Final Project

Oilivia Wagner| Weijia Xiong | Wurongyan Zhang | Yiling Yang

Contents

Abstract

Introduction

In our project, we examined the whether gender discrimination existed in setting salaries for people in the academia or higher education institutions.

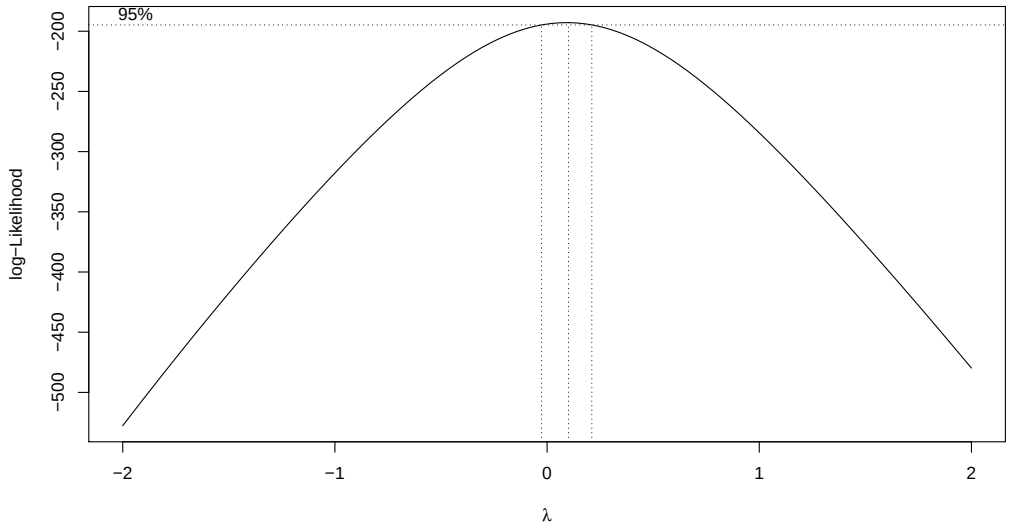
Exploratory data analysis

Data used in our studies were collected from 261 individuals who work in the academia or higher education institutions. The raw dataset contains following features.

- **Dept**: 1= Biochemistry/Molecular Biology 2= Physiology 3= Genetics 4= Pediatrics 5= Medicine 6= Surgery
- **Gender**: 1= Male, 0= Female
- **Clin**: 1= Primarily clinical emphasis, 0= Primarily research emphasis
- **Cert**: 1= Board certified, 0= not certified
- **Prate**: Publication rate (# publications on cv) / (# years between CV date and MD date)
- **Exper**: # years since obtaining MD
- **Rank**: 1= Assistant, 2= Associate, 3= Full professor (a proxy for productivity)
- **Sal94**: Salary in academic year 1994
- **Sal95**: Salary after increment to Sal94

Table 1 contains the summary of variables in the dataset. Fortunately, there are no missing values in our dataset. We then need to examine each interested variable against the main effect and main interest.

Salary



Since $\lambda = 0$, we use log transformation.

Department

Fig 1.1. shows that the gender ratios in the department of Genetics and Physiology are very balanced, compared with department of Medicine and Surgery, which are very imbalanced. The differences in department of Biochemistry and Pediatrics are moderate. Fig 1.2 includes our main interest the salary into account. It seems that across all department, male earn more than female do. However, before further analysis, we cannot tell whether those difference are significant.

Fig.1.1 Gender Distribution of Department

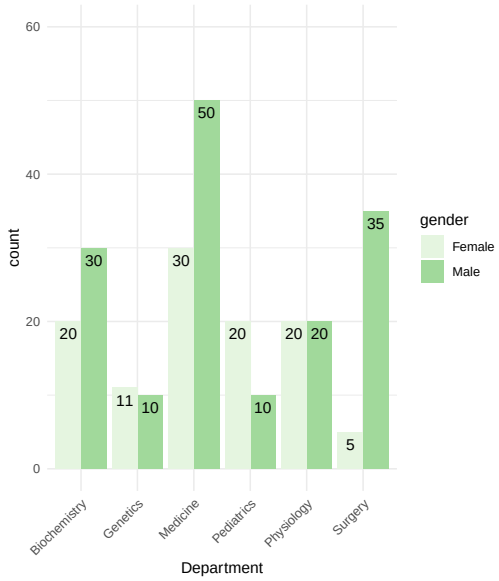
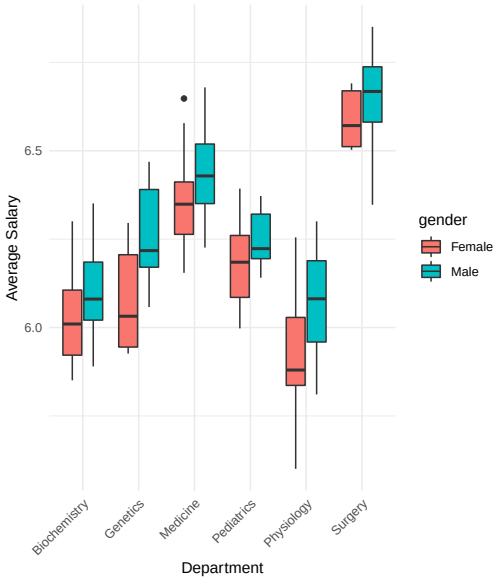


Fig.1.2 Average Salary across Department & Gender



Clinical or Research Division

Fig 2.1 shows that in either division, there are more male than female, especially the clinical division. Regarding salar, again, male in either division earned more than female do.

Fig.2.1 Gender Distribution across Division

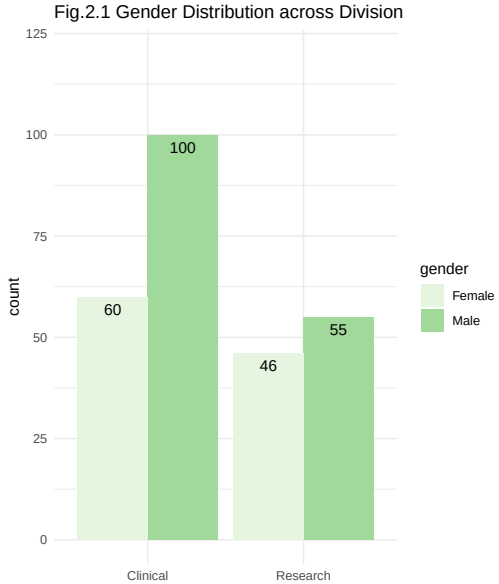
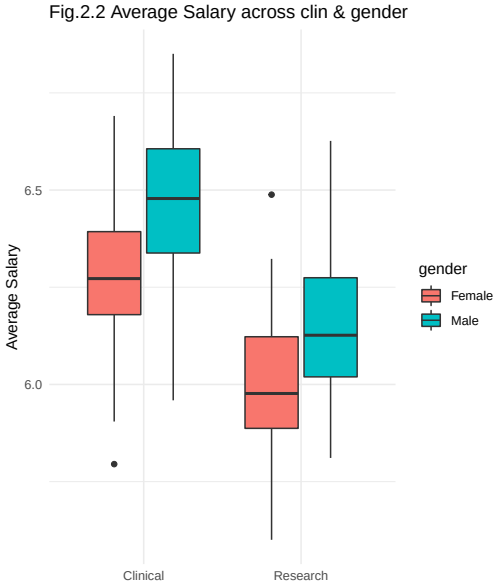


Fig.2.2 Average Salary across clin & gender



Certfication status

Fig.3. shows that the amount of certified male outnumbers surely the amount of certified female. However, for those without certification, theose are just even. For the salar, male again earn more than female do regardless of his or her certification status.

