Course Code: 23ES4T10

BONAM VENKATA CHALAMAYYA INSTITUTE OF TECHNOLOGY & SCIENCE

(AUTONOMOUS)

II-B. Tech II-Semester Regular Examinations (BR23), Apr/May - 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-A and Part-B***

PART–A (10X2 = 20M)

		Marks	CO	BL
1. a)	Mention any two advantages and disadvantages of data visualization.	(2M)	CO1	BL1
b)	If a random variable has a Poisson distribution such that P(1) = P(2), find mean of the distribution.	(2M)	CO1	BL1
c)	What are the large sample tests?	(2M)	CO2	BL1
d)	List the steps involved in test of hypothesis.	(2M)	CO2	BL1
e)	Write the normal equations in method of least squares to fit a straight line.	(2M)	CO3	BL1
f)	Convert an exponential curve into linear curve and write the normal equation by the method of least squares.	(2M)	CO3	BL1
g)	What are the different methods to find secular trend.	(2M)	CO4	BL1
h)	List the different methods to find seasonal variations.	(2M)	CO4	BL1
i)	What is sigmoid function?	(2M)	CO5	BL1
j)	What are the advantages and disadvantages of logistic regression.	(2M)	CO5	BL1

PART–B (5X10 = 50M)

2a.	What do you understand by data visualization? List out the advantages of data advantages and disadvantages.	5(M)	CO1	BL2
b.	In a sample of 1000 cases, the mean of a certain test is 14 and standard deviation is 2.5. Assuming the distribution to be normal, find (i) how many students score between 12 and 15? (ii) How many score above 18.	5(M)		BL3
	(OR)			
3a.	The frequency distribution of marks obtained by 60 students of a class in a college XYZ is given below. Marks : 30 – 34 35 – 39 40 – 44 45 – 49 50 – 54 55 – 59 60 - 64 No. of Students: 3 5 12 18 14 6 2 Draw histogram for the distribution and hence find the modal value.	5(M)	CO1	BL2
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.	5(M)		BL3
4a.	What is the maximum error you can expect to make with probability 0.99when using the mean of a random sample of size 64 to estimate the mean of the population $\sigma^2 = 2.56$?	5(M)	CO2	BL3

b. In a random sample of 60 workers the average time taken by them to get to work is 33.8 minutes with a S.D of 6.1 minutes. Can we reject the null hypothesis in favor of alternative hypothesis $\mu > 32.6$ at α=1% LOS. 5(M) BL3

(OR)

5. To examine the hypothesis that the husbands are more intelligent than the wives, an investigator took a sample of 10 couples and administered them a test which measures the I.Q. The results are as follows: 10(M) CO2 BL3

Husbands	117	105	97	105	123	109	86	78	103	107
Wife's	106	98	87	104	116	95	90	69	108	85

Test the hypothesis with a reasonable test at the L.O.S 0.05.

6a. Price indices of cotton and wool are given below for the 12 months of a year. Obtain the equations of lines of regression between the indices. 5(M) CO3 BL3

X	78	77	85	88	87	82	81	77	76	83	97	93
Y	84	82	82	85	89	90	88	92	83	89	98	99

b. The results of measurement of electric resistance R of a copper bar at various temperatures t^oC are listed below: 5(M) BL3

t:	19	25	30	36	40	45	50
R:	76	77	79	80	82	83	85

Find a relation R=a+bt when a and b are constants to be determined by you.

(OR)

7. Find the multiple regression equation of Y on X_1 and X_2 from the data given below 10(M) CO3 BL3

Y	1	2	3	5
X_1	1	3	4	2
X_2	7	18	25	23

8. Find seasonal Indices using ratio trend method from the data given below. 10(M) CO4 BL4

Year/ Quarterly	Q_1	Q_2	Q_3	Q_4
2004	60	80	72	68
2005	68	104	100	88
2006	80	116	108	96

2007	108	152	136	124
2008	160	184	172	164

(OR)

9 Calculate seasonal Indices by ratio to moving averages method from the following data. 10(M) CO4 BL4

Year/Quarterly	Q_1	Q_2	Q_3	Q_4
2005	75	60	54	59
2006	86	65	63	80
2007	90	72	66	85
2008	110	78	72	93

10a. Explain the different types of classification problems with suitable examples. 5(M) CO5 BL2

b. What are null and residue deviance in the context of logistic regression? 5(M) BL2

(OR)

11a. How do you interpret the coefficients of logistic regression model? 5(M) CO5 BL2

b. Explain the different models in logistic regression model using suitable examples. 5(M) BL2

****