# DA2_Assignment2

Fatima Arshad

## Introduction

Do female health care workers earn more their male counterparts? Do employees with higher educational degrees earn more than the less qualified ones? This study is conducted to answer these questions by examining the gender pay gap and pay gap due to levels of education in the nursing, psychiatric, and home help aides. We will not only find the general pattern for the data but also forecast these differences which is achieved by generalizing the data set using statistical inference and external validity principles. We use the **Current Population Survey (CPS) dataset** for the US. For the purpose of this study we will choose the ***Nursing, Psychiatric, and Home Help Aides*** occupation with ***3600*** census occupation code. All other occupation codes can be accessed **here**.

## Data Transformation and Filtering

We are interested in the gender and age pay gap for health care workers. For this analysis, we will create the following four new variables: 1. A binary variable ***female*** with a binary value of ***1*** 2. hourly wage ***(hourly_wage)*** calculated by dividing the weekly earnings (earnwke) by the number of hours (uhours) 3. log of hourly wage ***(ln_hourly_wage)*** 4. A binary character variable ***gender*** using sex variable with male and female.

## Introductory Findings

The calculations in ***Table 1*** shows that our sample is varied with high number of females (445) as compared to males (63). Table 2 shows a descriptive summary of the variables of interest for this analysis. It is evident from the table that the sample distribution is right skewed as the mean lies on the right of the median. Multiple factors in the data contribute to the skewness. Firstly, the distribution is positively skewed because of the relatively high maximum hourly wage of USD 72 per hour. Secondly, the average weekly work hours is 36, however, for some workers it is significantly higher with a maximum of 82 hours per week. Moreover, the presence of extreme values such as a minimum wage of USD 2.50 also play a role in the skewness.

## Analysis

Initially, we visualize the distribution of the hourly wage using ggplot. ***Figure 1*** illustrates that the wage distribution (hourly_wage) is positively skewed. Subsequently, for this analysis we will use the natural logarithm of wage. ***Figure 2*** illustrates the distribution for ln(hourly_wage). Now we are all set to begin the regression analysis to uncover the difference in hourly wages with respect to gender and levels of education.

Firstly, we will analyze the wage gap with respect to gender. A summary of the unconditional gender gap is presented in ***Table 3***. We calculate following two regressions: 1. ***level-level regression:*** It describes that an average female earns USD 2.8 less than their male counterparts. This wage gap is significant at 1% significant level. 2. ***log-level regression:*** It shows that an average female earns 12% less than males. This coefficient is significant with more than 99.9% confidence level.

Table 1:

| data | N |
|---|---|
| Female | 445 |
| Male | 63 |

Table 2:

| | Mean | SD | Min | Max | Median | P95 | N |
|---|---|---|---|---|---|---|---|
| Weekly earnings | 536.24 | 364.27 | 42.50 | 2884.61 | 466.50 | 1157.84 | 508 |
| Weekly hours worked | 36.08 | 9.58 | 5.00 | 82.00 | 40.00 | 45.00 | 508 |
| hourly_wage | 14.63 | 8.46 | 2.50 | 72.12 | 12.50 | 28.94 | 508 |
| ln_hourly_wage | 2.57 | 0.44 | 0.92 | 4.28 | 2.53 | 3.37 | 508 |

Secondly, we will analyze the wage gap with respect to different levels of education. ***Table 4*** describes the results of multivariate regression. ***Model 2*** represents the findings of conditioning gender gap on education. It shows that on average, females earn 9.4 % less than males and this coefficient is significant at 1% significant level. ***Model 3*** compares employees of same gender with different education levels using Associate-vocational degree as the baseline variable. We select values with more than 99 % confidence level. Results show that same-gender Bachelor's degree employees on average, earn 12.9% more than employees with Associate- vocational degree. Next, the regression with interaction terms using same base variables is illustrated in ***Table 5***.The results show that there is still a gender gap for women with higher educational levels with 99% confidence interval. A women with a Masters degree earns on an average 32.5 % less than males. In conclusion, it is evident that females earn less than men even with the same level of education. Moreover, a 1000 iterations of bootstrap simulation yielded similar results which validates the conclusion.

Table 3:

| | LM - Hourly Wage | LM - Log Hourly Wage |
|---|---|---|
| (Intercept) | 17.0926** | 2.6808** |
| | (1.0602) | (0.0549) |
| female | −2.8148* | −0.1212* |
| | (1.1328) | (0.0587) |
| Num.Obs. | 508 | 508 |
| R2 | 0.012 | 0.008 |

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

Table 4:

| | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| (Intercept) | 2.6808** | −0.2134 | 2.6114** |
| | (0.0672) | (0.8974) | (0.0707) |
| female | −0.1212 | −0.0947 | −0.0866 |
| | (0.0701) | (0.0681) | (0.0670) |
| grade92 | | 0.0681** | |
| | | (0.0212) | |
| ed_Associate_ap | | | −0.0585 |
| | | | (0.0453) |
| ed_BA | | | 0.1290** |
| | | | (0.0473) |
| ed_MA | | | 0.1373 |
| | | | (0.1135) |
| ed_Profess | | | 0.3899 |
| | | | (0.4178) |
| ed_PhD | | | 0.5012** |
| | | | (0.1414) |
| Num.Obs. | 508 | 508 | 508 |
| R2 | 0.008 | 0.031 | 0.048 |

