61
## 253585 84871809302 4.00 2013 1303 6 2.84
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28652 -0.24116 0.00419 0.23147 2.84362
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28652 0.00615 209.36 <2e-16 \*\*\*
## LastAuthorFemale1 -0.01498 0.00288 -5.20 2e-07 \*\*\*
## Year1997 -0.01371 0.00888 -1.54 0.12
## Year1998 -0.07403 0.00829 -8.93 <2e-16 \*\*\*
## Year1999 -0.12780 0.00763 -16.76 <2e-16 \*\*\*
## Year2000 -0.13197 0.00807 -16.36 <2e-16 \*\*\*
## Year2001 -0.14114 0.00804 -17.56 <2e-16 \*\*\*
## Year2002 -0.14863 0.00738 -20.14 <2e-16 \*\*\*
## Year2003 -0.16680 0.00749 -22.26 <2e-16 \*\*\*
## Year2004 -0.16112 0.00736 -21.90 <2e-16 \*\*\*
## Year2005 -0.16629 0.00747 -22.25 <2e-16 \*\*\*
## Year2006 -0.17173 0.00749 -22.93 <2e-16 \*\*\*
## Year2007 -0.14756 0.00762 -19.36 <2e-16 \*\*\*
## Year2008 -0.14336 0.00755 -18.98 <2e-16 \*\*\*
## Year2009 -0.13957 0.00762 -18.33 <2e-16 \*\*\*
## Year2010 -0.13925 0.00764 -18.23 <2e-16 \*\*\*
## Year2011 -0.13944 0.00764 -18.25 <2e-16 \*\*\*
## Year2012 -0.14496 0.00791 -18.33 <2e-16 \*\*\*
## Year2013 -0.13214 0.00812 -16.28 <2e-16 \*\*\*
## Year2014 -0.12400 0.00855 -14.50 <2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.353
## Multiple R-squared: 0.0169, Adjusted R-squared: 0.0167
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 35 observations c(1531,5850,6115,7919,19539,20586,36590,43261,44926,51165,60019,68414,69355,72907,82858,88052,89379,90517,92684,94109,95291,96853,97334,101548,101798,102174,102831,104323,108419,109806,111811,111818,112703,114713,115251)
## are outliers with |weight| <= 5.3e-07 ( < 8.7e-07);
## 9892 weights are ~= 1. The remaining 105359 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.862 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 8.67e-07 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 115286"
## [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
## 2898 2791 2758 2721 2848 2460 2427 2427 2417 3373 2742 3810 2772 2644 2833
## 2011 2012 2013 2014
## 2941 2883 2830 2737
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1716 1659 1636 1631 1540 1285 1504 1464 1422 1478 1638 1652 1713 1578 1762
## 2011 2012 2013 2014
## 1888 1821 1862 1808
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1428 1394 1373 1341 1252 1090 1271 1209 1194 1212 1335 1356 1417 1312 1456
## 2011 2012 2013 2014
## 1573 1509 1539 1477
## [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 = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 30, df = 1, p-value = 4e-08



## [1] "Female first author team size geometric mean: 3.37483876665783"
## [1] "Male first author team size geometric mean: 3.07108347512473"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 7e+07, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.12158843418149"
## [1] "Male last author team size geometric mean: 3.16222361244947"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 5e+07, 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 1.02 1 1.01
## LastAuthorFemale 1.01 1 1.01
## UniqueAuthors 1.07 4 1.01
## Year 1.06 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26348 -0.28061 0.00115 0.28335 2.47570
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97129 0.01693 57.37 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.01949 0.00594 -3.28 0.00105 \*\*
## LastAuthorFemale1 -0.02703 0.00740 -3.65 0.00026 \*\*\*
## UniqueAuthors2 0.17936 0.01336 13.43 < 2e-16 \*\*\*
## UniqueAuthors3 0.28701 0.01546 18.56 < 2e-16 \*\*\*
## UniqueAuthors4 0.25246 0.01393 18.13 < 2e-16 \*\*\*
## UniqueAuthors5 0.29220 0.01341 21.79 < 2e-16 \*\*\*
## Year1997 -0.01165 0.01774 -0.66 0.51155
## Year1998 -0.02764 0.01676 -1.65 0.09916 .
## Year1999 -0.09407 0.01661 -5.66 1.5e-08 \*\*\*
## Year2000 -0.06591 0.01757 -3.75 0.00018 \*\*\*
## Year2001 -0.07034 0.01729 -4.07 4.8e-05 \*\*\*
## Year2002 -0.12102 0.01735 -6.97 3.2e-12 \*\*\*
## Year2003 -0.09189 0.01732 -5.30 1.1e-07 \*\*\*
## Year2004 -0.09469 0.01671 -5.67 1.5e-08 \*\*\*
## Year2005 -0.06434 0.01730 -3.72 0.00020 \*\*\*
## Year2006 -0.07134 0.01696 -4.21 2.6e-05 \*\*\*
## Year2007 -0.04204 0.01716 -2.45 0.01432 \*
## Year2008 -0.08263 0.01642 -5.03 4.9e-07 \*\*\*
## Year2009 -0.08315 0.01694 -4.91 9.3e-07 \*\*\*
## Year2010 -0.06369 0.01666 -3.82 0.00013 \*\*\*
## Year2011 -0.09619 0.01671 -5.76 8.7e-09 \*\*\*
## Year2012 -0.10888 0.01708 -6.37 1.9e-10 \*\*\*
## Year2013 -0.12527 0.01682 -7.45 9.7e-14 \*\*\*
## Year2014 -0.09047 0.01816 -4.98 6.3e-07 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.416
## Multiple R-squared: 0.0431, Adjusted R-squared: 0.0422
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 4 observations c(1118,1327,6424,21414)
## are outliers with |weight| = 0 ( < 3.9e-06);
## 2139 weights are ~= 1. The remaining 23595 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.006 0.869 0.950 0.902 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 3.89e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.01 1 1
## LastAuthorFemale 1.01 1 1
## Year 1.01 18 1



## [1] "List of 1 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 2463 0029878720 3.73 1996 1303 2 2.56
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17479 -0.27953 0.00253 0.28394 2.55921
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17479 0.01251 93.88 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.00403 0.00599 -0.67 0.50076
## LastAuthorFemale1 -0.02714 0.00751 -3.61 0.00030 \*\*\*
## Year1997 -0.01179 0.01795 -0.66 0.51125
## Year1998 -0.02793 0.01690 -1.65 0.09851 .
## Year1999 -0.09276 0.01666 -5.57 2.6e-08 \*\*\*
## Year2000 -0.05245 0.01767 -2.97 0.00300 \*\*
## Year2001 -0.06330 0.01736 -3.65 0.00027 \*\*\*
## Year2002 -0.10860 0.01744 -6.23 4.8e-10 \*\*\*
## Year2003 -0.07932 0.01757 -4.51 6.4e-06 \*\*\*
## Year2004 -0.08222 0.01685 -4.88 1.1e-06 \*\*\*
## Year2005 -0.04050 0.01743 -2.32 0.02017 \*
## Year2006 -0.05004 0.01711 -2.92 0.00346 \*\*
## Year2007 -0.02217 0.01727 -1.28 0.19910
## Year2008 -0.06094 0.01664 -3.66 0.00025 \*\*\*
## Year2009 -0.06100 0.01707 -3.57 0.00035 \*\*\*
## Year2010 -0.03950 0.01683 -2.35 0.01895 \*
## Year2011 -0.06648 0.01691 -3.93 8.5e-05 \*\*\*
## Year2012 -0.08053 0.01733 -4.65 3.4e-06 \*\*\*
## Year2013 -0.09006 0.01700 -5.30 1.2e-07 \*\*\*
## Year2014 -0.05590 0.01833 -3.05 0.00229 \*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.42
## Multiple R-squared: 0.00489, Adjusted R-squared: 0.00411
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1118,21414) are outliers with |weight| = 0 ( < 3.9e-06);
## 2199 weights are ~= 1. The remaining 23537 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.033 0.868 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 3.89e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1 18 1



## [1] "List of 1 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 2463 0029878720 3.73 1996 1303 2 2.56
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17156 -0.27896 0.00322 0.28450 2.56244
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17156 0.01248 93.89 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.00625 0.00599 -1.04 0.29700
## Year1997 -0.01161 0.01795 -0.65 0.51774
## Year1998 -0.02816 0.01689 -1.67 0.09562 .
## Year1999 -0.09275 0.01666 -5.57 2.6e-08 \*\*\*
## Year2000 -0.05235 0.01767 -2.96 0.00305 \*\*
## Year2001 -0.06444 0.01736 -3.71 0.00021 \*\*\*
## Year2002 -0.10939 0.01744 -6.27 3.6e-10 \*\*\*
## Year2003 -0.07943 0.01758 -4.52 6.2e-06 \*\*\*
## Year2004 -0.08216 0.01685 -4.88 1.1e-06 \*\*\*
## Year2005 -0.04066 0.01744 -2.33 0.01974 \*
## Year2006 -0.05072 0.01711 -2.96 0.00304 \*\*
## Year2007 -0.02259 0.01727 -1.31 0.19098
## Year2008 -0.06157 0.01664 -3.70 0.00022 \*\*\*
## Year2009 -0.06236 0.01708 -3.65 0.00026 \*\*\*
## Year2010 -0.04028 0.01683 -2.39 0.01669 \*
## Year2011 -0.06762 0.01691 -4.00 6.4e-05 \*\*\*
## Year2012 -0.08178 0.01732 -4.72 2.3e-06 \*\*\*
## Year2013 -0.09106 0.01699 -5.36 8.4e-08 \*\*\*
## Year2014 -0.05716 0.01832 -3.12 0.00181 \*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.42
## Multiple R-squared: 0.00435, Adjusted R-squared: 0.00362
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1118,21414) are outliers with |weight| = 0 ( < 3.9e-06);
## 2169 weights are ~= 1. The remaining 23567 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.032 0.868 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 3.89e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## LastAuthorFemale 1 1 1
## Year 1 18 1



## [1] "List of 1 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 2463 0029878720 3.73 1996 1303 2 2.56
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17383 -0.27925 0.00233 0.28400 2.56017
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17383 0.01240 94.66 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.02764 0.00752 -3.68 0.00024 \*\*\*
## Year1997 -0.01188 0.01795 -0.66 0.50820
## Year1998 -0.02794 0.01691 -1.65 0.09843 .
## Year1999 -0.09278 0.01666 -5.57 2.6e-08 \*\*\*
## Year2000 -0.05258 0.01768 -2.97 0.00294 \*\*
## Year2001 -0.06336 0.01736 -3.65 0.00026 \*\*\*
## Year2002 -0.10866 0.01745 -6.23 4.8e-10 \*\*\*
## Year2003 -0.07942 0.01757 -4.52 6.2e-06 \*\*\*
## Year2004 -0.08239 0.01685 -4.89 1.0e-06 \*\*\*
## Year2005 -0.04068 0.01743 -2.33 0.01964 \*
## Year2006 -0.05021 0.01712 -2.93 0.00336 \*\*
## Year2007 -0.02224 0.01727 -1.29 0.19778
## Year2008 -0.06113 0.01664 -3.67 0.00024 \*\*\*
## Year2009 -0.06124 0.01707 -3.59 0.00034 \*\*\*
## Year2010 -0.03973 0.01684 -2.36 0.01832 \*
## Year2011 -0.06665 0.01692 -3.94 8.2e-05 \*\*\*
## Year2012 -0.08085 0.01732 -4.67 3.1e-06 \*\*\*
## Year2013 -0.09024 0.01700 -5.31 1.1e-07 \*\*\*
## Year2014 -0.05615 0.01833 -3.06 0.00218 \*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.42
## Multiple R-squared: 0.00487, Adjusted R-squared: 0.00413
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1118,21414) are outliers with |weight| = 0 ( < 3.9e-06);
## 2197 weights are ~= 1. The remaining 23539 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.033 0.868 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 3.89e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 25738"
## [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
## 4722 4045 3004 1642 1736 1719 1858 1683 1920 2303 2358 2520 2680 2780 2933
## 2011 2012 2013 2014
## 3124 3079 2959 3052
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1369 1586 1190 915 731 676 1038 994 1158 1392 1380 1576 1642 1733 1776
## 2011 2012 2013 2014
## 1912 1850 1835 1930
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1148 1363 1017 786 625 574 892 847 997 1151 1166 1308 1356 1450 1452
## 2011 2012 2013 2014
## 1569 1513 1488 1578
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 2000, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 70, df = 1, p-value <2e-16



## [1] "Female first author team size geometric mean: 3.45551370883479"
## [1] "Male first author team size geometric mean: 3.08528429187163"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 6e+07, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.1427316949259"
## [1] "Male last author team size geometric mean: 3.2057709131541"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 4e+07, 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 1.03 1 1.01
## LastAuthorFemale 1.03 1 1.01
## UniqueAuthors 1.12 4 1.01
## Year 1.13 18 1.00



## [1] "List of 3 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 3533 0001506104 4.15 1996 1300 2 2.9
## 4853 0029670262 3.67 1996 1305 6 2.6
## 8326 0031022694 3.79 1997 1300 2 2.8
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4000 -0.3284 0.0183 0.3368 2.9009
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.786275 0.034938 22.50 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.014137 0.007638 -1.85 0.064 .
## LastAuthorFemale1 -0.015405 0.008907 -1.73 0.084 .
## UniqueAuthors2 0.289822 0.016168 17.93 < 2e-16 \*\*\*
## UniqueAuthors3 0.461785 0.018024 25.62 < 2e-16 \*\*\*
## UniqueAuthors4 0.398048 0.016870 23.60 < 2e-16 \*\*\*
## UniqueAuthors5 0.501679 0.015686 31.98 < 2e-16 \*\*\*
## Year1997 -0.191102 0.047543 -4.02 5.9e-05 \*\*\*
## Year1998 0.003537 0.042472 0.08 0.934
## Year1999 0.112037 0.038900 2.88 0.004 \*\*
## Year2000 0.089889 0.039563 2.27 0.023 \*
## Year2001 0.064333 0.039703 1.62 0.105
## Year2002 0.011178 0.037741 0.30 0.767
## Year2003 -0.007693 0.037145 -0.21 0.836
## Year2004 0.007231 0.036761 0.20 0.844
## Year2005 0.000737 0.036402 0.02 0.984
## Year2006 0.000446 0.036331 0.01 0.990
## Year2007 0.035842 0.036155 0.99 0.322
## Year2008 0.007284 0.036452 0.20 0.842
## Year2009 0.022102 0.036265 0.61 0.542
## Year2010 0.033993 0.036128 0.94 0.347
## Year2011 0.032419 0.036400 0.89 0.373
## Year2012 0.040867 0.036102 1.13 0.258
## Year2013 0.066678 0.036493 1.83 0.068 .
## Year2014 0.047740 0.036666 1.30 0.193
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.481
## Multiple R-squared: 0.102, Adjusted R-squared: 0.101
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 6 observations c(558,635,1082,2091,2099,18711)
## are outliers with |weight| = 0 ( < 4.5e-06);
## 1856 weights are ~= 1. The remaining 20418 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.853 0.948 0.891 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 4.49e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.02 1 1.01
## LastAuthorFemale 1.02 1 1.01
## Year 1.02 18 1.00



