RAproject

Jennifer and Henri

2022-12-03

Questions: is there a difference in inflation and salary inflation?

#Description

The 2008-09 nine-month academic salary for Assistant Professors, Associate Professors and Professors in a college in the U.S. The data were collected as part of the on-going effort of the college's administration to monitor salary differences between male and female faculty members.

#Descriptive Stats

```
library(carData)
#View(Salaries)

#Rank  Prof, AssocPRoc, Asst
#A- theoretical B - Applies
#years since phd
#years of service
#gender
#salary = updated to salary in 2022
head(Salaries)
```

```
##        rank discipline yrs.since.phd yrs.service  sex salary
## 1      Prof          B            19          18 Male 139750
## 2      Prof          B            20          16 Male 173200
## 3  AsstProf          B             4           3 Male  79750
## 4      Prof          B            45          39 Male 115000
## 5      Prof          B            40          41 Male 141500
## 6 AssocProf          B             6           6 Male  97000
```

```
summary(Salaries$rank)
```

```
##  AsstProf AssocProf      Prof
##        67        64       266
```

```
summary(Salaries$discipline)
```

```
##   A   B
## 181 216
```

```
summary(Salaries$yrs.since.phd)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   12.00   21.00   22.31   32.00   56.00
```

```
summary(Salaries$yrs.service)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    7.00   16.00   17.61   27.00   60.00
```

```
summary(Salaries$sex)
```

```
## Female   Male
##     39    358
```

Now lets adjust for inflation

```
#Accounted for inflation since 2009
#AIER.org - cost of living calculator
Salaries[,6] <- Salaries[,6]*1.35


summary(Salaries$salary)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   78030  122850  144855  153504  181150  312586
```

```
head(Salaries)
```

```
##        rank discipline yrs.since.phd yrs.service  sex   salary
## 1      Prof          B            19          18 Male 188662.5
## 2      Prof          B            20          16 Male 233820.0
## 3  AsstProf          B             4           3 Male 107662.5
## 4      Prof          B            45          39 Male 155250.0
## 5      Prof          B            40          41 Male 191025.0
## 6 AssocProf          B             6           6 Male 130950.0
```

Now making full model with no interaction

```
full_model <- lm(salary~. , data=Salaries)
summary(full_model)
```

```
##
## Call:
## lm(formula = salary ~ ., data = Salaries)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -88085 -17835  -2396  14018 134449
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    89039.6     6194.6  14.374  < 2e-16 ***
## rankAssocProf  17425.2     5596.1   3.114  0.00198 **
## rankProf       60839.1     5720.7  10.635  < 2e-16 ***
## disciplineB    19463.8     3162.9   6.154 1.88e-09 ***
## yrs.since.phd    722.3      325.3   2.220  0.02698 *
## yrs.service     -660.8      286.1  -2.310  0.02143 *
## sexMale         6457.7     5209.2   1.240  0.21584
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30430 on 390 degrees of freedom
## Multiple R-squared:  0.4547, Adjusted R-squared:  0.4463
## F-statistic:  54.2 on 6 and 390 DF,  p-value: < 2.2e-16
```

Lets check what is the best model

```
olsrr_all <- olsrr::ols_step_all_possible(full_model)
#Best based off of R squared
(olsrr_all[which.max(olsrr_all$rsquare),])
```

```
##    Index N                                    Predictors  R-Square
## 31    31 5 rank discipline yrs.since.phd yrs.service sex 0.4546766
##    Adj. R-Square Mallow's Cp
## 31      0.446287           5
```

```
#Best based off of Adj R squared
(olsrr_all[which.max(olsrr_all$adjr),])
```

```
##    Index N                                    Predictors  R-Square
## 31    31 5 rank discipline yrs.since.phd yrs.service sex 0.4546766
##    Adj. R-Square Mallow's Cp
## 31      0.446287           5
```

```
#Best based off of Mallows CP
(olsrr_all[which.min(olsrr_all$cp),])
```

```
##    Index N                                 Predictors  R-Square Adj. R-Square
## 26    26 4 rank discipline yrs.since.phd yrs.service 0.4525278     0.4455269
##    Mallow's Cp
## 26    4.536793
```

```
olsrr::ols_step_both_aic(full_model)
```

```
##
##
##                                        Stepwise Summary
## -------------------------------------------------------------------------------------------------
## Variable       Method       AIC             RSS               Sum Sq          R-Sq      Adj. R-Sq
```

```
## ------------------------------------------------------------------------------------------------
## rank          addition    9365.828    401075528012.785    261039893053.842    0.39425      0.39118
## discipline    addition    9333.106          3.67487e+11    294628440452.807    0.44498      0.44074
## ------------------------------------------------------------------------------------------------
```

confirmed that full model is the best model according to adj r squared and we don't want to take out sex because that is important factor

#Now recoding Salaries so that it includes everything as factor

```
Salaries2 <- Salaries
Salaries2 <- cbind(Salaries2, Salaries$rank)

library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
Salaries2$rank <- recode_factor(Salaries2$rank, "AssocProf" = 0, "AsstProf"= 1,"Prof"= 0)
Salaries2$`Salaries$rank` <-  recode_factor(Salaries2$`Salaries$rank`, "AssocProf" = 0, "Prof"= 1,"Asstl
Salaries2$sex <- recode_factor(Salaries2$sex, "Male"=0, "Female" = 1)
Salaries2$discipline <- recode_factor(Salaries2$discipline, "A"=0, "B" = 1)

Salaries2 <- Salaries2[, c(1,7,2,3,4,5,6)]

names(Salaries2)[2] <- "FRank"

head(Salaries2)
```

```
##   rank FRank discipline yrs.since.phd yrs.service sex   salary
## 1    0     1          1            19          18   0 188662.5
## 2    0     1          1            20          16   0 233820.0
## 3    1     0          1             4           3   0 107662.5
## 4    0     1          1            45          39   0 155250.0
## 5    0     1          1            40          41   0 191025.0
## 6    0     0          1             6           6   0 130950.0
```

#initial model

This model will include an interaction term

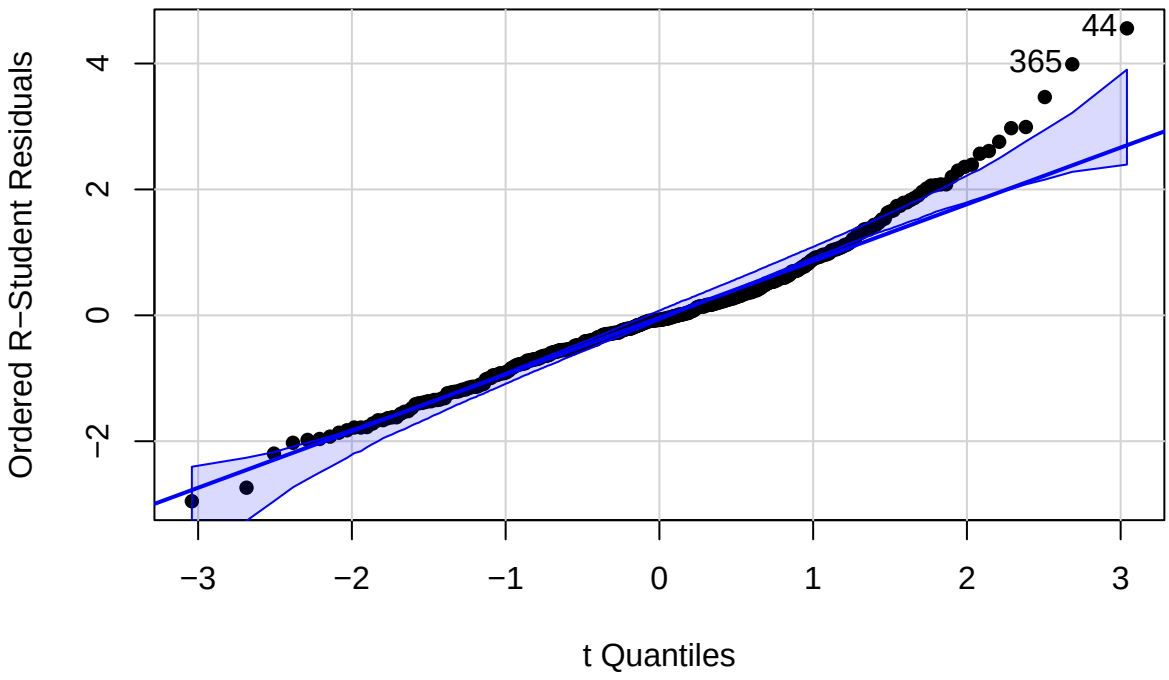
```
full <- lm(salary~ discipline*rank + discipline*Salaries2$FRank +yrs.service + sex + yrs.since.phd, data
summary(full)
```

```
##
## Call:
## lm(formula = salary ~ discipline * rank + discipline * Salaries2$FRank +
##     yrs.service + sex + yrs.since.phd, data = Salaries2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -88114 -17495  -2172  14281 134482
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   109055.9     6862.2  15.892  < 2e-16 ***
## discipline1                    25843.7     7792.0   3.317 0.000997 ***
## rank1                         -10102.3     8879.3  -1.138 0.255929
## Salaries2$FRank1               47169.7     6905.7   6.831 3.28e-11 ***
## yrs.service                     -666.0      286.5  -2.325 0.020598 *
## sex1                           -6775.2     5233.8  -1.295 0.196256
## yrs.since.phd                    734.6      326.0   2.253 0.024798 *
## discipline1:rank1             -11762.6    11005.8  -1.069 0.285841
## discipline1:Salaries2$FRank1   -6572.4     8606.4  -0.764 0.445528
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30460 on 388 degrees of freedom
## Multiple R-squared:  0.4563, Adjusted R-squared:  0.4451
## F-statistic:  40.7 on 8 and 388 DF,  p-value: < 2.2e-16
```

