Reg. No.															

Question Paper Code	14756

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2026

Third Semester

CSE with Artificial Intelligence and Machine Learning

(Common to Artificial Intelligence and Data Science)

24BSMA303 - PROBABILITY AND STATISTICAL MODELING

Regulations - 2024

(Use of *Statistical Tables* is permitted)

Duration: 3 Hours Max. Marks: 100

PART – A (i) (MCQ)
(16 × 1 = 16 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	The conditional probability of an event A, assuming that the event B has happened $(P(B) \neq 0)$, is defined as a) $P(A/B) = \frac{P(A \cup B)}{P(B)}$ b) $P(A/B) = \frac{P(A \cap B)}{P(B)}$ c) $P(A/B) = \frac{P(A \cap B)}{P(A)}$ d) $P(A/B) = \frac{P(A \cup B)}{P(A)}$	1	K1	CO1
2.	If f(x) is a probability density function of a random variable 'X' which is defined in the interval (a, b) then Mean is a) $\int_a^b f(x)dx$ b) $\int_a^b x^2 f(x)dx$ c) $\int_a^b xf(x)dx$ d) $\int_a^b \frac{1}{x} f(x)dx$	1	K2	CO1
3.	Mean of the Binomial distribution is a) npq b) np c) q/p d) q/p^2	1	K3	CO1
4.	Marginal density function of 'X' is $f_X(x) =$ a) $\int_{-\infty}^{\infty} f(x,y)dx$ b) $\int_{-\infty}^{\infty} f(x,y)dy$ c) $\int_{-\infty}^{\infty} f(x)dx$ d) $\int_{-\infty}^{\infty} f(y)dy$	1	K1	CO2
5.	If the random variable X and Y are independent, then a) cov(x,y) = 0 b) cov(x,y) =1 c) cov(x,y) = -1 d) cov(x,y) = a	1	K2	CO2
6.	Correlation coefficient between two random variables 'X' and 'Y' is r(x,y) = a) $\frac{\sigma_x^2 \sigma_y^2}{\text{cov}(x,y)}$ b) $\frac{\text{cov}(x,y)}{\sigma_x^2 \sigma_y^2}$ c) $\frac{\sigma_x \sigma_y}{\text{cov}(x,y)}$ d) $\frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$	1	K1	CO2

7.	The multiple correlation coefficient lies between a) 0 and 1 b) -1 and 0 c) -1 and 1 d) -2 and 2	1	K2	CO3
8.	If calculated F > tabulated F, then a) reject H_0 b) accept H_0 c) reject H_1 d) none of these	1	K2	CO3
9.	The test which is used to determine whether the two samples have been drawn from the same population or not is a) Kolmogorov-Smirnov Test b) One Sample Sign Test c) Kruskal-Wallis Test d) Mann-Whitney Test	1	K3	CO3
10.	The distance between an estimate and the estimated parameter is called a) Sampling error b) Standard error c) Bias d) Error of estimation	1	K3	CO4
11.	Estimation is possible only in case of a) Parameter b) Sample c) Random Sample d) Population	1	K3	CO4
12.	The process of using sample data to estimate the values of unknown population parameter is called a) Estimate b) Estimator c) Estimation d) Interval Estimation	1	K2	CO4
13.	The point estimator of population mean μ is a) Sample mean b) Sample Variance c) Sample Standard deviation d) Sample size	1	K3	CO5
14.	In an ARIMA(1,1,1) model, the first '1' represent a) Number of moving average terms b) Order of differencing c) Constant term d) Number of autoregressive terms	1	K3	CO5
15.	In ARIMA, 'I' stand for a) Independence b) Integration c) Interaction d) Imputation	1	K1	CO6
16.	The methods is commonly used for time series forecasting is a) Moving averages b) Cluster analysis c) Regression analysis d) Monte Carlo simulation	1	K2	CO6

PART – A (ii) (Short Answer Questions)
(6 × 3 = 18 Marks)
Answer ALL Questions

17.	A card is drawn from a pack of 52 cards. Find the probability of drawing a red or a king card.	3	K3	CO1
18.	State Central limit theorem.	3	K4	CO2
19.	State Neyman Pearson Lemma. Write the procedures of testing.	3	K4	CO3
20.	Write any two advantages and disadvantages of non-parametric methods.	3	K4	CO4
21.	Find the Likelihood function for the distribution $f(x,\theta) = \begin{cases} \frac{1}{\theta}x^{\theta}, 0 < x < 1 \\ 0\,,\ otherwise \end{cases}$.	3	K3	CO5
22.	Write the expansion of ARIMA and define forecasting of time series.	3	K3	CO6