## [1] "List of 3 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 3533 0001506104 4.15 1996 1300 2 3.09
## 4853 0029670262 3.67 1996 1305 6 2.62
## 8326 0031022694 3.79 1997 1300 2 2.94
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2463 -0.3409 0.0286 0.3452 3.0947
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05432 0.03517 29.98 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.00862 0.00787 1.10 0.27350
## LastAuthorFemale1 -0.02303 0.00928 -2.48 0.01311 \*
## Year1997 -0.19723 0.04939 -3.99 6.5e-05 \*\*\*
## Year1998 0.04254 0.04420 0.96 0.33582
## Year1999 0.16415 0.04023 4.08 4.5e-05 \*\*\*
## Year2000 0.14628 0.04093 3.57 0.00035 \*\*\*
## Year2001 0.12097 0.04108 2.94 0.00324 \*\*
## Year2002 0.09441 0.03919 2.41 0.01599 \*
## Year2003 0.08968 0.03835 2.34 0.01937 \*
## Year2004 0.10408 0.03794 2.74 0.00609 \*\*
## Year2005 0.10273 0.03760 2.73 0.00630 \*\*
## Year2006 0.11012 0.03747 2.94 0.00330 \*\*
## Year2007 0.14645 0.03721 3.94 8.3e-05 \*\*\*
## Year2008 0.11910 0.03760 3.17 0.00154 \*\*
## Year2009 0.12967 0.03753 3.45 0.00055 \*\*\*
## Year2010 0.15491 0.03730 4.15 3.3e-05 \*\*\*
## Year2011 0.15075 0.03758 4.01 6.1e-05 \*\*\*
## Year2012 0.15619 0.03727 4.19 2.8e-05 \*\*\*
## Year2013 0.18340 0.03763 4.87 1.1e-06 \*\*\*
## Year2014 0.17698 0.03792 4.67 3.1e-06 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.49
## Multiple R-squared: 0.0251, Adjusted R-squared: 0.0242
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 6 observations c(558,635,1082,2091,18711,19552)
## are outliers with |weight| = 0 ( < 4.5e-06);
## 1808 weights are ~= 1. The remaining 20466 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.844 0.947 0.887 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 4.49e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1.01 18 1



## [1] "List of 3 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 3533 0001506104 4.15 1996 1300 2 3.09
## 4853 0029670262 3.67 1996 1305 6 2.62
## 8326 0031022694 3.79 1997 1300 2 2.94
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2387 -0.3417 0.0286 0.3449 3.0978
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05121 0.03440 30.56 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.00581 0.00786 0.74 0.45951
## Year1997 -0.19412 0.04755 -4.08 4.5e-05 \*\*\*
## Year1998 0.04058 0.04372 0.93 0.35331
## Year1999 0.16292 0.03966 4.11 4.0e-05 \*\*\*
## Year2000 0.14670 0.04032 3.64 0.00027 \*\*\*
## Year2001 0.12022 0.04049 2.97 0.00299 \*\*
## Year2002 0.09321 0.03858 2.42 0.01571 \*
## Year2003 0.08909 0.03771 2.36 0.01816 \*
## Year2004 0.10294 0.03730 2.76 0.00579 \*\*
## Year2005 0.10171 0.03698 2.75 0.00595 \*\*
## Year2006 0.10930 0.03681 2.97 0.00299 \*\*
## Year2007 0.14535 0.03655 3.98 7.0e-05 \*\*\*
## Year2008 0.11809 0.03695 3.20 0.00139 \*\*
## Year2009 0.12809 0.03691 3.47 0.00052 \*\*\*
## Year2010 0.15336 0.03668 4.18 2.9e-05 \*\*\*
## Year2011 0.14919 0.03696 4.04 5.5e-05 \*\*\*
## Year2012 0.15499 0.03659 4.24 2.3e-05 \*\*\*
## Year2013 0.18169 0.03695 4.92 8.8e-07 \*\*\*
## Year2014 0.17498 0.03725 4.70 2.7e-06 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.499
## Multiple R-squared: 0.0243, Adjusted R-squared: 0.0234
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 6 observations c(558,635,1082,2091,18711,19552)
## are outliers with |weight| = 0 ( < 4.5e-06);
## 1846 weights are ~= 1. The remaining 20428 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.003 0.849 0.948 0.891 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 4.49e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.01 18 1.00



## [1] "List of 3 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 3533 0001506104 4.15 1996 1300 2 3.09
## 4853 0029670262 3.67 1996 1305 6 2.62
## 8326 0031022694 3.79 1997 1300 2 2.94
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.240 -0.341 0.029 0.346 3.093
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05571 0.03429 30.78 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.02207 0.00925 -2.38 0.01712 \*
## Year1997 -0.19348 0.04810 -4.02 5.8e-05 \*\*\*
## Year1998 0.04059 0.04319 0.94 0.34736
## Year1999 0.16362 0.03946 4.15 3.4e-05 \*\*\*
## Year2000 0.14650 0.04017 3.65 0.00027 \*\*\*
## Year2001 0.12135 0.04033 3.01 0.00262 \*\*
## Year2002 0.09451 0.03840 2.46 0.01384 \*
## Year2003 0.09014 0.03755 2.40 0.01639 \*
## Year2004 0.10484 0.03713 2.82 0.00475 \*\*
## Year2005 0.10331 0.03678 2.81 0.00497 \*\*
## Year2006 0.11073 0.03665 3.02 0.00252 \*\*
## Year2007 0.14728 0.03638 4.05 5.2e-05 \*\*\*
## Year2008 0.11984 0.03677 3.26 0.00112 \*\*
## Year2009 0.13005 0.03670 3.54 0.00040 \*\*\*
## Year2010 0.15514 0.03648 4.25 2.1e-05 \*\*\*
## Year2011 0.15098 0.03675 4.11 4.0e-05 \*\*\*
## Year2012 0.15701 0.03644 4.31 1.6e-05 \*\*\*
## Year2013 0.18423 0.03680 5.01 5.6e-07 \*\*\*
## Year2014 0.17773 0.03709 4.79 1.7e-06 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared: 0.0244, Adjusted R-squared: 0.0236
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 5 observations c(558,635,1082,2091,19552)
## are outliers with |weight| = 0 ( < 4.5e-06);
## 1825 weights are ~= 1. The remaining 20450 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.850 0.949 0.891 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 4.49e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 22280"
## [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
## 3820 3613 3796 3660 3761 4071 4161 3895 4058 4477 4654 4841 5028 4850 4774
## 2011 2012 2013 2014
## 5057 5014 5054 4983
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2173 2136 2352 2290 2048 1812 2790 2669 2684 2989 2953 3094 3273 3157 3005
## 2011 2012 2013 2014
## 3305 3316 3258 3311
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1852 1852 2022 1959 1719 1528 2357 2245 2230 2521 2494 2603 2751 2652 2553
## 2011 2012 2013 2014
## 2775 2831 2735 2785
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 800, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 100, df = 1, p-value <2e-16



## [1] "Female first author team size geometric mean: 5.35113968034439"
## [1] "Male first author team size geometric mean: 5.11963054964782"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 2e+08, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 5.01330812393877"
## [1] "Male last author team size geometric mean: 5.27074138141416"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 2e+08, p-value = 2e-14
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.03 1 1.01
## LastAuthorFemale 1.03 1 1.01
## UniqueAuthors 1.04 4 1.01
## Year 1.06 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.42440 -0.26450 0.00855 0.26298 1.96950
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.93026 0.01822 51.05 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.01176 0.00404 -2.91 0.00357 \*\*
## LastAuthorFemale1 -0.01417 0.00462 -3.07 0.00217 \*\*
## UniqueAuthors2 0.25800 0.01508 17.11 < 2e-16 \*\*\*
## UniqueAuthors3 0.37275 0.01578 23.63 < 2e-16 \*\*\*
## UniqueAuthors4 0.35531 0.01463 24.28 < 2e-16 \*\*\*
## UniqueAuthors5 0.49414 0.01370 36.07 < 2e-16 \*\*\*
## Year1997 -0.00880 0.01608 -0.55 0.58401
## Year1998 -0.05931 0.01551 -3.82 0.00013 \*\*\*
## Year1999 -0.09535 0.01540 -6.19 6.0e-10 \*\*\*
## Year2000 -0.11191 0.01575 -7.11 1.2e-12 \*\*\*
## Year2001 -0.10528 0.01687 -6.24 4.4e-10 \*\*\*
## Year2002 -0.09099 0.01499 -6.07 1.3e-09 \*\*\*
## Year2003 -0.12889 0.01430 -9.01 < 2e-16 \*\*\*
## Year2004 -0.16028 0.01445 -11.09 < 2e-16 \*\*\*
## Year2005 -0.13581 0.01444 -9.41 < 2e-16 \*\*\*
## Year2006 -0.11963 0.01445 -8.28 < 2e-16 \*\*\*
## Year2007 -0.08775 0.01446 -6.07 1.3e-09 \*\*\*
## Year2008 -0.08971 0.01434 -6.25 4.0e-10 \*\*\*
## Year2009 -0.08714 0.01449 -6.01 1.8e-09 \*\*\*
## Year2010 -0.11013 0.01464 -7.52 5.5e-14 \*\*\*
## Year2011 -0.12068 0.01461 -8.26 < 2e-16 \*\*\*
## Year2012 -0.11760 0.01471 -8.00 1.3e-15 \*\*\*
## Year2013 -0.11648 0.01486 -7.84 4.8e-15 \*\*\*
## Year2014 -0.10543 0.01532 -6.88 6.0e-12 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared: 0.0972, Adjusted R-squared: 0.0967
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1371,41948) are outliers with |weight| = 0 ( < 2.2e-06);
## 3706 weights are ~= 1. The remaining 40756 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.002 0.865 0.951 0.897 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.25e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.02 1 1.01
## LastAuthorFemale 1.02 1 1.01
## Year 1.02 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31945 -0.27161 0.00824 0.26701 1.88882
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31945 0.01268 104.07 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.00116 0.00414 -0.28 0.77979
## LastAuthorFemale1 -0.02309 0.00474 -4.87 1.1e-06 \*\*\*
## Year1997 -0.00880 0.01649 -0.53 0.59385
## Year1998 -0.05422 0.01599 -3.39 0.00070 \*\*\*
## Year1999 -0.08584 0.01574 -5.45 4.9e-08 \*\*\*
## Year2000 -0.09028 0.01612 -5.60 2.2e-08 \*\*\*
## Year2001 -0.08226 0.01723 -4.77 1.8e-06 \*\*\*
## Year2002 -0.06623 0.01534 -4.32 1.6e-05 \*\*\*
## Year2003 -0.09453 0.01475 -6.41 1.5e-10 \*\*\*
## Year2004 -0.12083 0.01496 -8.08 6.7e-16 \*\*\*
## Year2005 -0.08770 0.01485 -5.91 3.5e-09 \*\*\*
## Year2006 -0.07736 0.01485 -5.21 1.9e-07 \*\*\*
## Year2007 -0.04484 0.01488 -3.01 0.00258 \*\*
## Year2008 -0.04007 0.01472 -2.72 0.00647 \*\*
## Year2009 -0.04192 0.01485 -2.82 0.00476 \*\*
## Year2010 -0.06354 0.01503 -4.23 2.4e-05 \*\*\*
## Year2011 -0.07690 0.01503 -5.12 3.1e-07 \*\*\*
## Year2012 -0.06101 0.01509 -4.04 5.3e-05 \*\*\*
## Year2013 -0.07491 0.01521 -4.93 8.5e-07 \*\*\*
## Year2014 -0.06011 0.01580 -3.81 0.00014 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared: 0.00469, Adjusted R-squared: 0.00425
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 41948 is an outlier with |weight| = 0 ( < 2.2e-06);
## 3719 weights are ~= 1. The remaining 40744 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.028 0.865 0.950 0.892 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.25e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.01 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31591 -0.27133 0.00809 0.26702 1.89720
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31591 0.01265 104.02 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.00390 0.00412 -0.95 0.34327
## Year1997 -0.00893 0.01649 -0.54 0.58821
## Year1998 -0.05468 0.01599 -3.42 0.00063 \*\*\*
## Year1999 -0.08560 0.01573 -5.44 5.3e-08 \*\*\*
## Year2000 -0.09046 0.01612 -5.61 2.0e-08 \*\*\*
## Year2001 -0.08212 0.01723 -4.77 1.9e-06 \*\*\*
## Year2002 -0.06652 0.01534 -4.34 1.4e-05 \*\*\*
## Year2003 -0.09499 0.01475 -6.44 1.2e-10 \*\*\*
## Year2004 -0.12134 0.01496 -8.11 5.1e-16 \*\*\*
## Year2005 -0.08862 0.01485 -5.97 2.4e-09 \*\*\*
## Year2006 -0.07808 0.01485 -5.26 1.5e-07 \*\*\*
## Year2007 -0.04599 0.01488 -3.09 0.00199 \*\*
## Year2008 -0.04078 0.01472 -2.77 0.00559 \*\*
## Year2009 -0.04319 0.01485 -2.91 0.00363 \*\*
## Year2010 -0.06475 0.01503 -4.31 1.6e-05 \*\*\*
## Year2011 -0.07874 0.01502 -5.24 1.6e-07 \*\*\*
## Year2012 -0.06302 0.01509 -4.18 3.0e-05 \*\*\*
## Year2013 -0.07704 0.01521 -5.07 4.1e-07 \*\*\*
## Year2014 -0.06221 0.01580 -3.94 8.2e-05 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared: 0.00414, Adjusted R-squared: 0.00371
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 41948 is an outlier with |weight| = 0 ( < 2.2e-06);
## 3757 weights are ~= 1. The remaining 40706 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.026 0.865 0.950 0.892 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.25e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.01 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31911 -0.27134 0.00794 0.26709 1.88813
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31911 0.01261 104.64 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.02327 0.00472 -4.93 8.2e-07 \*\*\*
## Year1997 -0.00881 0.01649 -0.53 0.59322
## Year1998 -0.05426 0.01599 -3.39 0.00069 \*\*\*
## Year1999 -0.08587 0.01574 -5.46 4.9e-08 \*\*\*
## Year2000 -0.09029 0.01612 -5.60 2.1e-08 \*\*\*
## Year2001 -0.08229 0.01723 -4.77 1.8e-06 \*\*\*
## Year2002 -0.06626 0.01534 -4.32 1.6e-05 \*\*\*
## Year2003 -0.09458 0.01475 -6.41 1.4e-10 \*\*\*
## Year2004 -0.12088 0.01496 -8.08 6.5e-16 \*\*\*
## Year2005 -0.08777 0.01484 -5.91 3.4e-09 \*\*\*
## Year2006 -0.07746 0.01485 -5.22 1.8e-07 \*\*\*
## Year2007 -0.04493 0.01488 -3.02 0.00253 \*\*
## Year2008 -0.04018 0.01471 -2.73 0.00631 \*\*
## Year2009 -0.04205 0.01485 -2.83 0.00462 \*\*
## Year2010 -0.06366 0.01502 -4.24 2.3e-05 \*\*\*
## Year2011 -0.07702 0.01502 -5.13 3.0e-07 \*\*\*
## Year2012 -0.06114 0.01508 -4.05 5.1e-05 \*\*\*
## Year2013 -0.07506 0.01520 -4.94 7.9e-07 \*\*\*
## Year2014 -0.06024 0.01578 -3.82 0.00014 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared: 0.00469, Adjusted R-squared: 0.00427
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 41948 is an outlier with |weight| = 0 ( < 2.2e-06);
## 3711 weights are ~= 1. The remaining 40752 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.028 0.865 0.951 0.892 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 2.25e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 44464"
## [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
## 11470 9750 8975 7078 8359 7167 7484 6326 6990 6859 6852 7128
## 2008 2009 2010 2011 2012 2013 2014
## 7282 7298 7223 7368 7147 6645 6157
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5613 5245 4961 4477 4249 3336 4702 3902 4260 4211 4180 4371 4515 4490 4458
## 2011 2012 2013 2014
## 4557 4499 4225 3954
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4801 4508 4296 3843 3621 2843 3988 3303 3570 3514 3502 3689 3788 3710 3684
## 2011 2012 2013 2014
## 3799 3710 3484 3300
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 5000, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 200, df = 1, p-value <2e-16



