Reg. No.												

Question Paper Code	12092

27 JUL 2023

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2023

Third Semester

Computer Science and Business Systems

20BSMA305 - COMPUTATIONAL STATISTICS

(Regulations 2020)

Duration: 3 Hours Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

Marks, K-Level, CO

1. Find $\mu_x, \mu_y, \sigma_x, \sigma_y$ and ρ_{xy} for the following $\frac{1}{2\pi} e^{-\frac{1[x^2+y^2+4x-6y+13]}{2}}$ *2,K3,CO1*
2. Let X be the distribution as $N_3(\mu, \Sigma)$ when $\mu = (1, -1, 2)$ and $\Sigma \begin{pmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{pmatrix}$. Check whether X_1 and X_2 are independent. *2,K3,CO1*
3. Write the formula to find the residual of Multivariate linear regression. *2,K2,CO2*
4. Write the formula for Median and Inter Quartile Range in outlier test for odd number of observations. *2,K2,CO2*
5. Define Discriminant Analysis and its application. *2,K1,CO3*
6. Define $SSCP_T$. *2,K1,CO3*
7. Convert the covariance matrix $\Sigma = \begin{pmatrix} 1 & 4 \\ 4 & 100 \end{pmatrix}$ into correlation matrix. *2,K3,CO4*
8. What is the covariance structure for the orthogonal factor model? *2,K3,CO4*
9. Write the four properties of a Metric or distance measure. *2,K2,CO5*
10. What is non-hierarchical clustering? *2,K1,CO5*

PART - B (5 × 16 = 80 Marks)

Answer ALL Questions

11. a) (i) The life of a tube X_1 and X_2 are distributed as BND (2000, 0.1, 2500, 0.01, 0.87). If a filament diameter is 0.098, what is the probability that the tube will last 1950 hrs? *6,K3,CO1*

(ii) The co-variance matrix of a 3-dimensional vector X= (X_1, X_2, X_3) is given by $\Sigma = \begin{pmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{pmatrix}$. Determine the correlation matrix and correlation between X_1 and $\frac{X_2}{3} + \frac{X_3}{3}$. *10,K3,CO1*

OR

b) (i) Find the maximum likelihood estimates of 2×1 mean vector μ and 2×2 covariance matrix Σ based on the random sample $X = \begin{pmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 7 & 7 \end{pmatrix}$ from a bivariate normal population. *8,K3,CO1*

(ii) Determine the conditional distribution of X_1 given that $X_2 = x_2$ for the joint distribution in $\mu_1 = 0, \mu_2 = 2, {\sigma_{11}}^2 = 2,$ ${\sigma_{22}}^2 = 1$ and $\rho_{12} = 5$. *8,K3,CO1*

12. a) (i) Find the Regression Co-efficient for the following *8,K3,CO2*

Y_1	10	12	11
Y_2	100	110	105
X_1	9	8	7
X_2	62	58	64

(ii) Consider the data 2, 4, 6, 8, 10, 12, 14. Perform outlier test and find the outliers if any. *8,K3,CO2*

OR

b) (i) If the response variables take the value $Y = \begin{pmatrix} 10 & 100 \\ 12 & 110 \\ 11 & 105 \end{pmatrix}$ and the design matrix $X = \begin{pmatrix} 1 & 9 & 62 \\ 1 & 8 & 58 \\ 1 & 7 & 64 \end{pmatrix}$ takes the value. *8,K3,CO2*

1) Calculate $\widehat{\beta}$.
2) Calculate $\hat{\varepsilon}$.
3) Fit a linear regression model.

(ii) Compute the correlation matrix from the covariance matrix. *8,K3,CO2*

$$\Sigma = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{pmatrix}.$$

13. a) Construct a MANOVA table for the following *16,K3,CO3*

$$\begin{pmatrix} \binom{9}{3} & \binom{6}{2} & \binom{9}{7} \\ \binom{0}{4} & \binom{2}{0} & \\ \binom{3}{8} & \binom{1}{9} & \binom{2}{7} \end{pmatrix}.$$

OR

b) Compute the Linear discriminant projection for the following set of data: $X_1 = \{(4,1), (2,4), (2,3), (3,6), (4,4)\}$
$X_2 = \{(9,10), (6,8), (9,5), (8,7), (10,8)\}$. *16,K3,CO3*

K1 – Remember; K2 – Understand; K3 – Apply; K4 – Analyze; K5 – Evaluate; K6 – Create

14. a) Find the Standardized Random variable Z_1, Z_2 and Z_3 generatd by *m-1* factor model for the covariance matrix $\Sigma = \begin{pmatrix} 1 & 0.63 & 0.45 \\ 0.63 & 1 & 0.35 \\ 0.45 & 0.35 & 1 \end{pmatrix}$. Also find the correlation matrix. *16,K3,CO4*

OR

b) Calculate the Principle components Y_1, Y_2 & Y_3 for the covariance matrix $\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Also find $\text{Var}(Y_1)$, $\text{Cov}(Y_1, Y_2)$ & show that $\sum_{i=1}^{3} \sigma_{ii} = \sum_{i=1}^{3} \lambda_i$. *16,K3,CO4*

15. a) Find the clusters using Single Linkage procedure. Use Euclidean distance and draw the dendogram. *16,K3,CO5*

Points	A	B	C	D	E
X	2	6	2	2	5
Y	5	5	4	2	4

OR

b) Suppose we measure two variables X_1, X_2 for each of four items A, B, C, D. *16,K3,CO5*

	X_1	X_2
A	9	7
B	3	5
C	5	2
D	1	2

Use k-means clustering technique to divide the items into $k = 2$ clusters.