no scaling needed

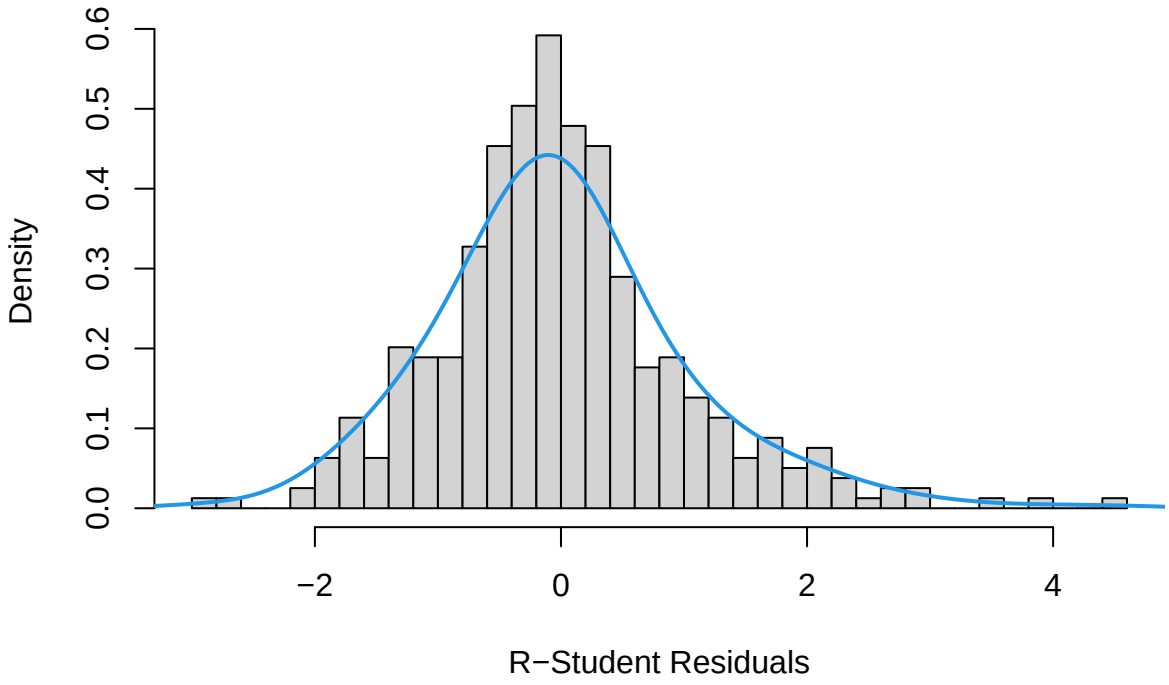
#Altering first model

```
#NORMALITY
car::qqPlot(full, id = TRUE, col.lines = "blue",
            reps = 1000, ylab = "Ordered R-Student Residuals", pch = 16)
```




```
## [1]  44 365
```

```
r_stud <- rstudent(full)

hist(r_stud, prob = TRUE, breaks = 40, xlab = "R-Student Residuals", main = "")
lines(density(r_stud, adjust = 2), col = 4, lwd = 2)
```



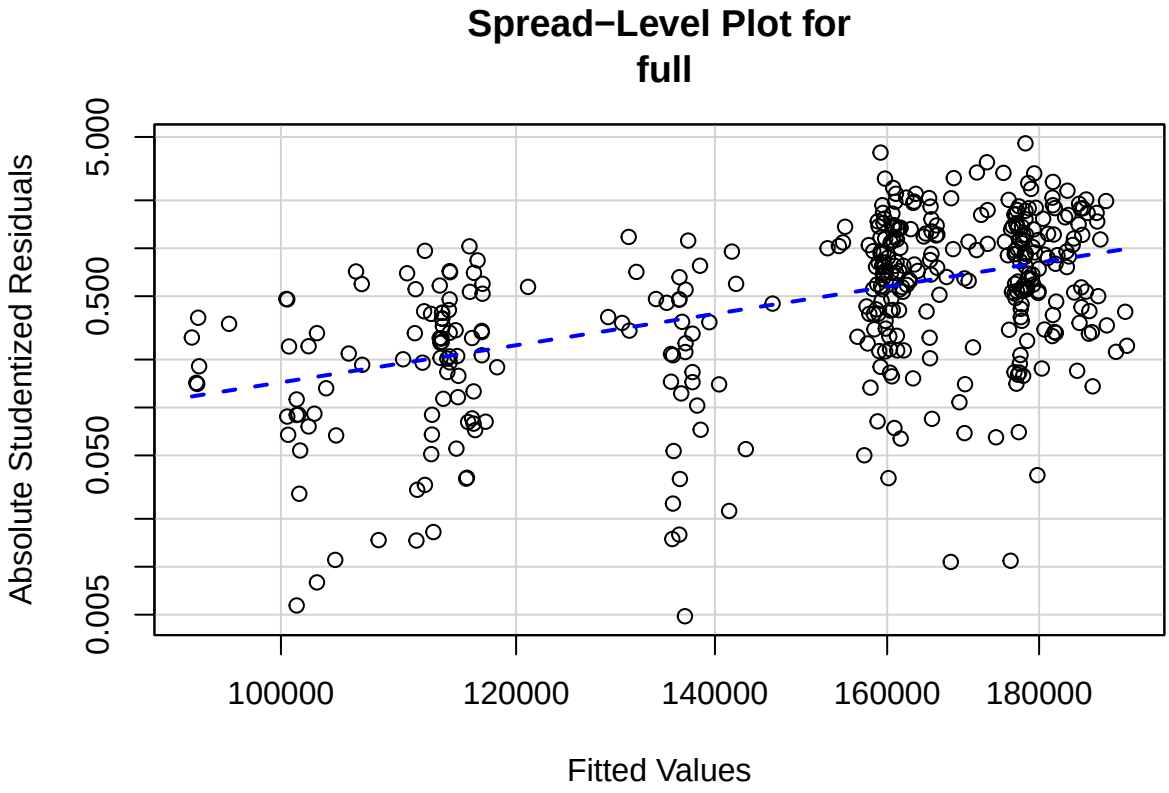

```
#could not do boxplot for best mod
carPT<-car::powerTransform(full, family = "bcPower")
summary(carPT)
```

```
## bcPower Transformation to Normality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1   -0.9056          -1       -1.215      -0.5961
##
## Likelihood ratio test that transformation parameter is equal to 0
##  (log transformation)
##                            LRT df       pval
## LR test, lambda = (0) 31.70877  1 1.7911e-08
##
## Likelihood ratio test that no transformation is needed
##                            LRT df       pval
## LR test, lambda = (1) 134.0308  1 < 2.22e-16
```

```
#no transformation
```

