**Multi Regression Analysis Assignment**

**State: New York**

**Name: Romario Medel-Estrada**

**Business Analytics Modeling of Disability Status on Wage Income**

Observed Personal Wage Income WAGP is a variable that reflects the choices made by  
individual and the employment needs of a firm. It is estimated using regression analysis method.  
To explain earned wage income, we use the life-cycle business model and other socio-economic  
factors.  
In this exercise, due to data limitations, we estimate typical wage income earned in a simplified  
model. The model assumes individuals maximize some utility (happiness) in the presence of  
budget constraints. We control for socio-economic, demographics, education, and disability  
status. We also address the cultural impact of household size (NP) on female labor force  
participation

**Introduction**

The state on which the data is being analyzed is New York. Some of the business questions we can answer from the data analysis include but are not limited to the following:

* We would like to closely estimate the least amount of salary an individual is accept based on factors such as age, sex, and education.
* How does an individual’s disability affect their wage income, and by how much.
* Should business that require low skill aim to recruit individuals with some type of disabilities.

**Table: Regression Results for Model A and Model B**

**OLS Multi-regression Models**

**Dependent Variable: Personal Wage Income**



**Model Assessment**

Goodness of fit in a multi-regression model are assessed by the Adjusted R-Square value and the F-Statistics. The closer to 1 is the Adjusted R-Square value, the better the ability of the model to explain variation in the dependent variable Wage Income (WAGP). The lower the p-value of the F-Statistics, the better the overall set of independent variables explain WAGP.

The F-Stats for model B is marginally better in model B (965 v 902.9). The Adjusted R-Square for model B is higher than in model A (0.1727 v 0.147).

However, because we use the hold-out method (training dataset of 60% of our data) in this multi regression, we assess the models using the RMSE value of the regression model on the remaining 40% of the data (validation set). The RMSE value for mode A is higher than B (202.57 v 177.83). Which suggest model overfitting in model A, so we select model B as the better model.

**Interpretation of Coefficients**

The multi-regression models estimate parametric coefficients using the Ordinary Least Square (OLS) method. There are several dummy variables in the models. The omitted dummies serve as benchmark to which the indicator variables are interpreted. The omitted dummy variables are working for family or self, non-American citizens, non-Married, Male, High School or less, and not disabled.

Using the results from the better model (B), I interpret the significant coefficients (p-value less than 5%, unless noted) as follow:

**Age:** Holding all relevant factors constant, if a respondent’s age goes up by one-year, typical wage income earned increases by $6524.44.

**Other Income**: holding all relevant factors constant, if a respondent’s other income goes up by $10,000, his/her wage income goes down by $8,360.

**Private employment:** holding all relevant factors constant, if a respondent works in the private sector, his/her wage income goes up by $26,429 compared to working for self or family.

**Public employment:** holding all relevant factors constant, if a respondent works in the public sector, his/her wage income goes up by $17,385 compared to working for self or family.

**Citizen:** holding all relevant factors constant, if the respondent is an American citizen, his/her wage income goes up by $8,886 compared to non-citizen.

**Married**: holding all relevant factors constant, if the respondent is married, his/her wage income goes up by $13,299 compared to non-married.

**Female**: holding all relevant factors constant, if the respondent is female, her wage income goes down $1,270 compared to male.

**Interaction Term Female\*Number of People**: holding all relevant factors constant, among females, as the number of people in the household goes up by one, the typical wage income earned goes down by $3,293.

**Bachelor’s degree**: holding all relevant factors constant, if a respondent holds a bachelor’s degree (or less but more than high school degree), his/her wage income goes up by $26,427, compared to someone with only a high school degree or less.

**Master’s degree or Higher**: holding all relevant factors constant, if a respondent holds a master’s degree (or higher), his/her wage income goes up by $69.700, compared to someone with only a high school degree or less.

**Physical disability**: holding all relevant factors constant, if a respondent has any physical disability, his/her wage income goes down by $19,272, compared to someone who does not have any disability.

**Ambulatory disability**: holding all relevant factors constant, if a respondent has an ambulatory disability, his/her wage income goes down by $5,592 compared to someone who does not have an ambulatory disability.

**Hearing disability**: holding all relevant factors constant, if a respondent has a hearing disability, his/her wage income goes down by $5,578 compared to someone who does not have a hearing disability.

**Visual disability**: holding all relevant factors constant, if a respondent has a visual disability, his/her wage income goes down by $4,416 compared to someone who does not have a visual disability.

**Economic Significance**

In model B, the largest three coefficients among the statistically significant variables (p-value<0.05) are: private employment ($26,429), Bachelor’s degree or less ($26,427), and Master’s degree or higher ($69,700).

**Predictions**

According to model B, an applicant who is 20 years old, with $0 other income, works in the private sector, one person in the household, American citizen, non-married, female, with a bachelor’s degree or less, and not disabled is predicted to earn a wage income of $16,933 in any given year.

According to model B, an applicant who is 40 years old, with $500 in other income, works in the public sector, four people in the household, American citizen, married, male, with high school or less in attained education level, and has physical disability is predicted to earn a wage income of $37,346 in any given year.

**New York vs Minnesota**

In Minnesota the 20-year-old is predicted to earn $28,939 vs $16,933 in New York all other variables equal. It seems that businesses can have less wage expenditure by hiring younger people in New York than in Minnesota. On the flip side, that 20-year-old is better off moving to Minnesota as they will earn more income with their same set of skills than in New York.

The 40-year-old in Minnesota is predicted to earn $23,205 vs $37,346 in New York. It looks like perhaps that this individual would earn more income if they moved to New York than if they stayed in Minesota. From a business perspective, targeting people like him in Minesota yields better results for businesses than it does in New York; keeping in mind that he does have a disability.

**Unbiasedness and Robustness**

The OLS regression models have not been tested for violations in model assumptions (multicollinearity, non-constant variance, omitted variable bias) or for model misspecifications.

**Appendix**

Attach screenshots of the R-studio syntaxes and output below.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated