

Quantitative Methods II

Assignment 2

September 4, 2011

This is the second homework assignment for the course. Its purpose is to continue increasing your proficiency in using a statistical program, producing presentation-worthy graphics, and writing to get your point across.

Remember that all statistics are an attempt to gather information about a process (or population) from a sample of data generated by the process. As such, we will never be able to answer a statistical question with absolute certainty, only with statements of confidence ranges and of expected Type I Error rates.

This assignment covers more tests for measures of central tendency, boxplots, and validating those tests using Monte Carlo techniques.

If you need assistance using R, *do not* hesitate to ask for it. To get the most out of such assistance, you will need to explicitly explain your issue, attach the **code** you have already written, and start asking earlier than Sunday.

When you hand in this assignment, you will email to me two separate files, your **typed solutions** to the questions asked in the homework and a separate **script file**. The script file allows me to check that you did the correct analysis. The solution file allows me to see that you can answer the questions in complete and coherent sentences, weaving in graphics and statistics appropriately.

The email must include, as its subject line:

POLS6123: Assignment 2

Note. *Make sure you include no code in the write-up. All code needs to be attached to the email in the separate script file.*

PROBLEM: VARIOUS QUESTIONS

[[5]]

Please answer the following questions in complete sentences. No code is needed for this section. You do need, however, to cite your source(s).

- (1) What is the most important assumption of the analysis of variance test?
- (2) This assumption is also required in the t-test; however, why is it more important in the analysis of variance test?
- (3) What are three things you can do if the main assumption of the analysis of variance test is unreasonable in the data?

PROBLEM: PREDICTING NATIONALIZED PETROLEUM

[[10]]

The data file `clf` contains data concerning petroleum producing States around the world. Campbell Faulkner's research question was "What factors influence the government's decision to nationalize its petroleum resources?" One factor that his literature review suggested was the quality of the State. From this, Faulker hypothesized that States which nationalized their petroleum production tended to have lower quality levels, as measured by the Failed State Index (fsi).

In other words, if we define μ_N as the average fsi measure for states with nationalized petroleum production and μ_P as the average fsi measure for states with privatized petroleum production, then Mr. Faulkner's research hypothesis was $H_R : \mu_N < \mu_P$. As this hypothesis does not contain the 'null', or 'equal', position the null hypothesis to test will be its opposite,

$$H_0 : \mu_N \geq \mu_P$$

Test Faulkner's hypothesis. Make sure you test the assumptions appropriately and use the appropriate test. In your write-up, provide the necessary graphs and statistics to support and illustrate your conclusion.

PROBLEM: BIOMES AND THE MEAN FIRE RETURN INTERVAL [[10]]

The data file `biome2` is pseudo data that contains measurements on 30 sites and five biomes located around the United States. Lindsey Rao's research question was "Does the biome type affect the mean fire return interval (mfri)?" Rao believes, and the literature review supports her, that the mfri varies across the five biomes. However, as the null hypothesis must always contain the 'null', or 'equal', relation, the null hypothesis is

$$H_0 : \mu_{STR} = \mu_{TBF} = \mu_{TCF} = \mu_{TS} = \mu_{XER}$$

Test this null hypothesis. Again, make sure you test the assumptions appropriately and use the appropriate test. In your write-up, you still need to provide the necessary graphs and statistics to support your conclusion.