DA2 - Assignment 1

Aftab Alam & Muhammad Talha Zahid

Introduction

This assignment has tried to find the gender gap between men and women in the *Financial Managers* occupation in the U.S. for 2014 based on the data set taken from the **OSF** website.

Data Changes

Before conducting any analysis on the data set, it has been filtered to suit our analysis. The focus is the occupation of Financial Managers who earn a weekly wage of more than 0 and have worked for more than 20 hours per week. We have also restricted the age between 17 and 64 years.

To be able to make an apples to apples comparison, we have calculated the wage per hour variable using the weekly age and number of hours worked. Moreover, a dummy variable (female) has been created for gender with a value of 1 for female.

Education levels of high school and higher have been taken from the data set, details available in the appendix of this document. In total, the education levels have been clubbed into 5 dummy variables for the purpose of parametric regressions with associate degrees and lower clubbed into 1 out of the total 5. However, for non-parametric regressions, the total 8 education levels have been used separately for a detailed representation.

Analysis

We first created a quick data summary for wage per hour and female variable to get the unconditional gender gap. As per the image 1 in appendix, on average females earned USD9.2 per hour less than their male counterparts in the Financial Managers occupation. This is also visible from regression 1 shown in the image 6 of the appendix.

We checked to see whether we should use the absolute values of wage per hour or log of wage per hour by creating density curves for both variables. Since, the absolute values of wage per hour showed a near normal distribution compared to its log distribution, we chose to use this variable for our regression analysis.

We have run 4 level-level regressions in total, accounting for heteroskedastic errors; regression table in appendix. Regression 1 gives us the unconditional gender gap; on average females earned around USD9.2 less than their male counterparts in the occupation where this coefficient is significant with more than 99.9% confidence. Whereas, when conditioned on education, regression 2 shows that females earned USD6.3 less than their male counterparts with the coefficient being significant at more than 99.9% confidence level. The rest of the education coefficients show that wage is higher for higher levels of education.

To gain a deeper understanding, we ran the last two regressions with interaction terms, the only difference between 3rd and 4th regression is that the 4th regression data does not include the 1 extreme value of USD120 per hour for ed_MA. This change does not largely impact the model apart from changing the ed_MA*female coefficient value and its significance. However, these regressions reveal that even for women with higher education levels, there exists a gender gap with a confidence level of 90%.

In terms of generalizing these results, we would exercise caution as the coefficients of interaction terms are significant only at 90% with large standard errors. Especially for ed_Profess and ed_PhD where the number of observations in the data set for the occupation are very limited. Moreover, since the non-parametric LOWESS regressions showed a non-linear relation between education and wage per hour (graphs in appendix), we could perhaps use a spline regression with interaction terms to get a better fit while having a parametric regression equation, however, that is beyond the analysis of this assignment.

as.factor(female)		Mean	SD	Min	Max	P25	P75	N
0	wagehour	38.42	16.47	0.00	115.38	25.48	50.00	606
1	wagehour	29.21	14.47	0.21	77.96	18.88	36.54	781