In summary for normality we do not need to adjust anything

```
#NON CONSTANT VARIANCE
car::spreadLevelPlot(full, smooth = FALSE)
```




```
##
## Suggested power transformation:  -1.944869
```

```
#slope is not close to one
```

```
car::ncvTest(full)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 67.89024, Df = 1, p = < 2.22e-16
```

```
#therefore not a constant variance
```

We infact do need to change our model! for a lambda > 1 the suggestion is -2 but -3 produces better results

Now checking and furthering results with transformed model

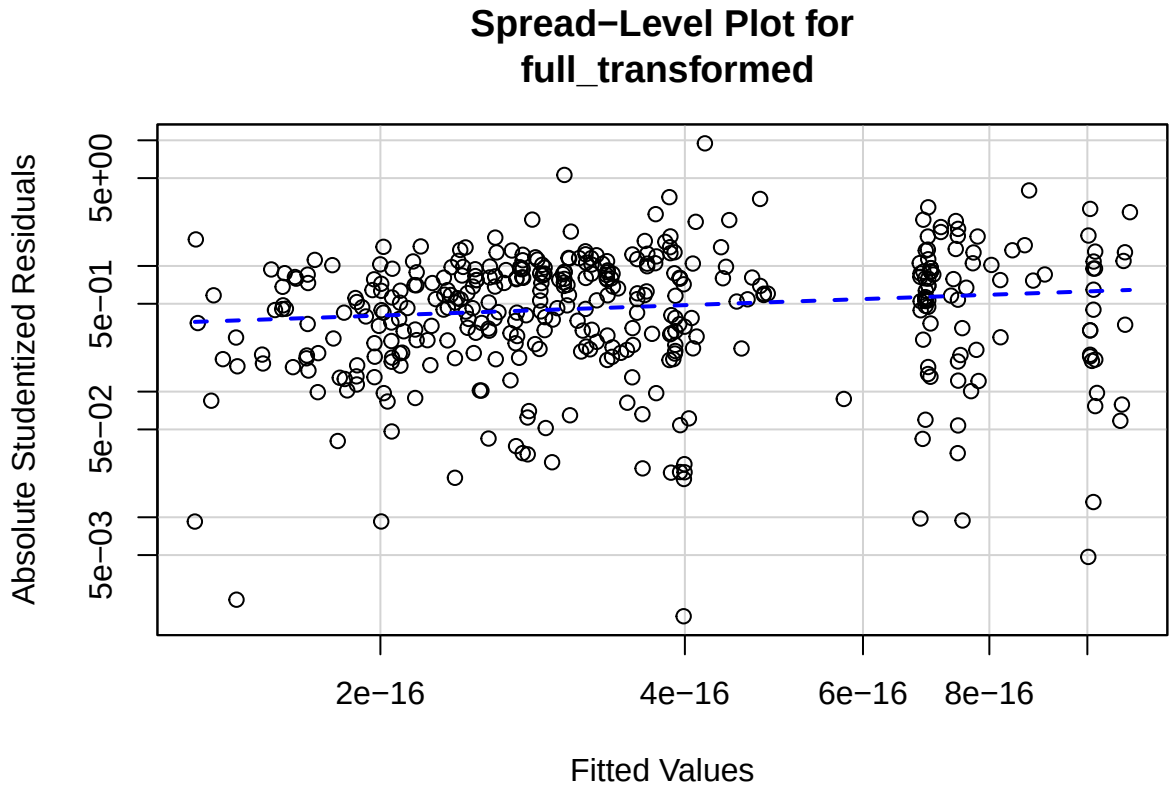
```
full_transformed <- lm((salary)^(-3)~ discipline*rank + discipline*Salaries2$FRank +yrs.service + sex +
summary(full_transformed)
```

```
##
## Call:
## lm(formula = (salary)^(-3) ~ discipline * rank + discipline *
##     Salaries2$FRank + yrs.service + sex + yrs.since.phd, data = Salaries2)
```

```
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -3.975e-16 -1.257e-16 -2.633e-17  7.622e-17  1.686e-15
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   7.139e-16  4.508e-17  15.836  < 2e-16 ***
## discipline1                  -3.497e-16  5.119e-17  -6.832 3.25e-11 ***
## rank1                         2.889e-16  5.833e-17   4.954 1.09e-06 ***
## Salaries2$FRank1             -4.968e-16  4.536e-17 -10.952  < 2e-16 ***
## yrs.service                   4.102e-18  1.882e-18   2.180   0.0299 *
## sex1                          7.544e-17  3.438e-17   2.194   0.0288 *
## yrs.since.phd                -1.481e-19  2.142e-18  -0.069   0.9449
## discipline1:rank1             3.132e-17  7.230e-17   0.433   0.6651
## discipline1:Salaries2$FRank1  2.590e-16  5.654e-17   4.581 6.25e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.001e-16 on 388 degrees of freedom
## Multiple R-squared:  0.6019, Adjusted R-squared:  0.5937
## F-statistic: 73.32 on 8 and 388 DF,  p-value: < 2.2e-16
```

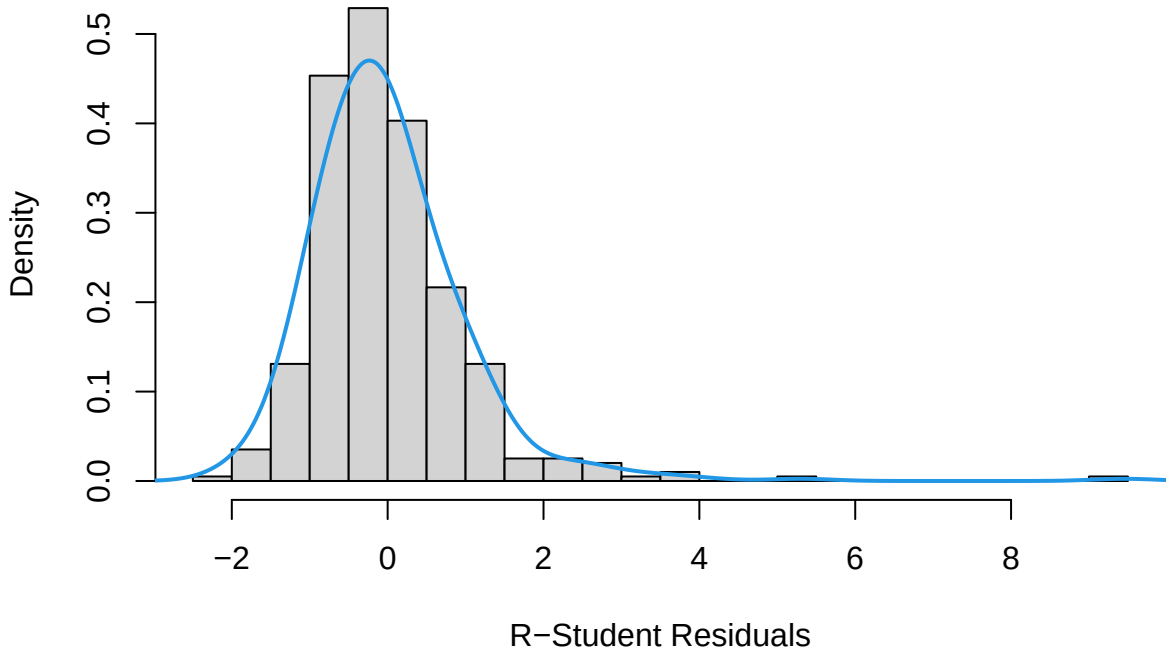
Checking to make sure what we already checked is still good

```
car::spreadLevelPlot(full_transformed, smooth = FALSE)
```



```
##
## Suggested power transformation:  0.725089
```

```
r_stud <- rstudent(full_transformed)

hist(r_stud, prob = TRUE, breaks = 40, xlab = "R-Student Residuals", main = "")
lines(density(r_stud, adjust = 2), col = 4, lwd = 2)
```



