

## CS544 Module 2 Assignment

© 2023, Suresh Kalathur, Boston University. All Rights Reserved.

The following document should not be disseminated outside the purview of its intended purpose.

### Part1) Probability - 25 points

Use the Bayes theorem to calculate the following probabilities. Show the **individual steps** of the Bayes theorem. You can use R for the calculations. Use the Word/PDF document for this part.

Suppose that in a particular state, among 10000 people surveyed, 4250 people are in the age group *18-34* years, 2850 people are in the age group *35-49* years, 1640 people are in the age group *50-64* years, and the remaining are *65 years & over*.

Out of those in the age group *18-34* years, 1062 people had a BMI of above 30. Of those in the age group *35-49* years, 1710 people had a BMI of above 30. Among those in the *50-64* years range, 656 people had a BMI of above 30. In the last age group, 189 people had a BMI of above 30.

- a) What is the probability that a randomly selected person in this survey will have a BMI of above 30?
- b) If a randomly selected person had a BMI of above 30, what is the probability of that person being in the age group *18-34* years?
- c) If a randomly selected person had a BMI of above 30, what is the probability of that person being in the age group *35-49* years?
- d) If a randomly selected person had a BMI of above 30, what is the probability of that person being in the age group *50-64* years?
- e) If a randomly selected person had a BMI of above 30, what is the probability of that person being in the *65 years & over*?

## Part2) Random Variables - 25 points

Consider a game which involves rolling **three** dice. Write the **R code** for the following.

Using the **rollDie** function from the **prob** library, setup the sample space for this experiment with the probability space.

For each of the following scenarios from **a)** through **e)**, show the corresponding **outcomes** and the **probability** of that event. The sample outputs for **b)** are shown as example.

**a)** The sum of the rolls is greater than 3 but less than 8.

**b)** All the three rolls are identical.

Sample Output for outcomes:

	X1	X2	X3	probs
1	1	1	1	0.00463
44	2	2	2	0.00463
87	3	3	3	0.00463
130	4	4	4	0.00463
173	5	5	5	0.00463
216	6	6	6	0.00463

Sample Output for probability:

```
[1] 0.02778
```

**c)** Only two of the three rolls are identical.

**d)** None of the three rolls are identical.

**e)** Only two of the three rolls are identical given that the sum of the rolls is greater than 3 but less than 8.

### Part3) Functions - 20 points

Using a **for** loop or a **while** loop, write your own **R function**,  
**sum\_of\_first\_N\_even\_squares** (*n*),  
that returns the sum of the squares of the first **n** positive even numbers.

For example, if  $n = 5$ , the first five even numbers are 2, 4, 6, 8, 10 and the required result is  $2^2 + 4^2 + 6^2 + 8^2 + 10^2 = 220$ .

Test your function as follows:

---

```
> sum_of_first_N_even_squares(2)
[1] 20
> sum_of_first_N_even_squares(5)
[1] 220
> sum_of_first_N_even_squares(10)
[1] 1540
```

#### Part4) R - 30 points

Initialize the Tesla stock data for the year 2022 using the read.csv function as shown below:

```
tsla <- read.csv("https://people.bu.edu/kalathur/datasets/TSLA2022.csv")
```

The first 6 rows of the dataset are as shown below:

```
> head(tsla)
      Date Open High Low Close  Volume
1  1/3/22  383  400 379   400 103931400
2  1/4/22  397  403 374   383 100248300
3  1/5/22  382  390 360   363  80119800
4  1/6/22  359  363 340   355  90336600
5  1/7/22  360  360 337   342  84164700
6 1/10/22  333  353 327   353  91815000
```

Provide the simplest R code and corresponding outputs for the following.

**The code should work for any given dataset.**

**a)** Store the result of the **summary** function for the *Close* attribute as the variable *sm*. Change the *names* of this variable so that the output appears as shown below.

```
> sm
      Min    Q1    Q2  Mean    Q3    Max
109.0 225.0 272.0 263.1 302.5 400.0
```

**b)** Produce the output for the minimum price of the Tesla *closing* value in the dataset as shown below:

```
[1] "The minimum Tesla value of 109 is at row 248 on 12/27/22"
```

**c)** Produce the output for the maximum price of the Tesla *closing* value in the dataset as shown below:

```
[1] "The maximum Tesla value of 400 is at row 1 on 1/3/22"
```

**d)** Based on the dataset, what is the probability that on any given day, the Tesla *closing* price would be greater than the *opening* price?

**e)** Based on the dataset, what is the probability that on any given day, the Tesla trading *volume* would be greater than 100 million shares?

**f)** Based on the dataset, what is the probability that on any given day, the Tesla *closing* price would be higher than the *opening* price, given that the trading *volume* is greater than 100 million shares?

**g)** Suppose you bought 1 share of Tesla stock on each day in the dataset at its *low* price on that day. How much gain/loss you would have if you sold all those shares on the last day in the dataset using that day's closing price.

## Submission:

When the term *lastName* is referenced, please replace it with your last name.

Provide all R code in a single file, **CS544\_HW2\_LastName.R**. Clearly mark each subpart of each question.

Provide the corresponding outputs from the R console in a single PDF document, **CS544\_HW2\_LastName.pdf**

Upload the two files to the Assignments section of Blackboard.

**Note: Only ONE submission is allowed. Please be sure that what you are submitting is your final submission.**