Portfolio 2
PA 606: Seminar in Quantitative Techniques
(Spring 2020)

25 points

Due: Saturday, April 4, 2020

Topics:

Bivariate Relationships – Analyzing the Difference in Means:
t-Test (t), Analysis of Variance (F)

Data Sources:

`GSS2014` (2014 General Social Survey data set located in the `vannstats` package)

- variable names and descriptions, as well as the codebook (linked), can also be found in the `vannstats` package information

Overview:

In this assignment, you'll be examining mean differences. Specifically, you'll be investigating whether or not groups have significantly different means. That is, is the mean of one variable different for different groups.

Answer each of the questions below, **in full sentences/paragraphs**, and show your tables/plots (if necessary) inline (interspersed within the text). For full credit, you must append (copy and paste) your R Script at the end of this portfolio as the final page(s).

Problems

Data Set Information

- Describe the data set, including the name and who administered the data/survey (e.g. which survey research firm), year administered, the unit of analysis, and the number of observations.

t-Test

In this test, you'll be examining mean differences in **occupational prestige** of a respondent's job (`sei10`) by their **sex** (`sex`).

- Describe the variables. Using the codebook and the list of variable descriptions for the `GSS2014` data set, for each variable, describe its text (e.g. the question asked for each variable), the level of measurement, and the values/categories within the variable.

- Define a research question for the variables (e.g. "*Is variation in X related to/associated with variation in Y; Is there a mean difference in Y by categories of X*").
- Define the null hypothesis (H_0) and the alternative hypothesis (H_1) for this test.
- Describe the various assumptions of the t-Test (t) and how you would assess them. Next, show all necessary tables/plots that demonstrate your assessment of whether or not you've met the assumptions of the test. (*Note: if showing tables, create them, do not simply copy and paste from your output.*)
- Run the test. Fully and correctly report the test. If you find significance, describe what you find in terms of mean differences by group (e.g. compare the group means).

ANOVA

In this test, you'll be examining mean differences in **occupational prestige** of a respondent's job (`sei10`) by their **type of occupation** (`workfor1`).

- Describe the variables. Using the codebook and the list of variable descriptions for the `GSS2014` data set, for each variable, describe its text (e.g. the question asked for each variable), the level of measurement, and the values/categories within the variable.
- Define a research question for the variables (e.g. "*Is there a mean difference in Y by categories of X*").
- Define the null hypothesis (H_0) and the alternative hypothesis (H_1) for this test.
- Describe the various assumptions of the ANOVA (F) and how you would assess them. Next, show all necessary tables/plots that demonstrate your assessment of whether or not you've met the assumptions of the test. (*Note: if showing tables, create them, do not simply copy and paste from your output.*)
- Run the test. Fully and correctly report the test. Whether or not you find significance, follow your results with a means plot to demonstrate the comparison of the means, and describe what you find in terms of mean differences by group (e.g. compare the group means).
- Describe when and why you would run a post-hoc test?

Extra Credit

- Run a post-hoc test for your ANOVA (specifically, Tukey's HSD) and describe your findings.