DA_2 Assignment 1

Hasan Mansoor Khan

2022-11-26

Data Analysis 2 - Assignment 1

Hasan Mansoor Khan

26/11/2022

Overview

The scope of my analysis pertains to a specific occupation group which is : General and operations managers. This is represented by occupational code 20. Considering Gender, Wages and the level of education completed, it is expected that a wage gap keeping in mind highest qualification earned can be witnessed. Using various regression models, my analysis will dive into the question regarding wage gaps and gender based on educational level achieved. Is the wage earned between males and females significantly different for General and operations managers and if so does it vary by educational qualification.

Firstly, I create a df importing all observations of the cps earnings dataset. I then filter data for occupational code 20 (General and operations managers). I then consider specific observations such as those having bachelors education or more, those 21 or more. This brings my narrowed df (df_2) to 501 observations. Of these 501, roughly 35% are female that is 174.

Secondly, a comparative analysis can be done by comparing differences in earnings by using the hourly wage rate for General and operations managers. Hence, filtering data for relevant variables, I include the hourly wage rate and the log of hourly wage rate.

Analysis & Interpretation

The first task is to identify the unconditional gender gap. For this purpose, it is imperative to overview the summary statistics which I done by using the data summary function. Then, I proceed to visualize using a histogram which shows a gap in wages between males and females in the narrowed data frame. A regression analysis of this can reveal further about the variation in wages and its magnitude. My goal is to perform the regression analysis and visualize it using a regression table and graphs.

For the regression analysis lm robust function is used for the variables. These exhibits the correlation between the variables, its regression and application of weights using a predefined procedure that takes into account 'influential observations'. For the analysis, I do not exclude any extreme values as there is no immediate reason to do so as far as my analysis goes. Excluding Y values without proper reasion is not good practice. After running 4 regressions using log-level analysis. Log (log of hourly wages) & level (education).

The unconditional gender wage gap is illustarted by a histogram. The histogram shows ln of hour wages on the x axis and number of respondents on y axis. A clear pattern is illustrated where at higher hour wage rates there are more men than women. Further regression analysis will incorporate education and wage gap.

sex		Mean	SD	Min	Max	N
Male	hour_wage	36.85	15.75	7.21	72.12	327
Female	hour_wage	33.05	14.88	2.60	72.12	174

sex		Mean	SD	Min	Max	N
Male	ln_hw	3.50	0.48	1.98	4.28	327
Female	ln_hw	3.39	0.50	0.96	4.28	174

Regression 1:

Assuming other factors remain unchanged and only wages and gender are considered, the first regression identifies that women earn 11.8 percent less than their male counterparts. This is depicted by the Coefficients in the regression. The intercept in regression one is at 3.5. The r square meanwhile stands at 0.013 which is value at the lower end.

Regression 2:

Regression 2, takes into account more factors in an attempt to secure a better R square and include more variables. It includes regression for educational level bachelors, masters & phd. The combined for all 3 categories, shows that women earn approximately 12% less while the specific qualification can cause for different results. For example education level at bachelors shows that women earn 25% less, women with masters earn 10% less and women with PhDs actually earn 20% more than their male counterparts.

Regression 3:

The third regression takes into account regressions by all educational categories including females with bachelors, masters, PhD's as well as professional. For females in general in the four categories, they tend to earn 18% less than their male general and operation managers. However, there is considerable difference between the specific categories. For example, those females with bachelors earn 24% less while those females with masters earn 13% less and PhDs earn 22% more! The r square for this model is 0.05 which again illustrates a low r square. This may be that factors other than gender are not taken into account.

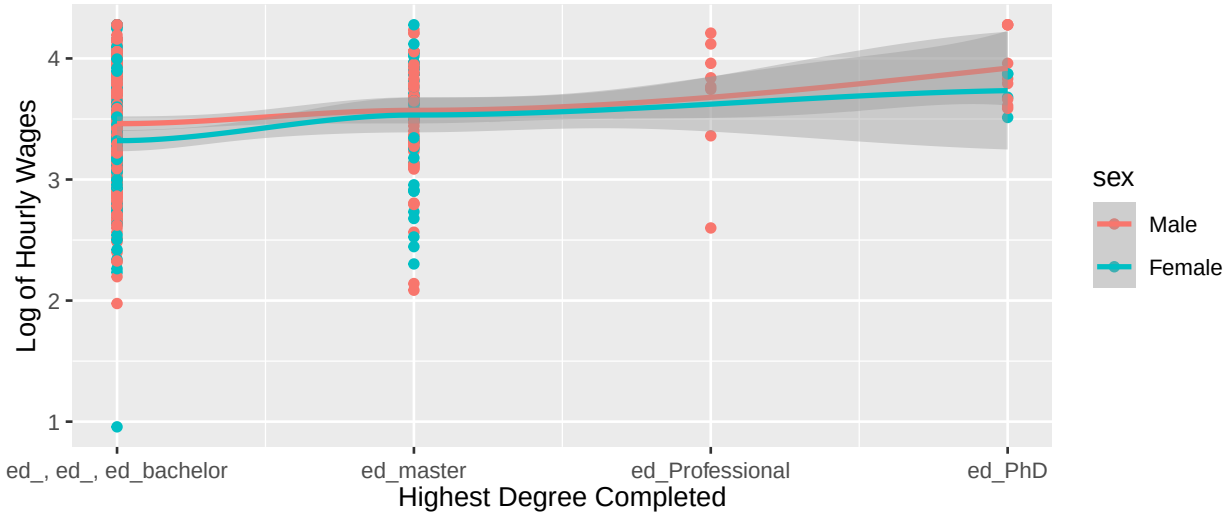
Regression 4:

