SOC 4015/5050: Lab-10 - Correlations in R

Christopher Prener, Ph.D.

Fall 2018

Directions

Please complete all steps below. All work should be uploaded to your GitHub assignment repository by 4:15pm on Monday, November 12th, 2018.

Analysis Development

Using RStudio and your operating system's file manager, create an R Project in the *existing* directory in your assignments repository named `Lab-10`. Add a `README.md` file, notebook, and all necessary folders before beginning.[1]

[1] This initial section follows the project workflow that is available in the `lecture-03` repo!

Data Preparation

1. Using the `auto17` data set from `testDriveR`, create a binary logical variable based on `guzzlerStr`, where `TRUE` is for vehicles that are gas guzzlers and `FALSE` is for all other vehicles.
2. Create a subset of the data that contains the following four variables: `combFE`, `fuelCost`, `cyl`, and your new binary guzzler variable. Write your analysis data to the `data/` subdirectory as a `.csv` file.
3. Conduct a quick missing data analysis - is missing data a problem in this data set? Use `knitr` to make sure this output is nicely formatted in your notebook.

Creating Scatterplots

For each plot below, make sure you have a version embedded in your notebook with a copy saved in your `results/` subdirectory.

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

5. 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.
6. 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.
7. Add linear model lines for both guzzler and non-guzzler vehicles, using both different colors and different patterns for each line.
8. Create a statistical scatterplot that also contains Pearson's *r* output for the relationship between combined fuel efficiency (in miles per gallon) and fuel cost.

Correlation in R

9. Create a correlation table for the four variables in your analysis data set. Use `knitr` to create a *well-formatted* version of this table in your notebook, and save a `.csv` version of this table in your `results/` subdirectory.
10. Provide an interpretation for the relationship between combined fuel efficiency (in miles per gallon) and fuel cost.
11. Provide an interpretation for the relationship between fuel cost and the number of cylinders in an engine.
12. Provide an interpretation for the relationship between combined fuel efficiency (in miles per gallon) and the number of cylinders in an engine.

Sample Size Estimates

13. 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?
14. 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?