Fig.3.1 Gender Distribution against Certification Status

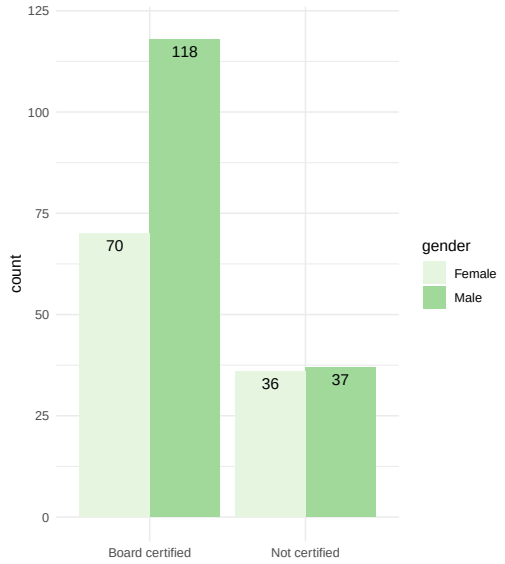
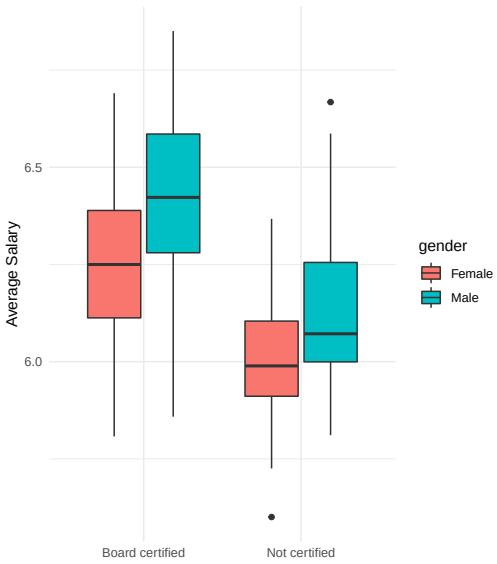


Fig.3.2 Average Salary across Cert & Gender



Prate

Fig 4.1 shows the density plot of publication rate for both male and female. No obvious difference was observed. Fig 4.2 implies that there might be linear trend between average salary and publication rate.However, it's hard to tell whetehr there is any difference on the effect on the salar regarding the gender.s

Fig.4.1 Density of Publication Rate against Gender

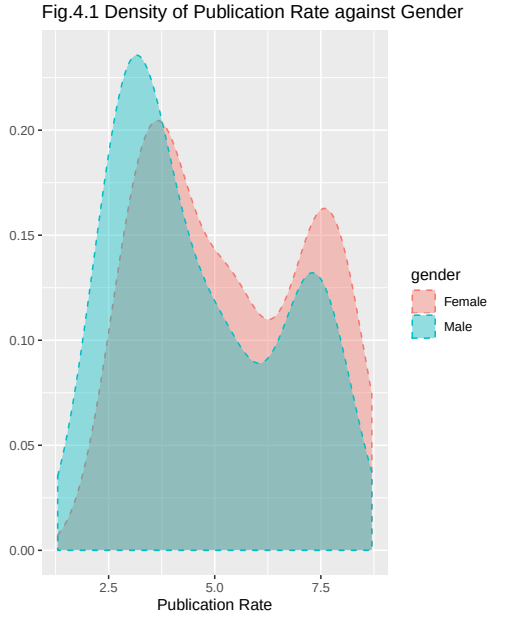
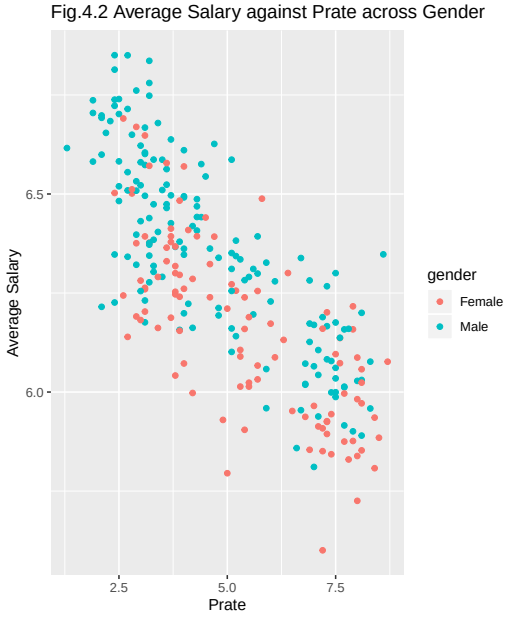


Fig.4.2 Average Salary against Prate across Gender



Exper

Fig.5.1 shows that the densities of experience year for both male and female are both very skewed, and seems that female have a more skewed trend. Although there are not obvious linear trend between salar and experience

year, the salar for male spread more widly than those for female and individual with high salar are dominately male.

Fig 5.1 Density of experience Year against Gender

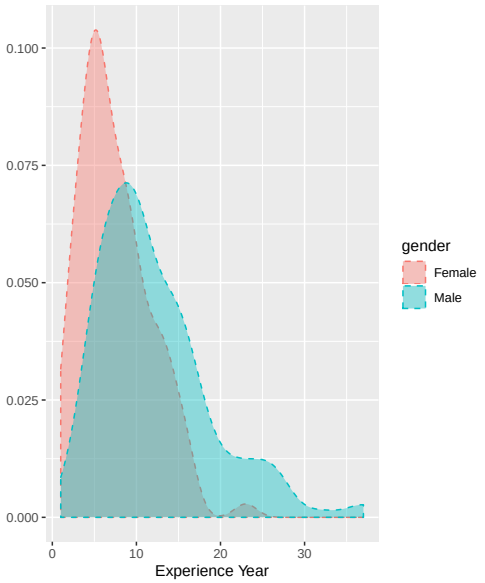
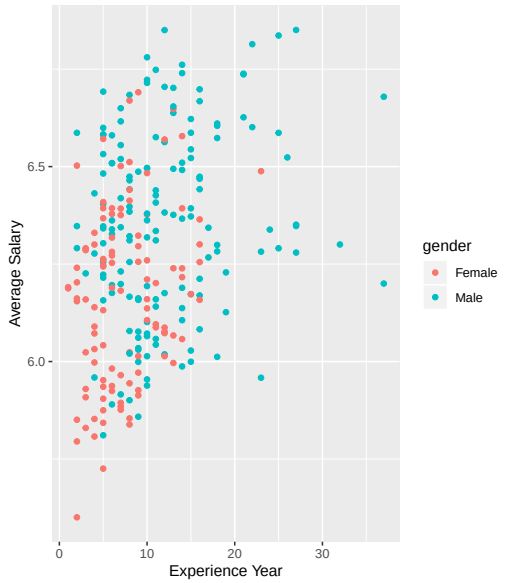


Fig 5.2 Average Salary against Experience Year across G



Rank

Fig 6.1 shows that there are many female assistant professor and less associate and full professor than male. Without surprising, in either rank, male earn more than female.

