

CS544 Module 2 Assignment

General Rules for Homework Assignments

- You are strongly encouraged to add comments for the code portions.
- You must work on your assignments individually. You are **not allowed** to copy the answers from the others.
- Each assignment has a strict deadline. However, you are still allowed to submit your assignment within 2 days after the deadline with a penalty. 15% of the credit will be deducted unless you made previous arrangements with your facilitator and professor. Assignments submitted 2 days after the deadline will not be graded.
- When the term *lastName* is referenced in an assignment, please replace it with your last name.

Part1) Probability - 20 points

a) From the Bayes' rule example given in Section 3.10, compute the probabilities that a randomly selected non-smoker i) has lung disease and ii) does not have lung disease. Show the calculations without using R. Then, verify with the *bayes* function provided in the code samples.

b) Suppose that in a particular state, among the registered voters, 40% are democrats, 50 % are republicans, and the rest are independents. Suppose that a ballot question is whether to impose sales tax on internet purchases or not. Suppose that 70% of democrats, 40% of republicans, and 20% of independents favor the sales tax. If a person is chosen at random that favors the sales tax, what is the probability that the person is i) a democrat? ii) a republican, iii) an independent. Show the solutions with the calculations without using R. Then, verify with the *bayes* function provided in the code samples.

Part2) Random Variables - 30 points

a) Consider the experiment of rolling a pair of dice. Using R, show how would you define a random variable for the absolute value of the difference of the two rolls, using a user-defined function.

b) Using the above result, what is the probability that the two rolls differ by exactly 2? What is the probability that the two rolls differ by at most 2? What is the probability that the two rolls differ by at least 3? Use the **Prob** function as shown in the code samples.

c) Show the marginal distribution of the above random variable (using R).

d) Using R, add another random variable to the above probability space using a user defined function. The random variable is TRUE if the sum of the two rolls is even, and FALSE otherwise. What is the probability that the sum of the two rolls is even? Show also the marginal distribution for this random variable.

Part3) Functions - 20 points

Using a for loop, write your own R function, **evensum**(*data*), that returns the sum of all the even values in the given numeric *data* vector.

Now, without using any loop, write your own R function, **evensum2**(*data*), that returns the sum of all the even values in the given numeric *data* vector.
Test both functions with sample data.

Sample output:

```
> evensum(c(15, 20, 25, 30, 35))
[1] 50
> evensum(1:10)
[1] 30
> evensum(seq(1,15, by = 2))
[1] 0
> evensum2(c(15, 20, 25, 30, 35))
[1] 50
> evensum2(1:10)
[1] 30
> evensum2(seq(1,15, by = 2))
[1] 0
```

Part4) R - 30 points

Initialize the Dow Jones Industrials daily closing data as shown below:

```
dow <- read.csv('http://kalathur.com/dow.csv', stringsAsFactors = FALSE)
```

Provide the simplest R code and output for all of the following. **The code should work for any given data.**

- Use the diff function to calculate the differences between consecutive values. Insert the value 0 at the beginning of these differences. Add this result as the DIFFS column of the data frame.
- How many days did the Dow close higher than its previous day value? How many days did the Dow close lower than its previous day value?
- Show the subset of the data where there was a gain of at least 400 points from its previous day value.

	DATE	VALUE	DIFFS
25	2/6/18	24913	567
29	2/12/18	24601	410
47	3/9/18	25336	441

- Provide the solution to compute the longest gaining streak of at least 100 points in the data. Show the data for that longest gaining streak. Hint: Use the **rle** function provided by R.

	DATE	VALUE	DIFFS	STREAK
36	2/22/18	24962	164	Y
37	2/23/18	25310	348	Y
38	2/26/18	25709	399	Y

Submission:

Create a folder, CS544_HW2_lastName and place the following files in this folder.

Provide the text and code part of the solutions and the corresponding output in a single Word document, **HW2_lastName.doc**.

For the code portions, provide the R file, **HW2_lastName.R**, with each portion of the code identified by comments.

Archive the folder (CS544_HW2_lastName.zip). Upload the zip file to the Assignments section of Blackboard.