

MET CS 555 A3 SPRG23 Assignment 1 (20 points)

SUBMISSION REQUIREMENTS: Please submit a single document (word or PDF) for submission. Your submission should contain a summary of your results (and answers to questions asked on the homework) as well as your R code used to generate your results (**please append your R code to the end of your submission**).

Homework submission filenames should take on the form **LASTNAME - Homework #.doc(x)/.pdf**.

No ZIP files please.

The data in the table below give the duration in days of hospital stays of patients admitted to the hospital with *C. Difficile*. Use the data on the following page to:

- (1) Save the data to a CSV file and read into R for analysis. **(2 points)**
- (2) Make a histogram of the duration of days of hospital stays. Ensure the histogram is labelled appropriately. Use a width of **1 day**. Describe the shape, center, and spread of the data. Are there any outliers? **(5 points)**
- (3) Find the mean, median, standard deviation, first and third quartiles, minimum and maximum of the durations of hospital stay in the sample. **Summarize these values in a table that you** create in EXCEL or WORD. In other words, do **not** simply copy and paste R output. You should be reporting a nicely labeled and formatted table.
 - Given the shape of the distribution, what is the best single number summary of the center of the distribution?
 - What is the best single number summary of the spread of the distribution? **(5 points)**
- (4) Assume that the literature on this topic suggests that the distribution of days of hospital stay are normally distributed with **a mean of 5** and **a standard deviation of 3**. Use R to determine the probabilities below based on the normal distribution described above (you should **not** be using the data set given on the following page):
 - (a) What percentage of patients are in the hospital for less than **10 days**? **(4 points)**
 - (b) Recent publications have indicated that hypervirulent strains of *C. Difficile* are on the rise. Such strains are associated with poor outcomes, including extended hospital stays. An investigator is interested in showing that the average hospital stay duration have increased versus published literature. He has a sample of **35 patients** from his hospital. If the published data are consistent with the truth, what is the probability that the sample mean in his sample will be greater than **6 days**? **(4 points)**

Data Set for Assignment 1

7	3	5	3	9	5	10	3	4	4
7	5	8	3	4	9	15	4	5	8
5	3	2	3	5	9	4	5	6	9
5	3	6	3	2	6	4	5	5	4
5	8	4	6	13	4	6	3	2	3
2	4	6	6	6	8	6	3	4	4
5	10	4	6	3	9	3	9	4	7
10	14	4	6	5	10	4	4	9	4
4	3	6	8	5	7	6	9	3	12
11	5	2	9	4	4	5	6	4	2

Extra Credit Question for Assignment 1 (4 point)

The amount that households pay service providers for access to the Internet varies quite a bit, but the mean monthly fee is \$38 and the standard deviation is \$10. The distribution is not Normal: many households pay a base rate for low-speed access, but some pay much more for faster connections. A sample survey asks an SRS of 500 households with Internet access how much they pay. Let \bar{x} be the mean amount paid.

- (a) Explain why you can't determine the probability that the amount a randomly selected household pays for access to the Internet exceeds \$39.
- (b) What are the mean and standard deviation of the sampling distribution of \bar{x} ?
- (c) What is the shape of the sampling distribution of \bar{x} ? Justify your answer.
- (d) Find the probability that the average fee paid by the sample of households exceeds \$39. Show your work.