**Chloe Nguyen – JSC270: Assignment 2, Report**

## Background

The chosen dataset takes a look at US incomes and related factors. Each row represents a worker and details their personal information (such as age, sex and race) as well as their income and occupational information (such as occupation and hours per week). This dataset was retrieved from the 1994 Census database from the United States Census Bureau.

## Motivations

The goal of this report is to determine whether or not men tend to work more hours per week than women. This is a question from the Regression section of the assignment.

## Methods

In order to determine whether or not men tend to work more hours per week than women, we must fit a linear regression model to estimate the hours per week a person works, using sex as the explanatory variable. Note that we discard observations with missing values in hours_per_week or sex in determining the regression model. After we fit the linear regression model, we plot the model and our data in order to visualize our findings.

## Results

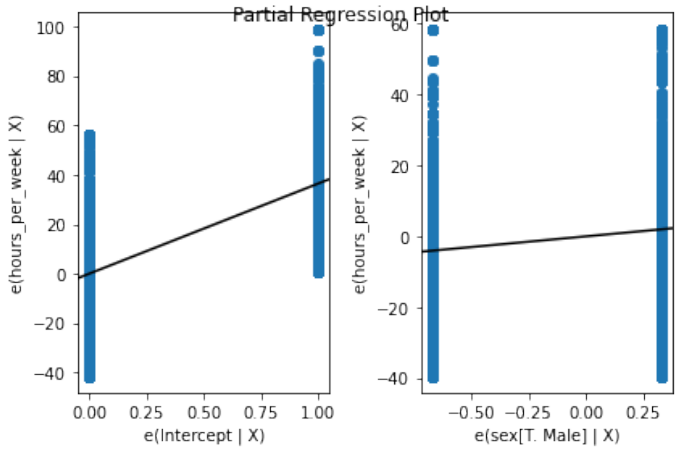
The fitted linear regression model (with hours_per_week as the dependent variable and sex as the independent variable):

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         hours_per_week   R-squared:                       0.053
Model:                            OLS   Adj. R-squared:                  0.053
Method:                 Least Squares   F-statistic:                     1807.
Date:                Sat, 06 Feb 2021   Prob (F-statistic):               0.00
Time:                        01:25:46   Log-Likelihood:            -1.2716e+05
No. Observations:               32561   AIC:                         2.543e+05
Df Residuals:                   32559   BIC:                         2.543e+05
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
```

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 36.4104 | 0.116 | 314.412 | 0.000 | 36.183 | 36.637 |
| sex[T. Male] | 6.0177 | 0.142 | 42.510 | 0.000 | 5.740 | 6.295 |

| | | | |
|---|---|---|---|
| Omnibus: | 2649.390 | Durbin-Watson: | 2.019 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 13090.867 |
| Skew: | 0.239 | Prob(JB): | 0.00 |
| Kurtosis: | 6.069 | Cond. No. | 3.24 |

Partial Regression Plot of Fitted Model




From the regression results, in the case where x=1 represents males, we can see that the p-value associated with the slope parameter is extremely small. This implies that there is significant evidence that supports the slope parameter not being zero- that is, there is significant evidence that males and females don't work the same amount of hours per week on average. Further, the estimated slope parameter is 6.0177 and the estimated intercept parameter is 36.4104. This implies that on average, males work 6.0177 hours more than females per week.

**Conclusion**

Our results imply that on average, males work more hours per week than females. However, it should be noted that our data was taken from the 1994 Census database and is over 20 years outdated. Thus, our findings may no longer be applicable to United States workers.