## Maryam Khan & Khawaja Abbas

### Overview

The purpose of this assignment is to analyze the unconditional gender wage gap between males and females for the occupation "Financial Analyst (0840)" and further analyze how the conditional gender wage gap varies with the level of education. The data used in this project is the Current Population Survey (CPS) by the Bureau of Labor Statistics.

### Data cleaning:

To make sure our analysis is more accurate we filtered and cleaned the data by calculating wage per hour by dividing the number of hours worked per week and the weekly wage. We also created a female binary variable where 1 stands for female and 0 stands for male. To conduct further analysis based on levels of education we also created binary variables for all levels of education.

### Analysis:

To examine the distribution of the wage data for the selected occupation, we made the density plot of the wage and log wage. According to the density plots the distribution of wage (level-level) is close to normal and the log transformation of wage causes the distribution to become skewed to the left. Therefore, we decided to run our regressions on a level-level model instead of using the log transformation.

**Unconditional gender gap:** To compare both models we ran regressions with the female binary variable for both wage and log wage and the result for the wage (level) regression was more statistically significant, hence, it made sense to go with the level-level regression.

log(wage) = 3.4093*** - 0.1202female

wage = 37.3321*** - 5.2147**female

According to the results of our regressions, we can say with 95% confidence that there exists an unconditional gender gap between male and female. Females that are financial analysts tend to earn 5.2147 USD less than males and males earn 37.33 USD per hour.

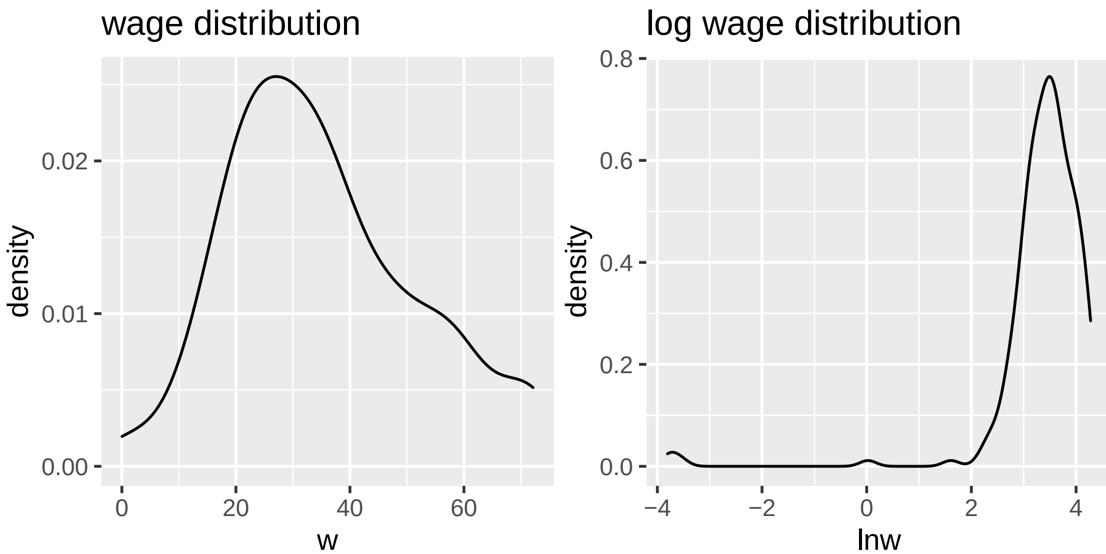
**Conditional gender gap with different levels of education:** In order to get a deeper understanding of how the gender wage gap changes with the level of education we ran regressions with levels of education above Associate level education grade where Associate was used as the base. A multi-variable regression was run and the female variable coefficient -5.5335 was statistically significant at the 1% level. It was also evident from the results that the wage increases as the level of education increases for Masters, Professional and PhD. However, there is no significant relation between bachelors and associate degrees. To further analyze, the wage difference between male and females for different education levels we ran a regression with interaction variables, even though the female coefficient was statistically significant most of the interaction variables showed no significance with very high standard errors. However, it was interesting that the female PhD interaction variable was significant at 95% confidence interval with a positive coefficient. Meaning that females with PhD degrees earn 20.99 dollars more than their male counterparts in this occupation. In conclusion, the multi-variable regression model's coefficients were more statistically significant as compared to the multi-variable regression model with interaction terms. Another point to note is that the standard errors of the multi-variable regression were significantly lower than the multi-variable regression with interaction terms.

