(TIME: 03 HOURS) **(MAX. MMARKS: 80)**

Note:

1. Question **No. 1** is compulsory.
2. Attempt **any three** questions out of remaining **five** questions.
3. Assume suitable data wherever necessary.
4. Figures to right indicate full marks.

			Marks
Q.1		Answer the following **(Any four)**	20
	a.	Differentiate between rectangular data and data frames.	
	b.	Explain Sample Mean and Population Mean with proper examples.	
	c.	What do you mean by A/B testing. What are its benefits?	
	d.	List some graphical visualization methods for summarizing data.	
	e.	Write a note on multiple linear regression.	
Q.2	a.	Using the following data, perform a one-way analysis of variance using α=.05.	10

Observations	A	B	C
1	35	30	50
2	40	450	70
3	45	45	65
4	50	50	65
5	60	55	60

	b.	The Sample Mean and Expected Mean value of the marks obtained by 25 students in a class is 510 and 490 respectively. What is the t-score if the standard deviation of the marks is 42. Determine if the result is statistically significant, assuming the significance level as 0.05.	10
Q.3	a.	Explain Binomial Distribution in detail. 50% of people who purchase iPhone are women. If 10 iPhone owners are randomly selected, find the probability that exactly 5 are women.	10
	b.	What is simple linear regression? Find the simple linear regression equation that fits the given data and coefficient of determination.	10

X	Y
3	4
5	7
7	9
9	10

X525Y93F77FX525Y93F77FX525Y93F77FX525Y93F77F

Q.4 a. What is Chi-Square test? A study was conducted, among 100 professors from 2 different divisions for the preference on beverages of 2 categories to test if there is any relationship between the field of teaching and preference of beverage and the results are summarized in the following contingency table. Use the Chi-Square test for independence to determine if there is a significant association between the profession and the beverage at a 5% significance level ($\alpha = 0.05$). 10

	Coffee	Tea	Cold drink	Total
Humanity	21	15	14	50
Engineering	18	20	12	50
Total	39	35	26	100

b. Explain Quantile-Quantile plots, Stem-and-Leaf plots in detail. 10

Q.5 a. Three different training methods, M1, M2, and M3, are to be compared to determine the effect on student performance. A sample of 12 students are divided into three groups and are evaluated based on these three training methods. The scores obtained by these students are as follows: 10

M1	M2	M3
70	80	75
60	56	50
55	60	68
61	74	59

Test that the median student's performance due to three methods is same by Kruskal Wallis test at 5% level of significance.

b. Construct a frequency distribution table for the following marks of 25 students using equal class intervals, one of them is 50-55 (55 not included). Also draw Histogram for the given data. 10
The marks are:

35, 45, 41, 38, 46, 48, 51,57, 53, 56, 63, 66, 69, 70, 77, 65, 76, 58, 79,49, 57, 35, 43, 56, 71.

Q.6 Write short notes on (ANY FOUR) 20

a) Friedman's Test
b) ANOVA
c) Type I and Type II error
d) Boot strapping V/s Re-sampling
e) Central Limit Theorem

****

X525Y93F77FX525Y93F77FX525Y93F77FX525Y93F77F