P8130 Final Project

Group 15: Jun Zhai (jz3181); Di Wu (dw2903)

Fangming Jin (fj2269); Xinru Wang (xw2676)

Abstract

Background Strong evidence suggested that women are less likely to earn as much as men both in industry and academia, and this association could be modified by age, rank, experience, marital status, etc.

Purpose Our project aimed to discover the association between salary and gender, as well as the potential confounders and interactions.

Methods A multiple linear regression model adjusting for confounders and interactions was built, then conducted by model diagnostics and stratified analyses.

Results The gender-pay gap was found significantly in multiple linear model, recognizing that gender was confounded by clinical emphasis, certification, department, rank, experience, and modified by rank and experience. After stratifying with rank and experience, we found gender's coefficient is only significant in rank stratification.

Conclusion/Discussion Gender-pay gap was significant in crude model and some situations after stratification, but we could not firmly conclude that it still existed after adjusting for all confoundings and interactions. Further study is needed to justify plaintiff's claim.

Introduction

The gender pay gap has persistently been a great issue in the United States since the 20th century. Back in the 1980s, full-time-working women only earned a median salary of 36 thousand dollars per year, which was only 65% of men's earnings. Although the gender earnings ratio has kept narrowing during the last decades, women's average wage is still 20% less than men's wages nowadays, and the gender gap of salary remained the same from 2003 to 2018.[1]

The same issue persists in academia as well. In unionized, public, liberal arts college, the 1998-1999 salary of male tenured faculty was 51 thousand dollars while female tenured faculty only earned 44 thousand dollars.[2] Also, the association between gender and salary could be modified by other factors such as marital status, experience, faculty rank, age and research productivity.[3,4] According to U.S. National Science Foundation[5], among PhD holders across all fields, the median salary of men is 92 thousand dollar, compared with women's 74 thousand dollar, which shows a 24% gap. The existing gender inequity in academia not only lowers the wage of female faculty, but also limits career promotion

and choices of path. Previous research implicated significant differences in promotion rates between male and female faculty in various fields[6,7], also higher quit rates, a.k.a. leaky pipeline, among women[8,9].

With more strong evidence showing inevitable structural career barriers for women, more and more researchers are getting aware of the gender disparity and call for immediate action to narrow the gender pay gap and advocate gender equity in universities and institutions. In our project, we will be analyzing the data of faculty earning in college, and discovering the association between annual salary of 1994/1995 and gender, with respect to other covariates such as department, certification, publication rate, job titles, work emphasis and experience, as well as the potential confounding and interactions between gender and other factors.

Methods

Salary Data

The data was provided by plaintiff in the case to show that the college has engaged in a pattern and practice of discrimination against women in giving promotions and setting salaries. It obtained 261 observations of 10 variables. Variables recorded detailed information related to salary, gender and career.

The first step of data analysis was cleaning, renaming variables, converting categorical variables into factor, and preparing for the following analyses. A new variable with mean of two years' salaries was built and considered to be our outcome. The method of linear regression was used to identify correlation between gender and salary. To achieve it, the assumption of linearity has to be achieved: the normality of outcome distribution. Through box-cox transformation, the non-normal dependent variables were transformed into a normal shape. To prevent collinearity, the correlation matrix was then computed to testify whether a strong association between different predictors exist after removing some unnecessary variables.

Statistical Analysis

We identified confounders and interactions of the main covariate of interest. Confounders are the variables that distribute unevenly between female group and male group, which have effects on salary and distort the real influence of gender on salary. Confounders are detected if the coefficient of gender in linear regression models differ by a relative magnitude of at least 10%.

The interactions exist when the effects of gender and other variables taken together, which are greater than the sum of their separate effects at the same doses. The magnitude of effect of gender on salary is modified by these variables.

According to previous literature mentioned above in the introduction part[3,4], we filtered several potential interactions. Through method of nested F-test, we testified if there were significant differences between models with and without interactions, adjusting for confounders. After recognizing all the confounders and interactions, they were applied in the model together with gender, obtaining the true effects of gender on salary. Upon stratification, we detected true effects of gender on salary over the presence of interactions.

Model Diagnostics and Influential Points

Residual vs fitted value plot, Q-Q plot, scale-location plot and residual vs leverage plot were applied to verify the assumptions made regarding residual. In addition, one influential observation found to be an outlier that caused major change on the slope of the regression line by two criterions, DFFITS and cook's distance, and then was diagnosed as a false interpretation of the population after refitting the model without it.