Fig.6.1 Gender Distribution against Rank

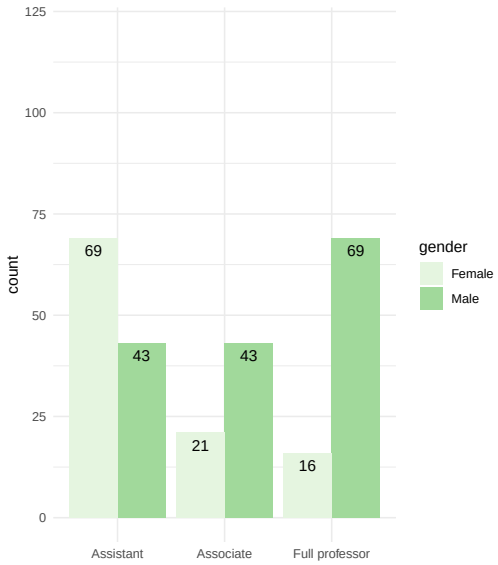
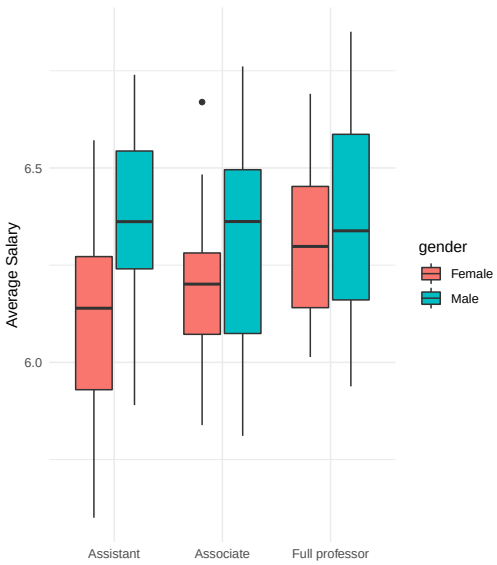
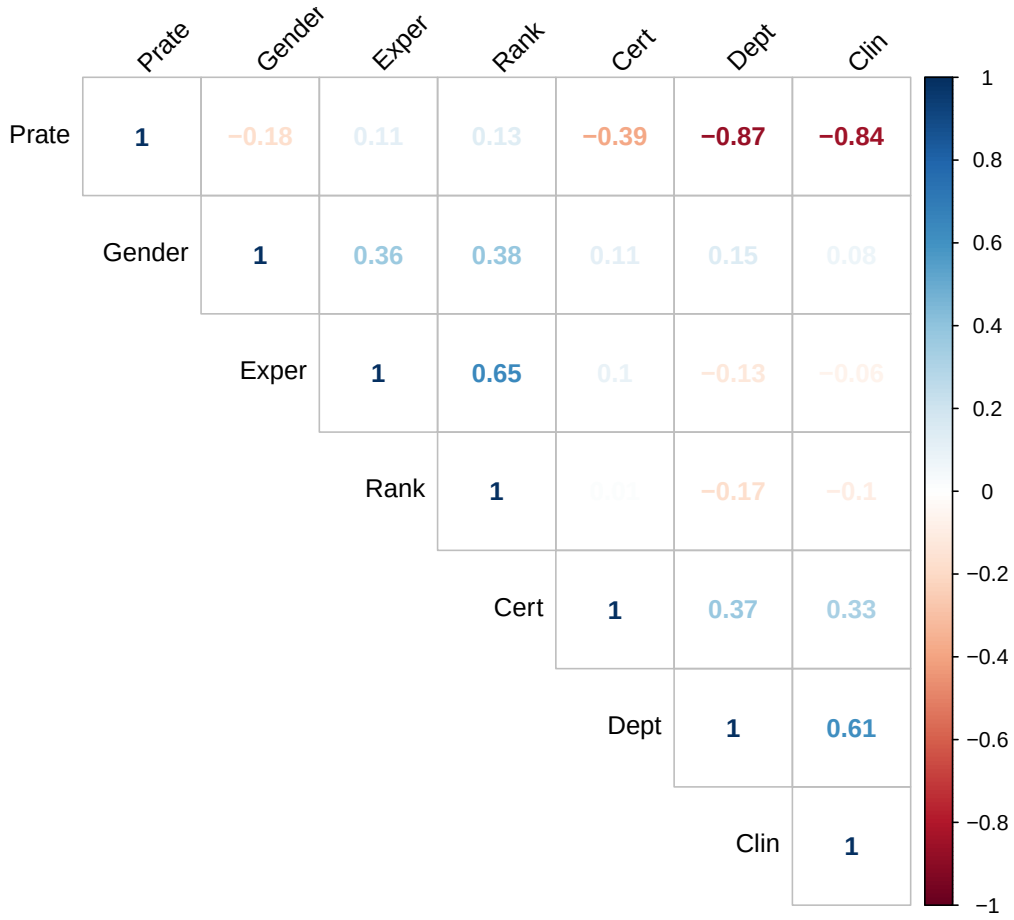


Fig.6.2 Average Salary across Rank & Gender



Lastly, after examing the correlation matrix, we can see that there are some highly related variables ($r > 80\%$). They are 1. Department (Dept) and Publish Rate (Prate) and 2. Clin and Prate. Some have morderate correlation such as 1. Experience Year (Exper) annd Cert (Certification status) and 2.Department (Dept) and Clin.



Those imply potential collinearities. As we go through those variables, there are some outliers. In later section, we will examine further about them.

