

Homework 2 of STAT 540

Section 001, Fall 2024

Due: Wednesday Sep 11, 2024 (before class)

Total Points: 75

Please hand in a hard copy of your homework (compiled pdf file from R markdown) in class and email your R code to Kaniz Fatema (KFATEMA@email.sc.edu). Please use the R markdown Homework template (HWtemplate.Rmd) to write your homework solutions. Work on the homework independently.

Problem 1. Use R to solve the following questions and present your results clearly. (5 points each)

- a. Read data into R from the text file dataP1HW2.txt.
- b. Calculate the sample mean and variance of X and Y separately.
- c. Calculate Pearson's correlation coefficient between X and Y .
- d. Explore the function "lm" in R by checking "help(lm)" and then use it to fit a simple linear regression with Y being the dependent variable and X the independent variable. Report the fitted linear regression line.
- e. Report the estimated regression coefficients in the form of a vector.
- f. Create a character variable readme for "This is a simple linear regression problem."
- g. Create a list object (with name listP1HW2) that contains the data containing X and Y in the form of a data frame with 2 columns, the estimated regression coefficients in the form of a vector, the correlation coefficients, and readme. Assign the following names c("data", "coef", "corr", "readme") to the objects in listP1HW2. Display listP1HW2 and check out the command listP1HW2\$Data\$X.

Problem 2. What is the sum of the first 100 positive integers? The formula for the sum of integers 1 through n is $n(n+1)/2$. Define $n=100$ and then perform the sum of

1 through 100 using the formula and a different approach in R. What is the sum?
(5 points)

Problem 3. Load the US murders dataset in the `dslabs` package. (5 points each)

Start by loading `library(dslabs)`

- a. Use the function `str` to examine the structure of the `murders` object.
- b. Use the accessor `$` to extract the state abbreviations and assign them to the object `a`. What is the class of this object?
- c. Now use the square brackets to extract the state abbreviations and assign them to the object `b`. Use the `identical` function to determine if `a` and `b` are the same.
- d. Use the `match` function to identify the states with abbreviations AK, MI, and IA. [Hint: start by defining an index of the entries of `murders$abb` that match the three abbreviations, then use the `[` operator to extract the states.]
- e. Compute the per 100,000 murder rate for each state and store it in an object called `murder_rate`. Then use logical operators to create a logical vector named `low` that tells us which entries of `murder_rate` are lower than 1. Use `R` to report the names of the states with murder rates lower than 1.
- f. Create a subset of the `murders` data (named `south`) to contain only the states in the south region and order by population size.
- g. 41. Use the `%in%` operator to create a logical vector that answers the question: which of the following are actual abbreviations: MA, ME, MI, MO, MU?