Unique Paper Code	:	62353607
Name of the Paper	:	SEC-Statistical Software-R
Name of the Course	:	B.A. Programme
Semester	:	VI
Duration	:	2 Hours
Maximum Marks	:	38 Marks

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.

2. Attempt any **two** subparts of each part in question 1. Each subpart is of 1 mark and whole question 1 is of 8 marks.

3. In questions 2 to 6, attempt any two parts out of four parts. Each part is of 3 marks and each question is of 6 marks.

----

1. Attempt any two subparts of each part. Each subpart is of 1 mark. This whole question carries 8 marks.

    (a)

    (i)   Write the R syntax for $\sin(30°)$.

    (ii) If x=c(2L, 5L), then class(x) will give _____ .

    (iii) _____ command is used to make scatter plot. (plot()/splot())

    (iv) _____ command is used to give a simple list of the colours available. (colours()/ colors())

    (b)

    (i)   _____ command is used to remove the objects beginning with 'b'. (rm(pattern= '^b') / rm(pattern= 'b'))

    (ii) Objects can be reordered using _____ command. (rank() / order())

    (iii) _____ command is used to convert an object to a matrix. (is.matrix() / as.matrix())

    (iv) Cumulative statistics can be obtained by _____ command. (cum() / cumsum())

    (c) State whether TRUE or FALSE

    (i)   ls(pattern= 'b') command is used to find all the defined variables beginning with 'b'.

(ii)   `getwd()` command is used to find the default path of the files saved.

(iii)   If we don't have any named object at all, then `ls()` command gives NA.

(iv)   List objects are two dimensional objects.

(d) State weather TRUE or FALSE

    (i)   Standard deviation of a data can be obtained by `stddev()` command.

    (ii)   `plot()` command is used to make bar charts.

    (iii) R language is key sensitive.

    (iv) `with()` command enables temporary access to an enclosing object.

2. Attempt any two parts. Each part is of 3 marks.

(a) Differentiate between data frame and matrix.
Write a command to convert the following data into integer:
Data: 7 13 15.5 12.9 10.1 8

(b) What is the difference between `summary()` and `str()` command.
Write a command to list the objects starting with **a** or ending with **r**.

(c) Write a command to remove all the variables defined ending with 'k'.
Explain the use of save and load command with example.

(d) What is `read.csv()` command. How it works. Expalin by giving an example.

3. Attempt any two parts. Each part is of 3 marks.

(a) Suppose that `Data` is the numerical vector (`3, 6, 7, 4, 1, 8`). Describe the output of the following commands:

   (i) `Data*Data`

   (ii) `Data[Data<4]`

   (iii)`Data[c(1,3,5)]`

(b) Use the `seq()` command to generate the arithmetic progression whose first term is 1, common difference is 3, and the last term is 100. How would you find the number of terms in the above arithmetic progression? How would you find the 20th term in the progression?

(c) What commands will you use to find the cumulative sum and cumulative maximum of a numerical vector? Illustrate the above using an example. What output is expected as cumulative sum if the numerical vector is replaced with a character vector?

(d) Let `X` be the numerical vector (`13, 15, 12, 14, 11`). Write down the output of the following commands: `sort(X)`, `order(X)`, `rank(X)`.

4. Attempt any two parts. Each part is of 3 marks.

a) Write down the commands to create the following two matrices as matrix objects, and construct a new 4x2 matrix using these matrices whose columns are the second columns of A and B, respectively.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}, B = \begin{bmatrix} 11 & 12 & 13 \\ 14 & 15 & 16 \\ 17 & 18 & 19 \\ 20 & 21 & 22 \end{bmatrix}.$$

b) Write the commands to store the data given below as a data frame, and then calculate the mean, median and standard deviation of the temperatures.

	Temperature	Humidity
10 am	32.0	54
11 am	33.2	53
12 noon	33.9	50
1 pm	34.2	51
2 pm	34.5	50

c) Write down the commands to store the data in the following table as a matrix named Marks. Explain what information will be provided in the output of the command summary(Marks).

	Mathematics	Physics
S-1	80	71
S-2	67	87
S-3	72	59
S-4	54	64
S-5	75	73

d) Name any four forms of data objects. If X is a data object, how would you determine the type of the object? Give an example of converting one type of data object to another.

5. Attempt any two parts. Each part is of 3 marks.

a) Explain "The Kolmogorov-Smirnov Test" with the help of an example.

b) Write the syntax to generate "n" random values of
  (i) Normal distribution
  (ii) Uniform distribution

c) Data5: 3 3 8 4 2 7 1 5 7 2 8 7
  Write the code in R for the following:
  (i) Create a histogram for the Data5;
  (ii) Specify the breaks of bars at nos. 2, 5, 6, 9;
  (iii)Color the bars.

d) Rearrange the data in the increasing order and draw a stem and leaf plot where data is:
$$X = 3, 5, 7, 5, 3, 2, 6, 8, 5, 6, 9.$$

6. Attempt any two parts. Each part is of 3 marks.

a) Consider the following data: Sun, Mon, Tue, Wed, Thru, Fri, Sat

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Rainfall (cm)	9	8	5	4	4	7	9

  Write the code in R for the following:
  (i) Plot the pie chart for the above data;
  (ii) Color the pie chart;
  (iii)Pie chart in clockwise direction.

b) The following is a data frame by the name "XYZ"

A	B
34	47
29	24
24	09
14	25
09	15
05	14
03	02

  Write the code in R for the following:
  (i) Plot the boxplot of "XYZ";
  (ii) Label the x-axis and y- axis as "Variable" and "Value" respectively;
  (iii)Give the name and range to the boxplot of "XYZ".

c) Using the data frame "XYZ" mentioned in above question, write the code in R for the following:
  (i) Plot the Scatter Plot of "XYZ";
  (ii) Add the line of best fit to Scatter Plot of "XYZ";
  (iii)Plot the Cleveland dot of "XYZ".

d) The following is a Matrix by the name **"MAT"**

	length	Speed	Algae	NO3	BOD
1	20	12	40	2.25	200
2	21	14	45	2.15	180
3	22	12	45	1.75	135
4	23	16	80	1.95	120
5	21	20	75	1.95	110
6	20	21	65	2.75	120

Write the code in R for the following:
- (i) Plot the pair plots for length, Speed and NO3 of **"MAT"**;
- (ii) Plot the pair plots for BOD, Algae and NO3 of **"MAT"** with plotting symbol **"*"**;
- (iii)Plot the pair plots for BOD, Algae and length of **"MAT"** with plotting symbol **"+"** having size double.