SOC 4015/5050: Lab-14 - ANOVA

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, December 10th, 2018. All data can be obtained from the `testDriveR` package's `auto17` data set.

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-14`. 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!

Part 1: Data Preparation and Plotting

1. Subset your data so that it contains only the `id`, `hwyFE`, and `driveStr2` variables.
2. Convert the variable `driveStr2` to a factor using `as.factor()` embedded in a `mutate()` call:

   ```
   > x <- mutate(x, aFac = as.factor(a))
   ```

3. Create a well-formatted violin plot *or* box plot of the differences in highway fuel efficiency between vehicles based on their drivetrain using your *factor* variable you created above.
4. Calculate the mean highway fuel efficiency for each group within the factor variable you created above representing drivetrain.[2]

[2] *Hint*: Use the `group_by()` and `summarize()` functions from `dplyr`!

Part 2: Assess Assumptions

Using the data created in Part 1, answer the following questions. Use highway fuel efficiency as your dependent variable and the factor variable you created above representing drivetrain type as your independent variable.

5. Check the homogeneity of variance assumption using the Bartlett Test.
6. Check the normality assumption using the standard techniques we've used this semester.

Part 3: Fit the ANOVA

7. Fit and interpret the results of an ANOVA using the highway fuel efficiency as your dependent variable and the factor variable you created above representing drivetrain type as your independent variable.
8. Use Tukey HSD values to report which of the permutations have statistically significant differences in means.

Part 4: Check for Outliers

9. Use the Bonferonni Test to identify any outliers in the model.