## [1] "Female first author team size geometric mean: 3.94822587854252"
## [1] "Male first author team size geometric mean: 3.7364162968257"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 6e+08, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.72766969022816"
## [1] "Male last author team size geometric mean: 3.84401855887855"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 4e+08, p-value = 5e-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 1.01 1 1.01
## LastAuthorFemale 1.01 1 1.01
## UniqueAuthors 1.04 4 1.00
## Year 1.05 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.44774 -0.23228 0.00238 0.22515 2.47260
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09861 0.01156 95.06 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.01595 0.00279 -5.71 1.1e-08 \*\*\*
## LastAuthorFemale1 -0.02344 0.00330 -7.11 1.2e-12 \*\*\*
## UniqueAuthors2 0.18933 0.00924 20.50 < 2e-16 \*\*\*
## UniqueAuthors3 0.24899 0.00964 25.84 < 2e-16 \*\*\*
## UniqueAuthors4 0.25406 0.00923 27.52 < 2e-16 \*\*\*
## UniqueAuthors5 0.34913 0.00892 39.12 < 2e-16 \*\*\*
## Year1997 -0.03155 0.01056 -2.99 0.0028 \*\*
## Year1998 -0.09378 0.01002 -9.36 < 2e-16 \*\*\*
## Year1999 -0.17317 0.00923 -18.76 < 2e-16 \*\*\*
## Year2000 -0.17754 0.01007 -17.63 < 2e-16 \*\*\*
## Year2001 -0.18537 0.00977 -18.97 < 2e-16 \*\*\*
## Year2002 -0.19021 0.00919 -20.69 < 2e-16 \*\*\*
## Year2003 -0.21560 0.00939 -22.95 < 2e-16 \*\*\*
## Year2004 -0.20738 0.00913 -22.71 < 2e-16 \*\*\*
## Year2005 -0.21494 0.00929 -23.13 < 2e-16 \*\*\*
## Year2006 -0.21700 0.00934 -23.22 < 2e-16 \*\*\*
## Year2007 -0.19966 0.00943 -21.17 < 2e-16 \*\*\*
## Year2008 -0.20110 0.00931 -21.60 < 2e-16 \*\*\*
## Year2009 -0.19647 0.00947 -20.74 < 2e-16 \*\*\*
## Year2010 -0.19319 0.00960 -20.12 < 2e-16 \*\*\*
## Year2011 -0.19548 0.00951 -20.55 < 2e-16 \*\*\*
## Year2012 -0.21554 0.00976 -22.08 < 2e-16 \*\*\*
## Year2013 -0.20712 0.01019 -20.32 < 2e-16 \*\*\*
## Year2014 -0.17906 0.01077 -16.63 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared: 0.0839, Adjusted R-squared: 0.0836
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 11 observations c(7071,52146,62942,64360,64752,66285,68919,69967,70543,70571,70737)
## are outliers with |weight| = 0 ( < 1.4e-06);
## 6038 weights are ~= 1. The remaining 64904 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.861 0.951 0.886 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 1.41e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.01 1 1.00
## LastAuthorFemale 1.01 1 1.01
## Year 1.01 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34353 -0.23787 0.00127 0.22759 2.39049
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.34353 0.00783 171.58 <2e-16 \*\*\*
## FirstAuthorFemale1 -0.00865 0.00284 -3.04 0.0023 \*\*
## LastAuthorFemale1 -0.02896 0.00334 -8.66 <2e-16 \*\*\*
## Year1997 -0.02395 0.01064 -2.25 0.0244 \*
## Year1998 -0.09013 0.01013 -8.90 <2e-16 \*\*\*
## Year1999 -0.16566 0.00934 -17.73 <2e-16 \*\*\*
## Year2000 -0.16951 0.01028 -16.49 <2e-16 \*\*\*
## Year2001 -0.16414 0.00995 -16.49 <2e-16 \*\*\*
## Year2002 -0.16758 0.00934 -17.94 <2e-16 \*\*\*
## Year2003 -0.19208 0.00953 -20.15 <2e-16 \*\*\*
## Year2004 -0.18329 0.00928 -19.76 <2e-16 \*\*\*
## Year2005 -0.18884 0.00946 -19.96 <2e-16 \*\*\*
## Year2006 -0.18807 0.00945 -19.91 <2e-16 \*\*\*
## Year2007 -0.16758 0.00957 -17.51 <2e-16 \*\*\*
## Year2008 -0.16670 0.00947 -17.60 <2e-16 \*\*\*
## Year2009 -0.16205 0.00964 -16.81 <2e-16 \*\*\*
## Year2010 -0.15642 0.00980 -15.95 <2e-16 \*\*\*
## Year2011 -0.15910 0.00973 -16.36 <2e-16 \*\*\*
## Year2012 -0.17518 0.00994 -17.62 <2e-16 \*\*\*
## Year2013 -0.17001 0.01042 -16.32 <2e-16 \*\*\*
## Year2014 -0.14341 0.01098 -13.06 <2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared: 0.024, Adjusted R-squared: 0.0237
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 9 observations c(52146,62942,64360,64752,66285,68436,68919,69967,70737)
## are outliers with |weight| = 0 ( < 1.4e-06);
## 6028 weights are ~= 1. The remaining 64916 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.862 0.951 0.884 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 1.41e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1 18 1



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.33901 -0.23897 0.00123 0.22713 2.39672
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.33901 0.00784 170.84 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.01089 0.00284 -3.83 0.00013 \*\*\*
## Year1997 -0.02407 0.01065 -2.26 0.02385 \*
## Year1998 -0.09024 0.01015 -8.89 < 2e-16 \*\*\*
## Year1999 -0.16644 0.00936 -17.79 < 2e-16 \*\*\*
## Year2000 -0.17006 0.01030 -16.51 < 2e-16 \*\*\*
## Year2001 -0.16474 0.00997 -16.53 < 2e-16 \*\*\*
## Year2002 -0.16817 0.00936 -17.97 < 2e-16 \*\*\*
## Year2003 -0.19271 0.00955 -20.19 < 2e-16 \*\*\*
## Year2004 -0.18438 0.00929 -19.84 < 2e-16 \*\*\*
## Year2005 -0.18965 0.00947 -20.03 < 2e-16 \*\*\*
## Year2006 -0.18920 0.00946 -20.00 < 2e-16 \*\*\*
## Year2007 -0.16913 0.00958 -17.66 < 2e-16 \*\*\*
## Year2008 -0.16826 0.00948 -17.74 < 2e-16 \*\*\*
## Year2009 -0.16350 0.00966 -16.93 < 2e-16 \*\*\*
## Year2010 -0.15814 0.00981 -16.12 < 2e-16 \*\*\*
## Year2011 -0.16116 0.00973 -16.56 < 2e-16 \*\*\*
## Year2012 -0.17720 0.00995 -17.81 < 2e-16 \*\*\*
## Year2013 -0.17173 0.01043 -16.46 < 2e-16 \*\*\*
## Year2014 -0.14583 0.01098 -13.28 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared: 0.023, Adjusted R-squared: 0.0227
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 10 observations c(52146,62942,64360,64752,66285,68434,68436,68919,69967,70737)
## are outliers with |weight| = 0 ( < 1.4e-06);
## 5987 weights are ~= 1. The remaining 64956 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.010 0.861 0.951 0.884 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 1.41e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1.01 18 1



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34048 -0.23788 0.00166 0.22766 2.39405
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.34048 0.00775 172.96 <2e-16 \*\*\*
## LastAuthorFemale1 -0.02989 0.00334 -8.95 <2e-16 \*\*\*
## Year1997 -0.02386 0.01064 -2.24 0.025 \*
## Year1998 -0.09010 0.01013 -8.89 <2e-16 \*\*\*
## Year1999 -0.16571 0.00934 -17.74 <2e-16 \*\*\*
## Year2000 -0.16958 0.01028 -16.50 <2e-16 \*\*\*
## Year2001 -0.16428 0.00995 -16.51 <2e-16 \*\*\*
## Year2002 -0.16757 0.00934 -17.94 <2e-16 \*\*\*
## Year2003 -0.19229 0.00954 -20.16 <2e-16 \*\*\*
## Year2004 -0.18362 0.00928 -19.78 <2e-16 \*\*\*
## Year2005 -0.18907 0.00946 -19.98 <2e-16 \*\*\*
## Year2006 -0.18831 0.00945 -19.93 <2e-16 \*\*\*
## Year2007 -0.16788 0.00957 -17.54 <2e-16 \*\*\*
## Year2008 -0.16714 0.00947 -17.64 <2e-16 \*\*\*
## Year2009 -0.16240 0.00964 -16.84 <2e-16 \*\*\*
## Year2010 -0.15694 0.00981 -16.00 <2e-16 \*\*\*
## Year2011 -0.15960 0.00973 -16.41 <2e-16 \*\*\*
## Year2012 -0.17558 0.00995 -17.65 <2e-16 \*\*\*
## Year2013 -0.17053 0.01042 -16.37 <2e-16 \*\*\*
## Year2014 -0.14399 0.01098 -13.12 <2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared: 0.0238, Adjusted R-squared: 0.0236
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 10 observations c(52146,62942,64360,64752,66285,68434,68436,68919,69967,70737)
## are outliers with |weight| = 0 ( < 1.4e-06);
## 6079 weights are ~= 1. The remaining 64864 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.010 0.861 0.951 0.884 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 1.41e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 70953"
## [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
## 2016 1866 1905 1863 1876 2110 2196 2078 2153 2131 2080 2317 2398 2233 2287
## 2011 2012 2013 2014
## 2334 2099 1940 1818
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1332 1188 1223 1205 1023 919 1434 1319 1374 1403 1321 1497 1571 1420 1491
## 2011 2012 2013 2014
## 1515 1404 1269 1256
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1166 1036 1062 1010 880 783 1242 1116 1147 1165 1107 1220 1322 1160 1202
## 2011 2012 2013 2014
## 1241 1150 1045 1056
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 400, df = 20, p-value <2e-16



##
## 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 geometric mean: 4.09828167184193"
## [1] "Male first author team size geometric mean: 4.10865404898422"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 5e+07, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.93831245048473"
## [1] "Male last author team size geometric mean: 4.15368130343593"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 4e+07, p-value = 1e-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 1.02 1 1.01
## LastAuthorFemale 1.01 1 1.01
## UniqueAuthors 1.08 4 1.01
## Year 1.09 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.25324 -0.25148 0.00505 0.25404 2.43368
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.832926 0.021748 38.30 <2e-16 \*\*\*
## FirstAuthorFemale1 0.019646 0.006006 3.27 0.0011 \*\*
## LastAuthorFemale1 0.005700 0.006970 0.82 0.4134
## UniqueAuthors2 0.167676 0.016613 10.09 <2e-16 \*\*\*
## UniqueAuthors3 0.244371 0.018458 13.24 <2e-16 \*\*\*
## UniqueAuthors4 0.262042 0.016585 15.80 <2e-16 \*\*\*
## UniqueAuthors5 0.345248 0.015540 22.22 <2e-16 \*\*\*
## Year1997 0.049722 0.023011 2.16 0.0307 \*
## Year1998 0.041475 0.021253 1.95 0.0510 .
## Year1999 0.004545 0.021651 0.21 0.8337
## Year2000 0.024767 0.021227 1.17 0.2433
## Year2001 0.020770 0.021013 0.99 0.3229
## Year2002 -0.030065 0.020041 -1.50 0.1336
## Year2003 -0.049678 0.020317 -2.45 0.0145 \*
## Year2004 -0.038252 0.019842 -1.93 0.0539 .
## Year2005 -0.055598 0.019976 -2.78 0.0054 \*\*
## Year2006 -0.056467 0.020040 -2.82 0.0048 \*\*
## Year2007 0.015920 0.020585 0.77 0.4393
## Year2008 0.007613 0.019736 0.39 0.6997
## Year2009 -0.000886 0.019864 -0.04 0.9644
## Year2010 -0.031225 0.019823 -1.58 0.1152
## Year2011 -0.022162 0.020002 -1.11 0.2679
## Year2012 -0.049536 0.020667 -2.40 0.0165 \*
## Year2013 -0.007316 0.021250 -0.34 0.7306
## Year2014 -0.010091 0.022086 -0.46 0.6478
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared: 0.0666, Adjusted R-squared: 0.0656
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 8 observations c(1239,8000,8386,11054,11502,12175,13935,18058)
## are outliers with |weight| = 0 ( < 4.7e-06);
## 1886 weights are ~= 1. The remaining 19216 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.858 0.950 0.890 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 4.74e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.01 1 1.01
## LastAuthorFemale 1.01 1 1.00
## Year 1.02 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15038 -0.25686 0.00759 0.25681 2.36633
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.067011 0.017420 61.25 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.023301 0.006116 3.81 0.00014 \*\*\*
## LastAuthorFemale1 -0.000929 0.007154 -0.13 0.89663
## Year1997 0.050663 0.023654 2.14 0.03222 \*
## Year1998 0.059178 0.022188 2.67 0.00766 \*\*
## Year1999 0.013453 0.022364 0.60 0.54748
## Year2000 0.036261 0.021915 1.65 0.09802 .
## Year2001 0.039350 0.021854 1.80 0.07178 .
## Year2002 -0.003207 0.020637 -0.16 0.87653
## Year2003 -0.021499 0.020988 -1.02 0.30569
## Year2004 -0.000361 0.020542 -0.02 0.98599
## Year2005 -0.016878 0.020652 -0.82 0.41378
## Year2006 -0.012246 0.020787 -0.59 0.55580
## Year2007 0.060071 0.021133 2.84 0.00448 \*\*
## Year2008 0.049402 0.020435 2.42 0.01563 \*
## Year2009 0.056257 0.020414 2.76 0.00586 \*\*
## Year2010 0.021620 0.020412 1.06 0.28954
## Year2011 0.020864 0.020617 1.01 0.31155
## Year2012 0.001344 0.021324 0.06 0.94976
## Year2013 0.050811 0.021668 2.34 0.01904 \*
## Year2014 0.035943 0.022645 1.59 0.11248
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared: 0.00525, Adjusted R-squared: 0.00431
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 6 observations c(1239,8386,11054,12175,13935,18058)
## are outliers with |weight| = 0 ( < 4.7e-06);
## 1885 weights are ~= 1. The remaining 19219 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.009 0.858 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 4.74e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.01 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15013 -0.25689 0.00757 0.25688 2.36648
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.066825 0.017359 61.46 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.023220 0.006117 3.80 0.00015 \*\*\*
## Year1997 0.050692 0.023653 2.14 0.03211 \*
## Year1998 0.059189 0.022188 2.67 0.00765 \*\*
## Year1999 0.013457 0.022364 0.60 0.54737
## Year2000 0.036271 0.021915 1.66 0.09792 .
## Year2001 0.039359 0.021854 1.80 0.07171 .
## Year2002 -0.003201 0.020637 -0.16 0.87674
## Year2003 -0.021452 0.020983 -1.02 0.30663
## Year2004 -0.000335 0.020539 -0.02 0.98700
## Year2005 -0.016856 0.020650 -0.82 0.41435
## Year2006 -0.012227 0.020786 -0.59 0.55639
## Year2007 0.060082 0.021133 2.84 0.00447 \*\*
## Year2008 0.049389 0.020434 2.42 0.01566 \*
## Year2009 0.056272 0.020413 2.76 0.00585 \*\*
## Year2010 0.021640 0.020411 1.06 0.28906
## Year2011 0.020847 0.020616 1.01 0.31194
## Year2012 0.001325 0.021322 0.06 0.95044
## Year2013 0.050793 0.021667 2.34 0.01907 \*
## Year2014 0.035916 0.022644 1.59 0.11274
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared: 0.00525, Adjusted R-squared: 0.00435
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 6 observations c(1239,8386,11054,12175,13935,18058)
## are outliers with |weight| = 0 ( < 4.7e-06);
## 1879 weights are ~= 1. The remaining 19225 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.009 0.858 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 4.74e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1.01 18 1



