SOC 5050: Problem Set 09

Christopher Prener, Ph.D.

November 21st, 2016

Directions

Please complete all steps below. Your final work by hand, table file, do-file, log-file, plots, and markdown file with answers should be uploaded to your GitHub assignment repository by 4:20pm on Monday, November 28th, 2016. This assignment uses the 2011 CPS dataest.

Initial Steps

1. Properly clean and recode the variables `HRNUMHOU` (family size), `PTDTRACE` (racial identity), and `PEHSPNON` (Hispanic identity).
2. Calculate descriptive statistics for each of these variables. Using a word processor of your choice, create a well-formatted table summarizing these descriptive statistics.
3. For each variable, create the appropriate descriptive plot.
4. Create box and whisker plots comparing family size with racial identity, and family size and Hispanic identity.
5. Conduct normality testing on your dependent variable to see if it meets the assumptions required for linear regression.
6. Finally, recode your race variable a second time so that it is a series of binary "dummy" variables that can be used for regression analysis.

Regression Model 1: Family Size and Hispanic Identity

6. Construct a hypothesis and null hypothesis for the relationship between family size and Hispanic identity.
7. Construct a regression equation modeling how Hispanic identity affects family size on a separate piece of paper. Scan this and turn it in with your repository.

8. Execute a bi-variate regression model that shows how Hispanic identity affects family size. For this model, use robust standard errors. Fully interpret the results of this model.

Regression Model 2: Family Size and Race

9. Construct a hypothesis and null hypothesis for the relationship between family size and race.
10. Construct a regression equation modeling how race affects family size on a separate piece of paper. Scan this and turn it in with your repository.[1]
11. Execute a bi-variate regression model that shows how race affects family size. For this model, use robust standard errors. Fully interpret the results of this model.[2]

[1] *Hint:* Be sure to include all dummy variables for race except for one reference category.

[2] *Hint:* Be sure to include all dummy variables for race except for one reference category.

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.

This work is licensed under a Creative Commons Attribution 4.0 International License.