Method

```
##
## Call:
## lm(formula = sal9495 ~ gender, data = lawsuit)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5628 -0.1829 -0.0039  0.1671  0.5275
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.16317    0.02298 268.246  < 2e-16 ***
## genderMale   0.19265    0.02981   6.462  5.1e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2366 on 259 degrees of freedom
## Multiple R-squared:  0.1388, Adjusted R-squared:  0.1355
## F-statistic: 41.75 on 1 and 259 DF,  p-value: 5.103e-10
```

```
##
## Call:
## lm(formula = sal9495 ~ gender + dept, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35272 -0.09716 -0.01250  0.08171  0.30390
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     6.01909    0.02055 292.901  < 2e-16 ***
## genderMale      0.10260    0.01678   6.116 3.60e-09 ***
## deptGenetics    0.10046    0.03301   3.044  0.00258 **
## deptMedicine    0.32471    0.02284  14.215  < 2e-16 ***
## deptPediatrics  0.15267    0.02960   5.158 5.02e-07 ***
## deptPhysiology -0.06597    0.02693  -2.450  0.01496 *
## deptSurgery     0.53834    0.02727  19.743  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.1267 on 254 degrees of freedom
## Multiple R-squared:  0.7578, Adjusted R-squared:  0.752
## F-statistic: 132.4 on 6 and 254 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = sal9495 ~ gender + clin, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50456 -0.12512 -0.00948  0.12282  0.49713
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.29514    0.02065 304.872  < 2e-16 ***
## genderMale    0.16859    0.02336   7.218 5.89e-12 ***
## clinResearch -0.30410    0.02355 -12.912  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1847 on 258 degrees of freedom
## Multiple R-squared:  0.4769, Adjusted R-squared:  0.4728
## F-statistic: 117.6 on 2 and 258 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = sal9495 ~ gender + cert, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.55933 -0.13835 -0.00961  0.15540  0.50971
```

```
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        6.25144    0.02228 280.648  < 2e-16 ***
## genderMale         0.16642    0.02617   6.360 9.11e-10 ***
## certNot certified -0.25991    0.02863  -9.078  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2063 on 258 degrees of freedom
## Multiple R-squared:  0.3473, Adjusted R-squared:  0.3423
## F-statistic: 68.65 on 2 and 258 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = sal9495 ~ gender + prate, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.40134 -0.09766  0.00195  0.09624  0.36769
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.670724   0.030122 221.457  < 2e-16 ***
## genderMale   0.125907   0.019411   6.486 4.46e-10 ***
## prate       -0.094870   0.004912 -19.313  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1516 on 258 degrees of freedom
```

```
## Multiple R-squared:  0.6479, Adjusted R-squared:  0.6452
## F-statistic: 237.4 on 2 and 258 DF,  p-value: < 2.2e-16
##
## Call:
## lm(formula = sal9495 ~ gender + exper, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.51649 -0.18686  0.02018  0.16638  0.51473
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.100037   0.029194 208.950  < 2e-16 ***
## genderMale  0.153775   0.031384   4.900  1.7e-06 ***
## exper       0.008428   0.002480   3.399 0.000784 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2319 on 258 degrees of freedom
## Multiple R-squared:  0.1757, Adjusted R-squared:  0.1693
## F-statistic:  27.5 on 2 and 258 DF,  p-value: 1.488e-11
##
## Call:
## lm(formula = sal9495 ~ gender + rank, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.55399 -0.18936  0.00746  0.17699  0.51766
##
```

```
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.154393   0.025482 241.520  < 2e-16 ***
## genderMale          0.174740   0.032210   5.425 1.34e-07 ***
## rankAssociate      -0.002612   0.038094  -0.069   0.9454
## rankFull professor  0.061583   0.036611   1.682   0.0938 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2358 on 257 degrees of freedom
## Multiple R-squared:  0.151,  Adjusted R-squared:  0.1411
## F-statistic: 15.23 on 3 and 257 DF,  p-value: 3.747e-09
```

Since the changes of estimate of genderMale are more than 10%, dept, clin, cert and exper are all confounders. So we need to add these covariates into the model.

```
##
## Call:
## lm(formula = sal9495 ~ gender + dept + clin + cert + exper, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18571 -0.05097  0.00044  0.04429  0.42863
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        6.0314905  0.0194750 309.705  < 2e-16 ***
## genderMale         0.0339569  0.0107443   3.160  0.00177 **
## deptGenetics       0.0929245  0.0206775   4.494 1.07e-05 ***
## deptMedicine       0.2615613  0.0166801  15.681  < 2e-16 ***
## deptPediatrics     0.0986776  0.0201405   4.899 1.72e-06 ***
```

```
## deptPhysiology    -0.0806445  0.0164703  -4.896 1.75e-06 ***
## deptSurgery        0.4516913  0.0200155  22.567  < 2e-16 ***
## clinResearch      -0.1039933  0.0123844  -8.397 3.40e-15 ***
## certNot certified -0.0915366  0.0119239  -7.677 3.63e-13 ***
## exper              0.0132859  0.0008423  15.774  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07574 on 251 degrees of freedom
## Multiple R-squared:  0.9144, Adjusted R-squared:  0.9114
## F-statistic: 298.1 on 9 and 251 DF,  p-value: < 2.2e-16
```

Then we add interaction

```
##
## Call:
## lm(formula = sal9495 ~ gender * exper + dept + clin + cert, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.17887 -0.04469  0.00158  0.04131  0.40799
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.974101   0.023092 258.710  < 2e-16 ***
## genderMale         0.105216   0.019578   5.374 1.76e-07 ***
## exper              0.020034   0.001770  11.319  < 2e-16 ***
## deptGenetics       0.095713   0.020005   4.784 2.93e-06 ***
## deptMedicine       0.267482   0.016188  16.524  < 2e-16 ***
## deptPediatrics     0.108945   0.019621   5.552 7.19e-08 ***
## deptPhysiology    -0.073740   0.016007  -4.607 6.52e-06 ***
```

```
## deptSurgery        0.457157   0.019396  23.569  < 2e-16 ***
## clinResearch      -0.104183   0.011975  -8.700 4.53e-16 ***
## certNot certified -0.087238   0.011573  -7.538 8.78e-13 ***
## genderMale:exper  -0.008377   0.001951  -4.294 2.51e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07324 on 250 degrees of freedom
## Multiple R-squared:  0.9203, Adjusted R-squared:  0.9171
## F-statistic: 288.8 on 10 and 250 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = sal9495 ~ gender * dept + exper + clin + cert, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18564 -0.04880  0.00124  0.04409  0.42850
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                6.0325227  0.0227388 265.297  < 2e-16 ***
## genderMale                 0.0302541  0.0223541   1.353 0.177168
## deptGenetics               0.0833594  0.0290913   2.865 0.004524 **
## deptMedicine               0.2637483  0.0240299  10.976  < 2e-16 ***
## deptPediatrics             0.0976654  0.0259728   3.760 0.000212 ***
## deptPhysiology            -0.0920384  0.0244470  -3.765 0.000209 ***
## deptSurgery                0.4716680  0.0400878  11.766  < 2e-16 ***
## exper                      0.0132652  0.0008508  15.591  < 2e-16 ***
## clinResearch              -0.1034518  0.0126267  -8.193 1.40e-14 ***
```

```
## certNot certified         -0.0895941  0.0121654  -7.365 2.67e-12 ***
## genderMale:deptGenetics    0.0190307  0.0406550   0.468 0.640125
## genderMale:deptMedicine   -0.0018252  0.0282104  -0.065 0.948465
## genderMale:deptPediatrics  0.0032796  0.0368694   0.089 0.929191
## genderMale:deptPhysiology  0.0225352  0.0328368   0.686 0.493183
## genderMale:deptSurgery    -0.0201824  0.0426198  -0.474 0.636244
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0763 on 246 degrees of freedom
## Multiple R-squared:  0.9149, Adjusted R-squared:  0.9101
## F-statistic: 188.9 on 14 and 246 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = sal9495 ~ gender * clin + exper + cert + dept, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19595 -0.04838 -0.00007  0.04304  0.41441
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              6.0430697  0.0196756 307.135  < 2e-16 ***
## genderMale               0.0106440  0.0135578   0.785  0.43315
## clinResearch            -0.1324656  0.0159951  -8.282 7.40e-15 ***
## exper                    0.0132387  0.0008316  15.921  < 2e-16 ***
## certNot certified       -0.0949870  0.0118359  -8.025 3.95e-14 ***
## deptGenetics             0.1019910  0.0206728   4.934 1.47e-06 ***
## deptMedicine             0.2643889  0.0164962  16.027  < 2e-16 ***
```

```
## deptPediatrics           0.0975262  0.0198844   4.905 1.69e-06 ***
## deptPhysiology          -0.0787295  0.0162721  -4.838 2.29e-06 ***
## deptSurgery              0.4605185  0.0200139  23.010  < 2e-16 ***
## genderMale:clinResearch  0.0551454  0.0199793   2.760  0.00621 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07476 on 250 degrees of freedom
## Multiple R-squared:  0.917,  Adjusted R-squared:  0.9137
## F-statistic: 276.1 on 10 and 250 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = sal9495 ~ gender * cert + exper + clin + dept, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18588 -0.05139  0.00078  0.04455  0.42827
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  6.031796   0.020117 299.842  < 2e-16 ***
## genderMale                   0.033539   0.012668   2.647  0.00863 **
## certNot certified           -0.092297   0.017038  -5.417 1.42e-07 ***
## exper                        0.013285   0.000844  15.740  < 2e-16 ***
## clinResearch                -0.104079   0.012485  -8.336 5.16e-15 ***
## deptGenetics                 0.093072   0.020853   4.463 1.22e-05 ***
## deptMedicine                 0.261518   0.016727  15.634  < 2e-16 ***
## deptPediatrics               0.098545   0.020291   4.856 2.11e-06 ***
## deptPhysiology              -0.080561   0.016556  -4.866 2.02e-06 ***
```