Results

Unadjusted Analysis

Sample characteristics of faculty at the school are presented in **Table 1.** This sample contains 261 faculties at Houston College of Medicine, of which 106 (40.6%) were women and 155 (59.4%) were men. Mean salary of females in this sample is 118871.3, which is much smaller than that of males in this sample as 177338.8. In addition, there are differences in years as obtaining MD between males and females. The proportion of assistant in female faculties and the distributions of departments among two genders in this sample also varies. The distribution of female salary in 1994 was left skewed compared with that of male (**Figure 1**). For example, there were more female earned less than $150000 annually, and less female earned more than $250000 compared with male.

In general, salary of male is distributed over that of female (**Figure 2**). In unadjusted gender salary gap estimation process, a model with only gender as its predictor and salary as its outcome was built. The p-value of gender dummy variable is significant (**Figure 4**), which means that male on average earn more than women on average in this sample when not controlling other variables.

Adjusted Analysis

The high correlation variable was firstly identified and deleted: prate (**Figure 3**). During the modeling process, dept, clin, cert, rank and exper were selected as confounders (**Figure 5**), and rank as well as exper were selected as interactions (**Table 2**). After recognizing and removing the influential point (**Figure 6, 7**), gender was still a significant predictor of salary in the model (**Table 3**), which implied that gender salary gap still exists adjusting for interactions and confounders. There are two interactions modifying association between salary and gender: rank and exper. After stratifying this dataset according to these two interactions (**Table 4**), gender salary gap showed a significant trend in associate professor group.

Conclusion/Discussion

Gender is a significant predictor of salary only in Associate Professor group after stratification according to rank and exper, indicating that more evidence is needed to be presented to support the plaintiffs' claim that female faculty in this College are experiencing discrimination in salary.

The rank variance between male and female group, which males tended to earn more salary in a higher rank compared with females, caused gender a non-significant predictor in stratified model. After stratifying according to rank, the gender salary gap disappeared, which proved this assumption. The same condition applied to experience factor. Therefore, further studies are needed to be done to explore the association between gender and rank as well as gender and experience in order to draw the conclusion: whether gender is a significant predictor of salary.

This study has several limitations: First, there may be other factors that can be a confounder or interact with gender, such as part-time and full-time status or marital status as mentioned in previous study.[4] Second, there is not enough data to support the plaintiffs' claim, given the sample size is not big enough after stratifying. In addition, whether this data is true and effective remains to be verified given that there may be other payments to faculties other than reported incomes (eg, incentive payments, clinical incomes).