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.13645 -0.25699 0.00897 0.25804 2.36030
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.072969 0.017328 61.92 <2e-16 \*\*\*
## LastAuthorFemale1 0.001687 0.007155 0.24 0.8137
## Year1997 0.050727 0.023663 2.14 0.0321 \*
## Year1998 0.059183 0.022185 2.67 0.0076 \*\*
## Year1999 0.014200 0.022336 0.64 0.5249
## Year2000 0.036822 0.021908 1.68 0.0928 .
## Year2001 0.040347 0.021856 1.85 0.0649 .
## Year2002 -0.002293 0.020633 -0.11 0.9115
## Year2003 -0.020246 0.020967 -0.97 0.3343
## Year2004 0.000202 0.020531 0.01 0.9921
## Year2005 -0.015622 0.020630 -0.76 0.4489
## Year2006 -0.010844 0.020771 -0.52 0.6016
## Year2007 0.061795 0.021111 2.93 0.0034 \*\*
## Year2008 0.051044 0.020426 2.50 0.0125 \*
## Year2009 0.058051 0.020420 2.84 0.0045 \*\*
## Year2010 0.023987 0.020392 1.18 0.2395
## Year2011 0.023026 0.020616 1.12 0.2640
## Year2012 0.003364 0.021318 0.16 0.8746
## Year2013 0.053670 0.021657 2.48 0.0132 \*
## Year2014 0.039074 0.022603 1.73 0.0839 .
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared: 0.00452, Adjusted R-squared: 0.00362
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 6 observations c(1239,8386,11054,12175,13935,18058)
## are outliers with |weight| = 0 ( < 4.7e-06);
## 1877 weights are ~= 1. The remaining 19227 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.011 0.858 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 4.74e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 21110"
## [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
## 2163 1879 1822 1752 1684 1596 1548 1618 1633 1685 1965 1987 2047 1932 1831
## 2011 2012 2013 2014
## 1719 1799 1668 1803
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1447 1230 1239 1197 1031 846 1049 1079 1120 1139 1343 1350 1380 1258 1245
## 2011 2012 2013 2014
## 1143 1215 1150 1243
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1270 1080 1097 1053 900 732 931 943 985 995 1170 1154 1192 1096 1084
## 2011 2012 2013 2014
## 988 1057 991 1077
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 300, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 40, df = 1, p-value = 3e-09



## [1] "Female first author team size geometric mean: 3.38988707482452"
## [1] "Male first author team size geometric mean: 3.19029968291301"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 5e+07, p-value = 1e-07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.21845800865087"
## [1] "Male last author team size geometric mean: 3.29752370863151"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 4e+07, 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 1.03 1 1.01
## LastAuthorFemale 1.02 1 1.01
## UniqueAuthors 1.09 4 1.01
## Year 1.10 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.314703 -0.231130 0.000879 0.227533 1.558679
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09813 0.01522 72.13 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.01022 0.00517 -1.98 0.0483 \*
## LastAuthorFemale1 -0.00881 0.00570 -1.55 0.1223
## UniqueAuthors2 0.07642 0.01064 7.19 6.9e-13 \*\*\*
## UniqueAuthors3 0.07732 0.01256 6.16 7.6e-10 \*\*\*
## UniqueAuthors4 0.11269 0.01131 9.96 < 2e-16 \*\*\*
## UniqueAuthors5 0.21657 0.01062 20.39 < 2e-16 \*\*\*
## Year1997 0.01592 0.01740 0.91 0.3603
## Year1998 -0.01858 0.01746 -1.06 0.2873
## Year1999 -0.04772 0.01650 -2.89 0.0038 \*\*
## Year2000 -0.08956 0.01719 -5.21 1.9e-07 \*\*\*
## Year2001 -0.12342 0.01762 -7.01 2.5e-12 \*\*\*
## Year2002 -0.08521 0.01739 -4.90 9.7e-07 \*\*\*
## Year2003 -0.11617 0.01652 -7.03 2.1e-12 \*\*\*
## Year2004 -0.11080 0.01659 -6.68 2.5e-11 \*\*\*
## Year2005 -0.12100 0.01647 -7.35 2.1e-13 \*\*\*
## Year2006 -0.12862 0.01596 -8.06 8.1e-16 \*\*\*
## Year2007 -0.10914 0.01578 -6.92 4.8e-12 \*\*\*
## Year2008 -0.10195 0.01590 -6.41 1.5e-10 \*\*\*
## Year2009 -0.11604 0.01639 -7.08 1.5e-12 \*\*\*
## Year2010 -0.10516 0.01682 -6.25 4.1e-10 \*\*\*
## Year2011 -0.07306 0.01664 -4.39 1.1e-05 \*\*\*
## Year2012 -0.10705 0.01670 -6.41 1.5e-10 \*\*\*
## Year2013 -0.10532 0.01769 -5.95 2.7e-09 \*\*\*
## Year2014 -0.07649 0.01753 -4.36 1.3e-05 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared: 0.0501, Adjusted R-squared: 0.0489
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 1703 weights are ~= 1. The remaining 18092 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.002 0.867 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 5.05e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.02 1 1.01
## LastAuthorFemale 1.02 1 1.01
## Year 1.02 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.215028 -0.233738 -0.000175 0.233496 1.630484
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19616 0.01289 92.79 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.00633 0.00526 -1.20 0.22876
## LastAuthorFemale1 -0.01488 0.00579 -2.57 0.01013 \*
## Year1997 0.01886 0.01764 1.07 0.28501
## Year1998 -0.01477 0.01761 -0.84 0.40171
## Year1999 -0.03733 0.01677 -2.23 0.02605 \*
## Year2000 -0.07234 0.01733 -4.17 3.0e-05 \*\*\*
## Year2001 -0.10686 0.01773 -6.03 1.7e-09 \*\*\*
## Year2002 -0.06763 0.01766 -3.83 0.00013 \*\*\*
## Year2003 -0.09661 0.01669 -5.79 7.2e-09 \*\*\*
## Year2004 -0.08386 0.01687 -4.97 6.7e-07 \*\*\*
## Year2005 -0.09758 0.01670 -5.84 5.2e-09 \*\*\*
## Year2006 -0.10299 0.01626 -6.34 2.4e-10 \*\*\*
## Year2007 -0.07959 0.01596 -4.99 6.2e-07 \*\*\*
## Year2008 -0.06710 0.01607 -4.17 3.0e-05 \*\*\*
## Year2009 -0.07792 0.01657 -4.70 2.6e-06 \*\*\*
## Year2010 -0.06232 0.01689 -3.69 0.00022 \*\*\*
## Year2011 -0.03093 0.01700 -1.82 0.06885 .
## Year2012 -0.06011 0.01695 -3.55 0.00039 \*\*\*
## Year2013 -0.05702 0.01782 -3.20 0.00138 \*\*
## Year2014 -0.02493 0.01771 -1.41 0.15933
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared: 0.0101, Adjusted R-squared: 0.00908
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 14680 is an outlier with |weight| = 0 ( < 5.1e-06);
## 1697 weights are ~= 1. The remaining 18097 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.025 0.868 0.950 0.898 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 5.05e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1.01 18 1



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.211944 -0.233554 -0.000989 0.233965 1.635489
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19383 0.01286 92.82 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.00805 0.00522 -1.54 0.12303
## Year1997 0.01812 0.01765 1.03 0.30471
## Year1998 -0.01534 0.01762 -0.87 0.38397
## Year1999 -0.03775 0.01677 -2.25 0.02442 \*
## Year2000 -0.07322 0.01734 -4.22 2.4e-05 \*\*\*
## Year2001 -0.10794 0.01772 -6.09 1.1e-09 \*\*\*
## Year2002 -0.06900 0.01766 -3.91 9.4e-05 \*\*\*
## Year2003 -0.09759 0.01669 -5.85 5.1e-09 \*\*\*
## Year2004 -0.08491 0.01686 -5.04 4.8e-07 \*\*\*
## Year2005 -0.09829 0.01669 -5.89 4.0e-09 \*\*\*
## Year2006 -0.10428 0.01624 -6.42 1.4e-10 \*\*\*
## Year2007 -0.08052 0.01596 -5.05 4.5e-07 \*\*\*
## Year2008 -0.06879 0.01605 -4.28 1.8e-05 \*\*\*
## Year2009 -0.07904 0.01656 -4.77 1.8e-06 \*\*\*
## Year2010 -0.06327 0.01688 -3.75 0.00018 \*\*\*
## Year2011 -0.03245 0.01698 -1.91 0.05605 .
## Year2012 -0.06152 0.01695 -3.63 0.00028 \*\*\*
## Year2013 -0.05857 0.01781 -3.29 0.00101 \*\*
## Year2014 -0.02670 0.01771 -1.51 0.13160
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared: 0.00973, Adjusted R-squared: 0.00878
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 14680 is an outlier with |weight| = 0 ( < 5.1e-06);
## 1713 weights are ~= 1. The remaining 18081 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.023 0.868 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 5.05e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1.01 18 1



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.212629 -0.234609 0.000354 0.233232 1.626834
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19384 0.01275 93.62 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.01578 0.00575 -2.75 0.00604 \*\*
## Year1997 0.01879 0.01764 1.07 0.28686
## Year1998 -0.01481 0.01761 -0.84 0.40030
## Year1999 -0.03741 0.01677 -2.23 0.02570 \*
## Year2000 -0.07221 0.01733 -4.17 3.1e-05 \*\*\*
## Year2001 -0.10693 0.01772 -6.03 1.6e-09 \*\*\*
## Year2002 -0.06783 0.01765 -3.84 0.00012 \*\*\*
## Year2003 -0.09693 0.01669 -5.81 6.4e-09 \*\*\*
## Year2004 -0.08409 0.01686 -4.99 6.2e-07 \*\*\*
## Year2005 -0.09790 0.01669 -5.87 4.5e-09 \*\*\*
## Year2006 -0.10296 0.01625 -6.33 2.4e-10 \*\*\*
## Year2007 -0.07985 0.01595 -5.01 5.6e-07 \*\*\*
## Year2008 -0.06738 0.01607 -4.19 2.8e-05 \*\*\*
## Year2009 -0.07837 0.01655 -4.73 2.2e-06 \*\*\*
## Year2010 -0.06268 0.01688 -3.71 0.00021 \*\*\*
## Year2011 -0.03126 0.01700 -1.84 0.06595 .
## Year2012 -0.06058 0.01695 -3.58 0.00035 \*\*\*
## Year2013 -0.05755 0.01781 -3.23 0.00123 \*\*
## Year2014 -0.02544 0.01770 -1.44 0.15075
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared: 0.01, Adjusted R-squared: 0.00905
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 1722 weights are ~= 1. The remaining 18073 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.868 0.950 0.898 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 5.05e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 19795"
## [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
## 2144 2732 2138 2028 2239 2361 2056 1737 1931 2023 2153 2194 2127 1982 2109
## 2011 2012 2013 2014
## 2132 2059 2070 1978
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1282 1296 1215 1143 1111 1050 1280 1072 1204 1309 1459 1473 1483 1351 1428
## 2011 2012 2013 2014
## 1458 1390 1404 1407
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1094 1145 1048 960 939 898 1086 909 1031 1087 1227 1239 1279 1166 1197
## 2011 2012 2013 2014
## 1272 1198 1212 1206
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 300, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 30, df = 1, p-value = 7e-08



## [1] "Female first author team size geometric mean: 3.9806509757914"
## [1] "Male first author team size geometric mean: 3.56982243050111"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 6e+07, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.69683772343223"
## [1] "Male last author team size geometric mean: 3.7670231423275"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 4e+07, 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 1.04 1 1.02
## LastAuthorFemale 1.04 1 1.02
## UniqueAuthors 1.07 4 1.01
## Year 1.12 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3887 -0.2517 0.0104 0.2445 1.5514
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05697 0.01722 61.40 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.02134 0.00547 3.90 9.6e-05 \*\*\*
## LastAuthorFemale1 0.00987 0.00600 1.65 0.09972 .
## UniqueAuthors2 0.03970 0.01300 3.05 0.00227 \*\*
## UniqueAuthors3 0.13359 0.01414 9.45 < 2e-16 \*\*\*
## UniqueAuthors4 0.13476 0.01298 10.38 < 2e-16 \*\*\*
## UniqueAuthors5 0.22230 0.01209 18.38 < 2e-16 \*\*\*
## Year1997 0.08805 0.01926 4.57 4.9e-06 \*\*\*
## Year1998 -0.00529 0.01802 -0.29 0.76896
## Year1999 -0.04092 0.01817 -2.25 0.02438 \*
## Year2000 -0.04405 0.01852 -2.38 0.01742 \*
## Year2001 -0.04270 0.01801 -2.37 0.01776 \*
## Year2002 -0.07118 0.01803 -3.95 7.9e-05 \*\*\*
## Year2003 -0.10941 0.01838 -5.95 2.7e-09 \*\*\*
## Year2004 -0.09132 0.01757 -5.20 2.0e-07 \*\*\*
## Year2005 -0.08303 0.01745 -4.76 2.0e-06 \*\*\*
## Year2006 -0.07823 0.01739 -4.50 6.9e-06 \*\*\*
## Year2007 -0.07021 0.01704 -4.12 3.8e-05 \*\*\*
## Year2008 -0.06765 0.01731 -3.91 9.3e-05 \*\*\*
## Year2009 -0.03349 0.01768 -1.89 0.05823 .
## Year2010 -0.07748 0.01730 -4.48 7.6e-06 \*\*\*
## Year2011 -0.06196 0.01759 -3.52 0.00043 \*\*\*
## Year2012 -0.08959 0.01808 -4.96 7.3e-07 \*\*\*
## Year2013 -0.10151 0.01857 -5.47 4.6e-08 \*\*\*
## Year2014 -0.08640 0.01929 -4.48 7.6e-06 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.368
## Multiple R-squared: 0.0542, Adjusted R-squared: 0.0531
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 1761 weights are ~= 1. The remaining 19432 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.036 0.866 0.950 0.896 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 4.72e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.04 1 1.02
## LastAuthorFemale 1.04 1 1.02
## Year 1.04 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30194 -0.25346 0.00779 0.24865 1.49625
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17475 0.01419 82.80 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.02929 0.00558 5.25 1.5e-07 \*\*\*
## LastAuthorFemale1 0.00226 0.00607 0.37 0.71024
## Year1997 0.09565 0.01967 4.86 1.2e-06 \*\*\*
## Year1998 0.00765 0.01843 0.41 0.67817
## Year1999 -0.02911 0.01862 -1.56 0.11803
## Year2000 -0.02890 0.01899 -1.52 0.12810
## Year2001 -0.02772 0.01842 -1.50 0.13245
## Year2002 -0.05340 0.01841 -2.90 0.00373 \*\*
## Year2003 -0.09024 0.01899 -4.75 2.0e-06 \*\*\*
## Year2004 -0.06108 0.01798 -3.40 0.00068 \*\*\*
## Year2005 -0.05338 0.01784 -2.99 0.00277 \*\*
## Year2006 -0.05429 0.01771 -3.06 0.00218 \*\*
## Year2007 -0.03678 0.01739 -2.11 0.03447 \*
## Year2008 -0.03344 0.01765 -1.89 0.05816 .
## Year2009 0.00563 0.01781 0.32 0.75196
## Year2010 -0.03887 0.01766 -2.20 0.02776 \*
## Year2011 -0.02924 0.01787 -1.64 0.10185
## Year2012 -0.04809 0.01844 -2.61 0.00913 \*\*
## Year2013 -0.05981 0.01899 -3.15 0.00164 \*\*
## Year2014 -0.04266 0.01949 -2.19 0.02864 \*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.374
## Multiple R-squared: 0.0104, Adjusted R-squared: 0.00947
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 1759 weights are ~= 1. The remaining 19434 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.073 0.866 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 4.72e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.02 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30041 -0.25310 0.00791 0.24843 1.49588
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17512 0.01416 83.01 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.02960 0.00553 5.35 8.8e-08 \*\*\*
## Year1997 0.09569 0.01966 4.87 1.1e-06 \*\*\*
## Year1998 0.00763 0.01843 0.41 0.67887
## Year1999 -0.02904 0.01862 -1.56 0.11883
## Year2000 -0.02894 0.01899 -1.52 0.12757
## Year2001 -0.02768 0.01842 -1.50 0.13295
## Year2002 -0.05327 0.01840 -2.89 0.00380 \*\*
## Year2003 -0.09014 0.01898 -4.75 2.1e-06 \*\*\*
## Year2004 -0.06098 0.01797 -3.39 0.00069 \*\*\*
## Year2005 -0.05332 0.01784 -2.99 0.00280 \*\*
## Year2006 -0.05415 0.01771 -3.06 0.00223 \*\*
## Year2007 -0.03661 0.01738 -2.11 0.03516 \*
## Year2008 -0.03330 0.01764 -1.89 0.05911 .
## Year2009 0.00580 0.01780 0.33 0.74449
## Year2010 -0.03865 0.01764 -2.19 0.02849 \*
## Year2011 -0.02903 0.01785 -1.63 0.10397
## Year2012 -0.04781 0.01842 -2.60 0.00943 \*\*
## Year2013 -0.05954 0.01896 -3.14 0.00169 \*\*
## Year2014 -0.04233 0.01945 -2.18 0.02955 \*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.374
## Multiple R-squared: 0.0104, Adjusted R-squared: 0.00951
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 1771 weights are ~= 1. The remaining 19422 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.074 0.866 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 4.72e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.02 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28790 -0.25398 0.00702 0.24888 1.48619
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18481 0.01402 84.53 < 2e-16 \*\*\*
## LastAuthorFemale1 0.00709 0.00603 1.18 0.2392
## Year1997 0.09599 0.01966 4.88 1.0e-06 \*\*\*
## Year1998 0.00706 0.01842 0.38 0.7015
## Year1999 -0.02981 0.01860 -1.60 0.1091
## Year2000 -0.02894 0.01897 -1.53 0.1273
## Year2001 -0.02809 0.01843 -1.52 0.1274
## Year2002 -0.05231 0.01839 -2.84 0.0045 \*\*
## Year2003 -0.08909 0.01897 -4.70 2.7e-06 \*\*\*
## Year2004 -0.06025 0.01798 -3.35 0.0008 \*\*\*
## Year2005 -0.05222 0.01784 -2.93 0.0034 \*\*
## Year2006 -0.05269 0.01770 -2.98 0.0029 \*\*
## Year2007 -0.03430 0.01738 -1.97 0.0485 \*
## Year2008 -0.03089 0.01765 -1.75 0.0800 .
## Year2009 0.00894 0.01779 0.50 0.6151
## Year2010 -0.03584 0.01763 -2.03 0.0421 \*
## Year2011 -0.02531 0.01785 -1.42 0.1563
## Year2012 -0.04405 0.01841 -2.39 0.0168 \*
## Year2013 -0.05694 0.01896 -3.00 0.0027 \*\*
## Year2014 -0.03769 0.01943 -1.94 0.0524 .
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.374
## Multiple R-squared: 0.00905, Adjusted R-squared: 0.00816
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 1792 weights are ~= 1. The remaining 19401 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.079 0.866 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 4.72e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 21193"
## [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
## 9546 8329 816 5876 702 6648 6380 6049 6736 6582 6666 6605 6277 6733 6292
## 2011 2012 2013 2014
## 6484 6532 6543 6912
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3947 3736 391 3284 340 2623 3788 3600 3999 3843 3847 3804 3743 4014 3745
## 2011 2012 2013 2014
## 3948 3962 3995 4393
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3378 3252 351 2837 301 2253 3278 3098 3421 3284 3295 3270 3203 3409 3218
## 2011 2012 2013 2014
## 3366 3398 3407 3758
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 2000, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 200, df = 1, p-value <2e-16



