Assignment 1

- Mohammad Ahmed - Shahana Ayobi

2022-11-25

Github: https://github.com/ayobishahana/Data-Analysis-2/tree/main/Assignment%201

1. Introduction

We are interested to analyze the gender pay gap among **Financial Managers (CC: 0120)** using the Current Population Survey (CPS) from 2014.

2. Data Cleaning

To continue with our analysis, we filtered the data to only Financial Managers occupation and restricted ages from 15-64 since these ages are classified as working age in the U.S.. We then generated a binary variable `female` which takes the value 1 if the observation is female and 0 otherwise, where we have 792 observations for women and 607 for men. Other transformations, such as creating `hourly wage`, and `logarithm of hourly wage`, restricting the `hourly wage` to the value 1 or more have been performed. The education variable (`grade92`) have been categorized into 5 groups: Associate, Bachelor's, Master's, Professional School, and Doctorate degrees.

3. Analysis

The unconditional gender pay gap can be observed from `Table 1` that provides a summary of `female` and `hourly wage` variables. On average, women earn USD9.4 equivalent to 29% less than men and this gap is more apparent in the higher hourly wage percentile. The same result can be seen from running simple linear regressions and OLS robust regressions (`Table 2`).

To determine the wage disparity according to various degrees of education, we ran 5 log-level multivariate regressions accounting for heteroskedastic errors and changed educational base category for each regression. While running the simple regression in `Table 3`, `model 1` shows that women earn 29.3% less than men on average, when conditioned on education levels taking Associate degree as the base category, this gap decreases to 19.9 % less earning than men on average (`Table 3, Model 2`), ceteris paribus. We also went on to investigate the discrepancy in pay between education levels. `Model 2` compares women with various levels of education, using the associate's degree as the base variable. For instance, female employees with bachelor's degree tend to earn 30.3% more on average compared to women with Associate degrees and it is significant at 1% significance level. The rest of the models follow a similar pattern where the coefficients show that higher levels of education result in a higher wage in females when changing the base categories for educational levels. However, `Table 3, Model 2` tends to be a better fit for our analysis when taking associate degree as the base where 19% variation in wage is explained by the explanatory variables in the model.

We now run a regression with interaction terms on the same base variable (`Table 4`). Although the `female` variable is still significant, the interactions are not, meaning that having higher levels of education for women

Table 1: Summary of Wages for Both Genders

as.factor(female)		Mean	SD	Min	Max	P25	P75	N
0	wage	38.44	16.43	2.88	115.38	25.48	50.00	607
1	wage	29.07	14.43	2.00	77.96	18.75	36.32	792

Table 2: Simple Regressions Result

	Model 1	Model 2	Model 3	Model 4
(Intercept)	3.5406**	3.5406**	38.4351**	38.4351**
	(0.0206)	(0.0203)	(0.6221)	(0.6668)
female	−0.2934**	−0.2934**	−9.3603**	−9.3603**
	(0.0274)	(0.0273)	(0.8269)	(0.8411)
Num.Obs.	1399	1399	1399	1399
R2	0.076	0.076	0.084	0.084
RMSE	0.51	0.51	15.32	15.32

* p < 0.05, ** p < 0.01

compared to men does not have effect on earnings. However, we will need higher number of observations and more relevant variables in the model to confirm this relationship.

In conclusion, women earn less than men in the Financial Managers sectors compared to men, higher education levels yield more earnings compared to lower education levels. However, the result of interaction term suggest while having the same education levels men and women's earnings do not differ which is to be investigated further to confirm.