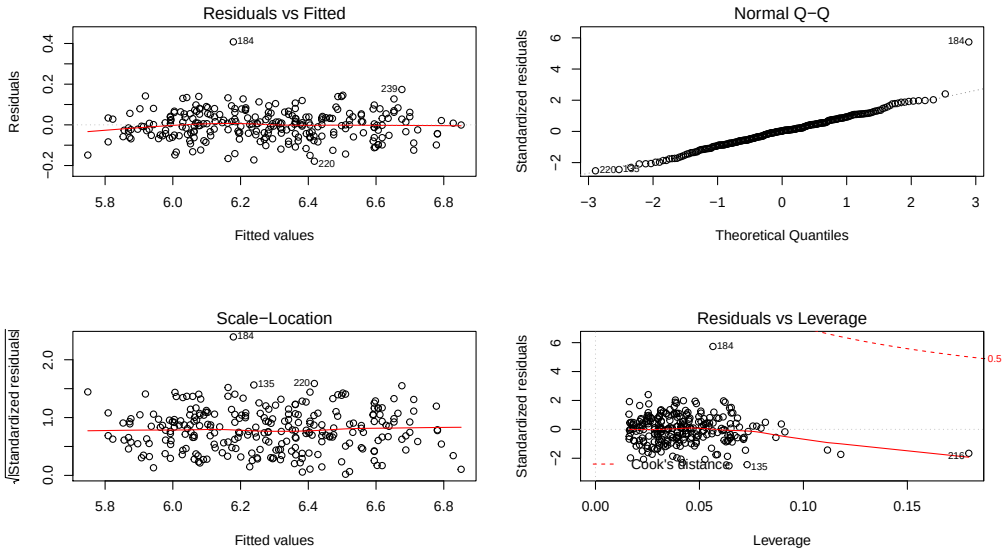
```
## deptSurgery                   0.451731   0.020066  22.513  < 2e-16 ***
## genderMale:certNot certified  0.001370   0.021885   0.063  0.95013
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07589 on 250 degrees of freedom
## Multiple R-squared:  0.9144, Adjusted R-squared:  0.911
## F-statistic: 267.2 on 10 and 250 DF,  p-value: < 2.2e-16
```

Interaction: gender * exper significant.

Final Model:

```
##
## Call:
## lm(formula = sal9495 ~ gender * exper + dept + clin + cert, data = lawsuit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.17887 -0.04469  0.00158  0.04131  0.40799
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       5.974101   0.023092 258.710  < 2e-16 ***
## genderMale        0.105216   0.019578   5.374 1.76e-07 ***
## exper             0.020034   0.001770  11.319  < 2e-16 ***
## deptGenetics      0.095713   0.020005   4.784 2.93e-06 ***
## deptMedicine      0.267482   0.016188  16.524  < 2e-16 ***
## deptPediatrics    0.108945   0.019621   5.552 7.19e-08 ***
## deptPhysiology   -0.073740   0.016007  -4.607 6.52e-06 ***
## deptSurgery       0.457157   0.019396  23.569  < 2e-16 ***
## clinResearch     -0.104183   0.011975  -8.700 4.53e-16 ***
```

```
## certNot certified -0.087238   0.011573  -7.538 8.78e-13 ***
## genderMale:exper  -0.008377   0.001951  -4.294 2.51e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07324 on 250 degrees of freedom
## Multiple R-squared:  0.9203, Adjusted R-squared:  0.9171
## F-statistic: 288.8 on 10 and 250 DF,  p-value: < 2.2e-16
```




```
## Potentially influential observations of
##   lm(formula = sal9495 ~ gender * exper + dept + clin + cert, data = lawsuit) :
##
##     dfb.1_ dfb.gndM dfb.expr dfb.dptG dfb.dptM dfb.dptPd dfb.dptPh
## 19    0.00    0.25    -0.01     0.10     0.05     0.01      0.13
## 39    0.00    0.01     0.01    -0.01    -0.01    -0.01     -0.01
## 58   -0.02   -0.04     0.00     0.02     0.02     0.02      0.05
## 91    0.01    0.01     0.00    -0.04    -0.01    -0.01      0.00
## 122  -0.19    0.17     0.16     0.04     0.08    -0.21      0.05
## 135   0.36   -0.18    -0.39    -0.11    -0.21    -0.48     -0.10
## 184  -0.65    0.63     0.13     0.14     0.84     0.45      0.20
```