PART - B (Descriptive Answer Questions)
(6 × 11 = 66 Marks)

Answer ALL Questions

23. a) A random variable 'X' has the following probability distribution. *11 K3 CO1*

x	0	1	2	3	4	5	6	7
P(x)	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2+k$

(i) Find the value of 'k' (ii) Find P(X < 6), P(X ≥ 6) and P(0 < X < 5) (iii) Find the distribution function of X (iv) If P(X ≤ k) > ½ , find the minimum value of k.

OR

b) In a test of 2000 electric bulbs, it was found that the life of a particular make was normally distributed with an average life of 2040 hours and standard deviation of 60 hours. Estimate the number of bulbs likely to burn for (i) More than 2150 hours (ii) Less than 1950 hours (iii) More than 1920 hours and less than 2160 hours. *11 K3 CO1*

24. a) The joint distribution of X and Y is given by $f(x,y) = \frac{x+y}{15}$, x = 0,1,2; y = 1,2 Find (i) the marginal distributions of X and Y (ii) Conditional distribution of X given Y = 1 and (iii) Conditional distribution of Y given X = 2 (iv) Covariance (x,y) *11 K3 CO2*

OR

b) If the joint p.d.f of a two dimensional random variable (x ,y) is given by $f(x,y) = \frac{xy}{96}$, 0< x < 4, 1 < y < 5, find the covariance between X and Y. *11 K3 CO2*

25. a) Heights of sons and fathers are given in centimeters. *11 K4 CO3*

X : Height of son	62	64	65	69	70	71	72	74
Y : Height of father	126	125	139	145	165	152	180	178

(i) Find the coefficient of correlation (ii) Find the two lines of regression
(iii) Find the height of the son when the height of the father is 154 cm.

OR

b) Four farmers each used four types of manures for the crop and obtained the yield (in quintals) as below. *11 K4 CO3*

		Farmers			
		1	2	3	4
Manures	A	22	16	21	12
	B	23	17	19	13
	C	21	14	18	11
	D	22	15	19	10

Is there any significant difference between the i) farmers ii) manures

26. a) Two classes of students are tested using a certain competitive exam. The scores of a sample of students from each class is given below *11 K3 CO4*

Class A	45	44	47	48	55	53	55	63			
Class B:	65	67	77	65	56	67	78	55	66	65	58

Using Mann-Whitney U-test, test whether both classes have similar scholastic levels.

OR

b) The following data shows the knowledge score and the duration of membership (In years) of 7 farmers. Compute Kendall's τ and test its significance. *11 K3 CO4*

Knowledge score	20	5	30	25	15	10	35
Duration of Membership	4	2	5	6	1	3	7

27. a) A random sample X_1, X_2, X_3, X_4, X_5 of size 5 is drawn from a normal population with unknown mean μ. Consider the following estimators to estimate μ: (i) $t_1 = \frac{X_1+X_2+X_3+X_4+X_5}{5}$ (ii) $t_2 = \frac{X_1+X_2}{2} + X_3$. Examine whether t_1 and t_2 are unbiased estimator of μ. *11 K4 CO5*

OR

b) In a random sampling from normal population $N(\mu, \sigma^2)$, find the maximum likelihood estimators for (i) μ when σ^2 is known (ii) σ^2 when μ is known. *11 K4 CO5*

28. a) Consider the time series model $y_t = 0.7 + 0.4y_{t-1} + 0.12y_{t-2} + e^t$, where e^t is an white noise process with variance σ^2. *11 K3 CO6*

i) Identify the ARIMA Model.
ii) Determine y_t is stationary process or not.
iii) Find the mean of the time series.
iv) Find the Auto correlation ρ_1, ρ_2 *and* ρ_3.

OR

b) For each of the following models: *11 K3 CO6*

$X_t = 0.3X_{t-1} + Z_t$
$X_t = Z_t - 1.3Z_{t-1} + 0.4Z_{t-2}$
$X_t = 0.5X_{t-1} + Z_t - 1.3Z_{t-1} + 0.4Z_{t-2}$
Express the model using B notation and determine whether the model is stationary and/or invertible.