## [1] "Female first author team size geometric mean: 3.62256062340099"
## [1] "Male first author team size geometric mean: 3.24400021893423"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 4e+08, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.40162710718271"
## [1] "Male last author team size geometric mean: 3.37108177529384"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 3e+08, 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 1.02 1 1.01
## LastAuthorFemale 1.02 1 1.01
## UniqueAuthors 1.05 4 1.01
## Year 1.05 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control = lmrob.control(fast.s.large.n = Inf))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.43688 -0.27477 0.00842 0.27597 2.40730
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19762 0.01153 103.83 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.03180 0.00377 -8.44 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.02925 0.00428 -6.83 8.6e-12 \*\*\*
## UniqueAuthors2 0.09748 0.00753 12.95 < 2e-16 \*\*\*
## UniqueAuthors3 0.15562 0.00875 17.78 < 2e-16 \*\*\*
## UniqueAuthors4 0.14173 0.00805 17.60 < 2e-16 \*\*\*
## UniqueAuthors5 0.23926 0.00726 32.94 < 2e-16 \*\*\*
## Year1997 -0.03495 0.01387 -2.52 0.012 \*
## Year1998 -0.00258 0.02888 -0.09 0.929
## Year1999 -0.13034 0.01243 -10.49 < 2e-16 \*\*\*
## Year2000 -0.01852 0.02483 -0.75 0.456
## Year2001 -0.11412 0.01336 -8.54 < 2e-16 \*\*\*
## Year2002 -0.14627 0.01231 -11.88 < 2e-16 \*\*\*
## Year2003 -0.17785 0.01216 -14.63 < 2e-16 \*\*\*
## Year2004 -0.19441 0.01203 -16.16 < 2e-16 \*\*\*
## Year2005 -0.18193 0.01203 -15.12 < 2e-16 \*\*\*
## Year2006 -0.17865 0.01202 -14.86 < 2e-16 \*\*\*
## Year2007 -0.17781 0.01207 -14.73 < 2e-16 \*\*\*
## Year2008 -0.15257 0.01232 -12.39 < 2e-16 \*\*\*
## Year2009 -0.20016 0.01227 -16.31 < 2e-16 \*\*\*
## Year2010 -0.17118 0.01228 -13.94 < 2e-16 \*\*\*
## Year2011 -0.17386 0.01231 -14.12 < 2e-16 \*\*\*
## Year2012 -0.19515 0.01239 -15.75 < 2e-16 \*\*\*
## Year2013 -0.19378 0.01266 -15.30 < 2e-16 \*\*\*
## Year2014 -0.18325 0.01284 -14.27 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.408
## Multiple R-squared: 0.0473, Adjusted R-squared: 0.0469
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 11 observations c(36,30069,37804,44582,45347,50659,52801,53184,53190,53426,55483)
## are outliers with |weight| = 0 ( < 1.8e-06);
## 4719 weights are ~= 1. The remaining 51047 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.867 0.950 0.896 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 1.79e-06 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
## 500 50 2 1 200 200
## trace.lev mts compute.rd
## 0 1000 0
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.02 1 1.01
## LastAuthorFemale 1.02 1 1.01
## Year 1.01 18 1.00



## [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))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3186 -0.2772 0.0074 0.2774 2.3979
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31858 0.01002 131.54 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.02086 0.00381 -5.48 4.3e-08 \*\*\*
## LastAuthorFemale1 -0.03142 0.00433 -7.25 4.1e-13 \*\*\*
## Year1997 -0.03198 0.01392 -2.30 0.022 \*
## Year1998 -0.00188 0.02903 -0.06 0.948
## Year1999 -0.11838 0.01245 -9.51 < 2e-16 \*\*\*
## Year2000 -0.01086 0.02512 -0.43 0.665
## Year2001 -0.08783 0.01333 -6.59 4.5e-11 \*\*\*
## Year2002 -0.12398 0.01238 -10.02 < 2e-16 \*\*\*
## Year2003 -0.15188 0.01224 -12.41 < 2e-16 \*\*\*
## Year2004 -0.16892 0.01210 -13.96 < 2e-16 \*\*\*
## Year2005 -0.15326 0.01204 -12.73 < 2e-16 \*\*\*
## Year2006 -0.14592 0.01203 -12.13 < 2e-16 \*\*\*
## Year2007 -0.14409 0.01209 -11.92 < 2e-16 \*\*\*
## Year2008 -0.11631 0.01237 -9.40 < 2e-16 \*\*\*
## Year2009 -0.16963 0.01237 -13.72 < 2e-16 \*\*\*
## Year2010 -0.13727 0.01232 -11.14 < 2e-16 \*\*\*
## Year2011 -0.13615 0.01238 -10.99 < 2e-16 \*\*\*
## Year2012 -0.15460 0.01246 -12.41 < 2e-16 \*\*\*
## Year2013 -0.15243 0.01271 -11.99 < 2e-16 \*\*\*
## Year2014 -0.13533 0.01291 -10.48 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared: 0.0135, Adjusted R-squared: 0.0131
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 16 observations c(36,995,1268,30069,37804,39529,43830,44582,45347,50659,51413,52801,53184,53190,53426,55483)
## are outliers with |weight| = 0 ( < 1.8e-06);
## 4625 weights are ~= 1. The remaining 51136 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.866 0.951 0.895 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 1.79e-06 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
## 500 50 2 1 200 200
## trace.lev mts compute.rd
## 0 1000 0
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1.01 18 1



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31343 -0.27733 0.00764 0.27650 2.40479
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31343 0.01000 131.39 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.02467 0.00379 -6.51 7.7e-11 \*\*\*
## Year1997 -0.03207 0.01393 -2.30 0.021 \*
## Year1998 -0.00217 0.02910 -0.07 0.941
## Year1999 -0.11834 0.01245 -9.51 < 2e-16 \*\*\*
## Year2000 -0.01178 0.02519 -0.47 0.640
## Year2001 -0.08847 0.01333 -6.64 3.3e-11 \*\*\*
## Year2002 -0.12425 0.01238 -10.04 < 2e-16 \*\*\*
## Year2003 -0.15213 0.01224 -12.43 < 2e-16 \*\*\*
## Year2004 -0.16904 0.01210 -13.97 < 2e-16 \*\*\*
## Year2005 -0.15327 0.01205 -12.72 < 2e-16 \*\*\*
## Year2006 -0.14646 0.01203 -12.18 < 2e-16 \*\*\*
## Year2007 -0.14453 0.01209 -11.95 < 2e-16 \*\*\*
## Year2008 -0.11745 0.01238 -9.49 < 2e-16 \*\*\*
## Year2009 -0.17123 0.01236 -13.86 < 2e-16 \*\*\*
## Year2010 -0.13851 0.01232 -11.24 < 2e-16 \*\*\*
## Year2011 -0.13791 0.01238 -11.14 < 2e-16 \*\*\*
## Year2012 -0.15609 0.01246 -12.53 < 2e-16 \*\*\*
## Year2013 -0.15422 0.01271 -12.13 < 2e-16 \*\*\*
## Year2014 -0.13742 0.01290 -10.65 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared: 0.0126, Adjusted R-squared: 0.0122
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 17 observations c(36,995,1268,30069,37804,39529,43830,44582,45347,50659,51413,52801,53178,53184,53190,53426,55483)
## are outliers with |weight| <= 2.7e-09 ( < 1.8e-06);
## 4610 weights are ~= 1. The remaining 51150 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.007 0.867 0.951 0.895 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 1.79e-06 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
## 500 50 2 1 200 200
## trace.lev mts compute.rd
## 0 1000 0
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1.01 18 1



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31241 -0.27766 0.00742 0.27693 2.40532
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31241 0.00995 131.95 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.03474 0.00431 -8.05 8.3e-16 \*\*\*
## Year1997 -0.03231 0.01391 -2.32 0.02 \*
## Year1998 -0.00143 0.02904 -0.05 0.96
## Year1999 -0.11847 0.01245 -9.52 < 2e-16 \*\*\*
## Year2000 -0.00940 0.02514 -0.37 0.71
## Year2001 -0.08839 0.01334 -6.63 3.4e-11 \*\*\*
## Year2002 -0.12432 0.01237 -10.05 < 2e-16 \*\*\*
## Year2003 -0.15271 0.01224 -12.48 < 2e-16 \*\*\*
## Year2004 -0.16965 0.01210 -14.02 < 2e-16 \*\*\*
## Year2005 -0.15429 0.01204 -12.81 < 2e-16 \*\*\*
## Year2006 -0.14685 0.01203 -12.21 < 2e-16 \*\*\*
## Year2007 -0.14521 0.01209 -12.01 < 2e-16 \*\*\*
## Year2008 -0.11767 0.01237 -9.51 < 2e-16 \*\*\*
## Year2009 -0.17101 0.01236 -13.83 < 2e-16 \*\*\*
## Year2010 -0.13860 0.01231 -11.26 < 2e-16 \*\*\*
## Year2011 -0.13759 0.01238 -11.11 < 2e-16 \*\*\*
## Year2012 -0.15609 0.01246 -12.53 < 2e-16 \*\*\*
## Year2013 -0.15373 0.01271 -12.09 < 2e-16 \*\*\*
## Year2014 -0.13703 0.01290 -10.62 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared: 0.0129, Adjusted R-squared: 0.0126
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 17 observations c(36,995,1268,30069,37804,39529,43830,44582,45347,50659,51413,52801,53178,53184,53190,53426,55483)
## are outliers with |weight| = 0 ( < 1.8e-06);
## 4636 weights are ~= 1. The remaining 51124 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.003 0.866 0.951 0.895 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 1.79e-06 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
## 500 50 2 1 200 200
## trace.lev mts compute.rd
## 0 1000 0
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 55777"
## [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
## 13414 11927 11138 9464 10165 9654 10173 9197 10477 10262 10752 10502
## 2008 2009 2010 2011 2012 2013 2014
## 10987 11733 11669 11740 10893 10655 10161
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7005 6586 5928 5935 5189 4548 6428 5782 6595 6329 6463 6393 6831 7289 7210
## 2011 2012 2013 2014
## 7258 6913 6750 6551
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6007 5625 5137 5048 4433 3893 5474 4930 5565 5266 5395 5378 5748 6101 5968
## 2011 2012 2013 2014
## 6044 5778 5598 5446
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 3000, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 300, df = 1, p-value <2e-16



## [1] "Female first author team size geometric mean: 3.78836234893218"
## [1] "Male first author team size geometric mean: 3.62912604627442"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 1e+09, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.64857967095046"
## [1] "Male last author team size geometric mean: 3.69747133227415"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 9e+08, p-value = 0.006
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.01 1 1.01
## LastAuthorFemale 1.01 1 1.01
## UniqueAuthors 1.03 4 1.00
## Year 1.04 18 1.00



## [1] "List of 5 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 40123 0031773680 3.96 1998 1303 6 2.77
## 71613 0035710746 3.68 2001 1300 2 2.60
## 165780 65449136284 3.76 2009 1303 7 2.69
## 222272 84875634162 3.75 2013 1312 2 2.63
## 225892 84871809302 4.00 2013 1303 6 2.78
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.39224 -0.23128 0.00449 0.22877 2.77772
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11994 0.00841 133.21 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.01240 0.00233 -5.31 1.1e-07 \*\*\*
## LastAuthorFemale1 -0.01685 0.00274 -6.14 8.3e-10 \*\*\*
## UniqueAuthors2 0.13905 0.00677 20.55 < 2e-16 \*\*\*
## UniqueAuthors3 0.19960 0.00735 27.16 < 2e-16 \*\*\*
## UniqueAuthors4 0.19615 0.00689 28.48 < 2e-16 \*\*\*
## UniqueAuthors5 0.27230 0.00659 41.34 < 2e-16 \*\*\*
## Year1997 -0.02458 0.00834 -2.95 0.0032 \*\*
## Year1998 -0.06529 0.00811 -8.05 8.3e-16 \*\*\*
## Year1999 -0.15042 0.00745 -20.20 < 2e-16 \*\*\*
## Year2000 -0.14853 0.00773 -19.21 < 2e-16 \*\*\*
## Year2001 -0.18269 0.00787 -23.20 < 2e-16 \*\*\*
## Year2002 -0.17950 0.00753 -23.85 < 2e-16 \*\*\*
## Year2003 -0.21283 0.00756 -28.14 < 2e-16 \*\*\*
## Year2004 -0.20616 0.00739 -27.91 < 2e-16 \*\*\*
## Year2005 -0.19276 0.00739 -26.08 < 2e-16 \*\*\*
## Year2006 -0.19373 0.00748 -25.89 < 2e-16 \*\*\*
## Year2007 -0.18879 0.00751 -25.14 < 2e-16 \*\*\*
## Year2008 -0.17408 0.00735 -23.69 < 2e-16 \*\*\*
## Year2009 -0.18590 0.00740 -25.13 < 2e-16 \*\*\*
## Year2010 -0.17769 0.00750 -23.70 < 2e-16 \*\*\*
## Year2011 -0.16681 0.00753 -22.14 < 2e-16 \*\*\*
## Year2012 -0.18967 0.00769 -24.65 < 2e-16 \*\*\*
## Year2013 -0.17196 0.00802 -21.45 < 2e-16 \*\*\*
## Year2014 -0.16681 0.00854 -19.53 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.343
## Multiple R-squared: 0.0627, Adjusted R-squared: 0.0625
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 43 observations c(16682,29369,30101,30102,35534,36926,41036,42387,49097,49359,50255,58130,58978,62161,71244,71487,76374,77650,78711,80911,82089,83281,84668,84670,85116,89344,89612,89897,89910,90503,92046,95019,95752,97330,98763,98770,99386,100675,101030,101586,101931,102454,102713)
## are outliers with |weight| = 0 ( < 9.7e-07);
## 8632 weights are ~= 1. The remaining 94159 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.865 0.951 0.892 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 9.72e-07 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.01 1 1.01
## LastAuthorFemale 1.01 1 1.00
## Year 1.01 18 1.00