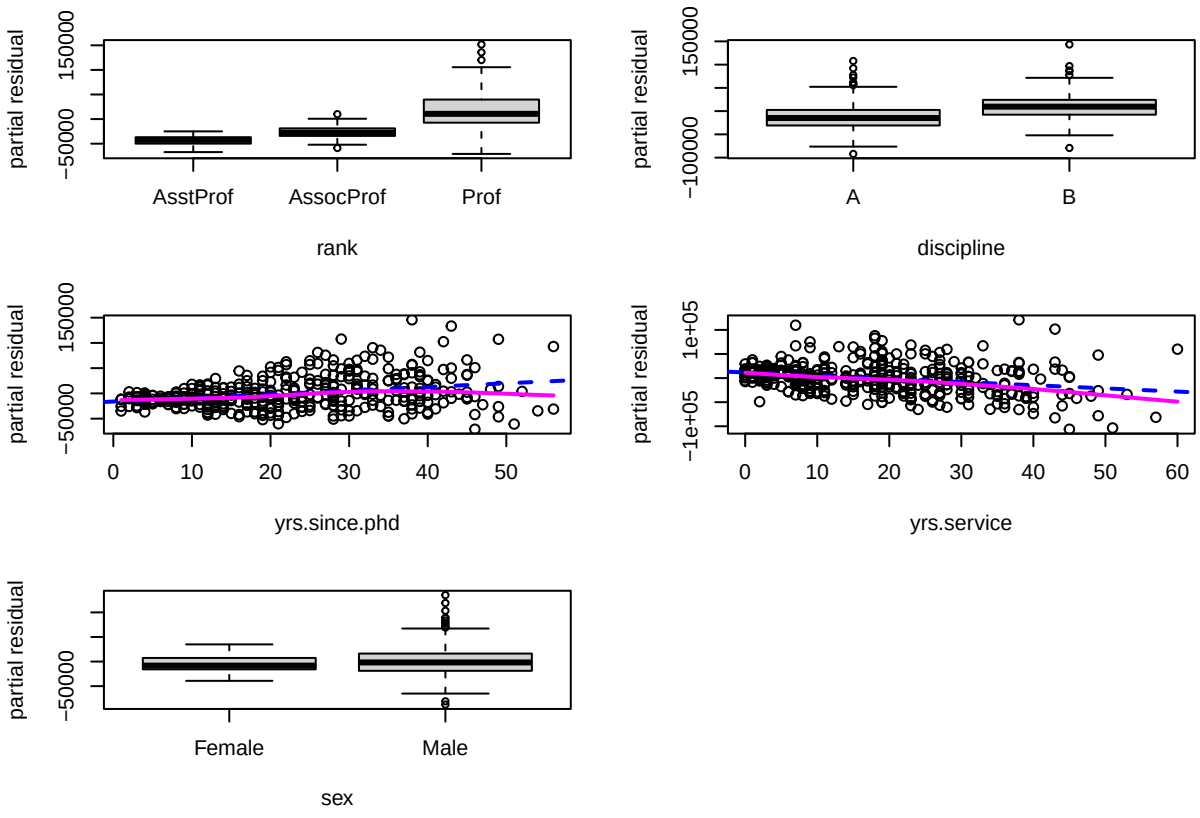

```
carPT<-car::powerTransform(full, family = "bcPower")
summary(carPT)
```

```
## bcPower Transformation to Normality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1   -0.9056          -1       -1.215      -0.5961
##
## Likelihood ratio test that transformation parameter is equal to 0
##  (log transformation)
##                            LRT df       pval
## LR test, lambda = (0) 31.70877  1 1.7911e-08
##
## Likelihood ratio test that no transformation is needed
##                            LRT df       pval
## LR test, lambda = (1) 134.0308  1 < 2.22e-16
```

```
#no transformation
```

and we are still good so lets continue

```
#NON LINEARITY
#we can only check this using a model with no interaction...
#so we do that and we are looking good as expected
car::crPlots(full_model, ylab = "partial residual", grid = FALSE, main = "")
```




```
#car::crPlots(full_transformed, ylab = "partial residual", grid = FALSE, main = "")
```

```
#Collinearity

(vif_all <- car::vif(full_transformed))
```

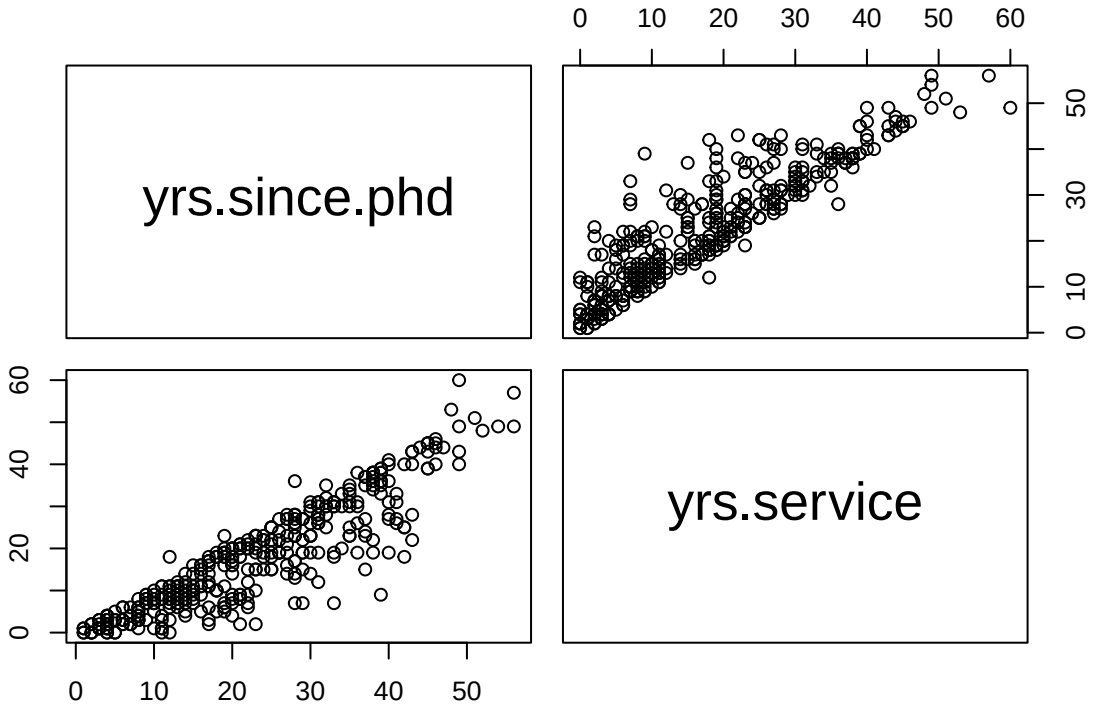
```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif

##                 discipline                       rank
##                   6.444235                   4.732461
##            Salaries2$FRank                yrs.service
##                   4.511405                   5.924700
##                        sex              yrs.since.phd
##                   1.038293                   7.533387
```

```
##             discipline:rank discipline:Salaries2$FRank
##                    5.005614                    7.112481
```

```
#none are greater than 10 so we are good!
```

Checking for multicollinearity

```
pairs(Salaries2[,c(-1, -2, -3, -6, -7)])
```



as expected yrs.since.phd and yrs.service are extremely correlated. Only way they wouldn't be is as if a professor changed universities So we will delete yrs.since.phd because yrs.service at a specific university will explain raises which explains salary

```
full_transformed_No_years <- lm((salary)^(-3)~ discipline*rank + discipline*FRank +yrs.service + sex , d
summary(full_transformed_No_years)
```

```
##
## Call:
## lm(formula = (salary)^(-3) ~ discipline * rank + discipline *
##     FRank + yrs.service + sex, data = Salaries2)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -3.977e-16 -1.253e-16 -2.742e-17  7.635e-17  1.686e-15
##
```

```
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          7.127e-16  4.187e-17  17.021  < 2e-16 ***
## discipline1         -3.494e-16  5.093e-17  -6.861 2.70e-11 ***
## rank1                2.895e-16  5.766e-17   5.022 7.82e-07 ***
## FRank1              -4.975e-16  4.429e-17 -11.233  < 2e-16 ***
## yrs.service          3.991e-18  9.860e-19   4.047 6.25e-05 ***
## sex1                 7.543e-17  3.434e-17   2.197   0.0286 *
## discipline1:rank1    3.113e-17  7.215e-17   0.431   0.6664
## discipline1:FRank1   2.590e-16  5.646e-17   4.586 6.09e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.998e-16 on 389 degrees of freedom
## Multiple R-squared:  0.6019, Adjusted R-squared:  0.5947
## F-statistic:    84 on 7 and 389 DF,  p-value: < 2.2e-16
```

```
#taking out years since phd
```

```
#Collinearity

(vif_all <- car::vif(full_transformed_No_years))
```

```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif

##       discipline             rank            FRank      yrs.service
##         6.395347         4.635618         4.311052         1.630687
##              sex  discipline:rank discipline:FRank
##         1.038267         4.998558         7.112193
```

```
#non are greater than 10 so we are good!
```