Figures and Tables

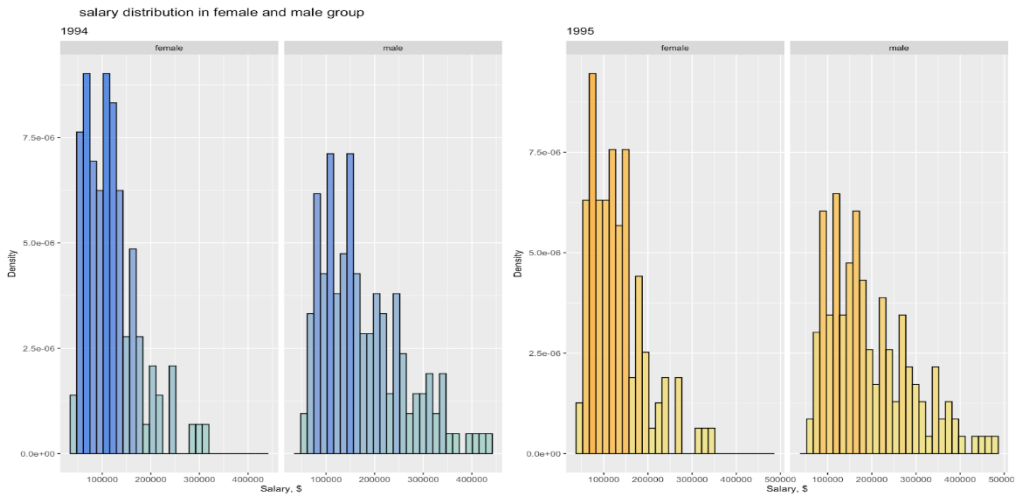

Figure 1: Histogram Plot of Salary Distribution in Female and Male

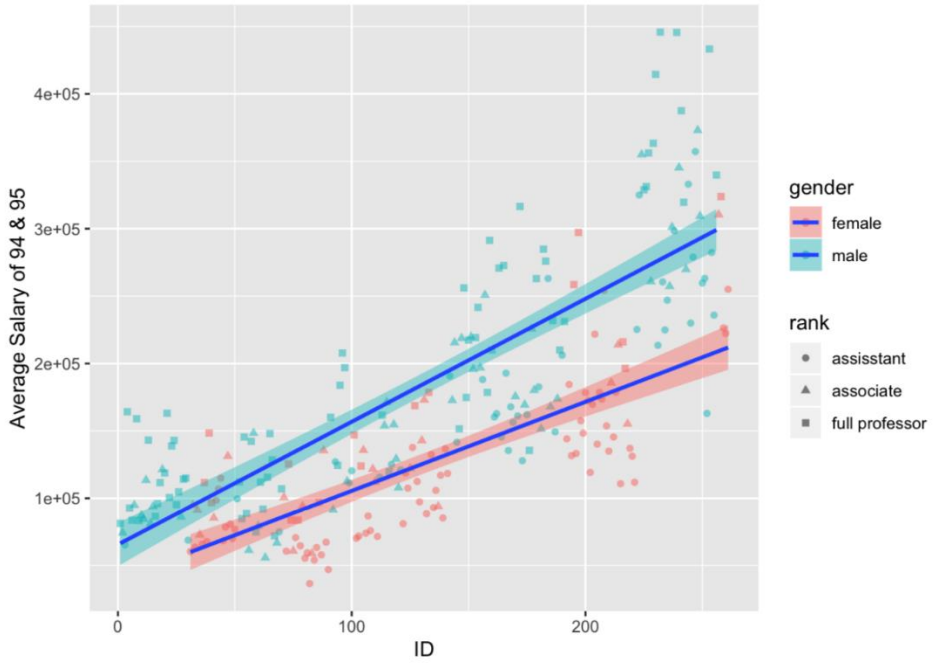

Figure 2: Scatter Plot of Salary Distribution in Female and Male

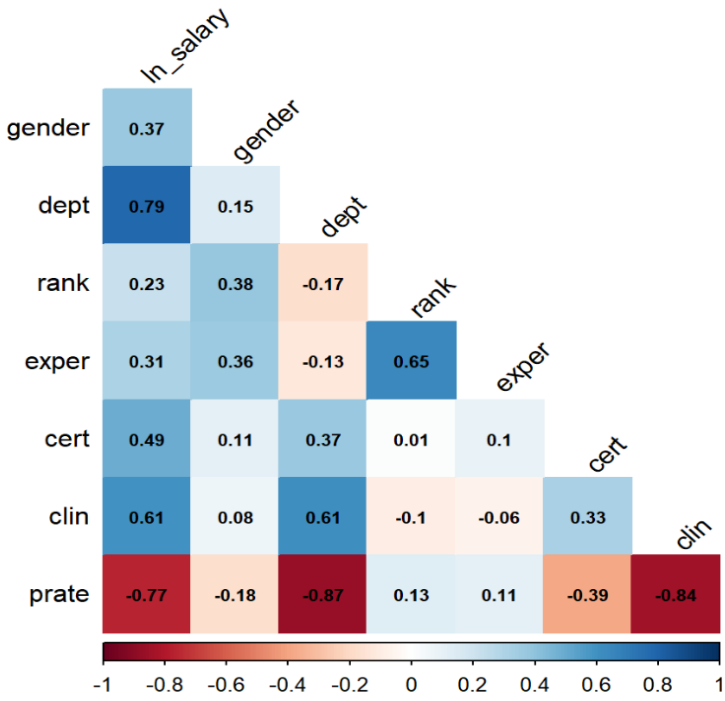

Figure 3: Correlation Plot

		ln salary	
Predictors	*Estimates*	*CI*	*p*
(Intercept)	11.63	11.54 – 11.72	**<0.001**
gender [male]	0.39	0.27 – 0.50	**<0.001**
Observations	261		
R^2 / R^2 adjusted	0.139 / 0.136		