## [1] "List of 5 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 40123 0031773680 3.96 1998 1303 6 2.72
## 71613 0035710746 3.68 2001 1300 2 2.54
## 165780 65449136284 3.76 2009 1303 7 2.62
## 222272 84875634162 3.75 2013 1312 2 2.61
## 225892 84871809302 4.00 2013 1303 6 2.84
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30029 -0.23456 0.00519 0.23022 2.83654
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30029 0.00606 214.41 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.00727 0.00237 -3.07 0.0022 \*\*
## LastAuthorFemale1 -0.01916 0.00278 -6.88 6.0e-12 \*\*\*
## Year1997 -0.02150 0.00849 -2.53 0.0113 \*
## Year1998 -0.06011 0.00820 -7.33 2.4e-13 \*\*\*
## Year1999 -0.13893 0.00757 -18.35 < 2e-16 \*\*\*
## Year2000 -0.13524 0.00785 -17.24 < 2e-16 \*\*\*
## Year2001 -0.16496 0.00800 -20.62 < 2e-16 \*\*\*
## Year2002 -0.16060 0.00764 -21.02 < 2e-16 \*\*\*
## Year2003 -0.19047 0.00771 -24.71 < 2e-16 \*\*\*
## Year2004 -0.18373 0.00753 -24.41 < 2e-16 \*\*\*
## Year2005 -0.16828 0.00752 -22.39 < 2e-16 \*\*\*
## Year2006 -0.16801 0.00757 -22.19 < 2e-16 \*\*\*
## Year2007 -0.16148 0.00762 -21.20 < 2e-16 \*\*\*
## Year2008 -0.14528 0.00745 -19.50 < 2e-16 \*\*\*
## Year2009 -0.16080 0.00753 -21.35 < 2e-16 \*\*\*
## Year2010 -0.14830 0.00762 -19.47 < 2e-16 \*\*\*
## Year2011 -0.13624 0.00767 -17.75 < 2e-16 \*\*\*
## Year2012 -0.15716 0.00781 -20.11 < 2e-16 \*\*\*
## Year2013 -0.13882 0.00812 -17.09 < 2e-16 \*\*\*
## Year2014 -0.13390 0.00869 -15.42 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared: 0.0202, Adjusted R-squared: 0.02
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 35 observations c(16682,29369,30101,30102,35534,36926,42387,50255,58130,58978,62161,69564,71487,76374,77650,78711,80911,82089,83281,84668,85116,89344,89612,89910,90503,92046,95752,97330,98763,98770,99386,100675,101931,102454,102713)
## are outliers with |weight| = 0 ( < 9.7e-07);
## 8571 weights are ~= 1. The remaining 94228 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.864 0.951 0.890 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 9.72e-07 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1 18 1



## [1] "List of 5 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 40123 0031773680 3.96 1998 1303 6 2.72
## 71613 0035710746 3.68 2001 1300 2 2.54
## 165780 65449136284 3.76 2009 1303 7 2.62
## 222272 84875634162 3.75 2013 1312 2 2.61
## 225892 84871809302 4.00 2013 1303 6 2.84
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29732 -0.23474 0.00503 0.23052 2.84054
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29732 0.00606 214.21 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.00896 0.00236 -3.79 0.00015 \*\*\*
## Year1997 -0.02162 0.00850 -2.54 0.01097 \*
## Year1998 -0.06031 0.00821 -7.34 2.1e-13 \*\*\*
## Year1999 -0.13913 0.00758 -18.36 < 2e-16 \*\*\*
## Year2000 -0.13534 0.00786 -17.23 < 2e-16 \*\*\*
## Year2001 -0.16536 0.00801 -20.65 < 2e-16 \*\*\*
## Year2002 -0.16093 0.00765 -21.05 < 2e-16 \*\*\*
## Year2003 -0.19070 0.00771 -24.73 < 2e-16 \*\*\*
## Year2004 -0.18411 0.00753 -24.44 < 2e-16 \*\*\*
## Year2005 -0.16853 0.00752 -22.41 < 2e-16 \*\*\*
## Year2006 -0.16867 0.00758 -22.26 < 2e-16 \*\*\*
## Year2007 -0.16227 0.00762 -21.29 < 2e-16 \*\*\*
## Year2008 -0.14619 0.00745 -19.61 < 2e-16 \*\*\*
## Year2009 -0.16157 0.00754 -21.43 < 2e-16 \*\*\*
## Year2010 -0.14888 0.00762 -19.53 < 2e-16 \*\*\*
## Year2011 -0.13735 0.00768 -17.89 < 2e-16 \*\*\*
## Year2012 -0.15797 0.00782 -20.20 < 2e-16 \*\*\*
## Year2013 -0.13986 0.00813 -17.21 < 2e-16 \*\*\*
## Year2014 -0.13492 0.00869 -15.53 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared: 0.0197, Adjusted R-squared: 0.0195
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 35 observations c(16682,29369,30101,30102,35534,36926,42387,50255,58130,58978,62161,69564,71487,76374,77650,78711,80911,82089,83281,84668,85116,89344,89612,89910,90503,92046,95752,97330,98763,98770,99386,100675,101931,102454,102713)
## are outliers with |weight| = 0 ( < 9.7e-07);
## 8677 weights are ~= 1. The remaining 94122 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.864 0.951 0.890 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 9.72e-07 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## LastAuthorFemale 1 1 1
## Year 1 18 1



## [1] "List of 5 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 40123 0031773680 3.96 1998 1303 6 2.72
## 71613 0035710746 3.68 2001 1300 2 2.54
## 165780 65449136284 3.76 2009 1303 7 2.62
## 222272 84875634162 3.75 2013 1312 2 2.61
## 225892 84871809302 4.00 2013 1303 6 2.84
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29795 -0.23433 0.00518 0.23037 2.83927
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29795 0.00601 215.98 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.02005 0.00278 -7.22 5.3e-13 \*\*\*
## Year1997 -0.02153 0.00849 -2.54 0.011 \*
## Year1998 -0.06010 0.00821 -7.32 2.4e-13 \*\*\*
## Year1999 -0.13897 0.00757 -18.35 < 2e-16 \*\*\*
## Year2000 -0.13541 0.00785 -17.26 < 2e-16 \*\*\*
## Year2001 -0.16496 0.00800 -20.62 < 2e-16 \*\*\*
## Year2002 -0.16071 0.00764 -21.04 < 2e-16 \*\*\*
## Year2003 -0.19067 0.00771 -24.74 < 2e-16 \*\*\*
## Year2004 -0.18392 0.00753 -24.43 < 2e-16 \*\*\*
## Year2005 -0.16839 0.00752 -22.41 < 2e-16 \*\*\*
## Year2006 -0.16828 0.00757 -22.23 < 2e-16 \*\*\*
## Year2007 -0.16179 0.00762 -21.24 < 2e-16 \*\*\*
## Year2008 -0.14561 0.00745 -19.55 < 2e-16 \*\*\*
## Year2009 -0.16116 0.00753 -21.40 < 2e-16 \*\*\*
## Year2010 -0.14863 0.00762 -19.51 < 2e-16 \*\*\*
## Year2011 -0.13662 0.00767 -17.80 < 2e-16 \*\*\*
## Year2012 -0.15762 0.00781 -20.17 < 2e-16 \*\*\*
## Year2013 -0.13922 0.00812 -17.15 < 2e-16 \*\*\*
## Year2014 -0.13443 0.00868 -15.48 < 2e-16 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared: 0.0201, Adjusted R-squared: 0.0199
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 35 observations c(16682,29369,30101,30102,35534,36926,42387,50255,58130,58978,62161,69564,71487,76374,77650,78711,80911,82089,83281,84668,85116,89344,89612,89910,90503,92046,95752,97330,98763,98770,99386,100675,101931,102454,102713)
## are outliers with |weight| = 0 ( < 9.7e-07);
## 8631 weights are ~= 1. The remaining 94168 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.864 0.951 0.890 0.986 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 9.72e-07 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 102834"
## [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
## 1865 1711 1698 1787 1793 2153 2148 2222 2418 2465 2744 2741 2916 3341 3499
## 2011 2012 2013 2014
## 3452 3468 3349 3239
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1215 1086 933 1127 875 1024 1300 1373 1485 1519 1631 1723 1823 2115 2204
## 2011 2012 2013 2014
## 2218 2160 2119 2112
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1037 935 796 932 752 867 1106 1150 1244 1210 1362 1401 1502 1729 1791
## 2011 2012 2013 2014
## 1806 1778 1726 1753
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 600, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7, df = 1, p-value = 0.009



## [1] "Female first author team size geometric mean: 4.79297110623185"
## [1] "Male first author team size geometric mean: 4.85868892391225"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 6e+07, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 4.72585533315949"
## [1] "Male last author team size geometric mean: 4.86467451335547"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 5e+07, p-value = 0.009
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.02 1 1.01
## LastAuthorFemale 1.01 1 1.00
## UniqueAuthors 1.05 4 1.01
## Year 1.06 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35190 -0.24651 0.00709 0.24630 2.46267
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96630 0.02269 42.59 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.00407 0.00550 0.74 0.45894
## LastAuthorFemale1 -0.01402 0.00653 -2.15 0.03190 \*
## UniqueAuthors2 0.24603 0.01836 13.40 < 2e-16 \*\*\*
## UniqueAuthors3 0.31474 0.02081 15.13 < 2e-16 \*\*\*
## UniqueAuthors4 0.31860 0.01842 17.30 < 2e-16 \*\*\*
## UniqueAuthors5 0.38560 0.01731 22.28 < 2e-16 \*\*\*
## Year1997 -0.02101 0.02326 -0.90 0.36648
## Year1998 -0.00128 0.02144 -0.06 0.95235
## Year1999 -0.06975 0.02043 -3.41 0.00064 \*\*\*
## Year2000 -0.05585 0.02090 -2.67 0.00754 \*\*
## Year2001 -0.13730 0.01996 -6.88 6.1e-12 \*\*\*
## Year2002 -0.13915 0.01928 -7.22 5.4e-13 \*\*\*
## Year2003 -0.15920 0.01849 -8.61 < 2e-16 \*\*\*
## Year2004 -0.17896 0.01835 -9.75 < 2e-16 \*\*\*
## Year2005 -0.16346 0.01879 -8.70 < 2e-16 \*\*\*
## Year2006 -0.16390 0.01830 -8.95 < 2e-16 \*\*\*
## Year2007 -0.15136 0.01835 -8.25 < 2e-16 \*\*\*
## Year2008 -0.12705 0.01836 -6.92 4.6e-12 \*\*\*
## Year2009 -0.12320 0.01781 -6.92 4.7e-12 \*\*\*
## Year2010 -0.15153 0.01788 -8.48 < 2e-16 \*\*\*
## Year2011 -0.12899 0.01811 -7.12 1.1e-12 \*\*\*
## Year2012 -0.14161 0.01847 -7.67 1.8e-14 \*\*\*
## Year2013 -0.10440 0.01869 -5.59 2.3e-08 \*\*\*
## Year2014 -0.10402 0.01919 -5.42 6.0e-08 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.37
## Multiple R-squared: 0.0713, Adjusted R-squared: 0.0704
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 9 observations c(873,900,4899,5295,21947,23214,24620,24662,24685)
## are outliers with |weight| = 0 ( < 4e-06);
## 2130 weights are ~= 1. The remaining 22738 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.002 0.862 0.951 0.891 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 4.02e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.01 1 1.01
## LastAuthorFemale 1.01 1 1.00
## Year 1.02 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2916 -0.2511 0.0093 0.2487 2.3931
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28193 0.01590 80.62 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.00842 0.00555 1.52 0.1294
## LastAuthorFemale1 -0.01584 0.00663 -2.39 0.0170 \*
## Year1997 -0.01930 0.02404 -0.80 0.4221
## Year1998 0.00121 0.02166 0.06 0.9553
## Year1999 -0.06044 0.02076 -2.91 0.0036 \*\*
## Year2000 -0.05251 0.02101 -2.50 0.0125 \*
## Year2001 -0.14559 0.02029 -7.17 7.5e-13 \*\*\*
## Year2002 -0.12934 0.01971 -6.56 5.4e-11 \*\*\*
## Year2003 -0.13625 0.01883 -7.23 4.8e-13 \*\*\*
## Year2004 -0.16128 0.01881 -8.57 < 2e-16 \*\*\*
## Year2005 -0.14952 0.01917 -7.80 6.5e-15 \*\*\*
## Year2006 -0.14243 0.01860 -7.66 2.0e-14 \*\*\*
## Year2007 -0.12591 0.01878 -6.71 2.0e-11 \*\*\*
## Year2008 -0.09771 0.01868 -5.23 1.7e-07 \*\*\*
## Year2009 -0.09885 0.01816 -5.44 5.3e-08 \*\*\*
## Year2010 -0.12618 0.01824 -6.92 4.7e-12 \*\*\*
## Year2011 -0.11073 0.01847 -6.00 2.0e-09 \*\*\*
## Year2012 -0.12439 0.01878 -6.62 3.5e-11 \*\*\*
## Year2013 -0.08366 0.01897 -4.41 1.0e-05 \*\*\*
## Year2014 -0.08190 0.01963 -4.17 3.0e-05 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.373
## Multiple R-squared: 0.0123, Adjusted R-squared: 0.0115
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 9 observations c(873,900,4899,5295,19346,21947,23214,24620,24662)
## are outliers with |weight| = 0 ( < 4e-06);
## 2087 weights are ~= 1. The remaining 22781 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.862 0.951 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 4.02e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.01 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2883 -0.2506 0.0097 0.2481 2.3952
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.279802 0.015889 80.55 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.007535 0.005561 1.35 0.1754
## Year1997 -0.019505 0.024053 -0.81 0.4174
## Year1998 0.000995 0.021673 0.05 0.9634
## Year1999 -0.060419 0.020764 -2.91 0.0036 \*\*
## Year2000 -0.052769 0.021015 -2.51 0.0120 \*
## Year2001 -0.146115 0.020307 -7.20 6.4e-13 \*\*\*
## Year2002 -0.129486 0.019718 -6.57 5.2e-11 \*\*\*
## Year2003 -0.136252 0.018838 -7.23 4.9e-13 \*\*\*
## Year2004 -0.161791 0.018815 -8.60 < 2e-16 \*\*\*
## Year2005 -0.149435 0.019182 -7.79 6.9e-15 \*\*\*
## Year2006 -0.143038 0.018617 -7.68 1.6e-14 \*\*\*
## Year2007 -0.126438 0.018781 -6.73 1.7e-11 \*\*\*
## Year2008 -0.098260 0.018692 -5.26 1.5e-07 \*\*\*
## Year2009 -0.099463 0.018170 -5.47 4.4e-08 \*\*\*
## Year2010 -0.126722 0.018256 -6.94 4.0e-12 \*\*\*
## Year2011 -0.111564 0.018474 -6.04 1.6e-09 \*\*\*
## Year2012 -0.125425 0.018775 -6.68 2.4e-11 \*\*\*
## Year2013 -0.084855 0.018977 -4.47 7.8e-06 \*\*\*
## Year2014 -0.082943 0.019632 -4.22 2.4e-05 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.373
## Multiple R-squared: 0.012, Adjusted R-squared: 0.0113
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 10 observations c(873,900,4899,5295,19346,21946,21947,23214,24620,24662)
## are outliers with |weight| = 0 ( < 4e-06);
## 2092 weights are ~= 1. The remaining 22775 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.001 0.862 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 4.02e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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
## Year 1.01 18 1



