STAT 6337
Advanced Statistical Methods I (Fall 2024)
Project 2

This project is individual work. So do not consult with anybody in or out of class. You can ask me or TA questions if something is not clear.

Complete this page below and attach with your project. Your project will not be graded without it.

This project is entirely my work. I have not discussed about this project with anybody in or out of class. I understand and have complied with the academic integrity policies written in the *Handbook of Operating Procedures* of UT Dallas https://policy.utdallas.edu/utdsp5003**.**

YOUR NAME _____

DATE _____

YOUR SIGNATURE (NOT just typed name) _____

Stat 6337: Advanced Statistical Methods (Fall 2024)

Project# 2

Notes:

- You are supposed to work on this project entirely on your own. So, do not consult with anyone within or outside the class.
- You are welcome to ask me or TA questions. However, first try to find the answer on your own. Don't be afraid to google! It is a necessary skill for becoming a successful programmer.

1. Two independent random samples with sample sizes n_1 and n_2 are drawn from two populations. To test the equality of the two population means, suppose we want to use two-sided t-test assuming equality of variances at 0.05 significance level. Before doing that, we would like to ensure that the true type I error rate is indeed 5% using Monte Carlo (MC) simulations. For this task, generate both samples from same distribution, which is to be varied as: (1) Normal(0, 1), (2) Uniform(0,1), (3) t-distribution with 1 degree of freedom. For sample sizes, consider the following settings: (1) $n_1 = n_2 = n = 10$, (1) $n_1 = n_2 = n = 30$, (1) $n_1 = 10, n_2 = 30$. For each combination of distribution and sample sizes, generate 1,000 replicates and report the estimated type I error rate. Compare the nine type I error rates (resulting from 3 distributions and 3 sample sizes) with the nominal level of 5% and comment on the reason(s) for differences, if any.

2. Consider question 1(f) from the last project, specifically the goodness of fit test. Estimate p-value of that test using MC simulations with 1,000 replicates. Compare it with p-value reported in the last project for the same test and also with p-value for the Shapiro Wilks test and comment.

3. Consider the Framingham Heart Study from the last project.

   (a) Fit a simple linear regression model of SBP on BMI. Report the fitted model, confidence interval for slope, and an unbiased estimate of σ^2.

(b) Report interval estimates for the mean response and an individual future response for BMI values equal to its average and first quartile. Also compute the corresponding confidence intervals using Working-Hotelling method. Compare the six intervals in terms of their widths and explain the reasons for differences.

(c) Compare MSR and MSE and explain when do we expect them to be similar. What is a formal test for making this comparison? Report the corresponding test statistic, p-value, and conclusion.

(d) Report coefficient of determination and interpret it. Show the correspondence between the coefficient of determination, F test statistic, and a t test statistic using the appropriate numbers from the output.

(e) Report simultaneous confidence sets for the regression coefficients using the two methods discussed in class. Plot and compare them. Also, compare with the confidence interval for slope in part (a).

Submit your report with the following components (in this order):

- at most 3 pages of typed answers;
- relevant parts of SAS output with relevant numbers highlighted (label each part of the question);
- your SAS code including brief typed comments of main steps (each part labelled).