Figure 4: p-value of Gender Variable

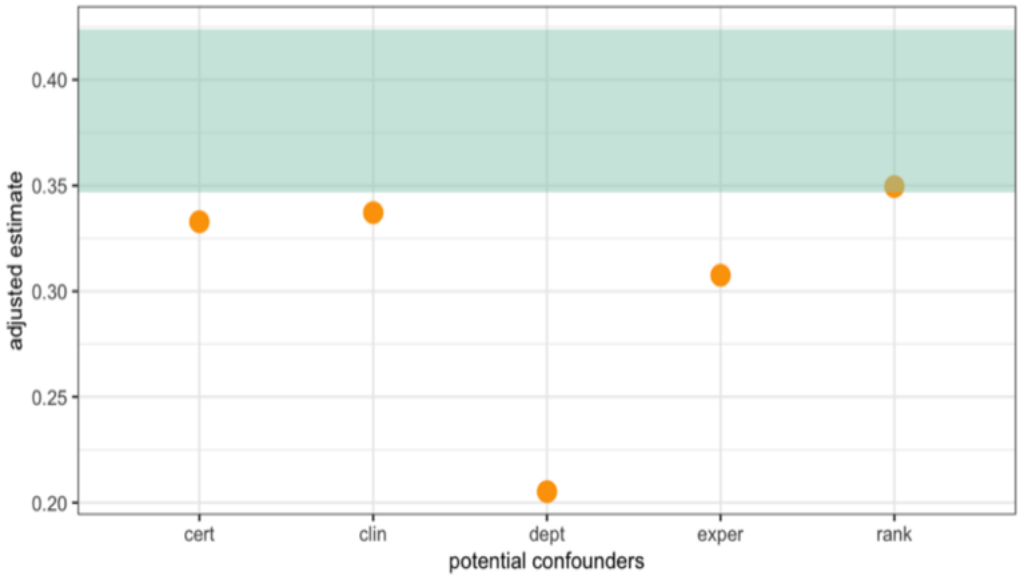

Figure 5: Identification of Confounders

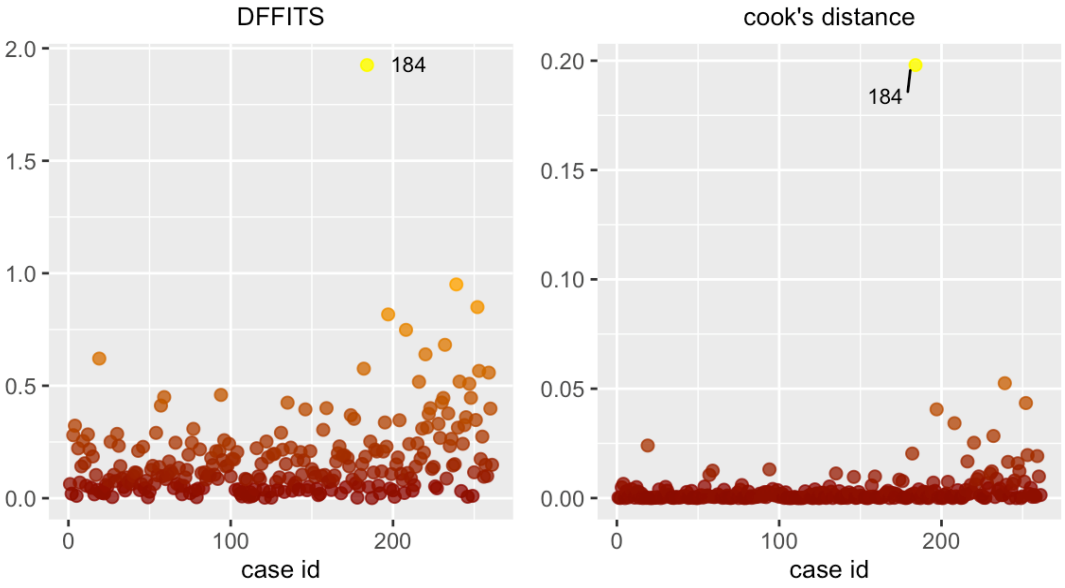

Figure 6: Identification of Outliers

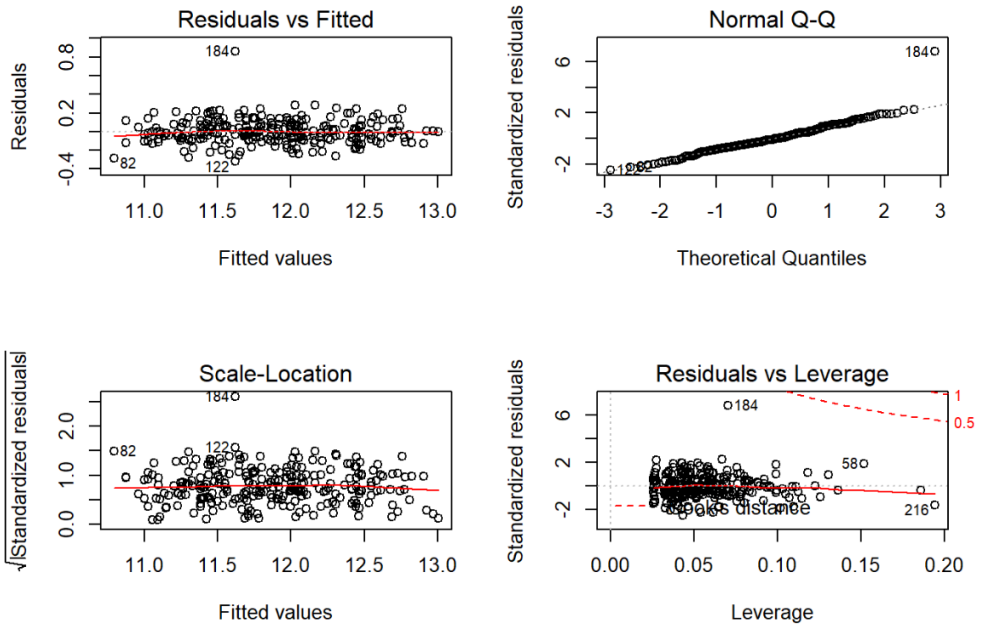

Figure 7: Model Diagnostics

Figure 8: Scatter Plot of Stratification Model by Gender

	female (n=106)	male (n=155)	Overall (n=261)
salary in 94			
Mean (SD)	118871.3 (± 56168.01)	177338.8 (± 85930.54)	153593.3 (± 80469.67)
Median	108457.0	155006.0	133284.0
Range	34514.00~308081.0	52582.00~428876.0	34514.00~428876.0
salary in 95			
Mean (SD)	130876.9 (± 62034.51)	194914.1 (± 94902.73)	168906.7 (± 88778.43)
Median	119135.0	170967.0	148117.0
Range	38675.00~339664.0	58923.00~472589.0	38675.00~472589.0
prate			
Mean (SD)	5.350000 (± 1.886052)	4.646452 (± 1.937909)	4.932184 (± 1.944430)
Median	5.250000	4.000000	4.400000
Range	2.400000~8.700000	1.300000~8.600000	1.300000~8.700000
exper			
Mean (SD)	7.490566 (± 4.166180)	12.10323 (± 6.703531)	10.22989 (± 6.227176)
Median	7.000000	10.00000	9.000000
