SOC 5050: Problem Set 07

Christopher Prener, Ph.D.

October 31st, 2016

Directions

Please complete all steps below. This lab requires that you use the 2013 NHIS dataset. Your final work by do-file, log-file, and markdown file with answers should be uploaded to your GitHub assignment repository by 4:20pm on Monday, November 7th, 2016.

Deep Cleaning Data

1. Create a new martial status variable (from the current variable `R_MARITL`) that (a) properly handles missing data and (b) reduces categories in a logical and consistent manner. Exclude all participants from this variable who are under 18 years of age.[1] Be sure that your new variable is properly labeled with well-formatted value labels. Using variable notes, document the recoding process.
2. Construct logic checks using the `assert` command to ensure that your recodes worked correctly.
3. Report descriptive statistics for this new variable that are appropriate given its level of measurement. Be sure to use local macros to insert values into your answers.
4. Produce a well-formatted plot for this new variable that shows its distribution of data. Select the plot type that is *most* appropriate given its level of measurement.
5. Create dummy variables of your new variable. Be sure that your new variables are properly labeled with well-formatted value labels. Using variable notes, document the recoding process.
6. Use a loop to create new versions of the following variables: `LCTIME1`, `LCTIME2`, `LCTIME3`, `LCTIME4`, `LCTIME5`, `LCTIME6`, `LCTIME7A`, `LCTIME8`, `LCTIME9`, `LCTIME10`, `LCTIME11`, `LCTIME12`, and `LCTIME13`. Each of these variables measures the length of time that a participant has had a particular medical condition. These new variables should be ordinal variables that have the following categories: "less than 1 year"[2], "1 to 4 years", "5 to 9 years", "10 to 14 years", and "15 to 17 years". If respondents have have the condition since

[1] *Hint*: Use the variable `AGE_P`.

[2] *Hint*: Use the corresponding `LCUNIT` variable (i.e. if you are working with `LCTIME1`, the corresponding variable would be `LCUNIT1`) to determine the units associated with the `LCTIME` variable.

birth, use their age[3] as the length of time they have had the given medical condition. Make sure that all values that indicate missing data are properly coded as missing in the new variables.

7. Drop all variables other than those that you have created in this lab.
8. Add a dataset label and dataset-level notes to your dataset.
9. Create data documentation files (a variable index and codebook).
10. In your analysis file, output the descriptive statistics for each of the new variables you created.
11. Use a loop to create well-designed and laid out plots for each of the variables you created above.

[3] *Hint*: Use the variable `AGE_P`.

Grading Rubric

Part 1 This section is worth 22 points (2 points per question).

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

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. Data are drawn from the ULCA Institute for Digital Research and Education.