Register Number:……..

KUMARAGURU
college of technology
character is life

B.E/B.TECH DEGREE EXAMINATIONS: NOV/DEC 2023

(Regulation 2018)

Second Semester

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

U18MAI2203: Probability and Statistics

COURSE OUTCOMES

CO1: Understand and apply the concept of probability and random variables and predict probabilities of events in models following normal distribution.

CO2: Apply the concepts of two-dimensional random variables, central limit theorem and estimation, which lay the foundation for Machine Learning and Data Science.

CO3: Perform hypothesis testing and interpret the results which will form the basis for Data Analysis.

CO4: Understand the principles of design of experiments and perform analysis of variance which will help in Data Analysis.

CO5: Learn and apply multivariate analysis necessary for Principal Component Analysis.

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

Answer all the Questions:-

PART A (10 x 2 = 20 Marks)

(Answer not more than 40 words)

1. Find the probability chance that a leap year selected at random will contain 53 Sundays. CO1 [K2]

2. 6 coins are tossed 6400 times. Using Poisson distribution, what is the approximate probability of getting 6 heads 10 times. CO1 [K_3]

3. In a correlation analysis the equations of the two regression lines are $3x + 12y = 19, 3y + 9x = 46$. Find the value of correlation co-efficient. CO2 [K_3]

4. Write any two properties of regression lines. CO2 [K_1]

5. Define Type-I and Type-II Error. CO3 [K_2]

6. A sample of 200 persons with a particular disease was selected. Out of these, 100 were given a drug and the others were not given any drug. The results are as follows: CO3 [K_3]

No. of persons	Drug	No drug
Cured	65	55
Not cured	35	45

Find the Chi-square value of the given data.

7. State the basic principles of experimental designs. CO4 [K_1]

8. Find the values of A and B. CO4 [K_3]

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F Ratio
Between Treatments	2	R_1 =5838.44	A	B
Error	15	R_2 =1126	75.067	
Total				

9. Define Random vectors and Random Matrices. CO5 [K_1]

10. Find the mean vector $E(x_1), E(x_2)$. CO5 [K_2]

$x_1 \backslash x_2$	0	1
-1	0.24	0.06
0	0.16	0.14
1	0.40	0.00

Answer any FIVE Questions:-
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11. a) In a bolt factory, machines A, B and C manufacture respectively 25%, 35% and 40% of the total. Of their output 5, 4, 2 percent are defective bolts. A bolt is drawn at random from the product and is found to be defective. What are the probabilities that it was manufactured by machines A, B and C? 8 CO1 [K_3]

b) Out of 800 families with 4 children each, how many families would be expected to have (i)2 boys and 2 girls (ii) at least 1 boy (iii) at-most 2 girls. Assume equal probabilities for boys and girls. 8 CO1 [K_3]

12. a) The joint probability density function of the random variable (x, y) is given by $f(x, y) = k\, xy\, e^{-(x^2+y^2)}, x > 0, y > 0$. 8 CO2 [$K_2$]
(i) Find k.
(ii) Check x and y are independent.

b) Find Karl Pearson' correlation coefficient for the following heights in inches of fathers (x) and their sons (y): 8 CO2 [K_2]

X:	65	66	67	67	68	69	70	72
Y:	67	68	65	68	72	72	69	71

13. a) Two random samples gave the following results: 10 CO3 [K_3]

Sample	Size	Sample mean	Sum of the square of deviations from the mean

1	10	15	90
2	12	14	108

Examine whether the samples come from the same normal population at 5% level of significance.

b) Two researchers adopted different sampling techniques while investigating the same group of students to find the number of students falling into different intelligence level. The results are as follows: 6 CO3 [K3]

Research	Below average	Average	Above average	Excellent	Total
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	300

Would you say that the sampling techniques adopted by the two researchers are significantly different?

14. A completely Randomized Design experiment with ten plots and three treatments gave the results given below. 16 CO4 [K4]

Plot No	1	2	3	4	5	6	7	8	9	10
Treatment	A	B	C	A	C	C	A	B	A	B
Yield	5	4	3	7	5	1	3	4	1	7

Analyze the results for the effects of treatments.

15. a) Compute the principal component analysis to the following co variance matrix 8 CO5 [K4]

$\Sigma = \begin{bmatrix} 1 & 4 \\ 4 & 100 \end{bmatrix}$.

b) Let X_1 and X_2 have the joint pmf 8 CO5 [K4]

$P(x_1, x_2) = \frac{x_1 + 2x_2}{18}, x_1 = 1,2; x_2 = 1,2$.

Find (i) Marginal pmf 's of x_1 and x_2. (ii) Mean Vector (iii) Variance- covariance matrix .

16. a) A discrete random variable has the following probability distribution 8 CO1 [K3]

x:	0	1	2	3	4	5	6	7	8
$P(x)$:	a	$3a$	$5a$	$7a$	$9a$	$11a$	$13a$	$15a$	$17a$

Find (i) the value of a (ii) $P(2 \leq X < 6)$ (iii)$P(X > 3)$ (iv) distribution function of X.

b) A simple sample of heights of 6400 Englishmen has a mean of 170 cm and a S.D. of 6.4 cm, while a simple sample of heights of 1600 Americans has a mean of 8 CO4 [K3]

172 cm and a S.D. of 6.3 cm. Do the data indicate that Americans are, on the average, taller than Englishmen?

****