Statistical Methods for Data Science
Mini Project 3

Exercise 1 (10 points)

Consider the dataset stored in the file bp.txt. This dataset contains one measurement of systolic blood pressure (in mmHg) made by each of two methods—a finger method and an arm method—from the same 200 patients.

(a) Perform an exploratory analysis of the data by examining the distributions of the measurements from the two methods using boxplots. Comment on what you see. Do the two distributions seem similar? Justify your answer.

(b) Use histograms and QQ plots to examine the shapes of the two distributions. Comment on what you see. Does the assumption of normality seem reasonable? Justify your answer.

(c) Construct an appropriate 95% confidence interval for the difference in the means of the two methods. Interpret your results. Can we conclude that the two methods have identical means? What assumptions, if any, did you make to construct the interval? Do the assumptions seem to hold? Justify all your answers.

Exercise 2 (10 points)

We know how to construct a large sample confidence interval for a population proportion p. How large n should be for this interval to have acceptable accuracy? Answer this question by computing the coverage probability of this interval using Monte Carlo simulation, and examining how close the probability is to the nominal confidence level. Take level of confidence to be 95% but use a variety of values for n and p, e.g., n = 5, 10, 30, 50, 100, and p = 0.05, 0.1, 0.25, 0.5, 0.9, 0.95. Summarize your results graphically. Comment on any patterns you see in the results. Based on your findings, what n would you recommend for the use of this confidence interval? Would your answer depend on p? Explain.

<u>Instructions:</u>

- Due date: Thursday, March 2.
- Total points = 20.
- Submit a typed report.

- You can work on the project either individually or in a group of no more than two students. In case of the latter, submit only one report for the group, and include a description of the contribution of each member.
- Do a good job.
- You must use the following template for your report:

  Mini Project #
  Name
  Names of group members (if applicable)
  Contribution of each group member
  Section 1. Answers to the specific questions asked.
  Section 2: R code. Your code must be annotated. No points may be given if a brief look at the code does not tell us what it is doing.