# *SOC 5050: Lab 13*

*Christopher Prener, Ph.D.*

*November 14th, 2016*

## *Directions*

Please complete all steps below. Your final do-file, log-file, plots, and markdown file with answers should be uploaded to your GitHub assignment repository by 4:20pm on Monday, November 21st, 2016. You can show your work in your do-file using the `display` command. The following analysis will utilize four source variables from the 2011 CPS dataset: `HEFAMINC`, `PESEX`, `PEHSPNON`, and `PEHRUSL1`. Before you begin, look up each of these variables in the codebook to learn about their meaning and structure.

## *Initial Steps*

1. For each of the variables, recode them to properly handle missing data. Be sure that your new variables are properly labeled with well-formatted value labels. Using variable notes, document the recoding process.
2. For each of the variables, calculate the appropriate baseline statistics.
3. For each of the continuous variables, assess whether these variables should be considered normally distributed.

## *Calculating and Interpreting Correlations*

4. Using the variables you have cleaned, construct a correlation matrix. For this assignment, you do not need to create a table - simply provide the output in your markdown file.
5. Provide interpretations of the correlations between the variable you created from `HEFAMINC` and the other three variables you created from the source variables.
6. Create three scatter plots of the relationship between the variable you created from `HEFAMINC` and the other three variables you created from the source variables. Make sure these are well-designed plots that use all of the appropriate options.

7. Create a factored scatter plot that shows variation in family income and hours worked between non-Hispanic and Hispanic respondents. Make sure this is a well-designed plot that uses all of the appropriate options.

8. Create a faceted scatter plot that shows variation in family income and hours worked between non-Hispanic and Hispanic respondents. Make sure this is a well-designed plot that uses all of the appropriate options.

## *Power Analysis*

9. What sample size would be needed to identify a moderate effect size of $r = 0.45$ with statistical power of 0.8?

10. What sample size would be needed to identify a weak effect size of $r = 0.25$ with statistical power of 0.9?

## *Cronbach's α*

11. Calculate and interpret the results of a Cronbach's $\alpha$ for the variables you created and cleaned above.

12. Do these variables make a meaningful (i.e. valid), reliable scale? Why or why not?

## *Document Details*

Document produced by Christopher Prener, Ph.D. for the Saint Louis University course SOC 5050 - QUANTITATIVE ANALYSIS: APPLIED INFERENTIAL STATISTICS. See the course wiki and the repository `README.md` file for additional details.