SOCI 385 - Social Statistics

Fall 2021 - Problem Set 2

This assignment is due via Canvas by 12:00 PM on Monday, November 15, 2021. Show all your code, write your responses in complete sentences, and label axes and plots as appropriate. Submit your knitted pdf and the R Markdown (.Rmd) file for your notebook. Remember to include your name in the header.

REMINDER: Problem sets must be completed individually. No collaboration with other students is allowed. You may use R Shortcuts in all questions except #1 where you should show all your work.

This problem set requires the `ps2.csv` file on Canvas which includes data from the 2010-2018 waves of the General Social Survey. Any missing values are coded as `NA`. The variables are:

`year`: year the survey was completed, either 2010, 2012, 2014, 2016, or 2018;

`age`: respondent's age, from 18-89;

`racehisp`: respondent's self-reported race and Hispanic identification, categorized as "Black", "Hispanic", "Other", or "White";

`degree`: respondent's highest educational category, coded as "Less than HS", "HS Diploma", "Some College", "Bachelor's Degree", "Grad/Prof Degree";

`eqwlth`: ***For a description of this variable, search the gssr package;***

`conlegis`: ***For a description of this variable, search the gssr package***

Questions

1. Without using any R shortcuts, find the 95% confidence interval for the mean of `eqwlth` in each of the following years: 2010, 2014, and 2018. Plot these intervals in a figure (with error bars), and use your figure to describe how the mean responses have changed over the survey years.

2. Create a new variable grouping the `age` variable into the following categories: 18-24, 25-39, 40-54, 55-64, 65+. Which (if any) age categories showed significant differences in mean `eqwlth` scores between the 2010 and 2018 surveys? What is a sociological explanation for these differences?

3. Does the proportion of respondents with "Hardly any" confidence in congress differ between respondents at the lowest and highest extremes of the `eqwlth` scale? What is an additional variable you would want to explain your result in more detail?

4. In a single paragraph, summarize the results of the following three tests of association. In addition to offering sociological interpretations of your findings, describe why you chose which statistical tests to use.

- Is there an association between `racehisp` and `eqwlth`?
- Among respondents with less than a high school diploma, is there an association between `racehisp` and `eqwlth`?
- Is there an association between age (using the categories you created in #2) and confidence in Congress?