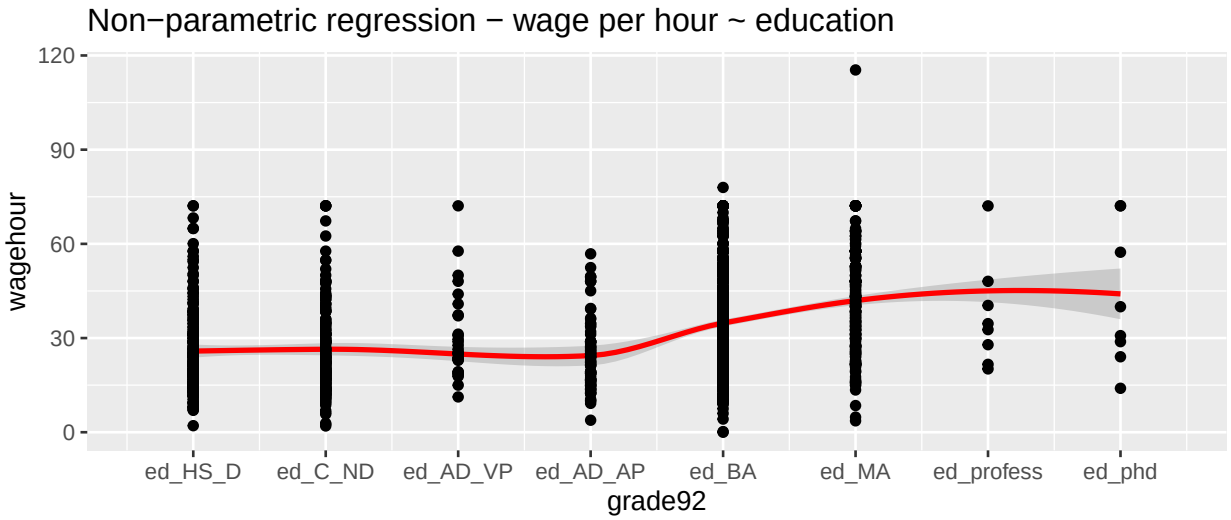
Appendix

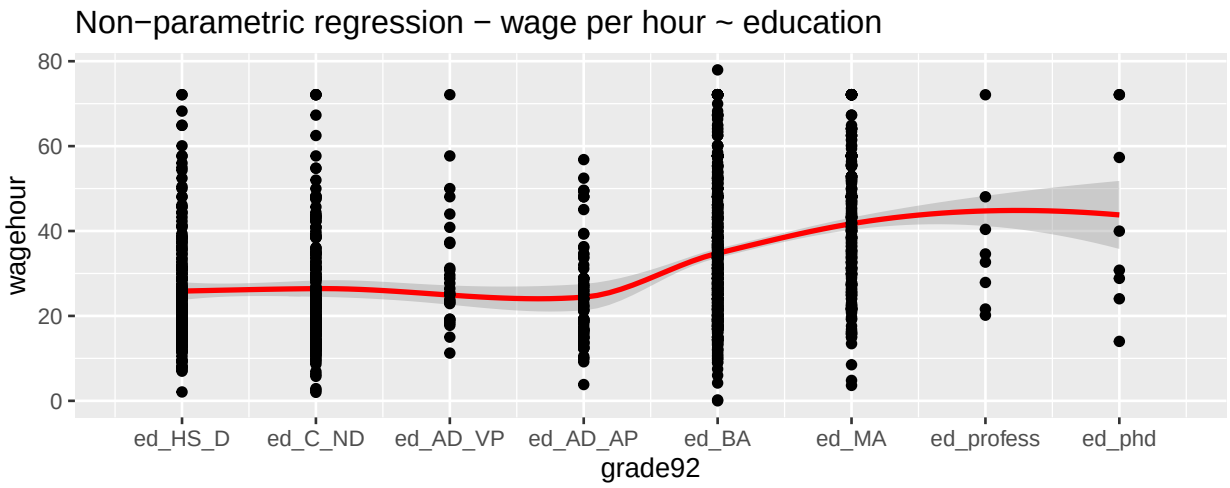
Assigning degree names to grade92 variable based on the information provided on page 25 of the cpsx document provided.

- ed_AD_less contains the individuals with education levels of associate degrees, some college with no degree, high school, diploma or GED.
- ed_BA contains the individuals with education levels of a Bachelor's degree (e.g.BA,AB,BS)
- ed_MA contains the individuals with education levels of a Master's or an equivalent 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)

Additional degree names used for non-parametric regressions that have been clubbed into ed_AD_less above

- ed_HS_D High school graduate, diploma or GED
- ed_C_ND Some college but no degree
- ed_AD_VP Associate degree – occupational/vocational
- ed_AD_AP Associate degree – academic program





The extreme value of $120 has been dropped from the ed_MA

	reg1, y = wage/hour	reg2, y = wage/hour	reg3, y = wage/hour	reg4, y = wage/hour
(Intercept)	38.420 ****	30.507 ****	28.695 ****	28.695 ****
	(0.669)	(0.864)	(1.200)	(1.200)
female	-9.208 ****	-6.325 ****	-3.815 ***	-3.815 ***
	(0.846)	(0.831)	(1.377)	(1.377)
ed_BA		7.772 ****	9.946 ****	9.946 ****
		(0.856)	(1.500)	(1.500)
ed_MA		14.882 ****	17.404 ****	16.965 ****
		(1.202)	(1.723)	(1.671)
ed_Profess		10.634 **	22.252 **	22.252 **
		(4.958)	(9.477)	(9.477)
ed_PhD		13.472 *	15.948 *	15.948 *
		(7.287)	(8.278)	(8.278)
female:ed_BA			-3.211 *	-3.211 *
			(1.826)	(1.826)
female:ed_MA			-4.430 *	-3.992
			(2.485)	(2.449)
female:ed_Profess			-18.199 *	-18.199 *
			(10.106)	(10.106)
female:ed_PhD			-5.163	-5.163
			(17.475)	(17.476)
N	1387	1387	1387	1386
R2	0.081	0.187	0.191	0.189

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