MET CS 555 Assignment 1 – 20 points

Spring, 2018

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 to the end of your submission).

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

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 excel or 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? (7 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. 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? (6 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:

(a) What percentage of patients are in the hospital for less than a week? (2 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 durations have increased versus published literature. He has a sample of 10 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 7 days? (3 points)

7	3	5	3	1	5	10	3	4	4
7	5	8	3	4	1	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	14	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	13	4	6	5	10	4	4	9	4
4	3	6	8	5	7	6	1	3	12
15	5	2	1	4	4	5	6	4	12