#Now checking unusual data

```
#all influential points
summary(influence.measures(full_transformed_No_years))
```

```
## Potentially influential observations of
##   lm(formula = (salary)^(-3) ~ discipline * rank + discipline *      FRank + yrs.service + sex, data
##
##     dfb.1_ dfb.dsc1 dfb.rnk1 dfb.FRn1 dfb.yrs. dfb.sex1 dfb.d1:1 dfb.d1:F
## 25    0.14  -0.12    -0.12    -0.13    -0.01     0.11     0.09     0.11
## 35    0.00   0.00     0.00     0.00     0.00     0.02     0.01     0.00
## 50    0.01   0.00     0.00     0.00    -0.02    -0.05     0.15     0.00
## 64    0.01  -0.06     0.00     0.00    -0.01    -0.09     0.04     0.06
## 65    0.01   0.00     0.00    -0.01     0.00    -0.06     0.19     0.00
## 107 -0.01   0.01     0.01     0.01     0.00     0.00    -0.01    -0.01
## 108 -0.02   0.02     0.02     0.02     0.00     0.00    -0.01    -0.02
## 109  0.01   0.00     0.00    -0.01     0.00     0.00     0.00     0.00
## 113  0.00   0.00     0.05     0.00     0.00    -0.01    -0.04     0.00
## 115  0.00   0.00    -0.01     0.02    -0.03     0.05     0.00    -0.01
```

```
## 119  0.00   0.00     0.05     0.00     0.00    -0.01    -0.04     0.00
## 120  0.01   0.00    -0.05     0.00     0.00    -0.05     0.04     0.00
## 124  0.61  -0.61    -0.54    -0.73     0.22     0.61     0.47     0.56
## 128  0.00   0.00    -0.01     0.00     0.00    -0.01     0.01     0.00
## 130  0.00   0.00     0.03     0.00     0.00    -0.01    -0.02     0.00
## 131 -0.04   0.03     0.03     0.03     0.00     0.01    -0.02    -0.03
## 133  0.05  -0.04    -0.04    -0.05     0.00     0.04     0.03     0.04
## 134  0.00   0.00    -0.01     0.00     0.00    -0.01     0.01     0.00
## 142  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 154  0.01  -0.06     0.00    -0.01     0.00    -0.09     0.04     0.06
## 187  0.01  -0.06     0.00    -0.01     0.00    -0.09     0.04     0.06
## 195  0.01  -0.01     0.00     0.00    -0.02     0.00     0.01     0.01
## 227  0.03   0.00     0.42     0.00    -0.03    -0.13    -0.34     0.00
## 228  0.46  -0.36    -0.32    -0.39    -0.08    -0.07     0.25     0.32
## 232  0.11  -0.12    -0.10    -0.14     0.05     0.12     0.09     0.11
## 236  0.03   0.00    -0.01     0.10    -0.09    -0.03     0.00    -0.06
## 238 -0.07   0.00     0.39     0.01     0.08     0.36    -0.30     0.00
## 239 -0.06   0.01     0.04     0.04     0.20    -0.01    -0.01    -0.07
## 256 -0.03   0.03     0.02     0.03     0.01     0.01    -0.02    -0.02
## 259  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 273  0.00   0.00     0.03     0.00     0.00    -0.01    -0.02     0.00
## 274  0.00   0.00    -0.01     0.00     0.00     0.00     0.01     0.00
## 283 -0.40   0.06     0.24     0.03     1.25_*   0.01    -0.04    -0.25
## 286 -0.12   0.13     0.09     0.19    -0.15     0.01    -0.09    -0.12
## 290  0.00   0.00    -0.01     0.00     0.00     0.00     0.01     0.00
## 299 -0.10   0.02     0.06     0.04     0.32    -0.01    -0.01    -0.09
## 307  0.00   0.00    -0.03     0.00     0.00     0.01     0.02     0.00
## 309  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 317 -0.07   0.36    -0.03     0.03     0.01     0.50    -0.22    -0.33
## 318 -0.20   0.03     0.12    -0.15     0.62    -0.01    -0.02     0.15
## 335  0.01  -0.06     0.01    -0.01     0.01    -0.08     0.04     0.05
## 364 -0.02   0.02     0.02     0.02     0.00     0.00    -0.01    -0.02
## 368 -0.43   0.32     0.30     0.33     0.13     0.07    -0.22    -0.28
## 377  0.00   0.00    -0.03     0.00     0.00     0.01     0.02     0.00
## 381  0.00   0.00    -0.07     0.00     0.00     0.02     0.05     0.00
##     dffit   cov.r   cook.d hat
## 25   0.19    1.07_*  0.00   0.06
## 35   0.02    1.07_*  0.00   0.05
## 50   0.36    0.93_*  0.02   0.02
## 64  -0.13    1.06_*  0.00   0.05
## 65   0.46_*  0.88_*  0.03   0.02
## 107 -0.01    1.06_*  0.00   0.04
## 108 -0.02    1.06_*  0.00   0.04
## 109  0.01    1.06_*  0.00   0.04
## 113  0.07    1.07_*  0.00   0.04
## 115  0.07    1.07_*  0.00   0.04
## 119  0.07    1.07_*  0.00   0.04
## 120 -0.08    1.08_*  0.00   0.06
## 124  1.03_*  0.79_*  0.13   0.06_*
## 128 -0.01    1.08_*  0.00   0.06
## 130  0.04    1.07_*  0.00   0.04
## 131 -0.04    1.06_*  0.00   0.04
## 133  0.07    1.08_*  0.00   0.06
## 134 -0.02    1.08_*  0.00   0.06
```

```
## 142  0.00    1.06_*  0.00   0.04
## 154 -0.13    1.06_*  0.00   0.05
## 187 -0.13    1.06_*  0.00   0.05
## 195 -0.02    1.10_*  0.00   0.07_*
## 227  0.61_*  0.90_*  0.05   0.04
## 228  0.47_*  0.96    0.03   0.04
## 232  0.20    1.08_*  0.00   0.06_*
## 236  0.22    0.92_*  0.01   0.01
## 238  0.67_*  0.94_*  0.06   0.06
## 239  0.30    0.90_*  0.01   0.01
## 256 -0.04    1.06_*  0.00   0.04
## 259  0.00    1.07_*  0.00   0.04
## 273  0.04    1.07_*  0.00   0.04
## 274 -0.02    1.07_*  0.00   0.04
## 283  1.51_*  0.20_*  0.23   0.02
## 286 -0.23    1.08_*  0.01   0.07_*
## 290 -0.02    1.07_*  0.00   0.04
## 299  0.45_*  0.81_*  0.02   0.02
## 307 -0.04    1.07_*  0.00   0.04
## 309  0.00    1.07_*  0.00   0.04
## 317  0.76_*  0.84_*  0.07   0.05
## 318  0.78_*  0.60_*  0.07   0.02
## 335 -0.12    1.06_*  0.00   0.05
## 364 -0.02    1.06_*  0.00   0.04
## 368 -0.43_*  0.98    0.02   0.04
## 377 -0.04    1.07_*  0.00   0.04
## 381 -0.10    1.06_*  0.00   0.04
```

That is alot! Lets take out all of those that are flagged not by cov.r or flagged by 2 or more

```
Salaries_noIP <- Salaries2[c(-65,-124,-132,-195, -232, -238, -283, -286, -299, -317, -318, -228, -368),]
noinfluential <- lm((salary)^(-3)~ discipline*rank + discipline*FRank +yrs.service + sex , data= Salaries_noIP)
summary(noinfluential)
```

```
##
## Call:
## lm(formula = (salary)^(-3) ~ discipline * rank + discipline *
##     FRank + yrs.service + sex, data = Salaries_noIP)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -3.659e-16 -1.038e-16 -1.669e-17  7.359e-17  5.780e-16
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         7.181e-16  3.419e-17  21.006  < 2e-16 ***
## discipline1        -3.336e-16  4.115e-17  -8.106 7.44e-15 ***
## rank1               2.815e-16  4.589e-17   6.134 2.17e-09 ***
## FRank1             -4.459e-16  3.653e-17 -12.207  < 2e-16 ***
## yrs.service         8.272e-19  8.030e-19   1.030    0.304
## sex1                1.702e-17  2.695e-17   0.632    0.528
## discipline1:rank1   1.856e-17  5.669e-17   0.327    0.744
## discipline1:FRank1  2.496e-16  4.509e-17   5.535 5.85e-08 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.497e-16 on 376 degrees of freedom
## Multiple R-squared:  0.7095, Adjusted R-squared:  0.704
## F-statistic: 131.2 on 7 and 376 DF,  p-value: < 2.2e-16
```

