September 13, 2020

Assignment 1

CS 624, Fall 2020

Due on September 20, 2020

Use R to perform all necessary calculations. Attach your code and output. Give interpretation and discuss all relevant statistical measures.

Problem 1. We are interested in comparing the efficiencies of two teaching methods. Six students were taught using each method and their performance scores are shown in the following table.

Method 1	Method 2
79	71
66	43
57	58
91	78
42	20
59	56

a) Assume that the twelve students are different. Perform the appropriate analysis regarding the comparison of the efficiencies of two teaching methods. In particular, test for normality, state the null and alternative hypothesis, conduct the test at alpha level 0.01, derive a symmetric 99% confidence interval for the parameters of interest. Give interpretations for all relevant quantities.

b) Assume that the same six students were tested twice after being taught each method. Perform the appropriate analysis regarding the comparison of the efficiencies of two teaching methods. In particular, test for normality, state the null and alternative hypothesis, conduct the test at alpha level 0.01, derive a symmetric 99% confidence interval for the parameters of interest. Give interpretations for all relevant quantities.

c) Test the hypothesis that the average performance score for students taught using Method 1 is 80. Construct a symmetric 90% confidence interval for the population average score.

d) Compare the results in parts a) and b). Comment on which one you would choose if the data collection process is not a problem. What would be an even better approach to collect data in an attempt to answer the question of interest?

Problem 2. We analyze the dataset wcgs.dta (available on Blackboard). This is a Stata file format and you need to load foreign package to read it into R (library(foreign)). We are interested in comparing the average weights of patients across the four coronary heart disease diagnosis groups (coded as 0, 1, 2, 3). Perform all the necessary data analyses steps to answer the research question of interest. These steps should include,

a) Test for normality within each group (using Shapiro-Wilk's test).

b) Test for homogeneity of variances (using Bartlett test).

c) Test for equality of means using ANOVA if appropriate (followed by Bonferroni adjusted pairwise t-tests or Tukey's honest significance difference test if the global test rejects the null hypothesis of no difference across the four group means).

d) Test for equality of medians using Kruskal-Wallis if ANOVA fails (followed by Bonferroni adjusted Wilcoxon signed rank test if the global test rejects the null hypothesis of no difference across the four group medians).

Problem 3. Four groups of people were subjected to different diets. The weight losses for all people were recorded after the completion of the diet plans. We would like to compare the effectiveness of these diets. The file diets.txt contains these data and is available on the Blackboard website.

a) State the null and alternative hypotheses.

b) Perform the necessary preliminary tests for ANOVA.

c) ANOVA using only the mean() and tapply() functions in R.

d) Perform all Bonferonni adjusted pairwise comparisons.

e) Perform the Tukey's HSD test.

f) Use the aov() function and compere with the results from part a.

g) Use the pairwise.t.test() function in R and compare with the results from part d.

Problem 4. Calculate the exact p-value of the Wald-Walfowitz runs test to check for randomness using the sequence aabbaaaabbbaababbaaabbbbbbababbbabbb.