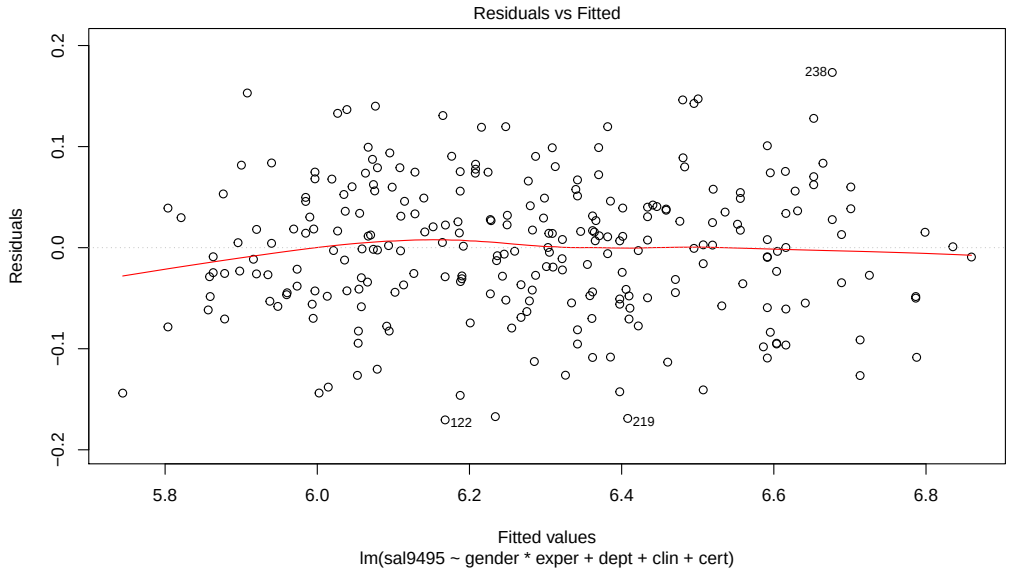
```
## 216  0.53  -0.37    -0.69    -0.11    -0.33    -0.19     -0.12
## 220  0.36  -0.15    -0.41    -0.12    -0.38    -0.16     -0.10
## 239 -0.01   0.01     0.01     0.00     0.00     0.00      0.00
##     dfb.dptS dfb.clnR dfb.crNc dfb.gnM: dffit   cov.r   cook.d hat
## 19   0.04    -0.02    -0.20    -0.25    -0.64_*  1.04    0.04   0.12
## 39  -0.01    -0.01     0.00    -0.01     0.02    1.14_*  0.00   0.08
## 58   0.02     0.03    -0.02     0.04     0.12    1.14_*  0.00   0.09
## 91  -0.01    -0.02     0.02    -0.01    -0.06    1.15_*  0.00   0.09
## 122  0.06     0.10     0.04    -0.15    -0.49    0.86_*  0.02   0.04
## 135 -0.22    -0.31     0.02     0.33    -0.69_*  0.86_*  0.04   0.07
## 184  0.46     0.73     0.66    -0.42     1.50_*  0.23_*  0.18   0.06
## 216 -0.23    -0.26    -0.01     0.61    -0.78_*  1.13    0.05   0.18_*
## 220 -0.25    -0.37     0.07     0.36    -0.67_*  0.84_*  0.04   0.06
## 239  0.22    -0.01    -0.01     0.00     0.39    0.83_*  0.01   0.03
```

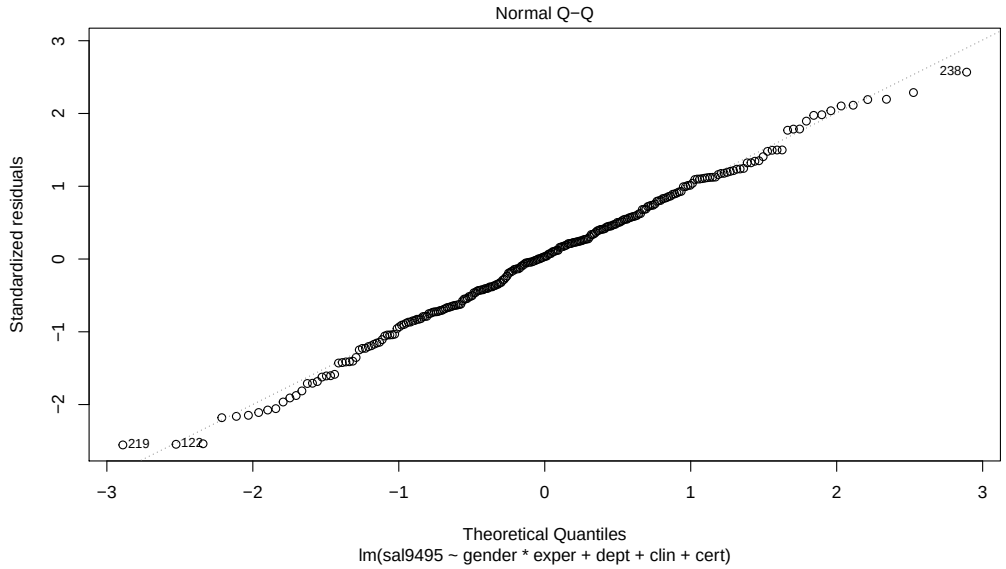
	dfb.1_	dfb.gndM	dfb.expr	dfb.dptG	dfb.dptM	dfb.dptPd	dfb.dptPh	dfb.dptS	dfb.clnR	dfb.crNc	dfb.gnM:	dffit	cov.r	cook.d	hat
19	0.0022	0.2483	-0.0053	0.1019	0.0534	0.0093	0.1347	0.0363	-0.0228	-0.2015	-0.2491	-0.6367	1.0372	0.0366	0.1180
39	0.0041	0.0073	0.0145	-0.0087	-0.0129	-0.0105	-0.0091	-0.0113	-0.0115	-0.0034	-0.0139	0.0245	1.1352	0.0001	0.0797
58	-0.0202	-0.0357	0.0018	0.0159	0.0197	0.0182	0.0523	0.0156	0.0284	-0.0172	0.0387	0.1155	1.1408	0.0012	0.0894
91	0.0080	0.0096	0.0017	-0.0408	-0.0088	-0.0070	-0.0040	-0.0068	-0.0200	0.0165	-0.0132	-0.0555	1.1483	0.0003	0.0911
122	-0.1873	0.1720	0.1610	0.0419	0.0786	-0.2098	0.0467	0.0628	0.1003	0.0399	-0.1524	-0.4872	0.8604	0.0212	0.0419
135	0.3633	-0.1842	-0.3857	-0.1101	-0.2134	-0.4827	-0.0987	-0.2173	-0.3129	0.0171	0.3297	-0.6929	0.8637	0.0428	0.0729
184	-0.6471	0.6344	0.1333	0.1445	0.8352	0.4492	0.1991	0.4609	0.7288	0.6551	-0.4150	1.5029	0.2347	0.1790	0.0565
216	0.5306	-0.3743	-0.6869	-0.1128	-0.3318	-0.1917	-0.1228	-0.2286	-0.2570	-0.0090	0.6050	-0.7786	1.1278	0.0547	0.1795
220	0.3636	-0.1548	-0.4125	-0.1235	-0.3818	-0.1584	-0.1024	-0.2476	-0.3676	0.0735	0.3592	-0.6687	0.8406	0.0398	0.0642
239	-0.0093	0.0148	0.0070	0.0028	-0.0043	0.0024	0.0016	0.2221	-0.0071	-0.0114	0.0016	0.3917	0.8291	0.0137	0.0254

