**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.**

The file "[**flowers.csv**](https://my.uopeople.edu/pluginfile.php/1524540/mod_workshop/instructauthors/flowers.csv?time=1574619130240)" file contains information on measurements of the iris flowers. Create an R data frame by the name "flower.data" that contains the data in the file.

The following R code shows an example of how to round a vector of numbers to zero decimal places and then calculate some statistics using the rounded numbers.  You might need some of the calculations for this assignment, but you might not need others. You would replace example$years with the name of the R object that you want to analyze (in other programming languages, you might call example$years a *variable*).    
  
> x <- round(example$years, 0)  
> freq <- table(x)  
> rel.freq <- freq/sum(freq)  
> cumsum(rel.freq)

**Cumulative Frequency Table for Petal Length**

Use the following table to answer tasks 2-4.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Value: | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Cumulative Relative Frequency: | .16 | .33 | .35 | .58 | .81 | .97 | 1.00 |

**Tasks**

1. Sometimes it is difficult to understand data if you do not know what the numbers represent. Provide short definitions of two words: *sepal*, and *petal* (be sure to cite your sources even if you paraphrase):

sepal:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

petal:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. There is a cumulative relative frequency table printed above for petal lengths (using rounded values for petal length).  Below the number 3 in that table is the number .35.  What does .35 represent? (multiple choice)

a. Three of the flowers had petal length of 0.35.  
b. There were 0.35 observations that had petal length of 3 (after rounding the petal lengths).  
c. Of all the flowers measured in this sample 35% had a petal length of 3 (after rounding the petal lengths).  
d. Of all the flowers measured in this sample 35% had a petal length of 3 or less (after rounding the petal lengths).  
e. A study of all flowers on the planet would show that about 35% had petal lengths of 3 or less (after rounding the petal lengths).

3. Using only the cumulative relative frequency table printed above combined with some simple paper-and-pencil calculations, which petal length occurs most frequently ?

4. Describe how you determined your answer to the previous question (describe the calculations that you used). Do not show R code for this task--it will not be counted as an answer.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

--------------------------------------------------------------------------------------------------------

5. Assuming that you read the flowers.csv file into an R object called flower.data, run the following R code (do not paste the ">” character into R) and paste both the command and the output into your answer (you should see five names, each of which should be enclosed in quotes--if you do not see this, try again or contact your instructor):

> names(flower.data)

6. The number of observations in the "flower.data" data frame is: \_\_\_\_.

7. List the variables in the data frame (you can do this by entering the name of the R object that holds that data that you read using the read.csv command--you should have called it flower.data).  If you do not see five columns of data, then there was a problem reading the input file--try again or contact your instructor.  For each variable identify the type of the variable (factor or numeric).  
    The name and type of the 1st variable:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
    The name and type of the 2nd variable:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
    The name and type of the 3nd variable:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
    The name and type of the 4nd variable:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
    The name and type of the 5nd variable: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

8. Round the data for the variable Sepal.Length so that it contains integers, then find the frequency of the value 7 (not the relative frequency): \_\_\_\_.

-----------------------------------------------------------------------------------------------------------

Assuming that you read the flowers.csv file into an R object called flower.data, run the following R code (do not paste the ">” character into R).  Note that we are not rounding the numbers here. Use the output for the next five tasks:

> table(flower.data$Sepal.Width)  
> plot(table(flower.data$Sepal.Width))

9. What is the sum of the first three frequencies in the frequency table for sepal width? \_\_\_\_\_

10. What does your answer to the previous question represent (in terms of sepal width and frequency and the percentage of all sepal measurements) \_\_\_\_

11. What is the sum of the last three frequencies in the frequency table for sepal width? \_\_\_\_\_

12. How many flowers in the sample had sepal widths less than 4 (do NOT round the sepal width numbers for this, but you can round your final answer to 3 decimal places)? \_\_\_\_\_\_\_\_\_

13. What does the tallest bar in the plot represent?  (multiple choice)

a. mean  
b. mode  
c. median

----------------------------------------------------------------------------------------------------------

14. Create a frequency table that shows the frequencies for each species of flower in the sample.  Paste your R command and output into your answer (do NOT display data from a data frame, display data using the table() command)\_\_\_\_\_\_\_\_\_

15. Explain two things about the table that you created for the previous task:

Why did the frequency table for flower species contain words in the first row as opposed to numbers?\_\_\_\_\_\_  
What is the meaning of the numbers in the second row of the table? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_