Table 1:

| fdata | N |
|---|---|
| Female | 87 |
| Male | 133 |

## Appendix

Summary showing difference in wage and log wage




Regression with wage and female

Table 2: Stats for wage and log wage

| | Mean | SD | Min | Max | Median | P95 | N |
|---|---|---|---|---|---|---|---|
| w | 35.27 | 16.66 | 0.02 | 72.12 | 32.80 | 70.32 | 220 |
| lnw | 3.36 | 0.99 | −3.81 | 4.28 | 3.49 | 4.25 | 220 |

Table 3: Unconditional Regression

| | Log Wage | Wage |
|---|---|---|
| (Intercept) | 37.3321*** | 3.4093*** |
| | (1.4306) | (0.0855) |
| female | −5.2147** | −0.1202 |
| | (2.2749) | (0.1359) |
| Num.Obs. | 220 | 220 |
| R2 | 0.024 | 0.004 |

** $p < 0.05$, *** $p < 0.01$

Table 4: Conditonal Regression

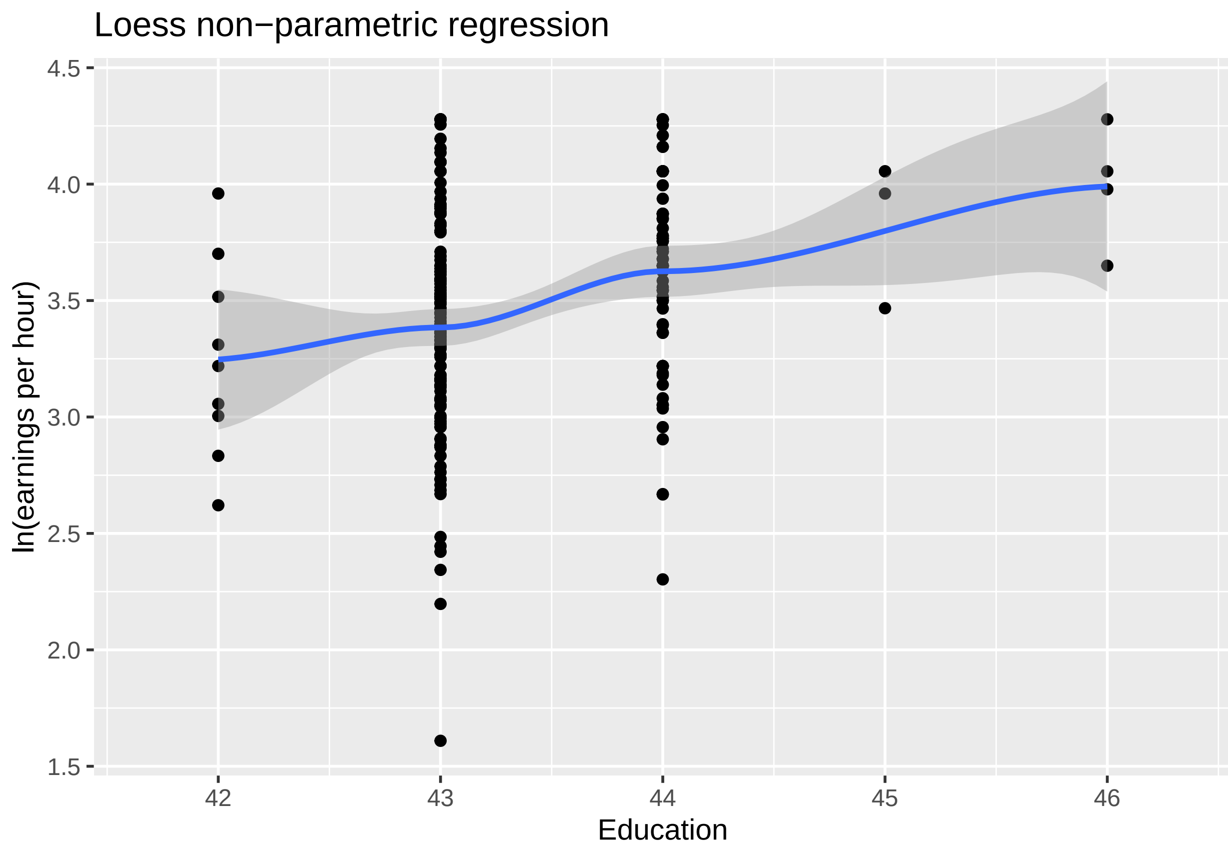
| | Wage gender | Wage (edu) | Wage (edu*female) |
|---|---|---|---|
| (Intercept) | 37.3321**** | 31.5963**** | 40.1123**** |
| | (1.5335) | (3.7220) | (6.0438) |
| female | −5.2147** | −5.5335*** | −18.3075*** |
| | (2.1835) | (2.0986) | (6.5997) |
| ed_Bachelor | | 2.8974 | −6.0413 |
| | | (3.6288) | (6.3727) |
| ed_MA | | 10.1119** | 1.8317 |
| | | (4.0236) | (6.5347) |
| ed_Profess | | 21.1367**** | 14.9520** |
| | | (5.9304) | (6.3360) |
| ed_PhD | | 27.9711**** | 13.2877** |
| | | (7.1363) | (6.0438) |
| female × ed_Bachelor | | | 13.8761* |
| | | | (7.0731) |
| female × ed_MA | | | 12.1138 |
| | | | (7.9164) |
| female × ed_Profess | | | 8.1116 |
| | | | (11.5462) |
| female × ed_PhD | | | 20.9972** |
| | | | (10.4844) |
| Num.Obs. | 220 | 220 | 220 |
| R2 | 0.024 | 0.116 | 0.123 |
| Std.Errors | HC1 | HC1 | HC1 |

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table 5: Table : 5

| | Women (log Wage) | Men (log Wage) | All (log Wage) |
|---|---|---|---|
| (Intercept) | 21.8048** | 40.1123** | 40.1123** |
| | (2.6680) | (6.0191) | (6.0438) |
| ed_Bachelor | 7.8349* | −6.0413 | −6.0413 |
| | (3.0883) | (6.3467) | (6.3727) |
| ed_MA | 13.9455** | 1.8317 | 1.8317 |
| | (4.4968) | (6.5080) | (6.5347) |
| ed_Profess | 23.0635* | 14.9520* | 14.9520* |
| | (9.7138) | (6.3101) | (6.3360) |
| ed_PhD | 34.2849** | 13.2877* | 13.2877* |
| | (8.6215) | (6.0191) | (6.0438) |
| female | | | −18.3075** |
| | | | (6.5997) |
| female × ed_Bachelor | | | 13.8761 |
| | | | (7.0731) |
| female × ed_MA | | | 12.1138 |
| | | | (7.9164) |
| female × ed_Profess | | | 8.1116 |
| | | | (11.5462) |
| female × ed_PhD | | | 20.9972* |
| | | | (10.4844) |
| Num.Obs. | 87 | 133 | 220 |
| R2 | 0.184 | 0.066 | 0.123 |

* p < 0.05, ** p < 0.01


Loess non−parametric regression
ln(earnings per hour)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
42
43
44
45
46
Education