Stat 123 Midterm 2 (Practice Test)

Monday, March 20, 2023

Duration: 4:30 pm to 5:20 pm + 10 minutes to upload submissions due by 5:20 pm.

- The test duration is 40 minutes+10 minutes for uploading. If you have a CAL time accommodation, you may add it to your 40-minute time limit.
- You can access all the required materials in the "Midterm 2" dropbox on Brightspace. Please follow the path: Brightspace-> Course Tools-> Assignments -> Midterm 2. [Please note that you will only be able to access this Dropbox folder 2 minutes before the start of the test on March 20.]
- Download the "Midterm2-Questions" file in PDF format and the "Houses.CSV" dataset.
- The lecture notes can be found in the "Midterm 2" dropbox.

0. Open a new R Markdown file.

Note: Your worksheet should be submitted as an R Markdown file (You MUST knit it to PDF on your computer, or you can knit it to Word and then convert to a PDF). You must have uploaded Your solutions in PDF format to the Brightspace drop box named “Midterm2” by no later than 5:20 pm unless you have a CAL time accommodation.

1. Load sample_data.csv into R and save it as df.

(a) Familiarize yourself with the dataframe. Notice how it has one column where the entries are characters and the rest of the columns are numeric values.

We want to perform mathematical operations on this dataframe in a variety of ways. We could, with a little fancy footwork, perform these operations on the dataframe itself, but to make our lives a little easier today we’re first going to create a matrix that contains only the numerical values. Create this matrix and name it nums.

Hint: We’ve done this exact thing in a previous lab.

(b) First, we are only going to deal with the 3rd column of nums. Save this column as its own vector and name it c3.

(c) Write a for-loop that adds up each element of c3 and save it to the name sc3. Print out sc3. Hint: You should set sc3 equal to zero before you write the for-loop.

2. Save columns 2 and 4 of nums as their own vectors and name them c2 and c4, respectively.

(a) Write a for-loop that adds each element of c2, c3, c4 together and save it to a vector named row sums. Note: row sums should be a vector where the first element is the sum of the first elements of c2, c3, and c4. Hint: Set up the row sums vector before you write the for-loop. You may use: row sums = rep(0, length(c2))

(b) Print out the row sums vector.

3. Use the built-in data set USArrests to answer this question.

(a) What is the variable Murder being measured in the data set?

(b) What type of variable is this?

(c) What is the most appropriate type of graph to visualize the distribution of this variable?

(d) Graph the distribution of the variable (using the type of graph that you identified in part (c)). Your graph should include:

- a main title.

- x-axis title.

- scales on the x and y-axis which fully extend from at least the min value to at least the max value.

4. Download and consider the gapminder data set. We will need this data set to answer this question.

(a) Either load the data set into R by typing in library(gapminder) or download the gapminder.csv file from Brightspace and read the data into R, saving it as gapminder.

(b) Suppose you are looking to explore the relationship between the population and Life Expectancy. What type of graph should you use to visualize this relationship?

(c) Create a graph which visualizes the relationship between these two variables. Put Life Expectancy is on the x-axis. This graph does not need any titles.