SOC 4930/5050: Lab-11 - Correlations in R

Christopher Prener, Ph.D.

November 6ᵗʰ, 2017

Directions

Please complete all steps below. Your your well formatted R Notebook (`.Rmd` file) and `html` output as well as your LATEX output should be uploaded to your GitHub assignment repository by 4:15pm on Monday, November 13ᵗʰ, 2017.

Correlation in R

1. Create a correlation table for the following three variables in the `auto17` data set from `testDriveR`: `combFE`, `fuelCost`, and `cyl`.

2. Provide an interpretation for the relationship between combined fuel efficiency (in miles per gallon) and fuel cost.

3. Provide an interpretation for the relationship between fuel cost and the number of cylinders in an engine.

4. Provide an interpretation for the relationship between combined fuel efficiency (in miles per gallon) and the number of cylinders in an engine.

5. Use `stargazer` to create a *well-formatted* LATEX version of this table.

Creating Scatterplots

6. Create a scatterplot of the relationship between combined fuel efficiency (in miles per gallon) and fuel cost.

7. Create a binary logical variable based on `guzzlerStr`, where `TRUE` is for vehicles that are gas guzzlers and `FALSE` is for all other vehicles.

8. Use the new guzzler logical variable as the grouping variable on a scatterplot of the relationship between combined fuel efficiency (in miles per gallon) and fuel cost.

9. Use the new guzzler logical variable as the faceting and grouping variables on a scatterplot of the relationship between combined fuel efficiency (in miles per gallon) and fuel cost.

10. Add best fit lines for both guzzler and non-guzzler vehicles, using both different colors and different patterns for each line.

Cronbach's alpha

11. Use the same three variables - `combFE`, `fuelCost`, and `cyl` - and evaluate them to see if they represent (potentially!) the same underlying construct of "fuel efficiency".

Sample Size Estimates

12. What sample size would be needed to detect a correlation coefficient of $r = .89$ with power of .8 in a two-sided test of significance?

13. What sample size would be needed to detect a correlation coefficient of $r = .29$ with power of .9 in a two-sided test of significance?