Assignment 2 Report

Relationship between working hours and education level & roles in family

---

Part 1: Background

This “Census Income” dataset was extracted by Barry Becker from the 1994 America census bureau database, and it contains data from 32561 individuals. It was originally used to predict whether a person’s income exceeds 50K/year based on factors like number of years of education, weekly working hours, sex, occupations, race, etc.

Part 2: Question of Interest

As we see in the assigned task that men tend to work for more hours per week than women, I am curious if this phenomenon further extends to subgroups like wife and husband. Thus, I want to investigate **if the number of weekly working hours is linearly related to number of years of education and roles in family (wife or husband)**.

Part 3: Description of my model

As my question of interest only concerns wives and husbands, I select only these people from the *relationship* variable in the original dataset. Other variables I used from the dataset are *hours_per_week* and *education_num*.

Then, I fit these 3 variables into a linear regression model where *education_num* (the number of hours of work per week) is the dependent variable, and the other two are independent variables.

Part 4: Interpretation

1. Effects of years of education

Based on the OLS regression results, **for both wives and husbands, weekly working hours increases as their years of education increase**. The estimated average change in weekly hours of work is 0.45, when years of education increases by 1 and roles in family holds constant. The low p-value which approaches 0 indicates there is statistically significant association between years of education and weekly hours of work.

We can interpret this result as no matter what a person's family role is, weekly working hours and years of education is positively related. Hence, **both men and women could benefit from being more educated, because work more may lead to gain more**. However, to draw a more concrete conclusion, we need to further consider their gross income.

2. **Effects of roles in family**

Wives generally work fewer hours per week than husbands, because the estimated average difference in weekly working hours between wives and husbands (wives' minus husbands') is -7.32 hrs, when years of education holds constant. And we are 95% confident that the population's average difference is between (-7.929, -6.712), which is a negative range.

Furthermore, because of the extremely low p-value which approaches 0, we tend to believe there is indeed some statistically significant linear association between weekly hours of work and roles in family.

Hence, when years of education holds constant, roles in family affects one's working hours. This might indicate **the division of responsibilities/work between wives and husbands, or more generally, between men and women in the current society.** It would be interesting to further investigate how weekly working hours for wives and husbands change across years or decades, and relate that to how society changes.

3. **The whole model**

Consider several statistics that evaluates how good my model fits the observed data:

(1) **RSE and RMSE**: these statistics are both **11.62**. As they indicate how large the standard error is, my model has relatively large standard error. However, when comparing with the 3 assigned models in part 2 of this assignment, this model has the lowest RSE and RMSE.

(2) **R-squared: 0.046**. Compared with the 3 assigned model, my model's R-squared is low, meaning it doesn't fit the observed data well, i.e. doesn't explain the variation well. However, as all coefficients described in part 4.1 and part 4.2 are all statistically significant, we can still gain insight into the average change of weekly working hours based on years of education and roles in family.

Hence, my model can still explain the association between weekly working hours and covariates.