This looks good! But is there any more influential points?

```
#all influential points
summary(influence.measures(noinfluential))
```

```
## Potentially influential observations of
##   lm(formula = (salary)^(-3) ~ discipline * rank + discipline *      FRank + yrs.service + sex, data
##
##     dfb.1_ dfb.dsc1 dfb.rnk1 dfb.FRn1 dfb.yrs. dfb.sex1 dfb.d1:1 dfb.d1:F
## 25    0.33  -0.29    -0.28    -0.31    -0.01     0.27     0.23     0.27
## 35   -0.01   0.00    -0.01     0.00     0.01     0.09     0.04     0.00
## 50    0.01   0.00     0.00     0.01    -0.03    -0.07     0.20     0.00
## 64    0.01  -0.03     0.00     0.00    -0.01    -0.05     0.02     0.03
## 76    0.00   0.00     0.00     0.00     0.00    -0.05     0.15     0.00
## 105 -0.06   0.04     0.04     0.05     0.00     0.00    -0.03    -0.04
## 107  0.01  -0.01    -0.01    -0.01     0.00     0.00     0.01     0.01
## 108 -0.01   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 109  0.04  -0.03    -0.03    -0.03     0.00     0.00     0.02     0.03
## 112  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 113  0.00   0.00     0.07     0.00     0.00    -0.02    -0.05     0.00
## 115  0.01   0.00    -0.01     0.03    -0.05     0.07     0.01    -0.01
## 119  0.00   0.00     0.07     0.00     0.00    -0.02    -0.05     0.00
## 120  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 126 -0.14   0.02     0.08    -0.01     0.50     0.01    -0.01    -0.09
## 128  0.00   0.00     0.05     0.00     0.00     0.05    -0.04     0.00
## 130  0.00   0.00     0.04     0.00     0.00    -0.01    -0.04     0.00
## 131 -0.02   0.02     0.02     0.02     0.00     0.00    -0.01    -0.02
## 133  0.19  -0.17    -0.16    -0.18     0.00     0.15     0.13     0.15
## 134  0.00   0.00     0.05     0.00     0.00     0.05    -0.04     0.00
## 139  0.44  -0.35    -0.32    -0.38    -0.06    -0.04     0.25     0.32
## 141 -0.49   0.39     0.36     0.43     0.05     0.04    -0.28    -0.35
## 142  0.04  -0.03    -0.03    -0.03     0.00     0.00     0.02     0.03
## 154  0.01  -0.03     0.00     0.00    -0.01    -0.05     0.02     0.03
## 187  0.01  -0.03     0.00     0.00    -0.01    -0.05     0.02     0.03
## 219  0.01  -0.07     0.01     0.00     0.00    -0.11     0.04     0.06
## 227  0.02   0.00     0.57     0.01    -0.04    -0.16    -0.47     0.00
## 236  0.04   0.00    -0.02     0.13    -0.12    -0.05     0.00    -0.08
## 239 -0.09   0.01     0.06     0.04     0.34    -0.01    -0.01    -0.10
## 245  0.01   0.00     0.00     0.09    -0.01    -0.03     0.00    -0.07
## 254  0.01   0.00    -0.14     0.00    -0.01    -0.15     0.11     0.00
## 255  0.00   0.00     0.01    -0.01     0.02    -0.04     0.00     0.01
## 256 -0.02   0.02     0.02     0.02     0.00     0.00    -0.01    -0.02
## 259  0.00   0.00     0.01     0.00     0.00     0.00    -0.01     0.00
## 273  0.00   0.00     0.04     0.00     0.00    -0.01    -0.04     0.00
## 274  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 275  0.02   0.00    -0.17     0.00    -0.02    -0.18     0.14     0.00
```

```
## 288 -0.02   0.00    -0.34    -0.01     0.04     0.10     0.28     0.00
## 290  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 294 -0.59   0.45     0.42     0.47     0.15     0.06    -0.33    -0.41
## 300  0.49  -0.48    -0.38    -0.67     0.42    -0.01     0.35     0.46
## 307  0.00   0.00    -0.02     0.00     0.00     0.01     0.02     0.00
## 309  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 335  0.00  -0.03     0.00     0.00     0.00    -0.05     0.02     0.03
## 364  0.03  -0.03    -0.03    -0.03     0.00     0.00     0.02     0.03
## 377  0.00   0.00    -0.03     0.00     0.00     0.01     0.03     0.00
## 380 -0.56   0.45     0.41     0.49     0.06     0.05    -0.32    -0.41
##     dffit   cov.r   cook.d hat
## 25   0.45_*  1.05    0.03   0.07_*
## 35   0.13    1.07_*  0.00   0.05
## 50   0.50_*  0.85_*  0.03   0.02
## 64  -0.07    1.08_*  0.00   0.05
## 76   0.37    0.93_*  0.02   0.02
## 105 -0.06    1.07_*  0.00   0.05
## 107  0.02    1.07_*  0.00   0.05
## 108 -0.01    1.07_*  0.00   0.05
## 109  0.04    1.07_*  0.00   0.05
## 112  0.00    1.07_*  0.00   0.05
## 113  0.10    1.07_*  0.00   0.05
## 115  0.10    1.07_*  0.00   0.05
## 119  0.10    1.07_*  0.00   0.05
## 120  0.01    1.09_*  0.00   0.06_*
## 126  0.60_*  0.79_*  0.04   0.03
## 128  0.09    1.09_*  0.00   0.06_*
## 130  0.06    1.07_*  0.00   0.05
## 131 -0.02    1.07_*  0.00   0.05
## 133  0.26    1.08_*  0.01   0.07_*
## 134  0.09    1.09_*  0.00   0.06_*
## 139  0.44_*  0.99    0.02   0.05
## 141 -0.50_*  0.97    0.03   0.05
## 142  0.04    1.07_*  0.00   0.05
## 154 -0.08    1.07_*  0.00   0.05
## 187 -0.07    1.08_*  0.00   0.05
## 219 -0.16    1.07_*  0.00   0.05
## 227  0.84_*  0.78_*  0.09   0.05
## 236  0.30    0.84_*  0.01   0.01
## 239  0.49_*  0.75_*  0.03   0.02
## 245  0.23    0.89_*  0.01   0.01
## 254 -0.26    1.07_*  0.01   0.06_*
## 255 -0.05    1.06_*  0.00   0.04
## 256 -0.02    1.07_*  0.00   0.05
## 259  0.01    1.07_*  0.00   0.04
## 273  0.07    1.07_*  0.00   0.05
## 274  0.00    1.07_*  0.00   0.05
## 275 -0.32    1.06    0.01   0.06_*
## 288 -0.51_*  0.95    0.03   0.05
## 290  0.00    1.07_*  0.00   0.05
## 294 -0.59_*  0.94    0.04   0.05
## 300  0.76_*  0.93_*  0.07   0.07_*
## 307 -0.03    1.07_*  0.00   0.05
## 309  0.01    1.07_*  0.00   0.05
```

```
## 335 -0.08    1.07_*  0.00   0.05
## 364  0.04    1.07_*  0.00   0.05
## 377 -0.05    1.07_*  0.00   0.05
## 380 -0.57_*  0.94    0.04   0.05
```

And there is even more soooooo... we are not gonna take nay more out because when we tried (aka whats below) it did not help

```
# p <- 7
# n <- length(Salaries_noIP[,1])
# covra <- covratio(noinfluential)
# covra[covra > (1 + 3*p/n) ]
# covra[covra < (1 - 3*p/n)]
```

