Name: _____

Due by 9:15 AM on Thursday (February 8, 2024)

YOU MUST WORK ON THE EXAM BY YOURSELF.
If you receive any help from anyone, that's cheating and, if caught, you will fail the
course. If you need clarification on any of the questions, please contact me via email.

1. (*5 pts*) **Fill in the blanks:**

   a) A subset of a population is called a _____.

   b) A numerical descriptive measure of a population is called a _____.

   c) The _____ is used as a measure of center in a population.

   d) A _____ skewed population distribution would have $\mu < \tilde{\mu}$.

   e) The systolic blood pressures of adults in a certain age group are normally distributed with a mean of 128 and a standard deviation of 12. <u>By the Empirical Rule (also known as the 68-95-99.7% Rule)</u>, _____% of adults have systolic blood pressures higher than 152.

Use R for the remaining questions (Questions 2 – 7)

Notes: ~~~~~~
- If you need help on an R function/command, type *?functionname* or *?commandname* and help for this function/command will appear in the Help window.

- Create a new folder on your desktop and name it **Exam1_*your name***. Set this as your working directory in RStudio.

- In RStudio, open a blank source file (R Script) to work in, and make sure all History entries are cleared before you start your work on the following questions.

- In the R Script, add <u>*Exam 1 (by **your name***)</u> as a comment line. In your script file, clearly mark your solution to each question and part using comments.
~~~~~~

2. (*21 pts.*) What is <u>the most efficient</u> R command used to

a) Write and run the R command that will generate the following vector when printed:

```
[1] "2"      "Dog"    "15"     "Horse" "TRUE"
```

b) Write and run the R command that will generate the following sequence:

```
[1] 5.0  5.5  6.0  6.5  7.0  7.5  8.0  8.5  9.0  9.5  10.0
```

c) Write and run the R command that will generate the following sequence:

```
[1] -12 -10  -8  -6  -4  -2   0   2   4   6   8  10  12
```

d) Write and run the R command that will generate the following <u>replicate</u> vector elements:

```
[1] 1 1 2 2 3 3 4 4 5 5 6 6 7 7 1 1 2 2 3 3 4 4 5 5 6 6 7 7
```

e) Write and run the R command that will generate the following <u>replicate</u> vector elements:

```
[1] 1 1 1 1 4 4 4 4 4 5 5 5 5 5 5 3 3
```

f) Write and run the R command that will generate the following <u>replicate</u> vector elements:

```
[1] "F" "F" "F" "F" "F" "M" "M" "M" "M" "M"
```

g) Write and run the R command that will generate the following <u>replicate</u> vector elements:

```
[1]  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE
```

3. (*12 pts.*) The exam scores for five students are given below:

```
 John  Mary Steve  Bill  Anna
   57    85    90    72    68
```

a) Write an R code that will result in the above output when run.

b) Use R to calculate the mean and the standard deviation of these scores.

c) Print the results from part (b) in the form of a sentence using the concatenate function:

```
The average score on the exam is _____ with a standard deviation of _____.
```

(*Hint: In the Console, the blanks must be filled with numerical values when the command is run.*)

d) Write an R command that will result in the output below when run:

```
 Mary  Bill
   85    72
```

4. (*20 pts.*) For `x <- c(16, 7, 9, 14, 4, 3, 10, 11, 9, 8)`, write and run the R command that will

a) select only the fourth element

b) select the first five elements

c) select the second and the sixth elements

d) select all the elements except the fifth element

e) select all the elements except the first and the fourth elements

f) select all the elements that are less than 5

g) select all the elements that are greater than or equal to 10

h) Transform `x` vector into a character vector assigning it to `x_char`.

i) Using relevant R command, check to see whether `x_char` actually is a character vector

j) Print `x_char`

5. (*24 pts.*) Write and run the R command that will create the following `y` matrix

```
> y
     [,1] [,2] [,3] [,4]
[1,]    0    2    4    6
[2,]    8   10    0    2
[3,]    4    6    8   10
```

a) What is the R command used to list the elements from the first two rows of the `y` matrix?

b) What is the R command used to list the elements in the third column of the `y` matrix?

c) What is the R command used to list the element in the second row and the fourth column of the `y` matrix?

d) What is the R command used to list the elements 6 and 8 from the last row of the `y` matrix?

e) Use **Jan, Feb, Mar** to name the rows; and **A, B, C, D** to name the columns of the `y` matrix. Then, print the matrix.

f) What is the R command used to bind the `Apr <- c(1, 3, 4, 2)` vector to the `y` matrix as the fourth row, name the new matrix `y2`, and print it.

g) What is the R command used to bind the `E <- c(1, 2, 3, 4)` vector to the `y2` matrix as the fifth column, name the new matrix `y3`, and print it.

6. (*10 pts.*) Go to the Moodle Course page. From **Data, Script, and other R Files** folder under **Course Documents**, download the Excel file *cars.xlsx* and save it into your **Exam1_*your name*** directory.

a) Import this dataset into R using Clipboard. Name it `cars`.

b) Write and run the R command that will find the proportion of `cars` that have <u>six</u> Cylinders and report it here: _____

c) Using the `apply()` function, calculate the median `MPG` and report it here: _____

d) Note that `Country` stored in the first column of `cars` is a categorical variable. Using this variable, find the frequencies for each country and report them below

France	Germany	Italy	Japan	Sweden	U.S.
_____	_____	_____	_____	_____	_____

7. (*8 pts.*) Suppose you performed a street questionnaire, and recorded the respondents' handedness using the letters "R" and "L". This information is stored in the following vector:

```
survey_vec <- c("R", "R", "L", "R", "L", "R", "R")
```

a) Create the above character vector, then convert it into a factor vector naming it `survey_fac`, with the levels "Right" and "Left" (not "R" and "L"!)

b) Print `survey_fac` to inspect its contents.

c) Have a look at the underlying structure of `survey_fac` using `str()`.

In R's internal representation, which element is stored using index number 1: _____ and which using index number 2: _____ .

d) Use an appropriate R command with the `survey_fac` to find the <u>percentage</u> of "Right" responses: _____ and "Left" responses: _____ .

<mark>Save your RScript naming it **RScript_*your name***. Then, email the file along with this exam file (with the blanks filled in and saved) to the professor at <u>sgazioglu@mtech.edu</u>.</mark>

<mark>Have '**Stat435 – Exam1**' in the subject line of the e-mail.</mark>