Range	1.000000~23.00000	2.000000~37.00000	1.000000~37.00000
rank			
assistant	69 (65 %)	43 (28 %)	112 (43 %)
associate	21 (20 %)	43 (28 %)	64 (25 %)
full professor	16 (15 %)	69 (45 %)	85 (33 %)
clin			
primarily research emphasis	46 (43 %)	55 (35 %)	101 (39 %)
primarily clinical emphasis	60 (57 %)	100 (65 %)	160 (61 %)
cert			
not certified	36 (34 %)	37 (24 %)	73 (28 %)
Board certified	70 (66 %)	118 (76 %)	188 (72 %)
dept			
biochemistry/molecular biology	20 (19 %)	30 (19 %)	50 (19 %)
physiology	20 (19 %)	20 (13 %)	40 (15 %)
genetics	11 (10 %)	10 (6 %)	21 (8 %)
pediatrics	20 (19 %)	10 (6 %)	30 (11 %)
medicine	30 (28 %)	50 (32 %)	80 (31 %)
surgery	5 (5 %)	35 (23 %)	40 (15 %)

Dept: Department; Gender: Male/Female; Clin: Primarily clinical emphasis/Primarily research emphasis; Cert: Board certified/ Not certified; Prate: Publication rate (# publications on cv)/ (# years between CV date and MD date); Exper: # years since obtaining MD; Rank: Proxy of productivity; Sal94: Salary in academic year 1994; Sal95: Salary in academic year 1995

