

Introduction

This assignment's objective was to figure out the gender wage gap in the *Chief Executives* occupation in the CPS survey for 2014 based data taken from the OSF website. (*morg-2014-emp.csv*).

Data Manipulation and Exploration.

To make the dataset suitable for my analysis, I applied several filters to the dataset. The first one being the occupation code that I wanted to work on (0010 – *Chief Executives*). The second filter was applied to get those individuals who earned a weekly wage of more than 0 and have at least worked for 20 hours per week. Individuals between 17 and 64 were included in the final filter to create my dataset used for carrying out the analysis.

To further modify the dataset, I created a variable which showed the hourly wage `wagehrly` of the individuals that was calculated by dividing the `earnwke` by `uhours`. Another variable was also created `lnwagehrly`, which is the log of the hourly wage variable. A dummy variable for female in gender was also created that takes a value of 1 when female and 0 when male.

Only the education levels of associate degree (occupational/vocational), and above have been included in the final dataset, with the associate degree(occupational/vocational) being the base for all the comparisons drawn. There are a total of 6 education levels included with a dummy variable created for each one, allowing for a detailed representation and insights for each education level.

Analysis

Whether to use the log of wage per hour or just use the absolute values before running regressions, I checked distributions of both variables using a plotly graph. From histograms, it was clear that the distribution of the absolute values resembled more of a normal distribution compared to the log values, hence I decide to use the `wagehrly` absolute values to perform all the regressions. (level-level regressions).

To get the unconditional gender wage gap, two methods were used which gave similar results. The first one being a regression for hourly wage and female variable. The second one was the same regression on hourly wage and female, but with *HCI*. Both showed that a female Chief Executive earns \$4.25 less than her male counterpart. To incorporate the education levels, I ran 2 regressions. The first regression (reg4) was conditional on education levels, while the second regression (reg5) was conditional on education and interaction terms were incorporated to control education levels of both genders. Note that all of the regressions accounted for heteroskedastic errors.

The first regression (reg3) gives us the unconditional gender gap; on average, a female Chief Executive earns USD 4.25 less than her male counterpart and this coefficient is significant at 99% confidence. Whereas, when conditioned on education, the regression (reg4) shows that females earned USD 3.75 less than their male counterparts with the coefficient being significant at 99% confidence level. The third regression (reg5), which was run with interactions terms of education levels and female, shows that while controlling for education, on average a female will earn approximately USD 2.9 less than her male counterpart with the coefficient being significant at 90% confidence level. In terms of generalizing the results, I am not confident because not all the coefficients of interactions terms are significant and even those which are significant are only at 95% confidence level. The second reason is that there are not enough observations in the data set to allow us to confidently say that they are representative of the entire population.

Notes: stargazer regression summaries comparing the three regressions are in code file. I will be adding a shortened version.

Figure 1: Comparison for selection of `wagehrly` or `lnwagehrly` as regression variable.

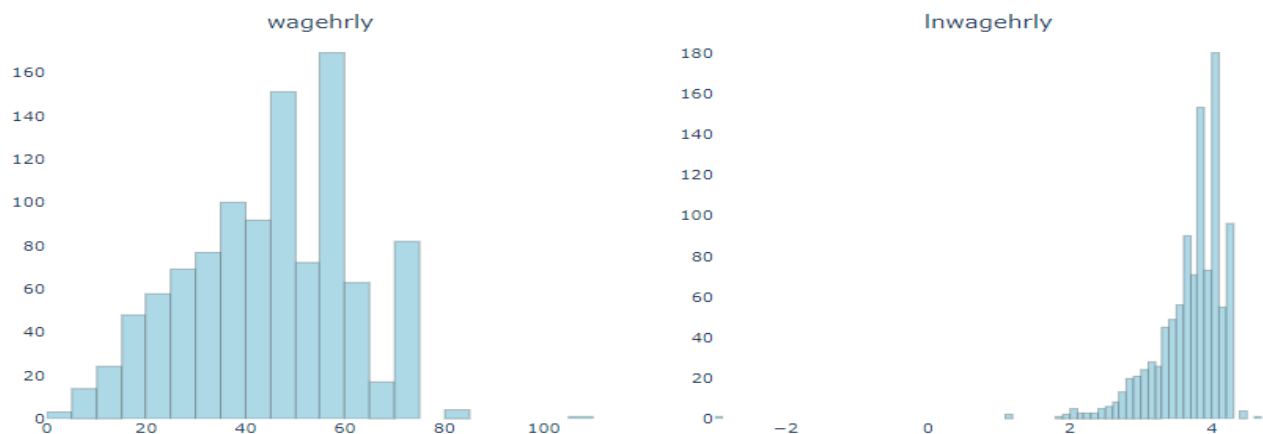


Figure 2: Non Paremetric regression of Hourly wage ~ Education Level

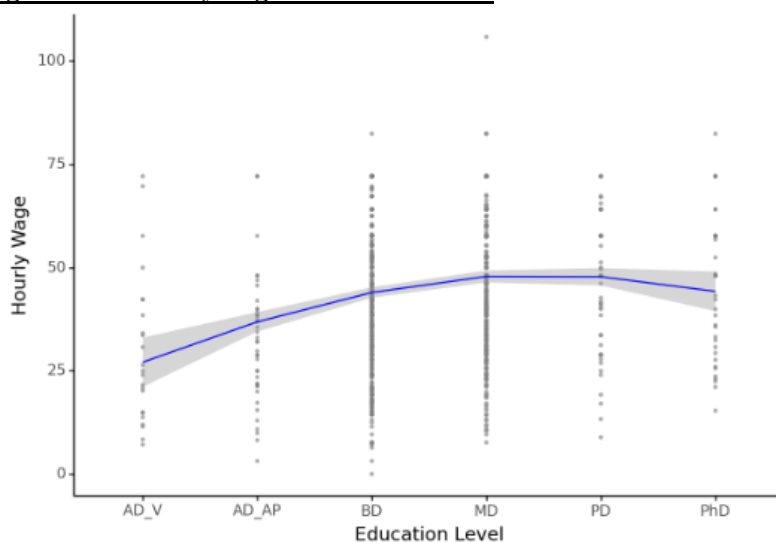


Figure 3: stargazer regression summaries

Dependent variable: wagehrly			
	Unconditional Gap	Education	Education with Interaction
	(1)	(2)	(3)
female	-4.256*** (1.229)	-3.750*** (1.185)	-2.883* (1.713)
Education		Yes	Yes
Education with Interaction			Yes
Observations	1044	1044	1044
R ²	0.013	0.063	0.071
Adjusted R ²	0.012	0.057	0.061
Residual Std. Error	16.446 (df=1042)	16.064 (df=1037)	16.028 (df=1032)
F Statistic	11.992*** (df=1; 1042)	1189.322*** (df=6; 1037)	698.974*** (df=11; 1032)
Note:		*p<0.1; **p<0.05; ***p<0.01	

Figure 3: Prediction graph (gender comparison for education levels)

Note: It says females with master's degree, as detailed stargazer view will show master's coefficient [interact] = 1.599

