KUMARAGURU
college of technology
character is life

M.E/M.TECH/MCA DEGREE EXAMINATIONS: NOV/ DEC 2024

(Regulation 2024)

First Semester

MASTER OF COMPUTER APPLICATIONS

24MAI504 - Probability And Statistics for Data Analysis

COURSE OUTCOMES

CO1: Analyse data using appropriate collection methods, central tendency measures, variation metrics, and graphical tools such as charts and box plots.

CO2: Apply correlation techniques (Karl Pearson's and Spearman's) to develop regression models for discrete data to predict outcomes.

CO3: Apply the concept of probability distributions for random variables, expectation and normal distribution to solve real-world problems.

CO4: Apply statistical methods such as large sample tests and chi-square tests (hypothesis testing) to make data-driven conclusions.

CO5: Analyse the effectiveness of experimental designs, including Completely Randomized Design (CRD), Randomized Block Design (RBD), and Latin Square Design (LSD), through Analysis of Variance (ANOVA).

CO6: Analyze, interpret and draw insights from data by applying statistical methods in R programming

Time: Three Hours   **Maximum Marks: 100**

PART A (4*20 = 80 Marks)

1.	a)	What are the essential characteristics of a good table? Give the advantages of tabulation.	5	CO1	[K$_2$]
	b)	What is the purpose of graphical data representation? State the difference between Bar chart and Pie chart.	5	CO1	[K$_1$]
	c)	Scenario: The yield and plant height of a paddy variety are given, the mean and standard deviation for yield are 50 kg and 10 kg respectively. The mean and standard deviation for plant height is 55 am and 5 cm respectively. Find the variability among yield and plant height.	6	CO1	[K$_4$]
	d)	Calculate the value of median and mode from the following data: 10,27,24,12,27,27,20,18,15,30	4	CO1	[K$_4$]
2.	a)	What is a scatter diagram? Mention its uses	2	CO2	[K$_1$]
	b)	How does rank correlation differ from Pearson's correlation coefficient? When might rank correlation be more appropriate?	2	CO2	[K$_2$]

c) Given two variables X and Y calculate the Pearson correlation coefficient. 4 CO2 [K₃]

X	10	15	20
Y	25	30	35

d) Calculate the rank correlation coefficient (Spearman's) for the dataset: 4 CO2 [K₃]

X	15	20	25
Y	30	40	50

e) Fit a linear regression line to the data points, and find the equation of the line. 8 CO2 [K₃]

X	2	4	6	8
Y	12	18	22	28

3. a) A random variable X has the following probability distribution 6 CO3 [K₃]

X	0	1	2	3	4	5	6	7
P(X)	0	K	2K	2K	3K	K^2	$2K^2$	$7K^2+K$

Find (i) K (ii) Find $P(X < 6)$ $iii)P(0 < X < 5)$

b) A Continuous random variable X has the p.d.f. $f(x) = kx^2 e^{-x}, x \geq 0$. Find k. Also find the mean and variance of X. 6 CO3 [K₃]

c) The saving bank account of a customer showed an average balance of Rs.150 and a standard deviation of Rs.50. Assuming that the account balances are normally distributed. (i) What percentage of account is over Rs. 200? (ii) What percentage of account is b/w Rs.120 and Rs.170? (ii) What percentage of account is less than Rs.75? 8 CO3 [K₃]

4. a) A light bulb company claims that the 40watt light bulb it sells has an average life of 1000 hours with a standard deviation of 75 hours. Sixty-four (64) new bulbs were allowed to burn out to test this claim. The average lifetime of these bulbs was found to be 975 hours. Does this indicate that the average life of a bulb is not 1000 hours? 7 CO4 [K₃]

b) A total no. of 3759 individuals were interviewed according to gender and decision in a public opinion survey on a political proposal with the results as in the following table. Use 0.05 level of significance and test the hypothesis that there is no association b/w gender and attitude. 7 CO4 [K₃]

Decision			
	Favoured	Opposed	Undecid
Male	1154	475	243
Female	1103	442	342

c) Suppose the probability mass function of the discrete random variable is

X=x	0	1	2	3
P(x)	0.2	0.1	0.4	0.3

6   CO3   [K₃]

What is the value of $E(3X + 2X^2)$?

Answer any ONE Question
PART B (1*20 = 20 Marks)

5. a) Find the mean, median.

x	1-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
f	7	10	16	30	24	17	10	5	1

10   CO1   [K₄]

b) From the following data find i) the two-regression equation ii) the coefficient of correlation between the marks in economics and statistics. iii)the most likely marks in statistics when marks in economics is 30.

10   CO3   [K₃]

Marks in eco X	25	28	35	32	31	36	29	38	34	32
Marks in stat Y	43	46	49	41	36	32	31	30	33	39

OR

6. a) Scenario: A farmer wishes to test the effect of 4 fertilizers A,B,C,D on the yield of wheat. The fertilizers are used in a LSD and the result are tabulated here. Perform an analysis of variance.

10   CO5   [K₄]

A18	C21	D25	B11
D22	B12	A15	C19
B15	A20	C23	D24
C22	D21	B10	A14

b)   A random sample is selected from each of 3 makes of ropes and their breaking
     strength are measured with the following results.

I	70	72	75	80	83		
II	100	110	108	112	113	120	107
III	60	65	57	84	87	73	

10   CO5   [K$_4$]

Test whether the breaking strength of the ropes differ significantly.

****