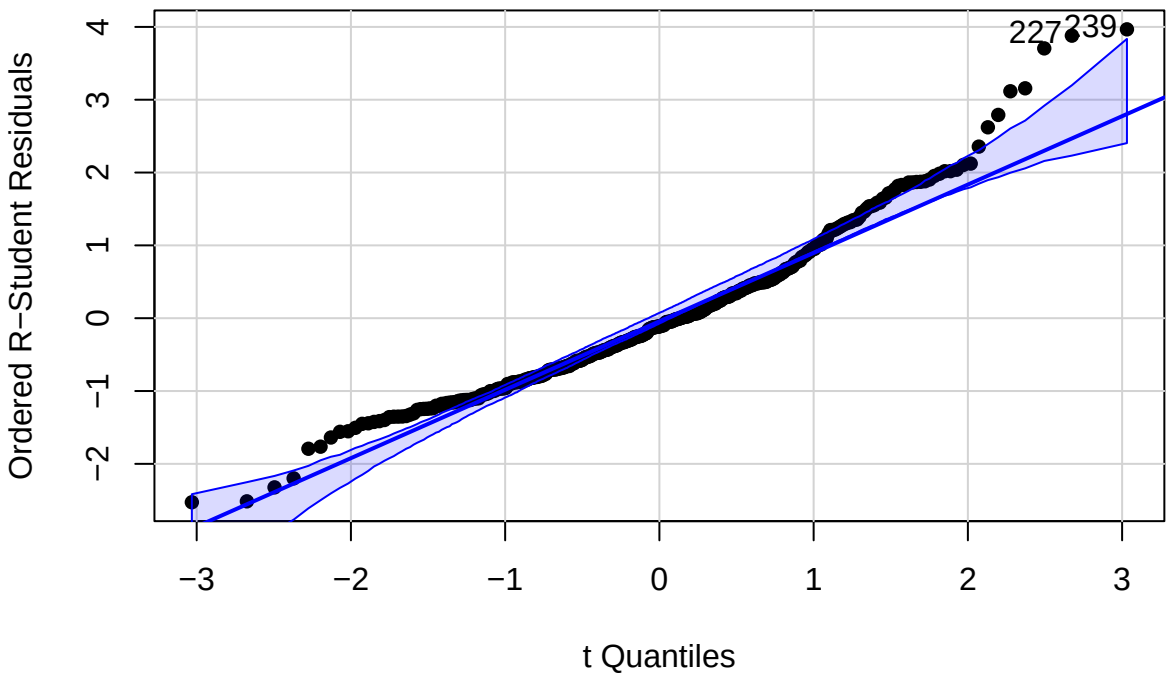
```
#Salaries3 <- Salaries_noIP[c(-10,-25,-35,-36,-64,-85,-104,-105,-107,-108,-109,-112,-113,-115,-119,-120
```

```
#noinfluential2 <- lm((salary)^(-2)~ discipline*rank + discipline*Salaries5$FRank +yrs.service + sex ,
#summary(noinfluential2)
```

#Checking to see if potential final model doesn't need any more transformations

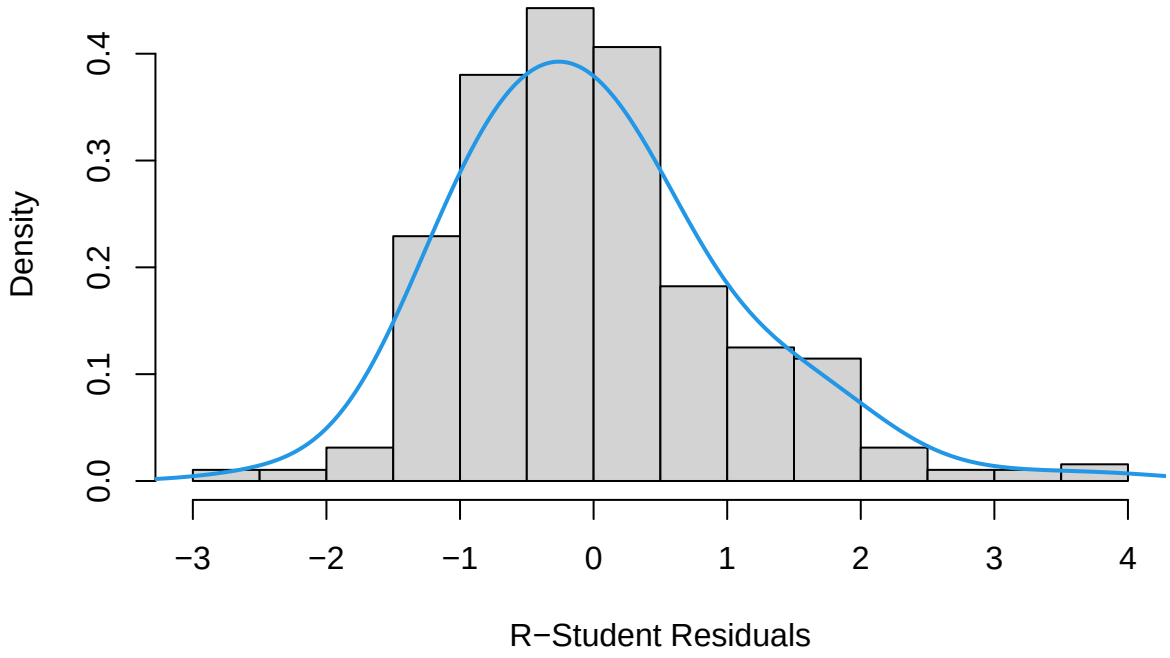
```
car::qqPlot(noinfluential, id = TRUE, col.lines = "blue",
            reps = 1000, ylab = "Ordered R-Student Residuals", pch = 16)
```

```
## Warning in rlm.default(x, y, weights, method = method, wt.method = wt.method, :
## 'rlm' failed to converge in 20 steps
```



```
## 227 239
## 223 232
```

```
r_stud <- rstudent(noinfluential)

hist(r_stud, prob = TRUE, breaks = 10, xlab = "R-Student Residuals", main = "")
lines(density(r_stud, adjust = 2), col = 4, lwd = 2)
```

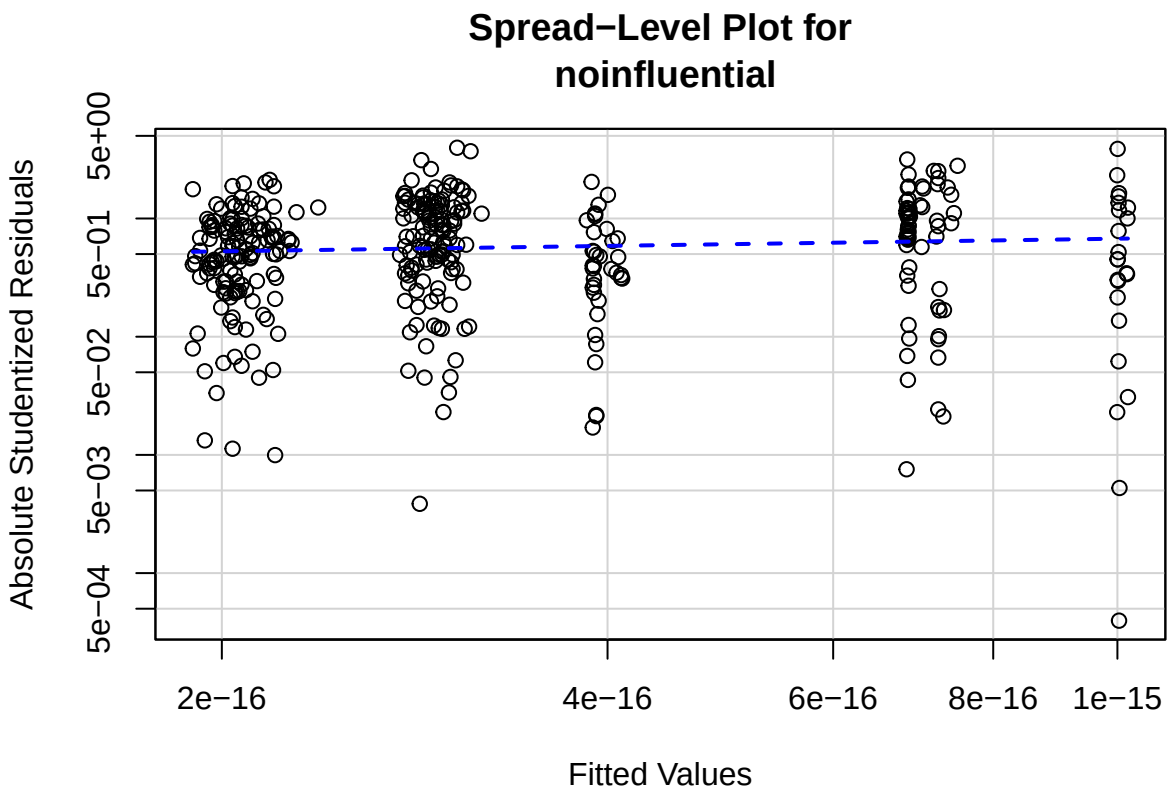



```
#could not do boxplot for best mod
carPT<-car::powerTransform(noinfluential, family = "bcPower")
summary(carPT)
```

```
## bcPower Transformation to Normality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1     0.524         0.5       0.4051       0.6429
##
## Likelihood ratio test that transformation parameter is equal to 0
##  (log transformation)
##                            LRT df       pval
## LR test, lambda = (0) 70.98555  1 < 2.22e-16
##
## Likelihood ratio test that no transformation is needed
##                            LRT df       pval
## LR test, lambda = (1) 2910.518  1 < 2.22e-16
```

```
#no transformation
```

```
#NON CONSTANT VARIANCE
car::spreadLevelPlot(noinfluential, smooth = FALSE)
```




```
##
## Suggested power transformation:  0.846546
```

```
#slope is not close to one
```

```
car::ncvTest(noinfluential)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 12.27298, Df = 1, p = 0.00045956
```

```
#therefore not a constant variance
```

```
#Collinearity

(vif_all <- car::vif(noinfluential))
```

```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
```

```
##       discipline             rank            FRank      yrs.service
##         7.175038         5.074360         4.956902         1.719434
##              sex  discipline:rank discipline:FRank
##         1.030711         5.364297         7.914792
```

```
#non are greater than 10 so we are good!
```

Looks good we will claim noinfluential as final model

```
final_mod <- noinfluential
summary(final_mod)
```

```
##
## Call:
## lm(formula = (salary)^(-3) ~ discipline * rank + discipline *
##     FRank + yrs.service + sex, data = Salaries_noIP)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -3.659e-16 -1.038e-16 -1.669e-17  7.359e-17  5.780e-16
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          7.181e-16  3.419e-17  21.006  < 2e-16 ***
## discipline1         -3.336e-16  4.115e-17  -8.106 7.44e-15 ***
## rank1                2.815e-16  4.589e-17   6.134 2.17e-09 ***
## FRank1              -4.459e-16  3.653e-17 -12.207  < 2e-16 ***
## yrs.service          8.272e-19  8.030e-19   1.030    0.304
## sex1                 1.702e-17  2.695e-17   0.632    0.528
## discipline1:rank1    1.856e-17  5.669e-17   0.327    0.744
## discipline1:FRank1   2.496e-16  4.509e-17   5.535 5.85e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.497e-16 on 376 degrees of freedom
## Multiple R-squared:  0.7095, Adjusted R-squared:  0.704
## F-statistic: 131.2 on 7 and 376 DF,  p-value: < 2.2e-16
```

#Prediction

```
yu <- data.frame(rank ="AsstProf",  discipline = "A", yrs.since.phd = 4, yrs.service = 2, sex = "Male")

predict(full_model, newdata=yu)
```

```
##       1
## 97064.9
```

#yu <- data.frame(rank =1, FRank = 0, discipline = 1, yrs.since.phd = 4, yrs.service = 2, sex = 0)

#yu$FRank <- -factor(yu$FRank) #yu <- yu %>% rename("Salaries$rank" = "Salaries.rank")

#yu <- Salaries4[1,-7] yu <- data.frame(rank ="AsstProf", discipline = "B", yrs.since.phd = 4, yrs.service = 2, sex = "Male")

yu2 <- cbind(yu, yu$rank)

#Salaries2 %>% rename(Rank2 = Salaries$FRank)

yu2$rank <- recode_factor(yu2$rank, "AssocProf" = 0, "AsstProf"= 1,"Prof"= 0) yu2$Salaries$rank<- recode_factor(yu2yurank, "AssocProf" = 0, "Prof" = 1, "AsstProf" = 0)yu2$sex <- recode_factor(yu2$sex, "Male" = 0, "Female" = 1)yu2$discipline <- recode_factor(yu2$discipline, "A"=0, "B" = 1)

yu2 <- yu2[, c(1,7,2,3,4,5,6)]

names(yu2)[2] <- "FRank"

predict(final_mod, newdata=yu2)

```
yu <- data.frame(rank =as.factor(1), FRank = as.factor(0),  discipline = as.factor(1), yrs.since.phd = 4



#Transforming the response back to original
(predict(final_mod, newdata=yu))^(2/3)/predict(final_mod, newdata=yu)
```

```
##        1
## 113370.5
```

```
(predict(final_mod, newdata=yu, interval="confidence", level = .95))^(2/3)/predict(final_mod, newdata=yu
```

```
##        fit    lwr      upr
## 1 113370.5 116015 110951.7
```

asst 10 prof 01 assoc 00

clough -> prof 1985 - 37, 37 years, female

spiller -> assoc prof, 2005 - 17, 14

sander -> prof, 2004- 18, 7 years

rowe -> prof, 1998- 24, 2014 - 8 (changed schools)

ongie -> asst prof, 2016 - 6, 2020 - 2

maadooliat -> assoc prof, 2011 - 11, 2013 - 9

hamilton -> assoc prof, 2012 - 10, 2014 - 8

pantone -> asst prof ruitenburg -> prof hamedani -> prof bansal -> prof

```
changediscipline <- 0

yu <- data.frame(rank =as.factor(1), FRank = as.factor(0),  discipline = as.factor(changediscipline), yr

clough <- data.frame(rank =as.factor(0), FRank = as.factor(1),  discipline = as.factor(changediscipline)


spiller <- data.frame(rank =as.factor(0), FRank = as.factor(0),  discipline = as.factor(changedisciplin


sander <- data.frame(rank =as.factor(0), FRank = as.factor(1),  discipline = as.factor(changediscipline)
```

```
rowe <- data.frame(rank =as.factor(0), FRank = as.factor(1),  discipline = as.factor(changediscipline),


ongie <- data.frame(rank =as.factor(1), FRank = as.factor(0),  discipline = as.factor(changediscipline)


maadooliat <- data.frame(rank =as.factor(0), FRank = as.factor(0),  discipline = as.factor(changedisci


hamilton <- data.frame(rank =as.factor(0), FRank = as.factor(0),  discipline = as.factor(changedisciplin
```

```
professors <- rbind(yu, clough, hamilton, ongie, maadooliat, rowe, sander, spiller)
professor_predict <- (predict(final_mod, newdata=professors, interval="confidence", level = .95))^(2/3),

professornames <- cbind(c("Yu", "Clough", "Hamilton", "Ongie", "Maadooliat", "Rowe", "Sanders", "Spille

names(professornames)[1] <- "Professor"
names(professornames)[2] <- "Rank = Asst Prof"
names(professornames)[3] <- "Rank = Full Prof"
names(professornames)[7] <- "Gender = Male"
names(professornames)[9] <- "Upper"
names(professornames)[10] <- "Lower"

professornames <- professornames[, c(1,2,3,4,5,6,7,8,10,9)]

professornames
```

```
##    Professor Rank = Asst Prof Rank = Full Prof discipline yrs.since.phd
## 1         Yu                1                0          0             4
## 2     Clough                0                1          0            37
## 3   Hamilton                0                0          0            10
## 4      Ongie                1                0          0             6
## 5 Maadooliat                0                0          0            11
## 6       Rowe                0                1          0            24
## 7    Sanders                0                1          0            18
## 8    Spiller                0                0          0            17
##   yrs.service Gender = Male       fit     Lower    Upper
## 1           2             0  99956.92  97960.92 102126.2
## 2          37             1 146229.24 137897.55 157027.0
## 3           8             1 110469.66 106761.27 114753.7
## 4           2             0  99956.92  97960.92 102126.2
## 5           9             0 111285.40 108165.11 114800.1
## 6           8             0 153073.39 146927.68 160396.2
## 7           7             1 150219.71 141249.22 162014.0
## 8          14             1 110224.37 106528.89 114492.6
```

"Ongie close but 112 was too high"