**Quizzes and graded assignments in MATH1280 are NOT group projects. Please answer the questions and do not share the questions or answers with others or accept advance access to questions or answers from others.**

**Data**  
For the tasks 1-8 (do not use R in tasks 3 and 5), consider the following data:

7.2, 1.2, 1.8, 2.8, 18, -1.9, -0.1, -1.5, 13.0, 3.2, -1.1, 7.0, 0.5, 3.9, 2.1, 4.1, 6.5  
  
In Tasks 1-8 you are asked to conduct some computations regarding this data. The computation should be carried out manually **with the exception of #3 and #5**. All the steps that go into the computation should be presented and explained. (You may use R in order to verify your computation, but not as a substitute for conducting the manual computations.)

**A Random Variable**  
In Tasks 9-18 you are asked to conduct some computations regarding a random variable.  Use the (incomplete) table below as the definition of this random variable (after you fill in the blank).  The sample space of a random variable is comprised of the integers 0, 1, 2, 3, 4, 5, and 6. The probabilities of each value are shown in the table below (with one missing value).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Value | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Probability | .10 | .15 | .25 | .10 |  | .10 | .15 |

**A Population**  
For Tasks 19-21, use the file called "pop3.csv" found [**here**](https://my.uopeople.edu/pluginfile.php/1524550/mod_book/chapter/311173/pop3.csv). That file contains information about time to failure of an entire production of some computer parts. The file contains two variables, "type" and "time", each measured over the 100,000 members of the population. The variable "type" is a factor, with three levels, "a", "b" and "c", and the variable "time" is numeric.  If the value of time is 4, that means that the part lasted 4 units of time (years?) before a failure occurred.  You should treat the content of this file as the information from an entire population.

Save the file on your computer and read the data stored in the file into R.  If you have trouble loading the data into R, email your instructor immediately—don't worry if you think you will find the answer 1 minute after sending the email—just send the email.  Tasks 19-21 refer to the information in the file.

**Submitting the Assignment**  
For the assignment, you should complete the following tasks. Tasks 1-8 refer to the sequence of 17 data values presented above, Tasks 9-18 refer to the random variable and Tasks 19-XXX21 refer to the information of a population of computer parts that is stored in the file "pop3.csv". Your answers should be short and clear.  
We recommend that you copy and paste the tasks below into the assignment submission area. You can then write your answers to the tasks in the designated positions that are marked in the text:

**Tasks**

**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:\_\_\_\_\_.

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

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):\_\_\_\_\_\_\_.

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

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: \_\_\_\_\_\_ and \_\_\_\_\_\_ (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.

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

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

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

**A Random Variable:**

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

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): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

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)?\_\_\_\_

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) \_\_\_\_\_\_\_\_\_

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

14.     What is the probability that a randomly selected value from the random variable would be exactly 1.5? \_\_\_\_\_\_\_\_\_\_\_ .

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

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)\_\_\_\_\_\_\_\_\_.

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):\_\_\_\_\_\_\_ (hint, you can not 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: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

**A Population:**

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

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

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