Reg. No.															

Question Paper Code	13545

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2025

Fifth Semester

Information Technology

20ITPW501 - STATISTICAL ANALYSIS USING R PROGRAMMING WITH LABORATORY

Regulations - 2020

Duration: 3 Hours | Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

		Marks	*K – Level*	*CO*
1.	What is the difference between a data frame and a matrix in R? (a) Data frames can only store numeric data (b) Matrices can only have rows, not columns (c) Data frames can hold different types of data in each column (d) Matrices are more memory efficient than data frames	*1*	*K1*	*CO1*
2.	What type of R object is used to represent categorical data? (a) Vector (b) Data Frame (c) Factor (d) Matrix	*1*	*K1*	*CO1*
3.	What is the purpose of a Q-Q plot in probability distributions? (a) To visualize histograms (b) To compare two distributions (c) To find random samples (d) To compute empirical distribution	*1*	*K1*	*CO2*
4.	Which of the following is NOT a discrete probability distribution? (a) Binomial (b) Poisson (c) Normal (d) Geometric	*1*	*K1*	*CO2*
5.	What does a parallel boxplot represent in grouped data analysis? (a) Distribution of a single variable (b) Frequency counts (c) Distribution comparisons between multiple groups (d) Categorical data comparison	*1*	*K1*	*CO3*
6.	What type of chart is best for visualizing the frequency distribution of a categorical variable? (a) Histogram (b) Dot chart (c) Pie chart (d) Strip chart	*1*	*K1*	*CO3*
7.	What do residuals in a regression analysis represent? (a) The actual data points (b) The predicted values (c) Difference between actual and predicted values (d) Correlation between variables	*1*	*K1*	*CO4*
8.	Which type of correlation is more robust to outliers? (a) Pearson correlation (b) Covariance (c) Spearman's ρ (d) Mean absolute deviation	*1*	*K1*	*CO4*
9.	Which of the following tests is used for checking the equality of variances across groups in ANOVA? (a) Kruskal–Wallis Test (b) Bartlett's Test (c) Friedman Test (d) Wilcoxon Test	*1*	*K1*	*CO5*
10.	What is the main purpose of using polynomial regression? (a) To model relationships with linear data (b) To model relationships with non-linear data (c) To classify categorical variables (d) To create dummy variables	*1*	*K1*	*CO6*

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

		Marks	*K – Level*	*CO*
11.	Convert the following multi-line operations to a single expression. Check that both approaches give the same result. w<- u + v w <- w / 2	*2*	*K2*	*CO1*
12.	State slicing operation in a matrix.	*2*	*K2*	*CO1*
13.	Summarize the use of any two built-in distribution functions in R.	*2*	*K2*	*CO2*

14.	Discuss how histograms are useful for visualizing distributions?	2	K2	CO2
15.	Compare the use of a one-sample t-test and a two-sample t-test.	2	K2	CO3
16.	Write an R-Script to compare variances between two samples using an F-test.	2	K2	CO3
17.	Differentiate Kendall's τ differ from Spearman's ρ in measuring correlation.	2	K2	CO4
18.	State outliers affect the Pearson correlation coefficient.	2	K2	CO4
19.	Explain the plot that is commonly used to visualize multivariate data.	2	K2	CO5
20.	Differentiate Pearson and Kendall Correlation Test.	2	K2	CO5
21.	Discuss the primary use of generalized linear models (GLMs) in statistical analysis.	2	K2	CO6
22.	What is polynomial regression and why is it used?	2	K1	CO6

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23.	a) (i)	Write an R program to get the first 10 Fibonacci numbers.	6	K3	CO1
	(ii)	Write an R program to perform the following operation in data frame: a. Access the element at 3rd column and 1st row in a matrix. b. Access only the second row c. Access the element at 2nd column and 4th row in a matrix	5	K3	CO1
		OR			
	b) (i)	Write an R program to count the number of NA values in a data frame column.	6	K3	CO1
	(ii)	Explain vector and how to create it, Create a vector A of elements 5, 2, -2, 6,7, 10,12,14,15 and from it create a vector Y containing elements of A>6.	5	K2	CO1
24.	a)	Write an R script that generates a boxplot for a given dataset, highlights outliers, and explains the significance of the median and quartiles in understanding data distribution.	11	K3	CO2
		OR			
	b)	Write an R program that simulates random numbers from a normal distribution, plots a histogram for these numbers.	11	K3	CO2
25.	a)	Explain the procedure and interpretation of the Wilcoxon Signed-Rank Test in R. Discuss a real-world scenario where this test would be applicable.	11	K2	CO3
		OR			
	b)	Compare and contrast the use of strip charts, dot charts, and bar plots in visualizing grouped data. Include sample code for each.	11	K2	CO3
26.	a)	Calculate the pearson correlation coefficient for the following data. X = 4, 8 ,12, 16 and Y = 5, 10, 15, 20.	11	K3	CO4
		OR			
	b)	A psychologist is studying the relationship between stress levels and sleep quality in a group of 10 individuals. The psychologist measures stress levels (on a scale of 1 to 10) and sleep quality (also on a scale of 1 to 10) but finds that the data contain tied ranks, so they decide to use Kendall's Tau correlation test. The dataset is as follows:	11	K3	CO4

Stress	4	7	5	9	6	3	8	5	6	2
Sleep	7	6	7	5	6	8	4	7	6	9

Perform a rank correlation test of x and y using Kendall test. Write the appropriate R code.

27. a) (i) Explain the steps of Bartlett Test used in Statistics. *6 K2 CO5*

(ii) Explain how multivariate data can be plotted by using R. *5 K2 CO5*

OR

b) A researcher wants to investigate whether there is a difference in median monthly sales performance among three sales teams (A, B, and C) within a company. The data on monthly sales (in thousands) collected from each team over six months is as follows: *11 K3 CO5*

Team	Monthly Sales (in thousands)
A	55, 60, 53, 58, 57, 61
B	48, 52, 49, 50, 51, 53
C	62, 65, 60, 63, 67, 64

Since the data is not normally distributed, the researcher decides to use the Kruskal-Wallis test.

28. a) Illustrate how to perform two-way ANOVA with replication. *11 K2 CO6*

OR

b) Explain the logistic regression model fitting process and interpret the resulting coefficients. *11 K2 CO6*