



**B.E/B.TECH DEGREE EXAMINATIONS: APRIL/ MAY 2024**

(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:-**

Statistical Tables are permitted

**PART A (10 x 2 = 20 Marks)**

**(Answer not more than 40 words)**

- |     |  |     |                   |
|-----|--|-----|-------------------|
| 1.  | A coin is tossed 3 times. What is the probability of getting exactly 2 heads?  | CO1 | [K <sub>3</sub> ] |
| 2.  | A normal distribution has a mean of 50 and a standard deviation of 10. What is the probability that a value is less than 40?   | CO1 | [K <sub>3</sub> ] |
| 3.  | State Central limit theorem.   | CO2 | [K <sub>3</sub> ] |
| 4.  | Given the regression equation $Y = 3X + 5$ , calculate Y when $X = 4$ .  | CO2 | [K <sub>3</sub> ] |
| 5.  | Define Type I and Type II errors in testing of hypothesis.   | CO3 | [K <sub>1</sub> ] |
| 6.  | A sample of size 100 is taken from a population with mean 50 and variance 25. What is the standard error of the mean?  | CO3 | [K <sub>3</sub> ] |
| 7.  | In an ANOVA test, the sum of squares between groups (SSB) is 30 and the sum of squares within groups (SSW) is 70. Calculate the F-ratio if there are 3 groups with 10 observations each. | CO4 | [K <sub>3</sub> ] |
| 8.  | Distinguish between CRD and RBD.   | CO4 | [K <sub>1</sub> ] |
| 9.  | Define random vector and random matrix   | CO5 | [K <sub>1</sub> ] |
| 10. | Given Covariance matrix, $\Sigma = \begin{pmatrix} 4 & 6 \\ 6 & 9 \end{pmatrix}$ . Compute standard deviation matrix   | CO5 | [K <sub>3</sub> ] |

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

11. Scenario: A manufacturer produces a large batch of resistors with an average resistance of 100 ohms and a standard deviation of 5 ohms.
- a) What is the probability that a randomly selected resistor has a resistance between 95 and 105 ohms? Using Normal distribution 7 CO1 [K<sub>3</sub>]
- b) If a sample of 50 resistors is taken, what is the probability that the sample mean resistance is between 99 and 101 ohms? Using central limit theorem 7 CO1 [K<sub>3</sub>]
- c) Discuss how this information could be used in quality control 2 CO1 [K<sub>2</sub>]
12. Scenario: A retailer analyzes monthly sales data for two products over a year. The monthly sales (in units) are:
- Product A: 20, 22, 21, 23, 20, 19, 24, 25, 22, 21, 20, 23
  - Product B: 18, 19, 20, 18, 17, 19, 20, 21, 19, 18, 20, 22
- a) Calculate the mean and standard deviation of monthly sales for each product. 7 CO3 [K<sub>3</sub>]
- b) Perform a t-test to determine if there is a significant difference in the average sales of the two products. 7 CO3 [K<sub>3</sub>]
- c) Discuss the implications of the t-test results for inventory management. 2 CO3 [K<sub>3</sub>]
13. a) Perform a correlation analysis on the following data: 7 CO2 [K<sub>4</sub>]
- |    |   |   |   |   |   |
|----|---|---|---|---|---|
| X: | 1 | 2 | 3 | 4 | 5 |
| Y: | 2 | 4 | 5 | 4 | 5 |
- Calculate the correlation coefficient.
- b) For the above data: Determine two regression lines Y on X and X on Y. 7 CO2 [K<sub>3</sub>]
- c) Discuss the limitations of correlation analysis. 2 CO2 [K<sub>4</sub>]
14. a) The joint probability function (X, Y) is given by 7 CO2 [K<sub>4</sub>]  
 $p(x, y) = k(2x + 3y)$ ,  $x = 0, 1, 2$ ;  $y = 1, 2, 3$ , find the value of k
- (i) Obtain the marginal distributions.
- (ii) Determine the probability distribution of (X+Y).

- b) Perform a chi-square test for independence using the following data: 7 CO3 [K<sub>3</sub>]
- Observed frequencies for two categories (A and B)  
A: 50, 30, 20  
B: 40, 35, 25
  - Calculate the chi-square statistic and determine if the categories are independent at 5% significance level.
- c) Discuss the importance of the chi-square test in statistical analysis. 2 CO3 [K<sub>3</sub>]
15. a) A manufacturing company is testing the effect of four different machines (M1, M2, M3, and M4) on the production time of a specific part. The company uses a Randomized Block Design (RBD) to account for variability among operators. Four operators are selected, and each operator uses all four machines, but in a different random order. The production time (in minutes) is recorded for each operator and each machine. The data is shown in the table below: 7 CO4 [K<sub>4</sub>]

Operator	Machine M1	Machine M2	Machine M3	Machine M4
1	12	15	14	13
2	16	18	17	15
3	14	16	15	14
4	15	17	16	15

State the hypotheses for the Randomized Block Design analysis and perform the ANOVA test at a 5% significance level.

- b) Calculate the sum of squares for treatments, blocks, and error. 7 CO4 [K<sub>4</sub>]
- c) Explain the advantage of using a Randomized Block Design in this experiment compared to a completely randomized design. 2 CO4 [K<sub>4</sub>]
16. a) If the covariance matrix of  $X = [X_1, X_2]^T$  is  $\Sigma = \begin{pmatrix} 1 & 2 \\ 2 & 16 \end{pmatrix}$ . Obtain the correlation coefficient matrix and find its eigen values and eigen vectors. 7 CO5 [K<sub>3</sub>]
- b) Determine the principal components of  $X_1, X_2$  using the standardized variables 7 CO5 [K<sub>3</sub>]
- c) Explain why principal components are often computed from standardized variable rather than original variables. 2 CO5 [K<sub>2</sub>]

\*\*\*\*\*