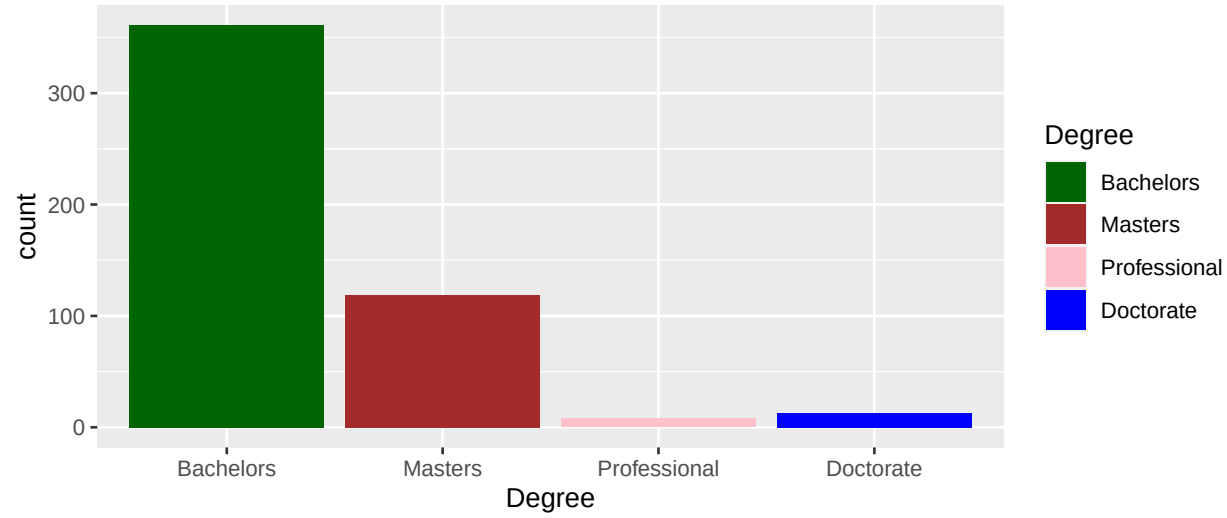
The fourth regression model includes age as well as educational qualifications the r square for this regression model is 0.11 which higher than the previous r square. It is roughly twice as much. This model showed females earning 10% less on average while those with bachelors earned 20% less and those with masters earned 8% less. This model also shows that females with PhDs earned 17% more. The coffecient for age is at 1.2% which is relatively insignificant when compared to educational qualification.

Conclusion:

All 4 regression models, focus on the correlation between female qualification and their hourly wage. The various models focus on different educational qualifications and added varibales such as age. The R squares signify that they are generally low. However, they are higher in regression model 4. All 4 models show that on avergae women earn less than men. However, there is variation when their qualification is taken into account. The gender gap is highest at bachelor level and lowest at PhDs where gender gap is minimal.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```





	(1)	(2)	(3)	(4)
(Intercept)	3.5041 ****	3.7003 ****	3.7003 ****	3.1426 ****
	(0.0263)	(0.1711)	(0.1714)	(0.1884)
female	-0.1177 **	-0.1158 **	-0.1855	-0.0997 **
	(0.0462)	(0.0456)	(0.1196)	(0.0445)
ed_bachelor		-0.2477	-0.2393	-0.2034
		(0.1736)	(0.1743)	(0.1646)
ed_master		-0.0991	-0.1290	-0.0831
		(0.1770)	(0.1792)	(0.1678)
ed_PhD		0.1981	0.2196	0.1716
		(0.1850)	(0.1948)	(0.1762)
ed_Professional				
female:ed_bachelor			0.0454	
			(0.1315)	
female:ed_master			0.1471	
			(0.1492)	
female:ed_Professional				
female:ed_PhD				
age				0.0118 ****
				(0.0020)
N	501	501	501	501
R2	0.0132	0.0503	0.0522	0.1120

**** p < 0.001; *** p < 0.01; ** p < 0.05; * p < 0.1.