DA2 - Assignment 1

Rauhan Nazir & Ali Hasnain Khan Sial

Introduction

This assignment has tried to find the gender wage gap in the *Chief Executives* occupation in the U.S. for 2014 based on the data set taken from the **OSF** website.

Data Manipulation

There were several filters applied to the dataset to make it suitable for our analysis. The first one being the occupation that we wanted to focus on (Chief Executives). The second one was choosing only those workers who earn a weekly wage of more than 0 and have at least worked for 20 hours per week. Finally, only the workers who were between 17 and 64 were included in the final dataset used for carrying out analysis.

We included a variable which showed the wage per hour of the employees that was calculated by dividing the weekly wage by the number of hours worked. A dummy variable for female in gender was also introduced that takes a value of 1 & 0 depending on the gender (1 when female), so that gender wise analyis can be performed.

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. For both parametric and non-parametric regressions there are In total 6 education levels with a dummy variable created for each one, allowing us to get a detailed representation and insights for each education level.

Analysis

To get the unconditional gender wage gap, two methods were used which gave the same results. The first one being a data summary for wage per hour and female variable. The second one was an unconditional regression on wage and female. Both showed that on average, a female Chief Executive earns $4.2 less than her male counterpart.

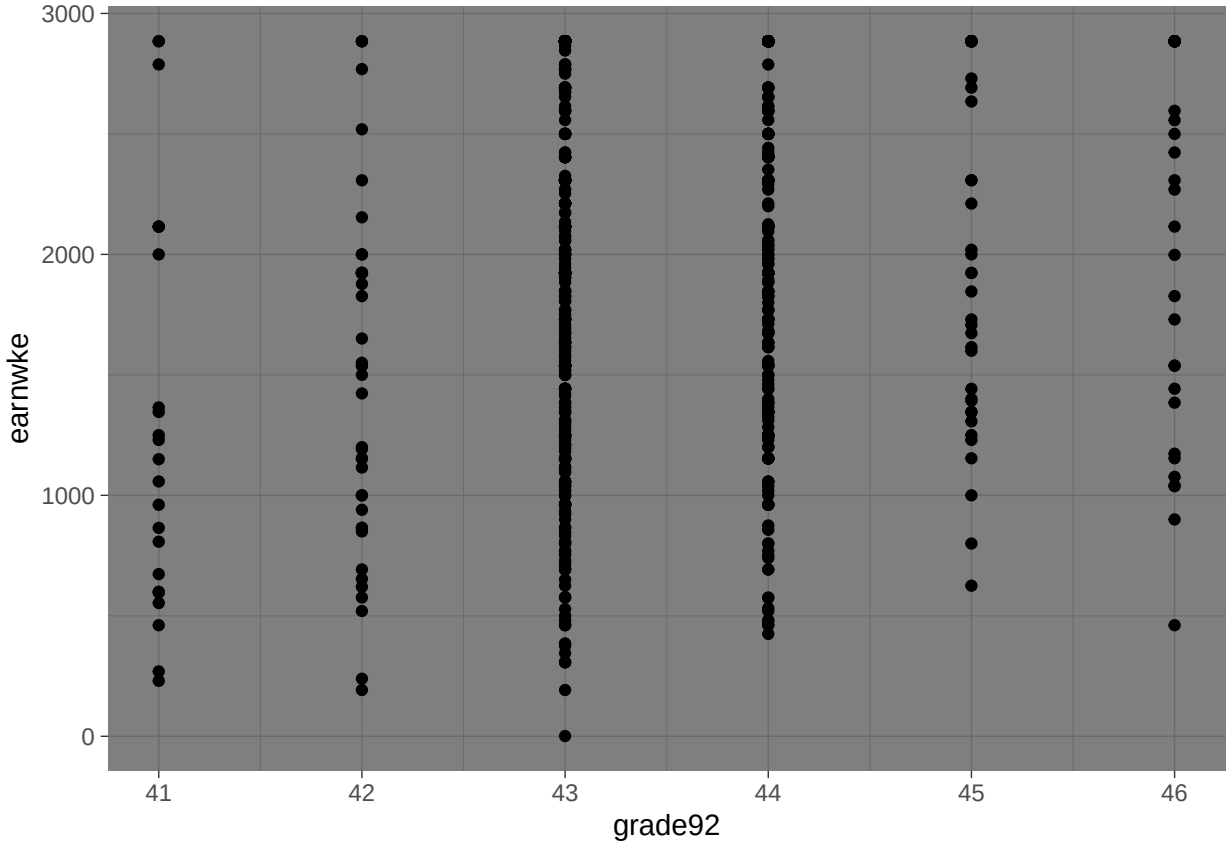
There was a decision that we had to make, whether to use the log of wage per hour or just use the absolute values before we run regressions. By creating density curves for both of the scenarios, we clearly saw that the distribution of the absolute values resembled more of a normal distribution compared to the log values, hence we decided to go with the absolute values to perform all the regression analysis.

We have run 3 level-level regressions in total, accounting for heteroskedastic errors. Regression 1 gives us the unconditional gender gap; on average, a female Chief Executive earns USD 4.2 less than her male counterpart and this coefficient is significant with more than 99.9% confidence. Whereas, when conditioned on education, regression 2 shows that females earned USD 3.7 less than their male counterparts with the coefficient being significant at more than 99% confidence level. Regression 3, 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 15.8 less than her male counter part with the coefficient being significant at more than 99% confidence, but the difference as per the interaction term will also be added to this coefficient. For instance when a female moves from Associate Degree(occupational/vocational) to a PHD degree, on average a positive $22 will be added to her wage, which is significant at a 90% confidence level (Just the education level change effect due to being a female).

