JSC270 –Assignment 2 Report

Cindy (Shih-Ting) Huang

Summary

Given a dataset of income & associated factors in 1994 America, a linear regression model is built to explore how *sex, education_num,* and *gross_income_group* (independent variables) affect *hours_per_week* (dependent variable). They have a strong positive relationship.

Background & Motivation

In assignment 2, I examined the factors affecting income in 1994 America. Variables include *education_num* (total number of years of education achieved)*, sex* (male/female)*, gross_income_group* (whether group has earned 50K+ or less annually)*, hours_per_week* (number of hours worked per week) and others. Each observation in the census data represents a group of people that share the same set of values for the variables and has a corresponding weight that indicates its share of the population. From this data, I wanted to investigate if there is a strong correlation between weekly work hours and a specific population group's gender, education level, and income category.

Method

To do so, I fitted a linear regression model for dependent variable *hours_per_week* and added one independent variable at a time (*sex, education_num, gross_income_group*) to determine their relationship and better tune the model.

Statistical Results

We see that with each additional independent variable, the R-squared value increased, suggesting that each time a larger proportion of variability in *hours_per_week* is addressed by

the independent variables in the built model. From the three models' summaries, I discovered that all of those independent variables have a fairly strong positive correlation with *hours_per_week*. This is evident in the scale of their respective coefficient in the linear model. Each coefficient measures how much change occurs in *hours_per_week* with one unit change in the corresponding variable, so we see that work hours change the most with respect to subject's gender and the least in relation to years of education.

Discussion

Although correlation does not imply causation, several trends can still be concluded from the regression model. In general, males tend to work more than females. Additionally, work hours increase proportionally with income and years of education. This result is quite surprising given that traditionally, leisure was associated with higher-income families or intellectuals since their jobs would enable them to earn more profit for a less amount of work. Our data illustrates that the reality in America was the opposite of conventional belief, that as one achieves higher level of education and earn more income, work hour count also rises.

Of course, the linear regression model is a preliminary one. There could be other factors not included in the model, such as marital status or nature of the occupation, that support this trend or present alternative patterns. Hence, it is beneficial to build more complex models in the future to accurately assess the association between characteristics.