* p < 0.05, ** p < 0.01

Figure 1

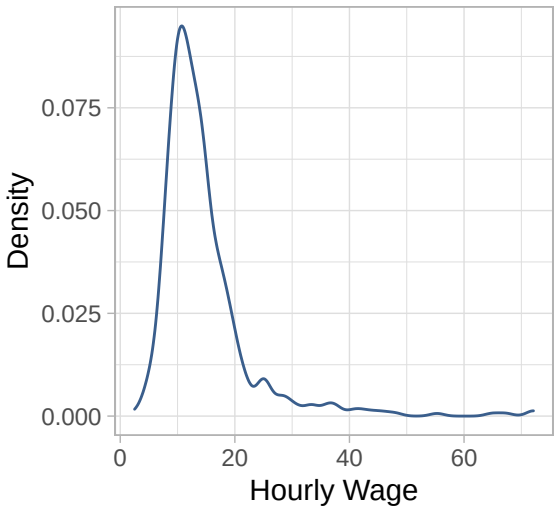



Figure 2

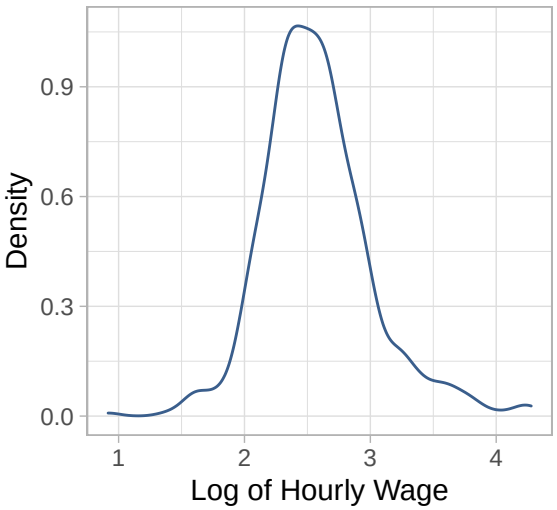
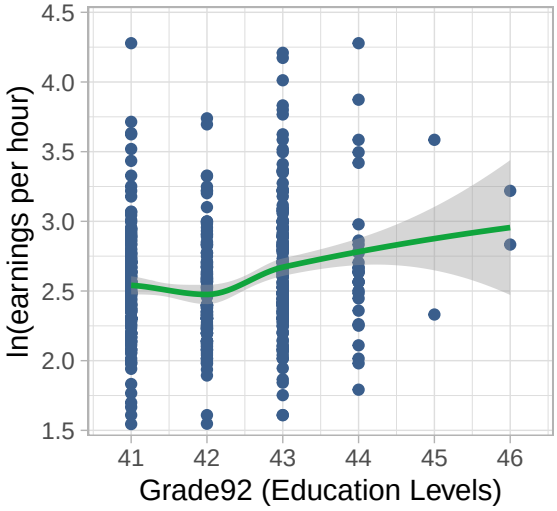


Table 5:

| | Women (log hourly Wage) | Men (log hourly Wage) | All (log hourly Wage) |
|---|---|---|---|
| grade92 | 0.0620** | 0.0585** | 0.0585** |
| | (0.0008) | (0.0028) | (0.0027) |
| ed_Associate_ap | −0.1315** | 0.0550 | 0.0550 |
| | (0.0476) | (0.1617) | (0.1568) |
| ed_BA | −0.0256 | 0.2858 | 0.2858 |
| | (0.0505) | (0.1557) | (0.1510) |
| ed_MA | −0.0738 | 0.2617 | 0.2617 |
| | (0.1222) | (0.3072) | (0.2980) |
| ed_Profess | −0.4571** | 0.9524** | 0.9524** |
| | (0.0379) | (0.1242) | (0.1205) |
| ed_PhD | 0.1758 | | 0.1931 |
| | (0.1426) | | (0.1428) |
| female | | | 0.1419 |
| | | | (0.1151) |
| female × ed_Associate_ap | | | −0.1831 |
| | | | (0.1619) |
| female × ed_BA | | | −0.3044 |
| | | | (0.1551) |
| female × ed_MA | | | −0.3251 |
| | | | (0.3192) |
| female × ed_Profess | | | −1.3956** |
| | | | (0.1151) |
| Num.Obs. | 445 | 63 | 508 |
| R2 | 0.034 | 0.149 | 0.063 |

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

Figure 3


4.5
4.0
3.5
3.0
2.5
2.0
1.5
ln(earnings per hour)
41
42
43
44
45
46
Grade92 (Education Levels)