Roll No. | | | | | | | | |

End Semester Examination 2024

Name of the Program: B.Tech (CSE)
Name of the Course: B.Tech (CSE)
Paper Name: Statistical Data Analysis with R
Time: 3 Hours

Semester: IV
Course Code: TCS 471

MM: 100

Note:

(i) All questions are compulsory.
(ii) Answer any two sub-sections among a, b, and c in each main question.
(iii) Total marks in each main question are twenty.
(iv) Each questions carry ten marks.

Q.1	**(10 × 2 = 20 Marks)**	CO-1
a)	What is the difference between discrete and continuous random variable?	
b)	The probabilities of X, Y and Z becoming managers are 4/9, 2/9 and 1/3 respectively. The probabilities that the Bonus Scheme will be introduced if X, Y and Z becomes managers are 3/10, 1/2 and 4/5 respectively. (i) What is the probability that Bonus Scheme will be introduced and (ii)If the Bonus Scheme has been introduced, what is the probability that the manager appointed was X?	
c)	Calculate the coefficient of correlation between X and Y for the following : X : 1 3 4 5 7 8 10 Y: 2 6 8 10 14 16 20	
Q.2	**(10 × 2 = 20 Marks)**	CO-2
a)	Find linear regression equation of Y on X for the following two sets of data: X : 2 4 6 8 Y: 3 7 5 10	
b)	Write the pdf/pmf of Normal, Geometric and Poisson distributions.	
c)	What is the probability mass function of Binomial distribution? Ten coins are thrown simultaneously. Find the probability of getting at least seven heads.	
Q.3	**(10 × 2 = 20 Marks)**	CO-3
a)	A random variable X has the following probability function: x : 0 1 2 3 4 5 6 7 Y: 0 k 2k 2k 3k k2 2k2 7k2 + k (i) Find k, (ii) Evaluate P(X < 6), P(X ≥ 6) and P(0 < X < 5)	
b)	A variable X has probability distribution function $f(x) = kx^3(4 - x)^2, 0 < x < 4$ Find the value of k, the mean and variance of the distribution.	
c)	Write an R script that: 1. Creates a 3x3 matrix with the numbers 1 to 9. 2. Computes the transpose of the matrix. 3. Computes the determinant of the matrix. 4. Multiplies the matrix by its transpose.	
Q.4	**(10 × 2 = 20 Marks)**	CO-4
a)	A) Write a R program to create a 5 × 4 matrix, 3 × 3 matrix with labels and fill the matrix by	

	rows and 2 × 2 matrix with labels and fill the matrix by columns.	
b)	A) Write a R program to create a Dataframes which contain details of 5 employees and display the details. B) Write a R program to create a simple bar plot of five subjects marks.	
c)	Write a R program to extract first 10 english letter in lower case and last 10 letters in upper case and extract letters between 22nd to 24th letters in upper case.	
Q.5	**(10 × 2 = 20 Marks)**	
a)	Write an R script that: 1. Creates a data frame with columns `Product` (character), `Category` (character), and `Sales` (numeric). 2. Adds six rows of data with at least two different categories. 3. Calculates the total sales for each category. 4. Finds the product with the highest sales in each category.	**CO-5**
b)	Write an R script that: 1. Creates a data frame with columns `ID` (integer), `Group` (character), and `Value` (numeric). 2. Adds eight rows of data with at least two different groups. 3. Applies a custom function to compute the square of each `Value`. 4. Adds a new column `SquaredValue` to the data frame with the computed squared values.	
c)	A sample of 100 dry battery cells tested to find the length of life produced the following results. Mean = μ = 12 hours, standard deviation = σ = 3 hours. Assuming that the data are normally distributed, what percentage of battery cells are expressed to have the life a) more than 15 hours, b) less than 6 hours, and c) between 10 hours and 14 hours.	