Written Assignment Unit 4

Data:

1. Using the list of 17 numbers at the top of the page, the median of this data, rounded to two decimal places, is:\_\_3.92\_\_\_.

*\*I calculated this by adding all the numbers up and then dividing by 17.*

2.    If you find the median using the original method (paper and pencil), you have to arrange the values into numeric order (True/False).\_\_\_\_\_\_\_\_\_False\_\_\_\_\_\_\_\_\_\_\_\_\_

*\*with addition the order does not matter.*

3.    (The calculations MUST be done manually, do not use R) The interquartile rang for this data is (round each value to 3 decimal places):\_\_\_6\_\_\_\_.

*\*calculations below by pen and paper.*

*Medians are calculated by adding up and dividing by total number of values added.*

*Median(middle) = 3.92*

*Q1 is median between -1.9 and 3.9 = 0.500*

*Q3 is median between 3.9 and 18 = 6.500*

*IQR is 6.500 – 0.500 = 6*

*-1.9 -1.5 -1.1 -0.1 0.5 1.2 1.8 2.1 2.8 3.2 3.9 4.1 6.5 7.0 7.2 13.0 18*

4.    The formula for calculating the interquartile range is\_\_\_\_\_Q3 minus Q1\_\_\_\_\_\_\_\_ (show the formula and a citation to the source that you used).

Reference: *Interquartile Range: Definition*. (2021). Stattrek. <https://stattrek.com/statistics/dictionary.aspx?definition=interquartile%20range>

5. (The calculations MUST be done manually, do not use R) Using techniques that we studied in this course, the upper and the lower cutoff points (rounded to three decimal places) for identifying outliers in the given data sample are: \_\_\_-8.500\_\_\_ and \_\_\_15.500\_\_\_ (this is not a request to show any outliers—just the cutoff points that would determine what constitutes an outlier.) You may round to three decimal places.

*IQR 6 \* 1.5 = 9*

*Q1 0.500 – 9 = -8.500*

*Q3 6.500 + 9 = 15.500*

6.    The summary() command shows a list of outliers, if there are any (True/False):\_\_\_\_\_\_\_\_\_False\_\_\_\_\_\_\_\_\_\_\_\_\_

7.    The list of outlier values is:\_\_\_\_\_\_\_\_18\_\_\_\_\_ (if there are none, write "NA").

8.    The standard deviation of the list of 17 numbers is (round to 3 decimal places): \_\_\_\_\_\_\_3.198\_\_\_\_\_\_\_

*\*calculations below:*

*Mean = -1.9 + -1.5 + -1.1 + -0.1 + 0.5 + 1.2 + 1.8 + 2.1 + 2.8 + 3.2 + 3.9 + 4.1 + 6.5 + 7.0 + 7.2 + 13.0 + 18 = 66.7/17 = 3.924*

*((-1.9 – 3.924)² + (-1.5 – 3.924)² + (-1.1 – 3.924)² + (-0.1 – 3.924)² + (0.5 – 3.924)² + (1.2 – 3.924)² + (1.8 – 3.924)² + (2.1 – 3.924)² + (2.8 – 3.924)² + (3.2 – 3.924)² + (3.9 – 3.924)² + (4.1 – 3.924)² + (6.5 – 3.924)² + (7.0 – 3.924)² + (7.2 – 3.924)² + (13.0 – 3.924)² + (18 – 3.924)²)…*

*((-5.824)² + (-5.424)² + (-5.024)² + (-4.024)² + (-3.424)² + (-2.724)² + (-2.124)² + (-1.824)² + (-1.124)² + (-0.724)² + (-0.024)² + (0.176)² + (2.576)² + (3.076)² + (3.276)² + (9.076)² + (14.076)²)…*

*((-33.918976) + (-29.419776) + (-25.240576) + (-16.192576) + (-11.723776) + (-7.420176) + (-4.511376) + (-3.326976) + (-1.263376) + (-0.524176) + (-0.000576) + (0.030976) + (6.635776) + (9.461776) + (10.732176) + (82.373776) + (198.133776)) = 173.82592/17 = √10.22505411764705 = 3.198*

A Random Variable:

9.    The missing probability value (under the number 4) in the random variable table above is:\_\_\_0.15\_\_\_\_

*\*Just subtract all the probability values from 1*

10.     The sum of the probabilites in the second row of any random variable table like the one above should equal (round to 3 decimal places): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_1.000\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

11.    Read section 4.4.1 in the book (Yakir, 2011).  Do the numbers in the table above (for the random variable) represent a data sample (Yes/No)?\_\_Yes\_\_

12.     In the random variable table shown above, the value in the second row represents the cumulative probability of the corresponding values in the first row (True/False) \_\_\_\_Yes\_\_\_\_\_

13.     The probability that a randomly selected value from this random value will be less than or equal to 3 is :\_\_\_0.6 or a 60% chance\_\_.

14.     What is the probability that a randomly selected value from the random variable would be exactly 1.5? \_\_\_\_\_0.2 or 20% chance\_\_\_\_\_\_ .

15.     Review section 4.4 in the book (Yakir, 2011), especially pages 57—58.  The expectation of the random variable is:\_\_\_15.857\_\_\_.

16.     To find the expectation of a random variable by using a relative frequency table, you can add the values in the first row of the table and divide by the number of columns in the table (True/False)\_\_\_\_True\_\_\_\_\_.

17.     Study Yakir (2011) pp. 57-59 and solved problems 4.1.6-4.1.8.  The (population) standard deviation of the random variable above is (round to 3 decimal places):\_\_\_??which random variable??\_\_\_\_ (hint, you cannot put values from the table into the sd() function because the sd() function does not adjust for the probabilities).

18.    If you have already calculated the standard deviation of a data sample, what is the next thing to do to find the variance: \_\_\_\_\_\_\_\_\_\_\_\_\_square the deviation\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

A Population:

19.    Determine how many observations in the pop3.csv file are of type a:  \_\_49949\_\_\_\_\_.

20. Using the appropriate R function with the default options, what is the median of the time column of pop3 (round to 3 decimal places): \_\_\_\_\_\_\_\_\_\_\_\_\_\_4.473\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

21. What is the variance of the time column of pop3 (rounded to three decimal places)? \_\_\_54.916\_\_\_\_