SOC 4930/5050: Lab-16 - Chi-Squared

Christopher Prener, Ph.D.

December 11th, 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 4:00pm on Monday, December 18th, 2017. Use the `auto17` data from `testDriveR` for this assignment.

Data Preparation

1. Create a vector named `germanCars` that contains the names of the four major German car manufacturers in the `auto17` data - BMW, Mercedes, Porsche, and Volkswagen.[1]
2. Create a new logical variable that is `TRUE` if the vehicle is a "German" vehicle and `FALSE` otherwise.[2]
3. Subset your data so that it contains only the `id`, your new logical variable, and the `driveStr` variables.

[1] *Hint*: Use the `c()` function!

[2] *Hint*: Use the `%in%` function as part of a `mutate()` call to see if a given vehicle's manufacturer is listed in the `germanCars` vector.

Create Tables

Using the data created in Part 1, answer the following questions.

4. Create a one-way table of the logical variable you created above using `janitor` that includes:
   (a) total row at the bottom and
   (b) property formatted percentages that display three decimal places.[3]
5. Create a two-way table of the logical variable you created above and `driveStr` using `janitor` that includes
   (a) a total row at the bottom and a total column,
   (b) properly formatted row percents that are display three decimal places,
   (c) and frequency values in the "front" position.

[3] *Hint:* You can skip the `adorn_percentages()` function here because a percentage column is automatically included in one-way tables.

Fit the Chi-Square and Check Assumptions

Using the data created in Part 1, answer the following questions.

6. Fit and interpret the results of a chi-squared test comparing the relationship of German vehicles to drivetrains. Is there a meaningful relationship between these two variables?

7. Does this model violate the Cochran conditions assumption?

8. Regardless of your answer above, fit a Fisher's Exact Test on these same data. Does this change your interpretation of question 6?