Analyzing the Gender Wage Gap

Background

The gender wage gap refers to the difference in earnings between females and males in the workforce. The Equal Pay Act of 1963 and the Civil Rights Act of 1964 guarantee protections against discrimination based on traits including age, race, and sex; however, many studies suggest that the gender wage gap has not been eliminated entirely. Let us investigate the gender wage gap using a dataset based on a portion of the Current Population Survey Uniform Extracts from the Center for Economic and Policy Research (Center, 2019).[1] For this analysis let us focus on *adult earners* who made $150,000 or less per year. The variables in the dataset are described in the table.

Variable	Description
Income	in USD
Age	in years
Sex	Male; Female
Race	White; Black; Hispanic; Asian; Other
Education	Less than high school; High school; Some college/Associates; Bachelors; PhD/MD/JD
MaritalStatus	Never married; Married; Separated; Divorced; Widowed
Employed	Yes ; No

Step 1: Ask a research question What is the research question?

Step 2: Design a study and collect data Use no more than two sentences to describe the data. Be sure to indicate how many observational units are in the analysis and what values the variables take on.

Your response.

Step 3: Explore the data

[1] All variable modifications are documented here.

a. Which variable is the explanatory variable? Which is the response variable?

b. Summarize income using a histogram and a boxplot. Do the mean and median income differ? Which is the more appropriate measure of central tendency (average)? Why?

c. Use boxplots to plot the conditional distributions of income for males and females. Consider whether there appears to be an association between the two variables. Can we conclude that there is a wage (income) gap? Explain.

d. Identify a potential confounding variable and explain how it may be confounding in this study.

e. Use an appropriate plot(s) to explore the association between income and the third variable. Is there evidence of a relationship between these two variables? Explain.

f. Describe the relationship between sex and third variable.

g. Create boxplots for income distributions conditional on the level of the third variable for males and females. Do the plots suggest that the third variable is confounding? Explain. *Your response.*

h. Does sex explain all the variation we observe in income? Explain using a Sources of Variation Diagram:[2]

Step 4: Draw inferences beyond the data Name one theory-based approach we could use to determine if there was a statistically significant difference in incomes between the sexes.

Step 5: Formulate conclusions Summarize the conclusions we could draw from this study. How far can we generalize our conclusions? Could we draw a cause-effect conclusion if we observed statistically significant differences in Step 4? Why or why not?

Step 6: Look back and ahead Identify limitations to our approach. What could we do to address them in a future study.

Further Investigation

Click here for an interactive, custom-designed applet that can be used to explore this dataset further.

References

Center for Economic and Policy Research. March CPS Uniform Extracts, Version 1.1. Washington, DC. https://ceprdata.org/cps-uniform-data-extracts/cps-outgoing-rotation-group/cps-org-data/. (2019). Accessed: 2020-08-12.

[2] A convenient table generator can be foundhere.