

## SOC 4930/5050: Lab-15 - ANOVA

Christopher Prener, Ph.D.

December 4<sup>th</sup>, 2017

### Directions

Please complete all steps below. Your your well formatted R Notebook (.Rmd file) and html output should be uploaded to your GitHub assignment repository by 4:15pm on Monday, December 11<sup>th</sup>, 2017. Use the auto17 data from testDriveR for this assignment.

### 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.<sup>1</sup>

<sup>1</sup> *Hint:* Use the `group_by()` and `summarize()` functions from `dplyr`!

### 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.

*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.

*Check for Outliers*

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