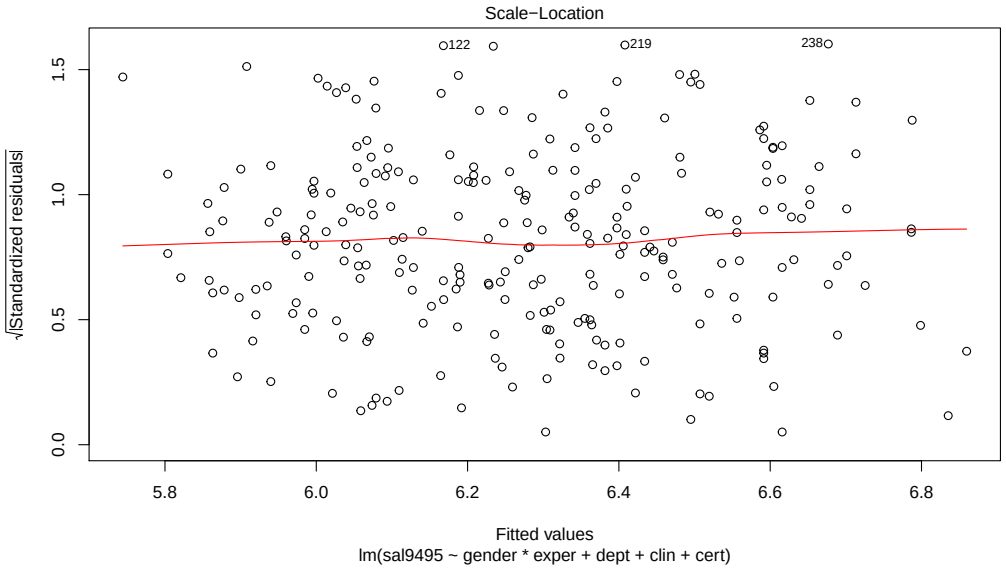
Remove outlier

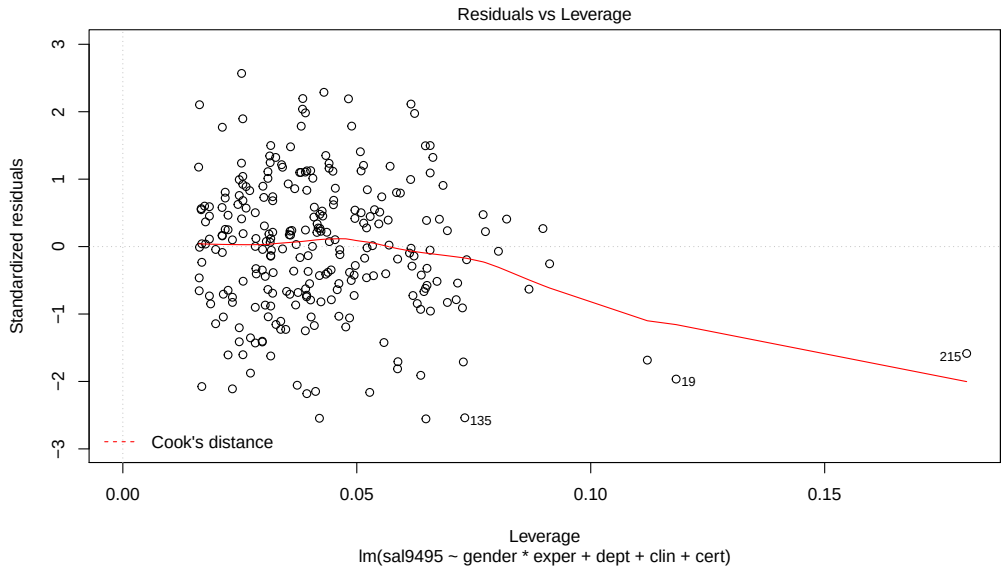
```
##
## Call:
## lm(formula = sal9495 ~ gender * exper + dept + clin + cert, data = newlawsuit)
##
## Residuals:
##       Min         1Q     Median         3Q        Max
## -0.170419 -0.044897  0.002285  0.045782  0.173369
##
## Coefficients:
```

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.988053   0.021682 276.180  < 2e-16 ***
## genderMale         0.093619   0.018378   5.094 6.93e-07 ***
## exper              0.019814   0.001653  11.986  < 2e-16 ***
## deptGenetics       0.093014   0.018685   4.978 1.20e-06 ***
## deptMedicine       0.254857   0.015255  16.707  < 2e-16 ***
## deptPediatrics     0.100715   0.018371   5.482 1.03e-07 ***
## deptPhysiology    -0.076716   0.014955  -5.130 5.84e-07 ***
## deptSurgery        0.448810   0.018162  24.711  < 2e-16 ***
## clinResearch      -0.112333   0.011261  -9.976  < 2e-16 ***
## certNot certified -0.094318   0.010868  -8.678 5.33e-16 ***
## genderMale:exper  -0.007621   0.001826  -4.174 4.13e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06839 on 249 degrees of freedom
## Multiple R-squared:  0.9304, Adjusted R-squared:  0.9276
## F-statistic: 332.9 on 10 and 249 DF,  p-value: < 2.2e-16
```



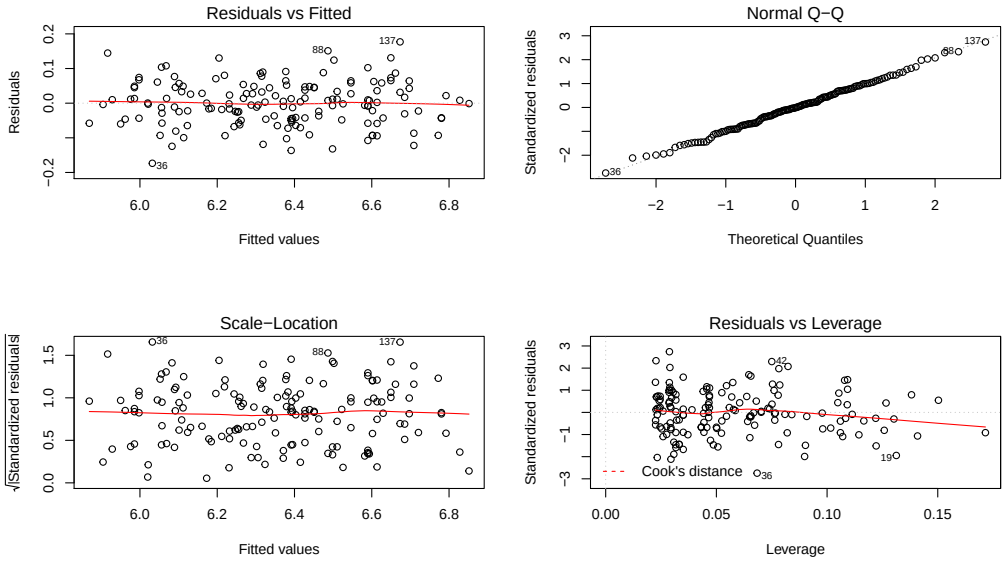

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(sal9495 ~ gender * exper + dept + clin + cert)
219
122
238


Scale−Location
√|Standardized residuals|
Fitted values
lm(sal9495 ~ gender * exper + dept + clin + cert)
122
219
238


Residuals vs Leverage
Standardized residuals
Leverage
lm(sal9495 ~ gender * exper + dept + clin + cert)
Cook's distance
215
19
135

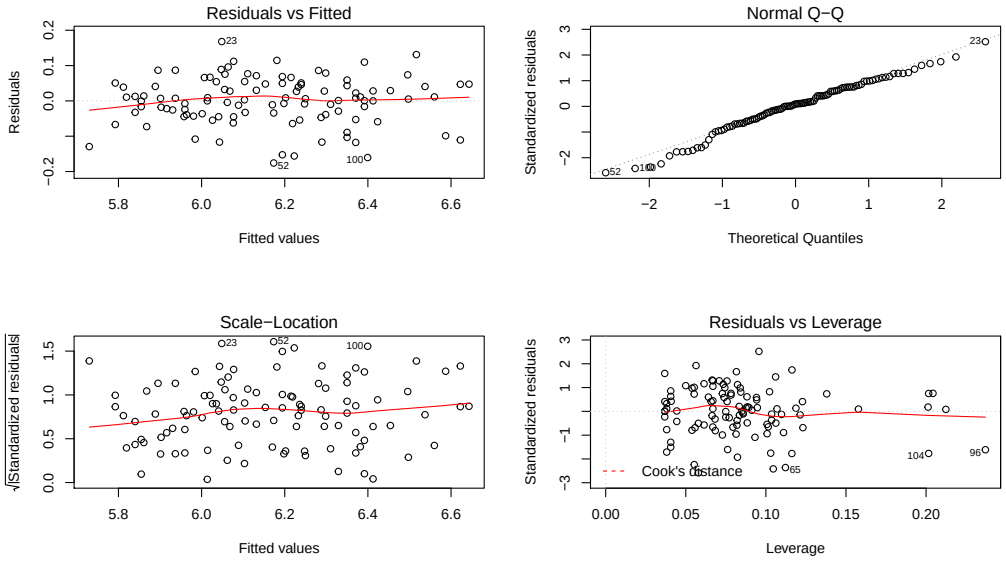
Stratification:

