Lab 13 - Chi square, ANOVA, & correlation

Your name here

November 21, 2017

Complete the following exercises below and include all code used to find the answers. Knit together the PDF document and commit both the Lab 13 RMD file and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

1. Select two categorical variables from your dataset whose association you're interested in and conduct a chi-square test. *If you only have continuous variables you will need to create categorical versions of these variables to make this work. You can do this using the* `cut` *function in mutate to add a new, categorical version of your variable to your dataset.*

a. Describe any modifications made to your data for the chi-square test and the composition of the variables used in the test (e.g., study time is measured using a three-category ordinal variable with categories indicating infrequent studying, medium studying, and frequent studying).

b. Does there appear to be an association between your two variables? Explain your reasoning.

c. What are the degrees of freedom for this test and how is this calculated?

d. What if the critical value for the test statistic? What is the obtained value for the test statistic?

e. How do you interpret the results of this test and the implications for your theoretical arguments about these two variables?

2. Select one continuous variable and one categorical variable from your dataset whose association you're interested in exploring. *Again, note that you'll need to create a categorical version of your independent variable to make this work.*

a. Describe any modifications made to your data for the ANOVA test and the composition of the variables used in the test (e.g., college rank is measured using a four-category variable with values indicating freshman, sophomore, junior, and senior class).

b. What are the degrees of freedom (both types) for this test and how are they calculated?

c. What is the obtained value of the test statistic?

d. What do the resuts tell you about the association between these two variables? What does this mean for your theoretical arguments about these variables?

3. Select two continuous variables from your dataset whos association you're interested in exploring.

a. What is the correlation between these two variables?

b. Create a scatterplot of the variables you selected. Does the correlation coefficient accurately represent the relationship between these two variables? Why or why not?

c. Create a correlation matrix of your data using the `ggcorr` function from the `GGally` package. Be sure to label each cell with the correlation coefficient.

d. What does this visual representation of correlation coefficients tell you about your data? Are there any relationships (or lack thereof) that are surprising to you? Why or why not?

e. What are the limitations of correlation coefficients? Can they ever be misleading? If so, in what ways?