SOC 5050: Problem Set 02

Christopher Prener, Ph.D.

September 5th, 2016

Directions

Please complete all steps below. Include your final stack of do-files, log-file, plots, and markdown file. Your markdown file should contain source code for PART 2 as well as your narrative, which should contain answers to the questions below. All requested documents should be uploaded to your GitHub assignment repository by 4:20pm on Monday, September 12th, 2016.

Part 1: Initial Data Cleaning

1. Beginning with the appropriate course template, complete a master do-file to trigger your `data.do` and `analysis.do` files.
2. Beginning with the appropriate course template, construct a data do-file that completes the following tasks:
   (a) Properly declare values as missing for the variable `PSPEDEIS`. If applicable, create missing values that differentiate between different reasons data may be missing. Be sure to use the full workflow for recoding variables.
   (b) Properly declare values as missing for the variable `PSPEDEM`. If applicable, create missing values that differentiate between different reasons data may be missing. Be sure to use the full workflow for recoding variables.
   (c) Starting with the variable `AGE_P`, create a new, recoded variable that is an *ordinal* measure. Select categories that you believe make the most sense with the goal of having five or six values in your new variable. Be sure to use the full workflow for recoding variables.
   (d) Starting with the variable `LCTIME1`, create a new, recoded variable that is an *ordinal* measure. Properly declare values as missing, but leave the category for "since birth" as is. For all other values, select categories that you believe make the most sense with the goal of having three or four values in your new variable (plus the category "since birth"). Be sure to use the full workflow for recoding variables.

(e) Starting with the variable `RACERPI2`, create a new, recoded variable that is a binary measure representing "non-white". Properly declare values as missing for the category "Race group not releasable", and include values for "multiple race" as "non-white". Be sure to use the full workflow for recoding variables.

Part 2: Missing Data and Descriptive Statistics

3. Beginning with the appropriate course template, construct a analysis do-file that completes the following tasks:

   (a) Conduct a missing data analysis on all of the variables you modified or created in PART 1. How much missing data is there? Are these data possibly MCAR or MAR? Are there threats to generalizability that are revealed by these tables?

   (b) Create a frequency table and obtain the appropriate descriptive statistics for the variable you created from `AGE_P`. Describe your findings and include a justification for why these are the appropriate statistics to report.

   (c) Create an appropriate plot for the variable you created from `AGE_P`. Interpret the plot and include a justification for why this is the appropriate plot.

   (d) Create a frequency table and obtain descriptive statistics for the variable you created from `LCTIME1`. Describe your findings and include a justification for why these are the appropriate statistics to report.

   (e) Create an appropriate plot for the variable you created from `LCTIME1`. Interpret the plot and include a justification for why this is the appropriate plot.

   (f) Create a frequency table and obtain descriptive statistics for the variable you created from `RACERPI2`. Describe your findings and include a justification for why these are the appropriate statistics to report.

   (g) Create an appropriate plot for the variable you created from `RACERPI2`. Interpret the plot and include a justification for why this is the appropriate plot.

Grading Rubric

Part 1 Completing question one is worth three points towards the thirty point total for the assignment. Each lettered element of question two is worth one point (five points total) towards the thirty point total for the assignment. These points are awarded based on correctly completing all of the workflow steps for recoding missing and valid data.

Part 2 Each lettered element of question three is worth one point (seven points total) towards the thirty point total for the assignment. A third of the credit comes from the proper use execution of the relevant Stata commands, a third comes from stating the correct answer, and a third comes from your accompanying justification.

Stata Do-File The overall quality of the Stata do-file stack is worth twelve points. This grade will be based on the clarity, organization, and layout of your do-files.

Design An additional three points are based on the layout and design of each of your figures. This grade will be based on the use of schemes as well as customization of the plots (titles, subtitles, and notes).

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.