Course Code: 23ES4T10

BONAM VENKATA CHALAMAYYA INSTITUTE OF TECHNOLOGY & SCIENCE

(AUTONOMOUS)

II-B. Tech II-Semester Supplementary Examinations (BR23), Aug - 2025

STATISTICAL METHODS FOR DATA SCIENCE

(CSE_AI&DS)

Time: 3 hours Max. Marks: 70

Question Paper consists of Part-A and Part-B

*Answer **ALL** the question in **Part-AandPart-B***

PART–A (10X2 = 20M)

		Marks	CO	BL
1. a)	What are data visualization techniques?	(2M)	CO1	BL1
b)	Tell any two conditions for Binomial distribution.	(2M)	CO1	BL1
c)	Define point estimation with one example.	(2M)	CO2	BL1
d)	Tell the statistic formulae for single mean and single proportion for large samples.	(2M)	CO2	BL1
e)	Write normal equations to fit a multiple linear regression of the form $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$.	(2M)	CO3	BL1
f)	What is polynomial regression?	(2M)	CO3	BL1
g)	List the components of time series analysis.	(2M)	CO4	BL1
h)	List the techniques of time series analysis.	(2M)	CO4	BL1
i)	What are key features of classification problem?	(2M)	CO5	BL1
j)	List the steps in logistic regression analysis.	(2M)	CO5	BL1

PART–B (5X10 = 50M)

2a. The frequency distribution of weights of a group of 60 salesmen (students) of a company (college) is given below: 5M CO1 BL2

Weights(in kg):	50 – 54	55 – 59	60 – 64	65 – 69	70 – 74	75 – 79	80 – 84
No. of Salesmen:	6	10	24	36	28	12	4

Draw histogram for the distribution and hence find the modal value.

b. If the probability that an individual suffers a bad reaction due to a certain injection is 0.001, determine the probability that out of 2000 individuals (i) exactly 3 (ii) more than 2 individuals will suffer a bad reaction. 5M BL3

(OR)

3a. The following table gives information on the favourite colours by the group of people 5M CO1 BL2

colours	yellow	pink	blue	green	orange
No. Of people	16	20	30	26	35

Draw the line graph for the information provided.

b. In a test on 1000 electric bulbs, it was found that the number of bulbs was normally distributed with an average life of 2040 hours and a standard deviation of 60 hours. How many bulbs are likely to be in usage for (a) more than 2150 hours (b) less than 1950 hours? 5M BL3

4a. An ambulance services claims that it takes on an average less than 10 minutes to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 min and variance 16 min to reach the destination in emergency calls. Test the claim at 5% level of significant. 5M CO2 BL3

b. In a sample of 500 people in Tamil Nadu 280 are tea drinkers and the rest are coffee drinkers. Can we assume that both coffee and tea are equally popular in the state at 1% LOS? 5M BL3

(OR)

5. A random sample from a company's very extensive files shows that the orders for a certain kind of machinery were filed, respectively in 10,12,19,14,15,18 11 and 13 days. Use the level of significance 0.01 to test the claim that on the average such orders are filed in 10.5 days. 10M CO2 BL3

6a. By the method of least square, fit a curve of the form $y = ax^b$ to the following data: 5M CO3 BL3

x	2	3	4	5
y	27.8	62.1	110	161

b. Fit a regression equation of Y on X from the data given below, taking deviations from actual means of X and Y. 5M BL3

Price(Rs.)	10	12	13	12	16	15
Amount Demanded	40	38	43	45	37	43

Estimate the likely demand when the price is Rs. 20.

(OR)

7. Find y when $x_1=10$ and $x_2=6$ from the least squares regression equation of y on x1 and x2 for the following data: 10M CO3 BL3

y	90	72	54	42	30	12
x_1	3	5	6	8	12	14
x_2	16	10	7	4	3	2

8. Find seasonal Indices using ratio to trend values. 10M CO4 BL4

Year/ Quarterly	Q_1	Q_2	Q_3	Q_4
2007	72	68	80	70
2008	76	70	82	74
2009	74	66	84	80

2010	76	74	84	78
2011	78	74	80	82

(OR)

9a.	Fit a straight line to the following data and calculate trend values. Time (t) : 2005 2006 2007 2008 2009 2010 Value (U_t): 16 22 24 28 32 34	5M	CO4	BL3
b.	What are the components of Time series and explain them.	5M		BL2
10a	How many years will it take for a bacteria population to reach 9000, it its growth is modelled by $f(t)=\frac{10000}{1+e^{-0.12(t-20)}}$ here, in t years?	5M	CO5	BL2
b.	Explain the advantages of logistic regression for classification problems.	5M		BL2

(OR)

11a	Explain key steps in comparing models.	5M	CO5	BL2
b.	What are significant characteristics of logistic regression and explain with suitable examples?	5M		BL2

****