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.

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

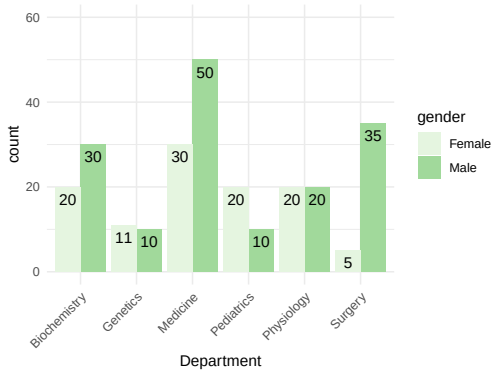
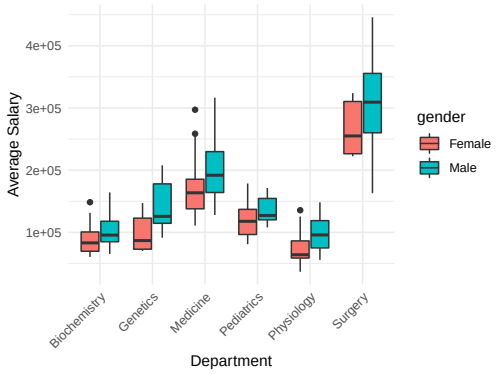


Fig.1.2 Average Salary across Department & Gender



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

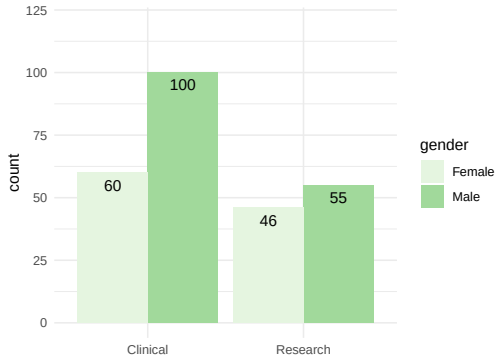
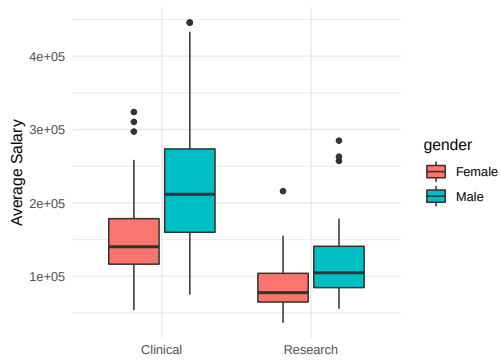


Fig.2.2 Average Salary across clin & gender



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

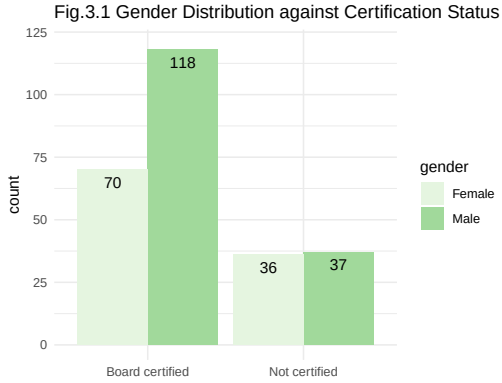
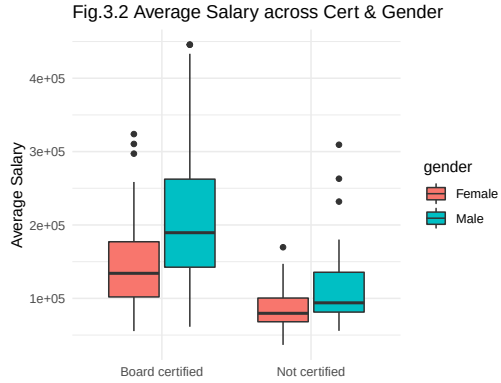


Fig.3.2 Average Salary across Cert & Gender



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

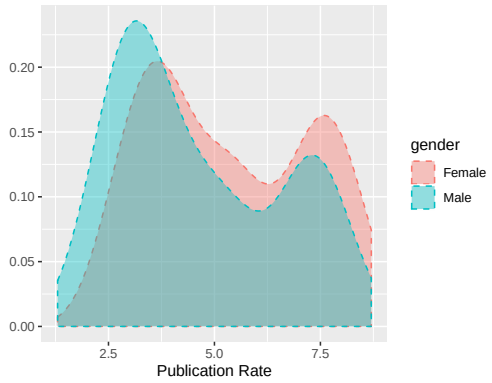
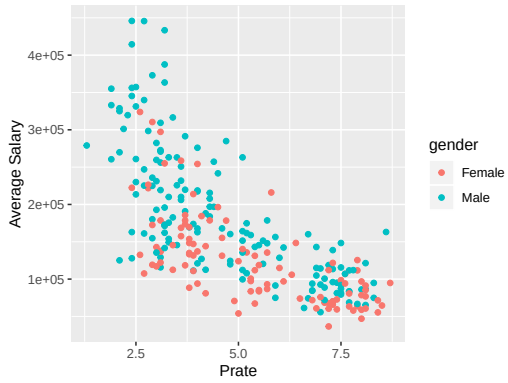