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28530 -0.25220 0.00943 0.24902 2.39125
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28375 0.01586 80.96 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.01518 0.00664 -2.28 0.0223 \*
## Year1997 -0.01900 0.02404 -0.79 0.4294
## Year1998 0.00154 0.02166 0.07 0.9431
## Year1999 -0.06038 0.02076 -2.91 0.0036 \*\*
## Year2000 -0.05228 0.02101 -2.49 0.0129 \*
## Year2001 -0.14514 0.02029 -7.15 8.8e-13 \*\*\*
## Year2002 -0.12894 0.01971 -6.54 6.2e-11 \*\*\*
## Year2003 -0.13563 0.01882 -7.20 6.0e-13 \*\*\*
## Year2004 -0.16102 0.01881 -8.56 < 2e-16 \*\*\*
## Year2005 -0.14925 0.01917 -7.78 7.3e-15 \*\*\*
## Year2006 -0.14179 0.01860 -7.62 2.6e-14 \*\*\*
## Year2007 -0.12552 0.01878 -6.68 2.4e-11 \*\*\*
## Year2008 -0.09681 0.01867 -5.19 2.2e-07 \*\*\*
## Year2009 -0.09808 0.01816 -5.40 6.6e-08 \*\*\*
## Year2010 -0.12530 0.01823 -6.87 6.4e-12 \*\*\*
## Year2011 -0.10977 0.01846 -5.95 2.8e-09 \*\*\*
## Year2012 -0.12350 0.01877 -6.58 4.8e-11 \*\*\*
## Year2013 -0.08251 0.01895 -4.35 1.3e-05 \*\*\*
## Year2014 -0.08071 0.01960 -4.12 3.8e-05 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.373
## Multiple R-squared: 0.0122, Adjusted R-squared: 0.0114
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 8 observations c(873,900,4899,5295,21947,23214,24620,24662)
## are outliers with |weight| = 0 ( < 4e-06);
## 2103 weights are ~= 1. The remaining 22766 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.861 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 4.02e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 24877"
## [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
## 4206 4954 4543 4184 4440 4071 4414 3972 4064 4354 4219 4702 4047 3829 3960
## 2011 2012 2013 2014
## 3945 3927 3791 3691
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2530 3101 2842 2577 2467 2005 2812 2555 2665 2836 2813 3082 2704 2530 2643
## 2011 2012 2013 2014
## 2705 2675 2607 2636
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2181 2705 2519 2239 2110 1736 2415 2181 2326 2421 2454 2666 2331 2168 2279
## 2011 2012 2013 2014
## 2289 2286 2248 2238
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 800, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 80, df = 1, p-value <2e-16



## [1] "Female first author team size geometric mean: 3.54557156857658"
## [1] "Male first author team size geometric mean: 3.31798389296315"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 2e+08, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.25969555865708"
## [1] "Male last author team size geometric mean: 3.41820502403128"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 1e+08, p-value = 1e-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 1.03 1 1.01
## LastAuthorFemale 1.03 1 1.01
## UniqueAuthors 1.08 4 1.01
## Year 1.10 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29500 -0.23375 0.00382 0.23175 1.80036
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08656 0.01086 100.09 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.00382 0.00375 -1.02 0.30754
## LastAuthorFemale1 -0.00755 0.00448 -1.69 0.09172 .
## UniqueAuthors2 0.07932 0.00773 10.26 < 2e-16 \*\*\*
## UniqueAuthors3 0.09560 0.00815 11.73 < 2e-16 \*\*\*
## UniqueAuthors4 0.12180 0.00788 15.47 < 2e-16 \*\*\*
## UniqueAuthors5 0.20205 0.00737 27.42 < 2e-16 \*\*\*
## Year1997 -0.01207 0.01229 -0.98 0.32629
## Year1998 0.00639 0.01201 0.53 0.59464
## Year1999 -0.02064 0.01188 -1.74 0.08232 .
## Year2000 -0.03961 0.01176 -3.37 0.00076 \*\*\*
## Year2001 -0.05414 0.01194 -4.53 5.8e-06 \*\*\*
## Year2002 -0.05971 0.01158 -5.16 2.5e-07 \*\*\*
## Year2003 -0.07423 0.01164 -6.38 1.8e-10 \*\*\*
## Year2004 -0.06050 0.01158 -5.22 1.8e-07 \*\*\*
## Year2005 -0.06842 0.01145 -5.97 2.3e-09 \*\*\*
## Year2006 -0.04399 0.01159 -3.79 0.00015 \*\*\*
## Year2007 -0.06402 0.01128 -5.67 1.4e-08 \*\*\*
## Year2008 -0.05145 0.01179 -4.36 1.3e-05 \*\*\*
## Year2009 -0.04119 0.01193 -3.45 0.00055 \*\*\*
## Year2010 -0.03968 0.01200 -3.31 0.00094 \*\*\*
## Year2011 -0.07236 0.01222 -5.92 3.3e-09 \*\*\*
## Year2012 -0.07576 0.01220 -6.21 5.3e-10 \*\*\*
## Year2013 -0.08855 0.01286 -6.88 5.9e-12 \*\*\*
## Year2014 -0.04865 0.01372 -3.55 0.00039 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.347
## Multiple R-squared: 0.0351, Adjusted R-squared: 0.0345
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 3 observations c(39896,42670,43189)
## are outliers with |weight| = 0 ( < 2.3e-06);
## 3757 weights are ~= 1. The remaining 40032 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.001 0.866 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 2.28e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.03 1 1.01
## LastAuthorFemale 1.02 1 1.01
## Year 1.02 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2030 -0.2372 0.0038 0.2342 1.6922
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.184058 0.009259 127.89 < 2e-16 \*\*\*
## FirstAuthorFemale1 0.000118 0.003794 0.03 0.97512
## LastAuthorFemale1 -0.014289 0.004543 -3.15 0.00166 \*\*
## Year1997 -0.003036 0.012351 -0.25 0.80580
## Year1998 0.018828 0.012049 1.56 0.11817
## Year1999 -0.004064 0.011903 -0.34 0.73280
## Year2000 -0.018988 0.011783 -1.61 0.10709
## Year2001 -0.029183 0.011958 -2.44 0.01467 \*
## Year2002 -0.032987 0.011627 -2.84 0.00455 \*\*
## Year2003 -0.045914 0.011713 -3.92 8.9e-05 \*\*\*
## Year2004 -0.029880 0.011601 -2.58 0.01001 \*
## Year2005 -0.034299 0.011487 -2.99 0.00283 \*\*
## Year2006 -0.011239 0.011643 -0.97 0.33439
## Year2007 -0.024815 0.011325 -2.19 0.02844 \*
## Year2008 -0.013350 0.011842 -1.13 0.25960
## Year2009 -0.002353 0.011945 -0.20 0.84384
## Year2010 0.001703 0.012044 0.14 0.88755
## Year2011 -0.031518 0.012212 -2.58 0.00986 \*\*
## Year2012 -0.032878 0.012241 -2.69 0.00724 \*\*
## Year2013 -0.042975 0.012888 -3.33 0.00085 \*\*\*
## Year2014 -0.004397 0.013765 -0.32 0.74938
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.351
## Multiple R-squared: 0.00252, Adjusted R-squared: 0.00206
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(39896,42670) are outliers with |weight| = 0 ( < 2.3e-06);
## 3862 weights are ~= 1. The remaining 39928 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.001 0.865 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 2.28e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.01 18 1.00



## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20109 -0.23712 0.00312 0.23464 1.69660
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18201 0.00925 127.84 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.00131 0.00377 -0.35 0.72752
## Year1997 -0.00294 0.01236 -0.24 0.81168
## Year1998 0.01908 0.01206 1.58 0.11356
## Year1999 -0.00398 0.01191 -0.33 0.73809
## Year2000 -0.01892 0.01179 -1.60 0.10864
## Year2001 -0.02914 0.01196 -2.44 0.01486 \*
## Year2002 -0.03317 0.01163 -2.85 0.00435 \*\*
## Year2003 -0.04595 0.01172 -3.92 8.9e-05 \*\*\*
## Year2004 -0.02993 0.01161 -2.58 0.00993 \*\*
## Year2005 -0.03432 0.01149 -2.99 0.00282 \*\*
## Year2006 -0.01150 0.01164 -0.99 0.32306
## Year2007 -0.02481 0.01133 -2.19 0.02851 \*
## Year2008 -0.01364 0.01185 -1.15 0.24963
## Year2009 -0.00296 0.01195 -0.25 0.80445
## Year2010 0.00132 0.01205 0.11 0.91289
## Year2011 -0.03231 0.01220 -2.65 0.00811 \*\*
## Year2012 -0.03313 0.01224 -2.71 0.00682 \*\*
## Year2013 -0.04365 0.01289 -3.39 0.00071 \*\*\*
## Year2014 -0.00530 0.01376 -0.39 0.70000
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.351
## Multiple R-squared: 0.00229, Adjusted R-squared: 0.00185
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(39896,42670) are outliers with |weight| = 0 ( < 2.3e-06);
## 3816 weights are ~= 1. The remaining 39974 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.001 0.865 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 2.28e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.01 18 1.00



## [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)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20291 -0.23712 0.00382 0.23421 1.69230
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18408 0.00921 128.55 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.01427 0.00451 -3.16 0.00157 \*\*
## Year1997 -0.00304 0.01235 -0.25 0.80572
## Year1998 0.01883 0.01205 1.56 0.11811
## Year1999 -0.00406 0.01190 -0.34 0.73291
## Year2000 -0.01898 0.01178 -1.61 0.10717
## Year2001 -0.02918 0.01196 -2.44 0.01469 \*
## Year2002 -0.03298 0.01163 -2.84 0.00456 \*\*
## Year2003 -0.04591 0.01171 -3.92 8.9e-05 \*\*\*
## Year2004 -0.02987 0.01160 -2.58 0.01000 \*
## Year2005 -0.03429 0.01148 -2.99 0.00283 \*\*
## Year2006 -0.01123 0.01164 -0.96 0.33459
## Year2007 -0.02480 0.01132 -2.19 0.02841 \*
## Year2008 -0.01334 0.01183 -1.13 0.25970
## Year2009 -0.00234 0.01194 -0.20 0.84444
## Year2010 0.00171 0.01204 0.14 0.88674
## Year2011 -0.03150 0.01220 -2.58 0.00982 \*\*
## Year2012 -0.03286 0.01223 -2.69 0.00722 \*\*
## Year2013 -0.04296 0.01287 -3.34 0.00084 \*\*\*
## Year2014 -0.00438 0.01375 -0.32 0.75003
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.351
## Multiple R-squared: 0.00252, Adjusted R-squared: 0.00209
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(39896,42670) are outliers with |weight| = 0 ( < 2.3e-06);
## 3858 weights are ~= 1. The remaining 39932 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.001 0.865 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 2.28e-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
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 43792"
## [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
## 876 808 882 909 903 859 835 900 1038 1095 1250 1258 1184 1109 1242
## 2011 2012 2013 2014
## 1201 1012 1026 1084
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 587 530 524 505 529 410 569 581 707 685 779 806 748 714 759
## 2011 2012 2013 2014
## 753 641 661 704
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 501 454 448 420 440 339 487 494 608 583 661 662 631 597 612
## 2011 2012 2013 2014
## 613 539 545 579
## [1] "Heteroscedasticity checks, confirming that there are problems with these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 200, df = 20, p-value <2e-16



##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 30, df = 1, p-value = 3e-08



## [1] "Female first author team size geometric mean: 3.47726127934999"
## [1] "Male first author team size geometric mean: 3.14730597432595"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 1e+07, p-value = 3e-12
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size geometric mean: 3.33228219806971"
## [1] "Male last author team size geometric mean: 3.21822036668838"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes and MaleTeamSizes
## W = 8e+06, 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 1.03 1 1.02
## LastAuthorFemale 1.03 1 1.02
## UniqueAuthors 1.09 4 1.01
## Year 1.11 18 1.00



## [1] "List of 1 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 17031 74549178560 4.21 2010 1315 2 3
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2588 -0.2884 0.0154 0.2901 2.9993
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05284 0.02554 41.22 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.02704 0.00965 -2.80 0.00508 \*\*
## LastAuthorFemale1 -0.01936 0.01132 -1.71 0.08731 .
## UniqueAuthors2 0.08613 0.01824 4.72 2.4e-06 \*\*\*
## UniqueAuthors3 0.15313 0.02141 7.15 9.0e-13 \*\*\*
## UniqueAuthors4 0.14430 0.01913 7.54 5.0e-14 \*\*\*
## UniqueAuthors5 0.20977 0.01802 11.64 < 2e-16 \*\*\*
## Year1997 0.02239 0.03033 0.74 0.46050
## Year1998 -0.00385 0.02929 -0.13 0.89548
## Year1999 -0.05401 0.02799 -1.93 0.05373 .
## Year2000 -0.05047 0.02872 -1.76 0.07891 .
## Year2001 -0.02280 0.02908 -0.78 0.43299
## Year2002 -0.08125 0.02871 -2.83 0.00466 \*\*
## Year2003 -0.09349 0.02870 -3.26 0.00113 \*\*
## Year2004 -0.08645 0.02738 -3.16 0.00160 \*\*
## Year2005 -0.08044 0.02861 -2.81 0.00494 \*\*
## Year2006 -0.01885 0.03032 -0.62 0.53419
## Year2007 -0.01860 0.03141 -0.59 0.55363
## Year2008 -0.06536 0.02920 -2.24 0.02520 \*
## Year2009 -0.06180 0.02743 -2.25 0.02430 \*
## Year2010 -0.05393 0.02678 -2.01 0.04403 \*
## Year2011 -0.06423 0.02785 -2.31 0.02109 \*
## Year2012 -0.10562 0.02950 -3.58 0.00035 \*\*\*
## Year2013 -0.15365 0.02978 -5.16 2.5e-07 \*\*\*
## Year2014 -0.10719 0.03064 -3.50 0.00047 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.426
## Multiple R-squared: 0.0289, Adjusted R-squared: 0.0266
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(1972,7856,8162)
## are outliers with |weight| = 0 ( < 9.8e-06);
## 866 weights are ~= 1. The remaining 9344 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.000 0.869 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 9.79e-06 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2\*Df))
## FirstAuthorFemale 1.03 1 1.01
## LastAuthorFemale 1.03 1 1.02
## Year 1.03 18 1.00



