Reg. No.															

Question Paper Code	14065

B.E / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2025

Fifth Semester

Information Technology

20ITPW501 - STATISTICAL ANALYSIS USING R PROGRAMMING WITH LABORATORY

Regulations - 2020

(Use of Student's T Table, F Table, Z Table, Chi Square Table permitted)

Duration: 3 Hours Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	Which function is used to create a repeating sequence of values in R? (a) repeat() (b) rep() (c) seq() (d) repeat_vec()	1	K1	CO1
2.	Consider a two-dimensional structure in R that can contain different data types in each column but same number of rows in each column. Name the structure: (a) Matrix (b) Array (c) Data Frame (d) Vector	1	K1	CO1
3.	The R function ecdf() is used to: (a) Compute the probability density function of a numeric variable (b) Compute the empirical cumulative distribution function of a numeric sample (c) Generate random numbers from a uniform distribution (d) Plot a Q-Q plot for normality	1	K1	CO2
4.	In R, which of the following functions returns the density estimate (kernel smoothed) of a numeric variable x? (a) hist(x) (b) density(x) (c) ecdf(x) (d) qqplot(x)	1	K1	CO2
5.	Histograms cover their purpose in (a) Showing data frequency distribution (b) Identifying data shape (c) Interpreting the meaning of bar heights and intervals (d) All the above	1	K1	CO3
6.	What is another name for a marginal table that displays frequencies for two categorical variables? (a) One-way table (b) Frequency table (c) Contingency table (d) Simple table	1	K2	CO3
7.	Which correlation coefficient is appropriate for measuring the strength and direction of a linear relationship between two normally distributed continuous variables? (a) Spearman's Rho (b) Kendall's Tau (c) Pearson's correlation coefficient (d) Point-biserial correlation	1	K2	CO4
8.	Which graphical display is best suited for comparing the overall distribution of a continuous variable across two or more different groups? (a) Bar plot (b) Pie chart (c) Parallel box plots (d) Strip chart	1	K1	CO4
9.	In R, which function is used to perform a Simple Linear Regression model? (a) regression() (b) lm() (c) anova() (d) model()	1	K2	CO5
10.	Polynomial Regression is used when: (a) The relationship between variables is linear (b) The relationship between variables is non-linear but continuous (c) The variables are categorical (d) The data has equal variances	1	K2	CO6

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

11.	Write the R command to compute the mean of vector v = c (4, NA, 7, 2, NA, 9) ignoring the missing values. Also, explain what the argument na.rm = TRUE does.	2	K2	CO1
12.	Compare matrix and array in R programming.	2	K2	CO1
13.	Give the R function to generate and plot 1000 normal random numbers.	2	K1	CO2

14.	Name the R function used to compute the number of combinations$\binom{n}{k}$.		2	K1	CO2
15.	Which R function would you use to generate 100 random samples from a standard uniform distribution?		2	K1	CO3
16.	Identify the use of par() function in R.		2	K2	CO3
17.	Compare and contrast positive, negative and no correlation relation between two variables.		2	K2	CO4
18.	Give the formula for Wilcoxon signed ranked test.		2	K1	CO4
19.	Can homogeneity of variance be tested through a statistical test?		2	K2	CO5
20.	Differentiate between one-way ANOVA and Kruskal Wallis Test.		2	K2	CO5
21.	Define Polynomial Regression.		2	K1	CO6
22.	State the usage of GLM.		2	K2	CO6

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23.	a)	Write a data frame in R named 'DF' containing the columns 'age, gender, income'. Write R code to filter 'DF' to include only rows where age > 30 AND gender = Male. Create a new column Income_Range which is set to "High" if income > 50000 else is set to "Low". Write the R code and explain each step in detail.	11	K2	CO1
		OR			
	b)	Summarize how interfacing to other environments help in accessing R studio.	11	K2	CO1
24.	a)	Summarize in detail about iterative and branching statements in R with appropriate example. Write R code to generate first n terms of a Fibonacci series.	11	K2	CO2
		OR			
	b)	Illustrate matrix addition, subtraction, multiplication and division using R. Explain usage of rbind() and cbind() in marix.	11	K2	CO2
25.	a)	Interpret the importance of boxplot, strip charts, pie chart and bar plots with R example.	11	K3	CO3
		OR			
	b)	Compute the 35th percentile, the 55th percentile, Q1, Q2 and Q3 for the following data. 16 28 29 13 17 20 11 34 32 27.-25 30 19 18 33. Plot the data using R function.	11	K3	CO3
26.	a)	Enumerate the concept of One Sample T Test? How to calculate t test statistic? Give an example to accept or reject the null hypotheses using t test. Write appropriate R code and output to support your answer.	11	K3	CO4
		OR			
	b)	Compare the assumptions and interpretations of Pearson, Spearman, and Kendall correlation coefficients in assessing linear versus monotonic relationships between variables.	11	K3	CO4
27.	a)	Interpret the use of interaction mean in a two-way ANOVA? Explain how Interaction is calculated in two-way ANOVA with example.	11	K3	CO5
		OR			
	b)	Heights(in cm) of father and son are given as follows Father(X): 150 152 155 157 160 161 164 165 Son(Y): 154 156 158 159 160 162 161 164 Fit a regression line predict the height of son given the height of father and provide a scatter-plot of the data points with the regression line superimposed. Comment briefly on whether a linear model seems reasonable.	11	K3	CO5

28. a) Articulate the steps and necessary R functions to implement polynomial regression and logistic regression on tabular data. *11 K3 CO6*

OR

b) Construct the design matrix for a multiple regression model in R. Explain how R encodes categorical predictors and interpret the structure of the design matrix. *11 K3 CO6*