

Statistical Methods for Data Science

Mini Project 3

Consider the trees data provided on eLearning in trees.csv file. The dataset contains measurements of three variables obtained from 31 felled black cherry trees. The variables are: girth (defined as diameter of the tree in inches measured at 4.5 feet above the ground); height (in feet); and volume of timber (in cubic feet). We use R to analyze these data. Follow the steps below.

1. For each variable,
 - make a histogram and a boxplot, calculate the 5-number summary, and comment on what you see regarding the data distribution;
 - report the appropriate measure of (center, spread) — (mean, SD) or (median, IQR) — and justify your choice; and
 - make a normal Q-Q plot, and comment on whether or not the normal distribution assumption for the data appears reasonable.
2. For each pair of variables, make a scatterplot, comment on whether or not the correlation is an appropriate measure of strength of relationship between the two variables, and compute the correlation if it is appropriate. [Note: You can use the R function 'pairs' to make a matrix of scatterplots.]
3. Perform a natural log transformation of the variables, and repeat steps 1 and 2 with the transformed data.
4. Should the data be analyzed on log scale or original scale? Justify your answer.

2 Bonus points will be given for good, neat work.

Instructions:

- Due date: Thursday, October 13.
- Total points = 10
- Submit a typed report and include all relevant plots.
- 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)

Contributions of group members

Answers and justifications for each exercise

Provide the R codes in an appendix. 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.