

Lab 0 Math Review and Intro to R

Submission Instructions:

Write all the solutions in a single Word document. Submit the Word document and R files.

Prepare your laptop: Follow the links in the slides to download and Install R and R Studio.

1. The following sample data shows the scores of 10 students in an exam:
45, 80, 83, 78, 75, 77, 79, 83, 83, 100

Part A. Show detailed steps for each of the following questions a) – c). Computation must be done BY HAND.

Part B. Compute a) – c) using R. For example: `mean(data)` is used to compute the mean for data vector.

- a) Compute the mean and median.
 - b) Compute the variance and the standard deviation.
 - c) Compute Q1, Q3 and the interquartile range.
2. Suppose we have the following summary of data of 5,000 patients in a health clinic. Answer questions a) to d).

Race	Diabetic	Non Diabetic	Total
American Indian	160	840	1,000
Non Hispanic Blacks	130	870	1,000
Hispanics	130	870	1,000
Asian Americans	90	910	1,000
Non Hispanic Whites	80	920	1,000
Total	590	4,410	5,000

- a) What is the probability that an individual has diabetes?
- b) What is the probability that a randomly selected patient is an American Indian who is not a diabetic?
- c) What is the probability that a randomly selected patient is an American Indian or a non diabetic?
- d) What is the probability of being diabetic among Hispanics?

3. Using the sample data as problem 1. Do the following:
- Using comparison operators, write the expression for exam scores less than the mean.
Sample output:

```
[1] TRUE FALSE FALSE TRUE TRUE TRUE FALSE FALSE FALSE  
FALSE
```
 - Output all the exam scores less than the mean.
Sample output:

```
[1] 45 78 75 77
```
 - Using seq to output every other element from the scores.
Sample output:

```
[1] 45 83 75 79 83
```
 - Create a matrix of size 2 x 5 using the exam scores data. The first five values belong to the first row of the matrix. Name the matrix, *scores.matrix*, and display the results.
Sample output:

```
[,1] [,2] [,3] [,4] [,5]  
[1,] 45 80 83 78 75  
[2,] 77 79 83 83 100
```
 - Without hardcoding the value 5 (the values 2, 3, 4, 5, etc. should not be in your code for this question; your code should work regardless of the number of columns in your matrix), show the code for displaying the first and last columns of the matrix.
Sample output:

```
[,1] [,2]  
[1,] 45 75  
[2,] 77 100
```
 - Assign row names for the *scores.matrix* as Student_1, Student_2,... and column names as Quiz_1, Quiz_2 The code should not hard code the values 2 and 5 (i.e. the values 2, 3, 4, 5, etc. should not be in your code for this question, your code should work regardless of the number of rows and columns in your matrix).
Sample output:

```
Quiz_1 Quiz_2 Quiz_3 Quiz_4 Quiz_5  
Student_1 45 80 83 78 75  
Student_2 77 79 83 83 100
```

4. Use the data frame `colleges.info` below and answer questions a) to c):

```
Name = c("Pomana", "Williams", "Stanford", "Princeton", "Yale")
State = c("CA", "MA", "CA", "NJ", "CT")
Cost = c(62632, 64020, 62801, 58965, 63970)
Population = c(1610, 2150, 18346, 8014, 12109)
colleges.info = data.frame(Name, State, Cost, Population)
```

- a) Show the summary for State and Cost.
- b) Show all rows of the sliced data frame whose Population is greater than 5000.
- c) Generate a new data frame called *colleges.info.2016* that includes the Name, State and Population columns from *colleges.info*; however, modify the Cost column to reflect a 5% increase in tuition costs. Round off the tuition costs to the nearest dollar. Display the new resulting data frame.