

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2024		
M. Tech.) Program: Artificial Intelligence Scheme I/II/IIB/III: II		
Regular Examination: FY Semester: I		
Course Code: PCEC103 and Course Name: Mathematical Foundations of Data Science		
Date of Exam: 13-01-2025	Duration: 2.5 Hours	Max. Marks: 60

Instructions:

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data, if necessary.

(5) Assume suitable data, if necessary.

Qu. No.	Question	Max. Marks	CO	BT level												
Qu-1	Solve any TWO questions out of three: (05 marks each)	10														
a)	Explain the difference between discrete and continuous random variables. If X is a random variable with $P(X=x)=x/15$ for $x=1,2,3,4,5$. find $P(X>2)$ and the expected value of X.	5	CO1	2												
b)	<p>A company tracks the number of advertising campaigns (A) and the corresponding monthly sales (S) in thousands of rupees for five months:</p> <table border="1"> <thead> <tr> <th>Advertising Campaigns (A)</th> <th>Monthly Sales (S)</th> </tr> </thead> <tbody> <tr> <td>3</td> <td>25</td> </tr> <tr> <td>5</td> <td>35</td> </tr> <tr> <td>6</td> <td>45</td> </tr> <tr> <td>8</td> <td>55</td> </tr> <tr> <td>9</td> <td>60</td> </tr> </tbody> </table> <p>i. Calculate the correlation coefficient between the number of advertising campaigns and monthly sales. ii. Interpret the strength and direction of the relationship between the two variables.</p>	Advertising Campaigns (A)	Monthly Sales (S)	3	25	5	35	6	45	8	55	9	60	5	CO3	3
Advertising Campaigns (A)	Monthly Sales (S)															
3	25															
5	35															
6	45															
8	55															
9	60															
c)	What is a confidence interval, and how is it interpreted in statistical analysis?	5	CO5	2												
Qu-2	Solve any TWO questions out of three: (05 marks each)	10														
a)	Explain the concept of Singular Value Decomposition (SVD) and describe its role in dimensionality reduction.	5	CO2	2												
b)	State the Central Limit Theorem (CLT). Discuss its significance in data science and demonstrate with an example where sampling is used to estimate the mean of a population.	5	CO4	2												
c)	How can Lift be used to identify and quantify the strength of correlation between two variables, and how does this differ from traditional correlation measures?	5	CO6	2												

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2024		
M. Tech.) Program: Artificial Intelligence Scheme I/II/IIB/III: II		
Regular Examination: FY Semester: I		
Course Code: PCEC103 and Course Name: Mathematical Foundations of Data Science		
Date of Exam: 13-01-2025	Duration: 2.5 Hours	Max. Marks: 60

Qu-3	Solve any TWO questions out of three. (10 marks each)	20		
a)	A factory has two machines, A and B. Machine A produces 60% of the items, while machine B produces 40%. 5% of the items from A are defective, and 10% from B are defective. If an item is selected at random and is defective, what is the probability it was produced by Machine A? (Use Bayes' theorem and show all steps).	10	CO1	3
b)	i. What do you mean by central tendency? What is the importance of measuring central tendency? ii. The mean age of a combined group of men and women is 30 years. If the mean age of the group of men is 32 and that of the group of women is 27 find out the percentage of the men and women in the group.	10	CO3	3
c)	i. Why is the Z test usually inappropriate as a test procedure when the sample size is small? ii. What assumptions are made when a Student's t test is employed to test a hypothesis involving a population mean?	10	CO5	2
Qu-4	Solve any TWO questions out of three. (10 marks each)	20		
a)	i) Determine the rank of matrix $A = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & -1 \\ 3 & 1 & -1 \end{bmatrix}$ ii) Find NULL space of the given matrix $A = \begin{bmatrix} -1 & 3 & 2 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$	10	CO2	3
b)	State and explain the key assumptions of the Binomial Distribution and use the same to solve the following problem: A fair coin is flipped 10 times. Let X represent the number of heads obtained. Assume X follows a binomial distribution. i. What is the probability of getting exactly 6 heads? ii. What is the probability of getting at most 4 heads?	10	CO4	3
c)	Explain the concepts of Precision, Recall, and Specificity in the context of a confusion matrix. How are these metrics calculated, and what do they indicate about the performance of a classification model? Provide an example of a binary classification problem and discuss how a trade-off between Precision and Recall might occur.	10	CO6	2
