Analyzing the Impact of Gender and Education on Working Hours

A Statistical Approach to Understanding Trends in Workforce Participation

Author: Omar Baayoun

Date: 25-11-2024

Introduction

Situation

Understanding working hours by gender and education is significant in analyzing workforce trends. Using statistical methods allows us to uncover disparities and their underlying causes. This report uses the UCI Adult Dataset to analyze working hours and their relationships with gender, education, and income.

Complication

There are notable disparities in working hours between genders, potentially influenced by structural societal differences. Education may play a critical role in shaping working hours. Analyzing these trends requires careful statistical modeling.

Question

1. Who works more hours—men or women?
2. Does controlling for education affect the gender trend in hours worked?
3. Is education a significant predictor of working hours?
4. Does including income improve the model explaining working hours?

Analysis and Results

1. Descriptive Statistics by Gender

Descriptive statistics highlight the average hours worked by males and females, as well as the variability within each group.

Sex	Count	Mean	Std	Min	25%	50%	75%	Max
Female	10771.0	36.41	11.81	1.0	30.0	40.0	40.0	99.0
Male	21790.0	42.43	12.12	1.0	40.0	40.0	49.0	99.0

Key Insight: Males work more hours on average than females and exhibit greater variability in hours worked.

2. Regression Models

The following models explore how working hours are influenced by gender, education, and income:

Model 1: Hours-per-Week as a Function of Sex

Intercept: 42.39
Coefficients: [-6.03844028]
Mean Squared Error (MSE): 146.33
R^2 Score: 0.05
Insight: Gender alone explains a small portion of the variance in working hours. Females work 6 hours less on average.

Model 2: Hours-per-Week as a Function of Sex and Education

Intercept: 35.54
Coefficients: [-5.97696086, 0.67727494]
Mean Squared Error (MSE): 142.41
R^2 Score: 0.08
Insight: Including education increases the explanatory power of the model. Education positively correlates with working hours.

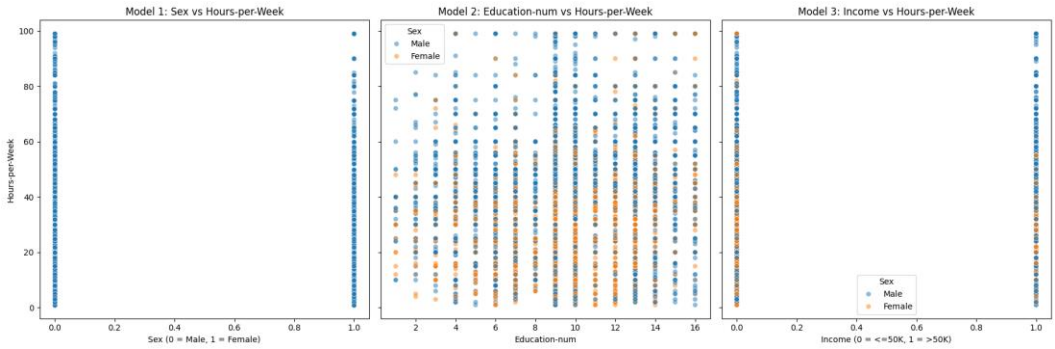
Model 3: Hours-per-Week as a Function of Sex, Education, and Income

Intercept: 36.67
Coefficients: [-5.10634826, 0.42860821, 4.54554483]
Mean Squared Error (MSE): 139.45
R^2 Score: 0.09
Insight: Adding income improves the model further. Higher-income individuals work more hours on average.

Diagnostics and Conclusion

Residual Analysis

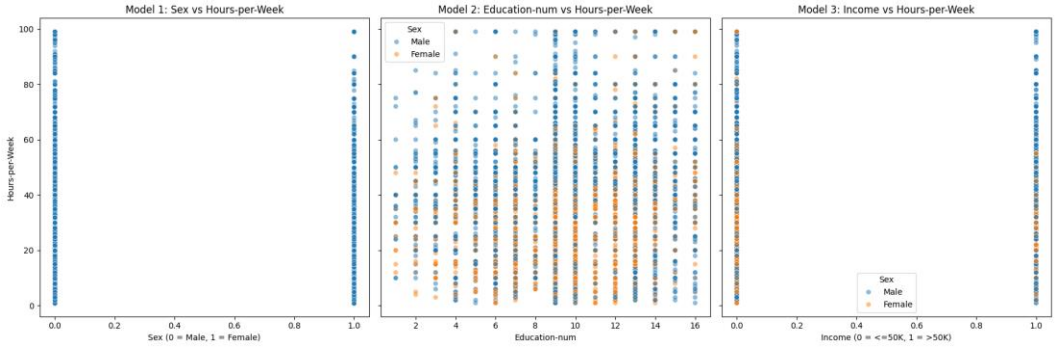
Residual analysis for all models shows approximate normality of residuals. Model 3 shows the best fit with the lowest MSE.



Caption: Residual Analysis for Models 1, 2, and 3

Influential Points (Cook's Distance)

Cook's Distance plots identify a few influential data points with high leverage. These points might disproportionately impact model results.



Caption: Leverage vs. Residuals for Models 1, 2, and 3

Conclusion

1. Males work more hours on average compared to females.
2. Education is a significant predictor of working hours and positively correlates with them.
3. Including income in the model further improves its explanatory power.
4. Model 3 is the best performing model based on MSE, though the R^2 value suggests more variance remains unexplained.

These findings shed light on the persistent gender differences in working hours, which are influenced by factors like education and income. However, it's important to recognize that this analysis doesn't capture the full picture—factors such as family responsibilities and other unaccounted variables could also play a significant role.