Table 1: Descriptive Statistics

term	interaction				
	dept	clin	cert	exper	rank
rss	4.4505618	4.4505618	4.4505618	4.4505618	4.4505618
df	-5.0000000	-1.0000000	-1.0000000	-1.0000000	-2.0000000
sumsq	-0.0465668	-0.0177407	-0.0022589	-0.1845845	-0.1117068
statistic	0.5159994	0.9925281	0.1259354	10.7307096	3.1795908
p.value	0.7640712	0.3200963	0.7229858	0.0012042	0.0433110

Table 2: Interaction Summary

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	10.9004715	0.0331431	328.8913294	0.0000000
factor(gender)male	0.1005267	0.0347250	2.8949406	0.0041345
factor(rank)associate	0.1382123	0.0330148	4.1863717	0.0000395
factor(rank)full professor	0.2136181	0.0443422	4.8174900	0.0000025
exper	0.0266056	0.0038870	6.8447732	0.0000000
factor(clin)primarily clinical emphasis	0.2259130	0.0203561	11.0980719	0.0000000
factor(dept)physiology	-0.1723819	0.0261820	-6.5839941	0.0000000
factor(dept)genetics	0.1836112	0.0324994	5.6496754	0.0000000
factor(dept)pediatrics	0.2001479	0.0322757	6.2011852	0.0000000
factor(dept)medicine	0.5205216	0.0266542	19.5286875	0.0000000
factor(dept)surgery	0.9228489	0.0318519	28.9730820	0.0000000
factor(cert)Board certified	0.1980533	0.0196805	10.0634088	0.0000000
factor(gender)male:factor(rank)associate	-0.0111982	0.0438880	-0.2551534	0.7988187
factor(gender)male:factor(rank)full professor	0.0021574	0.0547227	0.0394251	0.9685836
factor(gender)male:exper	-0.0096758	0.0042651	-2.2685735	0.0241650