## [1] "List of 1 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 17031 74549178560 4.21 2010 1315 2 3.07
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1928 -0.2919 0.0122 0.2907 3.0688
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.161001 0.021972 52.84 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.017305 0.009722 -1.78 0.075 .
## LastAuthorFemale1 -0.017542 0.011419 -1.54 0.125
## Year1997 0.031752 0.030542 1.04 0.299
## Year1998 0.000473 0.029244 0.02 0.987
## Year1999 -0.044370 0.028027 -1.58 0.113
## Year2000 -0.034003 0.028854 -1.18 0.239
## Year2001 0.001270 0.029340 0.04 0.965
## Year2002 -0.052844 0.028881 -1.83 0.067 .
## Year2003 -0.066833 0.028925 -2.31 0.021 \*
## Year2004 -0.057887 0.027578 -2.10 0.036 \*
## Year2005 -0.048449 0.028551 -1.70 0.090 .
## Year2006 0.012783 0.030172 0.42 0.672
## Year2007 0.004515 0.031617 0.14 0.886
## Year2008 -0.031657 0.029456 -1.07 0.283
## Year2009 -0.027614 0.027470 -1.01 0.315
## Year2010 -0.021777 0.026826 -0.81 0.417
## Year2011 -0.031526 0.027927 -1.13 0.259
## Year2012 -0.075186 0.029677 -2.53 0.011 \*
## Year2013 -0.128371 0.030183 -4.25 2.1e-05 \*\*\*
## Year2014 -0.070841 0.030874 -2.29 0.022 \*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared: 0.00761, Adjusted R-squared: 0.00566
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(1972,7856,8162)
## are outliers with |weight| = 0 ( < 9.8e-06);
## 884 weights are ~= 1. The remaining 9326 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.009 0.868 0.949 0.900 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 9.79e-06 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.01 18 1.00



## [1] "List of 1 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 17031 74549178560 4.21 2010 1315 2 3.07
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1910 -0.2923 0.0116 0.2920 3.0710
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.158699 0.021950 52.79 < 2e-16 \*\*\*
## FirstAuthorFemale1 -0.019110 0.009652 -1.98 0.048 \*
## Year1997 0.032301 0.030551 1.06 0.290
## Year1998 0.000349 0.029264 0.01 0.990
## Year1999 -0.044082 0.028032 -1.57 0.116
## Year2000 -0.034097 0.028846 -1.18 0.237
## Year2001 0.001117 0.029349 0.04 0.970
## Year2002 -0.053053 0.028867 -1.84 0.066 .
## Year2003 -0.066584 0.028943 -2.30 0.021 \*
## Year2004 -0.058302 0.027578 -2.11 0.035 \*
## Year2005 -0.048765 0.028563 -1.71 0.088 .
## Year2006 0.012289 0.030168 0.41 0.684
## Year2007 0.003709 0.031568 0.12 0.906
## Year2008 -0.032328 0.029441 -1.10 0.272
## Year2009 -0.028405 0.027472 -1.03 0.301
## Year2010 -0.021730 0.026846 -0.81 0.418
## Year2011 -0.031923 0.027936 -1.14 0.253
## Year2012 -0.075853 0.029699 -2.55 0.011 \*
## Year2013 -0.129305 0.030176 -4.29 1.8e-05 \*\*\*
## Year2014 -0.071555 0.030881 -2.32 0.021 \*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared: 0.00738, Adjusted R-squared: 0.00553
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(1972,7856,8162)
## are outliers with |weight| = 0 ( < 9.8e-06);
## 878 weights are ~= 1. The remaining 9332 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.009 0.867 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 9.79e-06 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## 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 1.01
## Year 1.02 18 1.00



## [1] "List of 1 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 17031 74549178560 4.21 2010 1315 2 3.07
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered)
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1881 -0.2929 0.0113 0.2911 3.0728
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.156750 0.021804 53.05 < 2e-16 \*\*\*
## LastAuthorFemale1 -0.020046 0.011341 -1.77 0.077 .
## Year1997 0.031328 0.030562 1.03 0.305
## Year1998 0.000690 0.029237 0.02 0.981
## Year1999 -0.044385 0.028050 -1.58 0.114
## Year2000 -0.035318 0.028852 -1.22 0.221
## Year2001 0.000465 0.029357 0.02 0.987
## Year2002 -0.052839 0.028882 -1.83 0.067 .
## Year2003 -0.066932 0.028933 -2.31 0.021 \*
## Year2004 -0.057932 0.027571 -2.10 0.036 \*
## Year2005 -0.048817 0.028562 -1.71 0.087 .
## Year2006 0.011907 0.030166 0.39 0.693
## Year2007 0.003522 0.031606 0.11 0.911
## Year2008 -0.032833 0.029462 -1.11 0.265
## Year2009 -0.027828 0.027480 -1.01 0.311
## Year2010 -0.021504 0.026827 -0.80 0.423
## Year2011 -0.031463 0.027926 -1.13 0.260
## Year2012 -0.075351 0.029679 -2.54 0.011 \*
## Year2013 -0.129145 0.030161 -4.28 1.9e-05 \*\*\*
## Year2014 -0.071478 0.030892 -2.31 0.021 \*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared: 0.00728, Adjusted R-squared: 0.00543
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(1972,7856,8162)
## are outliers with |weight| = 0 ( < 9.8e-06);
## 882 weights are ~= 1. The remaining 9328 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.009 0.868 0.949 0.900 0.985 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 9.79e-06 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max
## 500 50 2 1 200
## maxit.scale trace.lev mts compute.rd fast.s.large.n
## 200 0 1000 0 2000
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 10213"

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(BasicStats)

## US only Gend 1st nth Fem1 96 Fem1 18 Chg FemN 96 FemN 18 Chg
## 1300 107551 55.2 28.2 41.4 13.19 16.8 28.2 11.39
## 1301 1316 55.2 NA 33.9 NA NA 29.6 NA
## 1302 9349 63.6 43.9 45.3 1.38 35.7 37.2 1.42
## 1303 267058 50.0 28.8 35.8 7.03 15.1 23.2 8.17
## 1304 62758 49.3 24.0 28.6 4.55 13.9 20.2 6.30
## 1305 61594 45.5 25.8 32.1 6.35 14.4 22.9 8.51
## 1306 102014 54.0 30.2 45.1 14.83 18.5 33.3 14.82
## 1307 164377 50.1 35.0 43.4 8.48 18.9 27.7 8.83
## 1308 45810 53.4 28.0 44.0 16.00 20.9 30.5 9.57
## 1309 40001 58.3 39.7 51.6 11.91 20.6 37.5 16.88
## 1310 47024 54.1 37.7 56.8 19.19 22.6 41.3 18.71
## 1311 141702 49.0 31.8 44.7 12.97 19.7 31.8 12.16
## 1312 240060 50.8 33.3 41.9 8.61 18.0 28.0 9.97
## 1313 60172 51.5 22.5 40.2 17.72 15.3 24.7 9.34
## 1314 93151 56.2 24.6 44.9 20.29 17.0 28.5 11.50
## 1315 22802 52.7 25.9 32.4 6.41 16.4 21.1 4.76

print(GenderTeamSize)

## Articles FirstF FirstM FirstP LastF LastM LastP
## 1300 39625 3.64 3.19 1.91e-65 3.35 3.35 8.38e-01
## 1301 302 2.66 2.79 4.98e-01 2.67 2.77 8.11e-01
## 1302 4659 3.36 3.27 4.13e-01 3.09 3.42 7.86e-07
## 1303 115286 3.76 3.48 1.89e-83 3.61 3.55 3.42e-04
## 1304 25738 3.37 3.07 4.22e-24 3.12 3.16 1.03e-01
## 1305 22280 3.46 3.09 2.17e-29 3.14 3.21 1.04e-01
## 1306 44464 5.35 5.12 1.10e-02 5.01 5.27 1.66e-14
## 1307 70953 3.95 3.74 4.76e-18 3.73 3.84 5.44e-08
## 1308 21110 4.10 4.11 8.73e-01 3.94 4.15 1.35e-05
## 1309 19795 3.39 3.19 1.02e-07 3.22 3.30 6.86e-03
## 1310 21193 3.98 3.57 1.10e-23 3.70 3.77 4.75e-02
## 1311 55777 3.62 3.24 1.28e-64 3.40 3.37 1.76e-01
## 1312 102834 3.79 3.63 2.82e-20 3.65 3.70 6.27e-03
## 1313 24877 4.79 4.86 2.20e-02 4.73 4.86 8.72e-03
## 1314 43792 3.55 3.32 3.20e-17 3.26 3.42 1.05e-09
## 1315 10213 3.48 3.15 3.44e-12 3.33 3.22 1.12e-02

print(RegCoef)

## FFA1 FLA1 2 3 4 5+ FFA2 FLA2
## 1300 -0.04348 -0.04352 0.2484 0.3744 0.381 0.530 -0.012167 -0.049187
## 1301 0.15062 -0.21860 0.3833 0.5348 0.303 0.609 0.161419 -0.266664
## 1302 0.00061 0.02537 0.0988 0.1896 0.211 0.286 0.005520 0.012093
## 1303 -0.01197 -0.01496 0.1667 0.2224 0.214 0.285 -0.002849 -0.014646
## 1304 -0.01949 -0.02703 0.1794 0.2870 0.252 0.292 -0.004032 -0.027143
## 1305 -0.01414 -0.01540 0.2898 0.4618 0.398 0.502 0.008620 -0.023027
## 1306 -0.01176 -0.01417 0.2580 0.3727 0.355 0.494 -0.001156 -0.023094
## 1307 -0.01595 -0.02344 0.1893 0.2490 0.254 0.349 -0.008653 -0.028955
## 1308 0.01965 0.00570 0.1677 0.2444 0.262 0.345 0.023301 -0.000929
## 1309 -0.01022 -0.00881 0.0764 0.0773 0.113 0.217 -0.006325 -0.014882
## 1310 0.02134 0.00987 0.0397 0.1336 0.135 0.222 0.029287 0.002257
## 1311 -0.03180 -0.02925 0.0975 0.1556 0.142 0.239 -0.020859 -0.031423
## 1312 -0.01240 -0.01685 0.1391 0.1996 0.196 0.272 -0.007271 -0.019159
## 1313 0.00407 -0.01402 0.2460 0.3147 0.319 0.386 0.008421 -0.015838
## 1314 -0.00382 -0.00755 0.0793 0.0956 0.122 0.202 0.000118 -0.014289
## 1315 -0.02704 -0.01936 0.0861 0.1531 0.144 0.210 -0.017305 -0.017542
## FFA3 FLA4
## 1300 -0.01847 -0.05120
## 1301 0.10534 -0.21936
## 1302 0.00751 0.01311
## 1303 -0.00399 -0.01498
## 1304 -0.00625 -0.02764
## 1305 0.00581 -0.02207
## 1306 -0.00390 -0.02327
## 1307 -0.01089 -0.02989
## 1308 0.02322 0.00169
## 1309 -0.00805 -0.01578
## 1310 0.02960 0.00709
## 1311 -0.02467 -0.03474
## 1312 -0.00896 -0.02005
## 1313 0.00754 -0.01518
## 1314 -0.00131 -0.01427
## 1315 -0.01911 -0.02005

print(RegP)

## FFA1p FLA1p 2p 3p 4p 5+p FFA2p
## 1300 4.28e-13 7.81e-11 8.45e-99 4.29e-153 6.88e-200 0.00e+00 4.87e-02
## 1301 7.42e-02 2.50e-02 8.00e-05 9.17e-03 6.32e-02 7.02e-10 5.00e-02
## 1302 9.59e-01 4.78e-02 2.92e-05 3.47e-11 2.03e-17 1.03e-38 6.53e-01
## 1303 2.60e-07 1.64e-07 3.05e-121 9.52e-181 3.72e-191 0.00e+00 2.26e-01
## 1304 1.05e-03 2.59e-04 5.58e-41 2.09e-76 5.59e-73 2.39e-104 5.01e-01
## 1305 6.42e-02 8.37e-02 2.34e-71 1.06e-142 1.34e-121 1.75e-219 2.73e-01
## 1306 3.57e-03 2.17e-03 2.06e-65 1.12e-122 2.00e-129 7.85e-281 7.80e-01
## 1307 1.10e-08 1.16e-12 4.25e-93 1.58e-146 8.42e-166 0.00e+00 2.35e-03
## 1308 1.07e-03 4.13e-01 6.72e-24 7.48e-40 6.48e-56 4.13e-108 1.39e-04
## 1309 4.83e-02 1.22e-01 6.94e-13 7.61e-10 2.53e-23 1.75e-91 2.29e-01
## 1310 9.62e-05 9.97e-02 2.27e-03 3.81e-21 3.50e-25 6.85e-75 1.52e-07
## 1311 3.20e-17 8.60e-12 2.61e-38 1.57e-70 4.10e-69 1.02e-235 4.34e-08
## 1312 1.07e-07 8.31e-10 1.25e-93 7.18e-162 1.02e-177 0.00e+00 2.15e-03
## 1313 4.59e-01 3.19e-02 8.27e-41 1.80e-51 1.15e-66 6.76e-109 1.29e-01
## 1314 3.08e-01 9.17e-02 1.09e-24 9.91e-32 8.09e-54 4.22e-164 9.75e-01
## 1315 5.08e-03 8.73e-02 2.36e-06 9.02e-13 5.01e-14 3.99e-31 7.51e-02
## FLA2p FFA3p FLA4p
## 1300 2.00e-12 2.85e-03 2.75e-13
## 1301 1.00e-02 2.07e-01 2.91e-02
## 1302 3.56e-01 5.40e-01 3.15e-01
## 1303 3.91e-07 8.90e-02 2.02e-07
## 1304 3.04e-04 2.97e-01 2.36e-04
## 1305 1.31e-02 4.60e-01 1.71e-02
## 1306 1.12e-06 3.43e-01 8.17e-07
## 1307 4.83e-18 1.26e-04 3.61e-19
## 1308 8.97e-01 1.48e-04 8.14e-01
## 1309 1.01e-02 1.23e-01 6.04e-03
## 1310 7.10e-01 8.80e-08 2.39e-01
## 1311 4.11e-13 7.72e-11 8.34e-16
## 1312 6.02e-12 1.50e-04 5.30e-13
## 1313 1.70e-02 1.75e-01 2.23e-02
## 1314 1.66e-03 7.28e-01 1.57e-03
## 1315 1.25e-01 4.77e-02 7.72e-02

print(RegStar)

## FFA1p FLA1p 2p 3p 4p 5+p FFA2p FLA2p FFA3p FLA4p
## 1300 "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*" "\*\*\*" "\*\*" "\*\*\*"
## 1301 "" "\*" "\*\*\*" "\*\*" "" "\*\*\*" "\*" "\*" "" "\*"
## 1302 "" "\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "" "" "" ""
## 1303 "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "" "\*\*\*" "" "\*\*\*"
## 1304 "\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "" "\*\*\*" "" "\*\*\*"
## 1305 "" "" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "" "\*" "" "\*"
## 1306 "\*\*" "\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "" "\*\*\*" "" "\*\*\*"
## 1307 "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*" "\*\*\*" "\*\*\*" "\*\*\*"
## 1308 "\*\*" "" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "" "\*\*\*" ""
## 1309 "\*" "" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "" "\*" "" "\*\*"
## 1310 "\*\*\*" "" "\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "" "\*\*\*" ""
## 1311 "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*"
## 1312 "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*" "\*\*\*" "\*\*\*" "\*\*\*"
## 1313 "" "\*" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "" "\*" "" "\*"
## 1314 "" "" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "" "\*\*" "" "\*\*"
## 1315 "\*\*" "" "\*\*\*" "\*\*\*" "\*\*\*" "\*\*\*" "" "" "\*" ""

print(GenderAnalysed)

## CitationSet Female1st FemaleLast
## 1300 39625 14322 9381
## 1301 302 85 58
## 1302 4659 2075 1517
## 1303 115286 37116 20977
## 1304 25738 7397 4256
## 1305 22280 6780 4288
## 1306 44464 17218 10622
## 1307 70953 27748 15762
## 1308 21110 7172 4662
## 1309 19795 8609 5437
## 1310 21193 9423 6013
## 1311 55777 20468 12840
## 1312 102834 37905 21853
## 1313 24877 7577 4622
## 1314 43792 13788 8181
## 1315 10213 2904 1779