In terms of generalizing the results, we shouldn't be so confident. And we say this due to a couple of reasons. The first one being that not all of the coefficients of interactions terms are significant and even those which are significant are only at 90% 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.

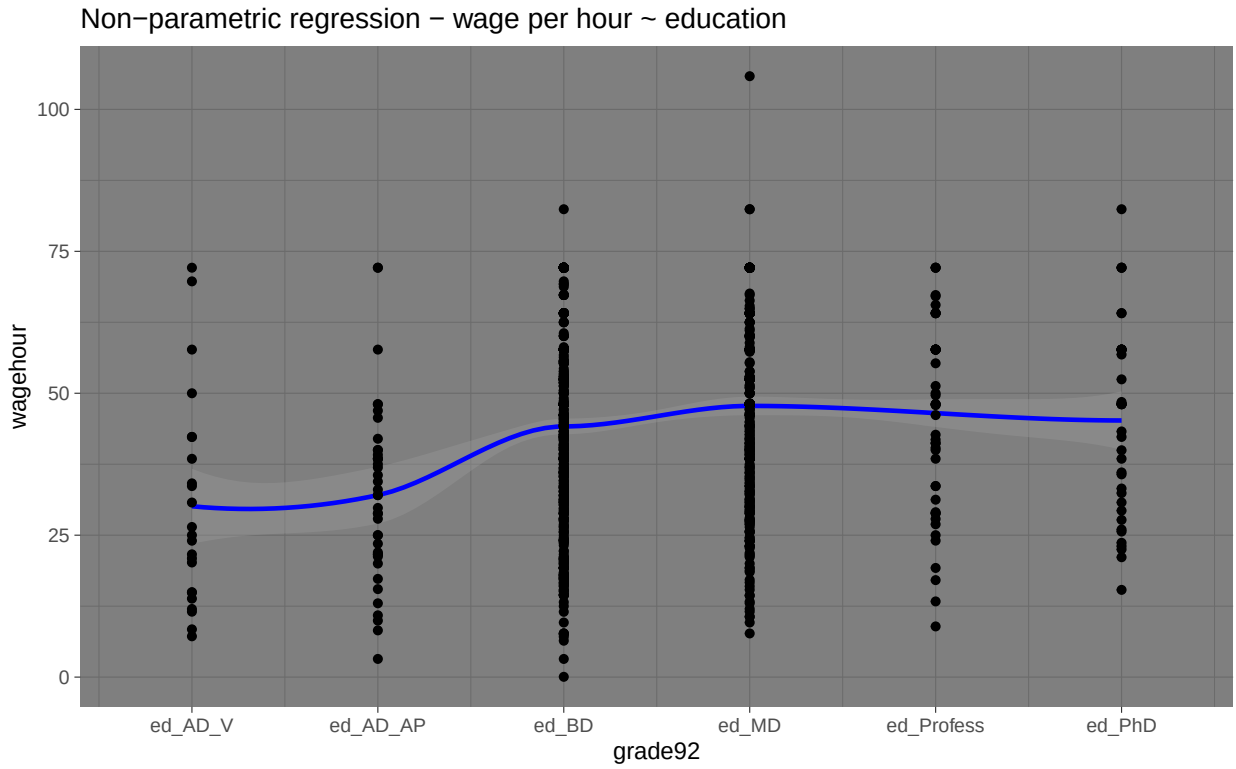
as.factor(female)		Mean	SD	Min	Max	P25	P75	N
0	wagehour	45.90	15.88	3.20	105.85	36.05	57.69	772
1	wagehour	41.64	17.96	0.05	82.42	27.59	57.69	272

Appendix



Assigning degree names to grade92 variable based on the information provided on page 25 of the cpsx document provided.

- ed_AD_V contains the individuals with education levels of associate degrees (Vocational/occupational)
- ed_AD_AP contains the individuals with education levels of associate degrees (Academic program)
- ed_BD contains the individuals with education levels of a Bachelor's degree (e.g.BA,AB,BS)
- ed_MD contains the individuals with education levels of a Masters degree(e.g.MA,MS,MEng,Med,MSW,MBA)
- ed_Profess contains the individuals with education of a professional degree (e.g.MD,DDS,DVM,LLB,JD)
- ed_PhD contains the individuals with education levels of a Doctorate degree (e.g.PhD,EdD)


Non−parametric regression − wage per hour ~ education
wagehour
100
75
50
25
0
ed_AD_V
ed_AD_AP
ed_BD
ed_MD
ed_Profess
ed_PhD
grade92

	reg1, y = wage/hour	reg2, y = wage/hour	reg3, y = wage/hour
(Intercept)	45.896 ****	31.070 ****	34.196 ****
	(0.572)	(3.718)	(4.642)
female	-4.256 ****	-3.750 ***	-15.733 ***
	(1.229)	(1.185)	(5.624)
ed_AD_AP		2.390	0.017
		(4.381)	(5.362)
ed_BD		14.132 ****	11.562 **
		(3.755)	(4.711)
ed_MD		17.535 ****	13.860 ***
		(3.781)	(4.732)
ed_Profess		15.807 ****	11.192 **
		(4.430)	(5.445)
ed_PhD		15.022 ***	9.421 *
		(4.608)	(5.475)
female:ed_AD_AP			9.975
			(7.681)
female:ed_BD			9.991 *
			(5.866)
female:ed_MD			14.449 **
			(5.964)
female:ed_Profess			17.366 **
			(7.861)
female:ed_PhD			22.432 **
			(8.918)
N	1044	1044	1044
R2	0.013	0.063	0.071

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