Table 3: Model Summary After Removing Outliers

Table 4: Stratified Model Summary

	ln salary full professor			ln salary associate			ln salary assistant			ln salary exper 1			ln salary exper 2			ln salary exper 3			ln salary exper 4		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	11.18	10.94 – 11.42	**<0.001**	10.93	10.76 – 11.09	**<0.001**	10.92	10.81 – 11.03	**<0.001**	10.98	10.77 – 11.18	**<0.001**	11.10	11.01 – 11.20	**<0.001**	11.22	11.12 – 11.32	**<0.001**	11.10	10.89 – 11.32	**<0.001**
gender [male]	0.03	-0.22 – 0.29	0.790	0.17	0.00 – 0.33	**0.049**	0.03	-0.09 – 0.15	0.618	0.05	-0.04 – 0.15	0.247	0.03	-0.05 – 0.12	0.399	0.02	-0.17 – 0.20	0.839	0.10	-0.17 – 0.38	0.456
exper	0.02	0.00 – 0.04	**0.023**	0.04	0.02 – 0.06	**<0.001**	0.03	0.02 – 0.04	**<0.001**												
dept [physiology]	-0.13	-0.20 – -0.05	**0.002**	-0.19	-0.27 – -0.10	**<0.001**	-0.22	-0.33 – -0.11	**<0.001**	-0.18	-0.40 – 0.04	0.101	-0.20	-0.33 – -0.08	**0.002**	-0.19	-0.29 – -0.09	**<0.001**	-0.08	-0.19 – 0.04	0.181
dept [genetics]	0.25	0.14 – 0.36	**<0.001**	0.18	0.07 – 0.28	**0.002**	0.14	0.03 – 0.26	**0.016**	0.25	0.00 – 0.50	**0.047**	0.14	-0.01 – 0.30	0.072	0.13	0.02 – 0.24	**0.026**	0.22	0.06 – 0.39	**0.007**
dept [pediatrics]	0.16	0.03 – 0.30	**0.017**	0.27	0.15 – 0.38	**<0.001**	0.20	0.08 – 0.31	**0.001**	0.25	0.01 – 0.48	**0.041**	0.25	0.10 – 0.40	**0.001**	0.17	0.04 – 0.29	**0.010**	0.17	-0.00 – 0.34	0.053
dept [medicine]	0.53	0.45 – 0.61	**<0.001**	0.53	0.44 – 0.63	**<0.001**	0.53	0.42 – 0.63	**<0.001**	0.55	0.32 – 0.79	**<0.001**	0.58	0.45 – 0.70	**<0.001**	0.53	0.42 – 0.64	**<0.001**	0.52	0.42 – 0.63	**<0.001**
dept [surgery]	0.95	0.85 – 1.05	**<0.001**	0.96	0.85 – 1.08	**<0.001**	0.89	0.77 – 1.01	**<0.001**	0.96	0.71 – 1.22	**<0.001**	0.95	0.81 – 1.10	**<0.001**	0.94	0.81 – 1.06	**<0.001**	0.97	0.83 – 1.11	**<0.001**
cert [Board ified]	0.25	0.18 – 0.32	**<0.001**	0.18	0.12 – 0.24	**<0.001**	0.16	0.09 – 0.23	**<0.001**	0.18	0.09 – 0.27	**<0.001**	0.17	0.07 – 0.27	**0.001**	0.18	0.10 – 0.27	**<0.001**	0.28	0.18 – 0.38	**<0.001**
clin [primarily ical emphasis]	0.19	0.12 – 0.25	**<0.001**	0.21	0.14 – 0.28	**<0.001**	0.25	0.17 – 0.32	**<0.001**	0.24	0.12 – 0.35	**<0.001**	0.18	0.07 – 0.29	**0.002**	0.28	0.20 – 0.36	**<0.001**	0.14	0.05 – 0.24	**0.002**
gender [male] * exper	-0.01	-0.02 – 0.01	0.544	-0.02	-0.04 – -0.00	**0.024**	0.00	-0.02 – 0.02	0.873												
rank [associate]										0.03	-0.19 – 0.26	0.764	0.15	0.04 – 0.26	**0.009**	0.12	-0.01 – 0.24	0.064	0.34	0.05 – 0.63	**0.021**
gender [male] : factor(rank)associate										0.05	-0.25 – 0.34	0.749	-0.01	-0.16 – 0.14	0.917	-0.06	-0.28 – 0.15	0.569	-0.12	-0.47 – 0.22	0.482
rank [full professor]													0.16	-0.08 – 0.39	0.180	0.16	0.02 – 0.31	**0.030**	0.41	0.18 – 0.64	**0.001**
gender [male] : factor(rank)full professor													0.03	-0.23 – 0.29	0.825	0.02	-0.21 – 0.25	0.860	-0.12	-0.42 – 0.18	0.421
Observations	85			64			111			63			57			74			66		
R^2 / R^2 adjusted	0.951 / 0.945			0.966 / 0.959			0.938 / 0.932			0.930 / 0.917			0.959 / 0.948			0.947 / 0.937			0.929 / 0.912		

Reference

1. Hegewisch A, DuMonthier A. The Gender Wage Gap: 2018 Earnings Differences by Race and Ethnicity. Institute for woman's policy research.

2. Burke K, Duncan K, Krall L, Spencer D. Gender differences in faculty pay and faculty salary compression. *Soc Sci J*. 2005. doi:10.1016/j.soscij.2005.03.006

3. Jena AB, Olenski AR, Blumenthal DM. Sex differences in physician salary in US public medical schools. *JAMA Intern Med*. 2016. doi:10.1001/jamainternmed.2016.3284

4. Toutkoushian RK, Bellas ML, Moore J V. The interaction effects of gender, race, and marital status on faculty salaries. *J Higher Educ*. 2007. doi:10.1353/jhe.2007.0031

5. National Science Foundation, National Center for Education Statistics. Women, minorities, and persons with disabilities in science and engineering. *Spec Rep NSF 17-310*. 2017. doi:Special Report NSF 17-310

6. Ginther DK, Hayes KJ. Gender differences in salary and promotion in the humanities. *Am Econ Rev*. 1999. doi:10.1257/aer.89.2.397

7. McDowell JM, Singell LD, Ziliak JP. Cracks in the glass ceiling: Gender and promotion in the economics profession. *Am Econ Rev*. 1999. doi:10.1257/aer.89.2.392

8. Gasser CE, Shaffer KS. Career Development of Women in Academia: Traversing the Leaky Pipeline. *Prof Couns*. 2014. doi:10.15241/ceg.4.4.332

9. Blickenstaff JC. Women and science careers: Leaky pipeline or gender filter? *Gend Educ*. 2005. doi:10.1080/09540250500145072