USN | | | | | | | | | | **BDS306C**

Third Semester B.E./B.Tech. Degree Examination, Dec.2024/Jan.2025
Data Analytics with R

Time: 3 hrs.   Max. Marks: 100

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.
2. M : Marks , L: Bloom's level , C: Course outcomes.

		Module – 1	M	L	C
Q.1	a.	List and explain basic data types in R programming.	10	L2	CO1
	b.	Explain the following with example with respect to vector: i) seq( ) ii) length( ) iii) which( ) iv) rep( ) v) seq-len( )	10	L2	CO1
		OR			
Q.2	a.	With suitable example and syntax discuss 3 kinds of loops used in R.	10	L2	CO1
	b.	Define vector. Discuss the components of vector with example	10	L2	CO1
		Module – 2			
Q.3	a.	Define lists. Illustrate the working of lists and its associated functions.	10	L2	CO2
	b.	Develop a R program to create two 3×3 matrices A and B and perform the following operations: i) Transpose of the matrix ii) Addition iii) Subtraction iv) Multiplication.	10	L3	CO2
		OR			
Q.4	a.	Explain the following with suitable example: i) substr( ) ii) strsplit( ) iii) path.expand( ) iv) basename( ) v) file.path( )	10	L2	CO2
	b.	Define dataframe. Describe the significance of any 5 functions of dataframes with suitable example.	10	L2	CO2
		Module – 3			
Q.5	a.	Design a data frame in R for storing about 10 employee details. Create a CSV file named "input.csv" that defines all the required information about the employee such as id, name, salary, start date, dept. Import into R and do the following analysis. i) Find the total number rows and columns ii) Find the maximum salary iii) Retrieve the details of the employee with maximum salary iv) Retrieve all the employees working in the IT department.	10	L3	CO3

	b.	Explain any 5 grouping functions used in R.	10	L2	CO3

<div align="center">OR</div>

Q.6	a.	Define data reshaping. Explain in detail with example.	6	L2	CO3
	b.	Discuss the 4 functions used to manipulate the strings in data cleaning process.	8	L3	CO3
	c.	Explain the different ways to access the DB in R.	6	L2	CO3

<div align="center">Module – 4</div>

Q.7	a.	Write a syntax of pie chart creation and create a pie chart for the given data show in the table below.	10	L2,3	CO4

<div align="center">

Name	Rose	Lotus	Lilly	Sunflower	Jasmine
Values	25	35	10	5	15

Fig.Q.7(a) data for pie chart
</div>

	b.	Demonstrate how lattice graphics can be used to create box plots.	6	L 3	CO4
	c.	Define EDA. List the steps used in iterative cycle of EDA.	4	L2	CO4

<div align="center">OR</div>

Q.8	a.	Define line plot. Discuss with syntax how line plots are created using base graphics and lattice graphics.	10	L2	CO4
	b.	Let us use built-in data set air quality which has daily air quality measurements in New York, May to September 1973. Develop R program to generate histogram by using appropriate arguments for the following statements. i) Aligning names, using the air quality data set. ii) Change colors of the histogram iii) Remove axis and add labels to histogram iv) Change axis limits of a histogram v) Add density curve to the histogram.	10	L3	CO4

<div align="center">Module – 5</div>

Q.9	a.	Discuss in detail built in functions used to generate binominal distribution.	10	L2	CO5
	b.	Explain the following with example: i) Mean ii) Standard deviation iii) Variance iv) Mode v) Median.	10	L2	CO5

<div align="center">OR</div>

Q.10	a.	Explain linear regression with suitable example.	8	L2	CO5
	b.	With respect to normal distribution, explain i) dnorm ii) pnorm iii) gnorm iv) rnorm	8	L2	CO5
	c.	Write the syntax of cor( ) and cor.test( ). Explain its parameters.	4	L3	CO5

<div align="center">* * * * *</div>