

Homework 8

INTRO TO DATA SCIENCE FALL 2023

DUE WEDNESDAY 11/13

NAME:

How to submit: You will submit your work as a .r file to the assignment which will be posted to Brightspace. For each block of code, you will `#comment` above the code which part of this assignment this is supposed to answer, and `#comment` below the code with the numerical answers (if applicable). For example:

- `#Part b)`
- `code`
- `# The mean is 14.3 minutes, etc.`

Problem 1 Consider the data set:

$-3, 5, 12, -40, 9, 23, 3, -23, 15, 15, -9.$

Write a program in R which uses a while loop to compute the average value of the positive numbers in this data set.

Problem 2 I spent 6 hours and 49 minutes collecting some data this weekend last year for you to analyze using R. I work very hard! I have attached in the email with this assignment a .csv file containing the length of time (in minutes) I spent running each mile from my trail marathon this past weekend. I want you to use R to analyze and visualize this data.

- Import the data to be used in R.
- Get the summary for the number of minutes for each mile, including mean, median, quartiles Q1 and Q3, maximum and minimum values. Then tell me what the mean, median, IQR, and the range for minutes is.
- Compute the sample standard deviation for the length of each mile.
- Plot a histogram for the length of miles using a 30 second binwidth, and make the histogram your favorite color.
- Find the correlation between the mile number and the length of the mile.
- Plot a scatter plot of X=mile vs Y=minutes, and include the regression line in the plot.
- Find the equation of the regression line using the mile as the predictor variable for the outcome variable minutes.
- Make a function in R using your answer from g.), and use it to predict the values of Y when X=6 miles, 15 miles, and 42 miles.