MCKV Institute of Engineering
243, G.T. Road (N), Liluah, Howrah-711204
Lab Assignment
Year- 2022

Subject : - Introduction to R Programming
Code : - PC-CS(D)593
Stream : - Computer Science and Engineering (Data Science)
Credit : - 1

Topic	Assignment No	Problem Statement
Introduction to R and R-Studio	1	Installation of R and R-Studio and customization of software. Familiarization with both software. Execution of R run R, use of R help files etc.
R Objects – Vectors	2	Create an integer vector of 5 elements and display the type. Print the 4th and 2nd element of the vector. Print all the items of the vector that are greater than 5. Sort the elements in descending order. Modify with 0 for all the elements that are greater than or equal to 4. Delete all the elements from the vector.
R Objects – Matrices	3	Write a R program to create three vectors a, b, c with 3 integers. Combine the three vectors to become a 3×3 matrix where each column represents a vector. Put the row name as A, B, C and column name as 1, b, c. Print the first and second row. Print the first and third column. Delete the 2nd row of the matrix. Again, create another vector d vector with 3 integers and concatenate it with the previous matrix. Now print the content of the matrix, diagonal elements and the transpose of the matrix. Finally check that it is a square matrix or not.
R Objects – List	4	Create a list by empId, empName and empSal for four employees. Print the name of all employees. Create another list of empAge. Concatenate two lists. Convert list to a matrix. Print the matrix and check the class of matrix.
R Objects –Data Frames	5	Create a Data Frame from an existing .csv file (consisting empId, empName, empSal, dept). Print the number of rows available in the Data Frame. Find the Maximum Salary. Print all the employee working in IT department. Print all the name of the employees of IT department whose salary is greater than 10000. Get the details of the person with max salary. Write into a .csv file.
R-Control Statements	6	Suppose an angle α is given as a positive real number of degrees. If $0 \leq \alpha < 90$ then it is quadrant 1. If $90 \leq \alpha < 180$ then it is quadrant 2. If $180 \leq \alpha < 270$ then it is quadrant 3. If $270 \leq \alpha < 360$ then it is quadrant 4. If $360 \leq \alpha < 450$ then it is quadrant 1. And so on. Write a program in R to find the quadrant of the angle α.

	7	Write a R program to print the name of the day of a week (using Switch statement)
	8	Create a vector "day". Write a program in R to display the days of a week. (use for loop)
	9	Create a character vector. Write a R program to print the number of characters in the vector element. (use while loop)

	10	Write a program in R to check whether a given matrix is symmetric or skew-symmetric or none of them. Also check the matrix is orthogonal or not. (Use switch in your program)
R Packages	**11**	Write a program in R to create an object from the existing ".xlsx" data set. Count the data. Short the data on "First Name" in descending order. Find the email id of "Abbot". Count the no of male and female employee in the database.
	12	Write a program in R to create an object from the existing "iris" data set. Count the data. Short the data on "Sepal Length" in descending order. Short the data groupwise on "Sepal Length" and "Sepal Width". Find all the data where Sepal Length is greater than 5. Create another object by selecting the 5th to 10th row from the original. Find the details of the maximum "Sepal Length".
	13	Write a R program to find the average age (mean) of the customers in CardioGoodFitness.csv. Find the Median of the age and find the age group with maximum number of customers. Count the total sample present in your data.
	14	Write a program in R to find the **correlation coefficient** of the following distribution: x 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 y 1.2 2.1 3.4 4.3 5.2 6.8 7.4 8.3 9.5
Linear Regression	**15**	Predict the fitness of a person when his/her average number of miles to walk/run is known to us.
Normal Distribution	**16**	The length of reproduction of a particular animal from conception to birth approximates a normal distribution with a mean of 266 days and a standard deviation of 16 days. What proportion of all reproductions will last between 240 and 270 days (roughly between 8 and 9 months)?
	17	The average number of acres burned by forest and range fires in a large New Mexico City is 4,300 acres per year, with a standard deviation of 750 acres. The distribution of the number of acres burned is normal. What is the probability that between 2,500 and 4,200 acres will be burned in any given year?

Decision Tree	**18**	Generate the decision tree for prediction of flower species from iris dataset.
R Graphics	**19**	Use standard graphics package to represent data visually using plot(), barplot(), pie() functions and save the graphs to files with different format like .pdf, .jpg . Use the dataset of your own choice and represent them visually.
	20	Install DMwR2 and ggplot2. (if it does not exist in your system else include the library in your program). Use the dataset "algae" and represent visually "Distribution of the Water Samples across Seasons" using ggplot() which are equivalent to plot(), pie(),barplot(). Create histogram for "Distribution of Algae a1" both using ggplot() and hist().

Signature of Faculty & Date   Signature of HOD & Date