Data Analysis Assignment 1

Analysis of geneder wage gap in IT management

Gyebnar Daniel

Goal definition and project scope

This project analyzes the differences in salary between man and woman in management level in the "Computer and Information System" sector , with taking educational level into account.

Project scope and key assumptions:

- Data source: CPS Annual Earnings from 2014 (<http://www2.nber.org/data/morg.html>), occ. code 0110
- Dependent variable: Hourly earnings of full time employees from 2014
- Explanatory variables: Gender and Education (no higher education, bachelors, masters, above masters)

Overview of the analysis and models used

The analysis included a preliminary analysis exploring key statistics and distributions, and followed by four linear regressions: (1) hourly earnings and gender, (2) hourly earnings and education, (3) hourly earnings and gender and education, (4) multilingual regression with interactions between gender and education.

Executive summary of key findings

- The conditional mean of hourly earnings is higher for males than for females (Figure 2), and the difference is statistically significant (Figure 3, (1))
- Females earn less than males with no degree, with bachelors and with a masters as well
- More obersvations are needed to lowerthe standard error of the regressions

Interpretation of exploratory data analysis

The distribution of hourly earnings is close to normal, and not skewed (Figure 1). Looking at conditional means per educational level and gender, the mean hourly earnings for males is higher in all key categories (no degree, bachelors, masters), and higher for females only with above masters (Figure 3)

Linear regression results (Figure 4)

(1): The mean of hourly wages for males is 41.051 USD, and for females is 4.106 USD lower

(2): The mean of hourly wages for employees with no degrees is 32.993 USD. The slope coefficient shows how much higher the wage is per education level, compared to people with no degree

(3): Females on average earn 4.397 USD less for all degree levels than males

(4): Females on avg. with no degree earn 2.082 USd less, with bachelors 2.082+1.651=3.733 less, with masters 2.082+5.885=7.967 less, and above masters |2.082-4.775|=2.693 more

The standard errors are high due to the small number of observations, which is especially visible for (4)

Major decision points

- I chose to analyse wages on management level, as it is highly competitive and promotions play a key role
- I focused on full-time employees only, as in management non-full-time positions are special occurrences
- I grouped education in 4 bins: No higher education, Bachelors, Masters, Above masters. In multilinear model I chose "no higher education" as a basis to see how degrees elevate earnings. I also grouped everything above masters together due to the low number of observations in these segments
- In modeling, for the dependent I used level hourly wage, as its distribution was close to normal
- I included a model with hourly earning and education only to see the mean effect of education
- I included interactions between education and gender to see the differences between each subgroup

Appendix

Figure 1: Distribution of hourly earnings:

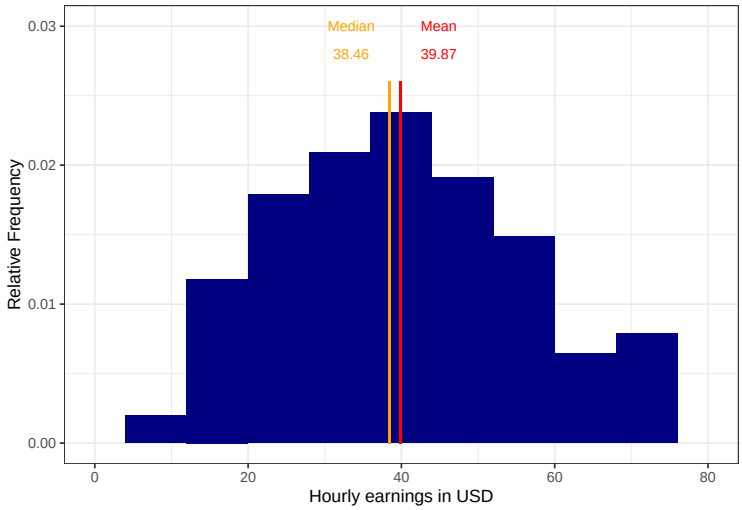
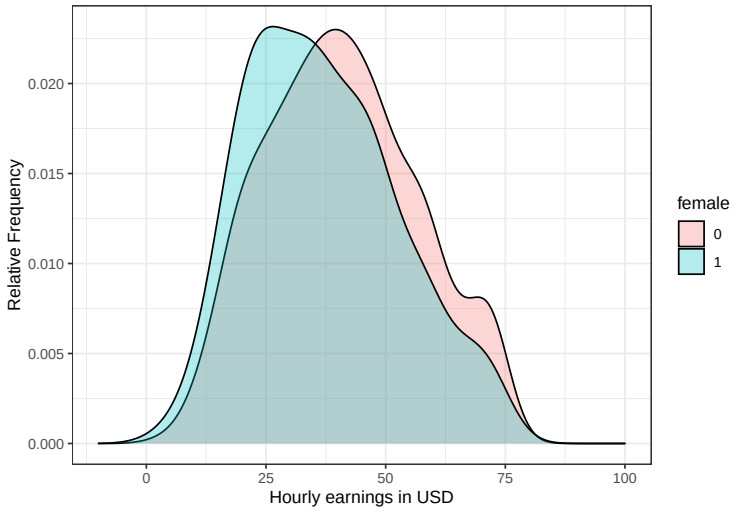


Figure 2: Distribution of hourly earning conditional on gender:



	eduation_bin	female	Mean	Median	SD	Min	Max	P25	P75	N	PercentMissing
w	No degree	0	33.55	30.38	15.23	7.50	72.12	22.57	41.43	128	0
		1	31.47	28.17	14.92	8.97	72.12	20.84	40.17	47	0
	Bachelors	0	42.12	42.30	15.23	3.84	72.12	31.13	52.45	228	0
		1	38.39	38.00	14.87	9.83	72.12	27.22	48.56	91	0
	Masters	0	46.56	48.08	14.88	12.24	72.12	36.01	57.69	129	0
		1	38.59	36.20	16.07	3.80	72.12	25.00	47.28	59	0
	Above masters	0	41.51	38.46	15.40	16.83	72.12	37.00	44.44	13	0
		1	44.20	41.53	20.99	21.62	72.12	34.24	51.48	4	0

Figure 3: Key statistics of wage conditional on gender and educational level:

	(1)	(2)	(3)	(4)
(Intercept)	41.051 ***	32.993 ***	34.174 ***	33.553 ***
	(0.710)	(1.144)	(1.199)	(1.349)
female1	-4.106 **		-4.397 ***	-2.082
	(1.306)		(1.278)	(2.551)
Bachelors		8.061 ***	8.134 ***	8.567 ***
		(1.426)	(1.425)	(1.686)
Masters		11.064 ***	11.263 ***	13.005 ***
		(1.617)	(1.599)	(1.882)
Above_masters		9.147 *	9.000 *	7.954
		(3.986)	(4.048)	(4.342)
Bachelors:female1				-1.651
				(3.156)
female1:Masters				-5.885
				(3.547)
female1:Above_masters				4.775
				(10.350)
N	699	699	699	699
R2	0.014	0.069	0.085	0.090

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

Figure 4: Summary of regression results: