

SOCI 385 - Fall 2023

Final Project

This assignment is due by 11:59 pm on Monday, December 18, 2023. Submit a rendered pdf of your report, a copy of your qmd file, *and a csv file of your dataset* via Canvas.

For this project, you will prepare a written report related to the research question that you proposed. Write in complete sentences and paragraphs and include a header for each new section of your report.

GUIDELINES

- Your project should be submitted individually, but you are allowed to ask other students enrolled in the course simple questions about coding and formatting.
- Clearly label any figures you include. This means all figures should have titles for axes, plots, and legends as appropriate.
- When possible, summarize your results so they can be interpreted by someone who is familiar with statistics but is not familiar with R output. Format your results in clean tables rather than dumping raw R output into your report. (Note: we will spend time during the last class going over how to present results.)
- There is no time limit (other than the submission deadline).
- You are free to consult lecture slides, notes, and previous homework assignments and problem sets.
- Late assignments will not be accepted without prior approval.

OUTLINE OF REPORT

Introduction (2 points)

State your research question and hypotheses. Why do you expect there to be an association between your dependent variable and your independent variable? How would you expect your control variable to relate to this association? (Note: no outside reading or research is required or expected.)

Explain the sample you will use for your analysis. Are you using the full set of observations in your dataset or are you restricting your sample to a subset (for example, GSS respondents from 2022 between the ages of 25-65)?

Define the intended audience for your report. For example, are you writing an academic paper, a policy memo, an article for a popular press, etc.?

Descriptive Statistics (10 points)

Briefly summarize your dependent variable, independent variable, and control variable using some of the key types of descriptive statistics (such as the distribution, center, spread, shape, etc.).

Describe the correlation between your dependent variable and your independent variable. (Note: this assumes your dependent variable and independent variable are continuous. If you were approved to use variables that are not continuous, you do not need to try a correlation.)

Include and interpret a figure that shows the mean of your dependent variable by each level of your categorical independent variable.

Create and summarize a new categorical variable based on your dependent variable and explain how you chose which values to put in which categories.

Create and summarize a new categorical variable based on your independent variable and explain how you chose which values to put in which categories.

Using the variables you created above, create a table showing the distribution of your categorical dependent variable and categorical independent variable holding your control variable constant. Interpret the results.

Inference (6 points)

Conduct *two* of the following three tests of statistical inference. Choose the appropriate tests based on the types of variables you are using and your sample sizes. State the null and alternative hypotheses, include each p-value, and interpret the results fully. You can do the tests in R.

- Test if the mean of your continuous dependent variable differs significantly at the .05 alpha level between two groups/levels of your categorical independent variable.
- Create a binary variable from your dependent variable. Test if the proportions with a 1 for this new binary variable differ significantly at the .05 alpha level between two groups/levels of your control variable.
- Test if your categorical dependent variable and your control variable are independent.

Regression (9 points)

Regress your dependent variable on your independent variable. Interpret the results in words, describe any coefficients that are significant, and use the coefficients to calculate two predicted values of the outcome.

Now add your control variable to the model you estimated above. Interpret the results, describe any coefficients that are significant, and use the coefficients to calculate the predicted values of the outcome for each level of your control variable.

Now add an interaction between your independent variable and your control variable. Interpret the results and describe any coefficients that are significant.

Conclusion (3 points)

Briefly summarize your main results and offer a sociological explanation for your findings. Were your hypotheses confirmed? Are there additional variables that you would want to include in other analyses? Are there limitations in the current sample design that should be addressed in a follow-up study? Include a few possible directions for future research based on your results.