Figure 1: Non−parametric Regression − Lnwage ~ Education Levels

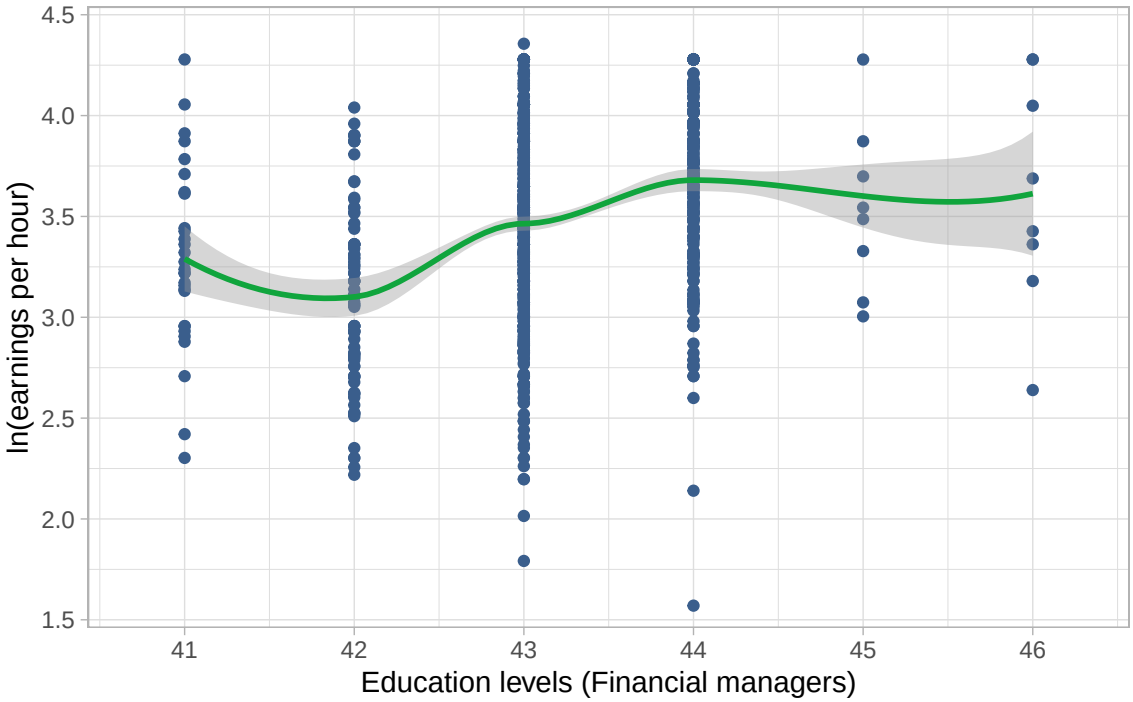

Table 3: Multivariate Regressions with Different Base Catagories for Education

	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	3.541 ***	3.258 ***	3.466 ***	3.489 ***	3.260 ***	3.263 ***
	(0.020)	(0.030)	(0.022)	(0.027)	(0.033)	(0.033)
female	-0.293 ***	-0.194 ***	-0.239 ***	-0.272 ***	-0.196 ***	-0.199 ***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
ed_BA		0.303 ***		0.113 ***	0.302 ***	0.301 ***
		(0.030)		(0.028)	(0.033)	(0.033)
ed_MA		0.489 ***	0.298 ***		0.488 ***	0.486 ***
		(0.037)	(0.032)		(0.040)	(0.040)
ed_Profess		0.399 **	0.219	0.217		0.398 **
		(0.121)	(0.116)	(0.115)		(0.122)
ed_PhD		0.403 *	0.206	0.192	0.401 *	
		(0.187)	(0.185)	(0.185)	(0.187)	
ed_Associate			-0.150 **	-0.147 **	0.023	0.023
			(0.048)	(0.050)	(0.052)	(0.052)
N	1399	1399	1399	1399	1399	1399
R2	0.076	0.190	0.133	0.098	0.187	0.187

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 4: Multivariate Regressions with Interaction Terms

	(1)	(2)	(3)
(Intercept)	3.077 ***	3.224 ***	3.224 ***
	(0.027)	(0.045)	(0.045)
ed_BA	0.281 ***	0.347 ***	0.347 ***
	(0.037)	(0.052)	(0.052)
ed_MA	0.463 ***	0.531 ***	0.531 ***
	(0.052)	(0.057)	(0.057)
ed_Profess	0.253 *	0.656 ***	0.656 ***
	(0.122)	(0.193)	(0.193)
ed_PhD	0.267	0.479 **	0.479 **
	(0.501)	(0.183)	(0.183)
female			-0.147 **
			(0.053)
ed_BA:female			-0.066
			(0.064)
ed_MA:female			-0.067
			(0.077)
ed_Profess:female			-0.403
			(0.228)
ed_PhD:female			-0.212
			(0.533)
N	792	607	1399
R2	0.110	0.147	0.192

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.