Fig.4.2 Average Salary against Prate across Gender

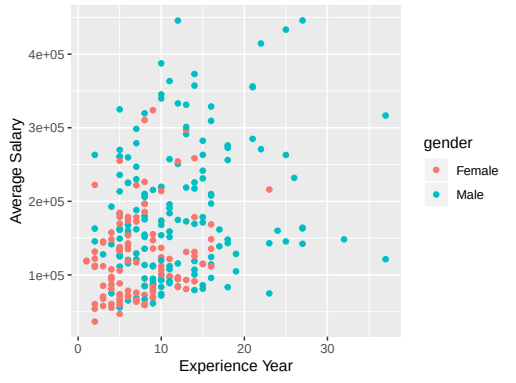


Exper

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 acros



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

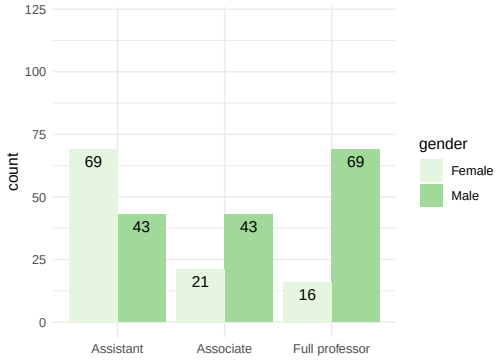
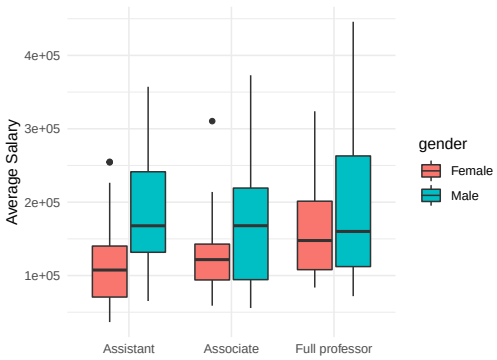
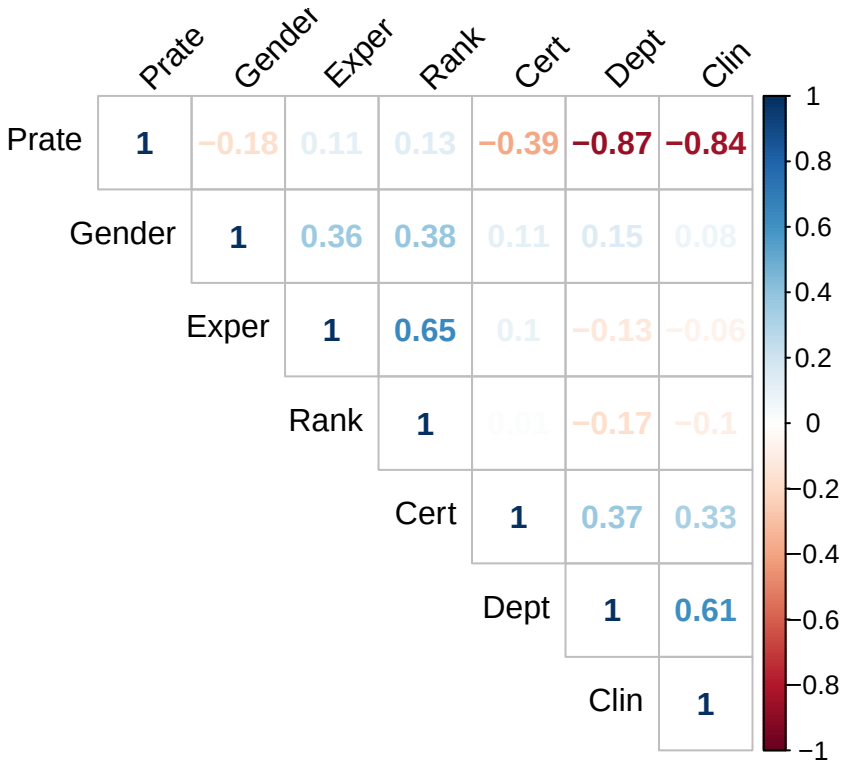


Fig.6.2 Average Salary across Rank & Gender



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.

Results

Conclusions

Discussion

Figures and tables

Table 1: Table 1

	Female (N=106)	Male (N=155)	Total (N=261)
dept			
- 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
clin			
- Clinical	60 (56.6%)	100 (64.5%)	160 (61.3%)
- Research	46 (43.4%)	55 (35.5%)	101 (38.7%)
- Missing	0	0	0
cert			
- 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
prate			
- 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
exper			
- Mean (SD)	7.491 (4.166)	12.103 (6.704)	10.230 (6.227)

	Female (N=106)	Male (N=155)	Total (N=261)
- 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			
- 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
sal94			
- 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
sal95			
- 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