```
##
## Call:
## lm(formula = sal9495 ~ exper + dept + clin + cert, data = male_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.173529 -0.043958 -0.000524  0.045339  0.176938
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        6.0738401  0.0222569 272.897  < 2e-16 ***
## exper              0.0118970  0.0008179  14.546  < 2e-16 ***
## deptGenetics       0.1132455  0.0262915   4.307 3.03e-05 ***
## deptMedicine       0.2580777  0.0191627  13.468  < 2e-16 ***
## deptPediatrics     0.0984634  0.0261797   3.761 0.000245 ***
## deptPhysiology    -0.0692456  0.0193867  -3.572 0.000482 ***
## deptSurgery        0.4566300  0.0214087  21.329  < 2e-16 ***
## clinResearch      -0.0796016  0.0156645  -5.082 1.14e-06 ***
## certNot certified -0.1156107  0.0140047  -8.255 8.55e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06552 on 145 degrees of freedom
## Multiple R-squared:  0.9315, Adjusted R-squared:  0.9277
## F-statistic: 246.5 on 8 and 145 DF,  p-value: < 2.2e-16
```




```
##
## Call:
## lm(formula = sal9495 ~ exper + dept + clin + cert, data = female_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.175835 -0.039097  0.005635  0.047279  0.167929
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.976891   0.030766 194.267  < 2e-16 ***
## exper              0.020835   0.001774  11.748  < 2e-16 ***
## deptGenetics       0.085369   0.027196   3.139 0.002246 **
## deptMedicine       0.268917   0.025088  10.719  < 2e-16 ***
## deptPediatrics     0.113209   0.027126   4.173 6.55e-05 ***
## deptPhysiology    -0.082054   0.023029  -3.563 0.000571 ***
## deptSurgery        0.478852   0.039957  11.984  < 2e-16 ***
## clinResearch      -0.138052   0.016656  -8.289 6.51e-13 ***
## certNot certified -0.068639   0.017731  -3.871 0.000197 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07011 on 97 degrees of freedom
## Multiple R-squared:  0.9109, Adjusted R-squared:  0.9035
## F-statistic: 123.9 on 8 and 97 DF,  p-value: < 2.2e-16
```



Results

Conclusions

Discussion

Figures and tables

	Female (N=106)	Male (N=155)	Total (N=261)	p value
Department				< 0.001
- Biochemistry	20 (18.9%)	30 (19.4%)	50 (19.2%)	
- Genetics	11 (10.4%)	10 (6.5%)	21 (8.0%)	
- Medicine	30 (28.3%)	50 (32.3%)	80 (30.7%)	
- Pediatrics	20 (18.9%)	10 (6.5%)	30 (11.5%)	
- Physiology	20 (18.9%)	20 (12.9%)	40 (15.3%)	
- Surgery	5 (4.7%)	35 (22.6%)	40 (15.3%)	
- Missing	0	0	0	
Clinical				0.197
- Clinical	60 (56.6%)	100 (64.5%)	160 (61.3%)	
- Research	46 (43.4%)	55 (35.5%)	101 (38.7%)	
- Missing	0	0	0	
Certified				0.074
- Board certified	70 (66.0%)	118 (76.1%)	188 (72.0%)	
- Not certified	36 (34.0%)	37 (23.9%)	73 (28.0%)	
- Missing	0	0	0	
Publication Rate				0.004
- Mean (SD)	5.350 (1.886)	4.646 (1.938)	4.932 (1.944)	
- Median (IQR)	5.250 (3.725, 7.275)	4.000 (3.100, 6.700)	4.400 (3.200, 6.900)	
- Min - Max	2.400 - 8.700	1.300 - 8.600	1.300 - 8.700	
- Missing	0	0	0	
Years since obtaining MD				< 0.001
- Mean (SD)	7.491 (4.166)	12.103 (6.704)	10.230 (6.227)	
- Median (IQR)	7.000 (5.000, 10.000)	10.000 (7.000, 15.000)	9.000 (6.000, 14.000)	
- Min - Max	1.000 - 23.000	2.000 - 37.000	1.000 - 37.000	
- Missing	0	0	0	
Rank				< 0.001

(continued)

	Female (N=106)	Male (N=155)	Total (N=261)	p value
- Assistant	69 (65.1%)	43 (27.7%)	112 (42.9%)	
- Associate	21 (19.8%)	43 (27.7%)	64 (24.5%)	
- Full professor	16 (15.1%)	69 (44.5%)	85 (32.6%)	
- Missing	0	0	0	
Salary in academic year 1994				< 0.001
- Mean (SD)	118871.274 (56168.006)	177338.761 (85930.540)	153593.345 (80469.667)	
- Median (IQR)	108457.000 (75774.500, 143096.000)	155006.000 (109687.000, 231501.500)	133284.000 (90771.000, 200543.000)	
- Min - Max	34514.000 - 308081.000	52582.000 - 428876.000	34514.000 - 428876.000	
- Missing	0	0	0	
Salary after increment to Sal94				< 0.001
- Mean (SD)	130876.915 (62034.507)	194914.090 (94902.728)	168906.655 (88778.425)	
- Median (IQR)	119135.000 (82345.250, 154170.500)	170967.000 (119952.500, 257163.000)	148117.000 (99972.000, 218955.000)	
- Min - Max	38675.000 - 339664.000	58923.000 - 472589.000	38675.000 - 472589.000	
- Missing	0	0	0	

References

Appendix