Reg. No.												

Question Paper Code	12453

B.E. / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2023

Third Semester

Computer Science and Business Systems

20BSMA305 – COMPUTATIONAL STATISTICS

(Regulations 2020)

(Use of Statistical Table is permitted)

Duration: 3 Hours — Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

		Marks, K-Level, CO
1.	Find the distribution of X_2 given $X_1 = x_1$ and $X_3 = x_3$, when $\mu = (-3,1,4)$ and $\Sigma \begin{pmatrix} 5 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$.	2,K2,CO1
2.	Let $X \sim N_5(\mu, \Sigma)$, find the distribution of $\begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$.	2,K2,CO1
3.	Write the formula for the regression coefficient $\hat{\beta}$.	2,K1,CO2
4.	Write the formula for Median and Inter Quartile Range in outlier test for odd no of observations.	2,K2,CO2
5.	Write the formula for Wilk's Lambda.	2,K1,CO3
6.	Define Fisher's Discriminant projection for two populations.	2,K1,CO3
7.	Define Principal Component Analysis.	2,K1,CO4
8.	What is a Residual matrix?	2,K1,CO4
9.	Define Cluster Analysis.	2,K1,CO5
10.	What is centroid linkage?	2,K1,CO5

PART - B (5 × 16 = 80 Marks)

Answer ALL Questions

11. a) (i) The random variables X and Y are described by a joint PDF of the form $f_{X,Y}(x,y) = c\, e^{-8x^2-6xy-18y^2}$. Evaluate the means, variances and the correlation coefficient of X and Y. Also find the value of the constant C. *8, K3,CO1*

(ii) If $X \sim N_3(\mu, \Sigma)$, $\mu = \begin{pmatrix} 5 \\ 3 \\ 7 \end{pmatrix}$ *and* $\Sigma = \begin{pmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ 1 & 0 & 2 \end{pmatrix}$, then find *8, K3,CO1*

(a) $P(5X_2 + 4X_3 > 70)$,

(b) $P(4X_1 - 3X_2 + 5X_3 < 80)$.

OR

b) Let $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ be a normal random vector with the following mean vector and covariance matrix $\mu = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $C = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$.Let also $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{pmatrix}$, $b = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$, $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = AX + b$ *16,K3,CO1*

(i) Find $P(0 \le X_2 \le 1)$

(ii) Find the expected value vector of Y, $\mu_y = E_y$

(iii)Find the covariance matrix of Y.

(iv)Find $P(Y_3 \le 4)$.

12. a) Given the data: *16,K3,CO2*

	0	1	2	3	4
Y_1	1	4	3	8	9
Y_2	-1	-1	2	3	2

(i) Fit the regression model.

(ii) Verify sum of the squared and cross product decomposition.

OR

b) (i) Compute the correlation matrix from the covariance matrix *8,K3,CO2*

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

(ii) 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*

(a) Calculate $\hat{\beta}$

(b) Calculate $\hat{\varepsilon}$

(c) Fit a linear regression model.

13. a) The School of business use GPA and GMAT score to decide admitting (π_1) or rejection (π_2) or to put in a borderline (π_3) a particular student. Let $\overline{X_1} = \begin{pmatrix} 3.40 \\ 561.23 \end{pmatrix}$,$\overline{X_2} = \begin{pmatrix} 2.48 \\ 447.07 \end{pmatrix}$ and $\overline{X_3} = \begin{pmatrix} 2.99 \\ 446.23 \end{pmatrix}$ are the mean values of the samples from population π_1, π_2 and π_3 respectively also given that $n_1 = 31, n_2 = 28, n_3 = 26$. Spooled $= \begin{pmatrix} 0.0361 & -2.0188 \\ -2.0188 & 3655.9011 \end{pmatrix}$. Based on the given data allocate a student with GPA = 3.21 and GMAT = 497 to any one of the population by using sample squared distance. *16,K3,CO3*

OR

b) Construct a MANOVA table for the following 16,K3,CO3

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

14. a) 16,K3,CO4

If $\Sigma = \begin{pmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{pmatrix}$ is a population covariance matrix and $L = \begin{pmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{pmatrix}$ is a factor loaded matrix then find its specific variance matrix Ψ and the communality of the variables X_1, X_2, X_3 and X_4.

OR

b) Find the Principle Components for the following sample data: $(2,1)(3,5)(4,3)(5,6)(6,7)(7,8)$ and derive the new observation obtained by the Principal Component. 16,K3,CO4

15. a) 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.

OR

b) Consider the matrix of distance 16,K3,CO5

A B C D E F

$$D = \begin{pmatrix} 0 & & & & & \\ 5 & 0 & & & & \\ 14 & 9 & 0 & & & \\ 12 & 20 & 13 & 0 & & \\ 18 & 16 & 7 & 3 & 0 & \\ 10 & 16 & 8 & 10 & 12 & 0 \end{pmatrix}$$

. Clusters the five items using the Average and Complete linkage hierarchical procedure. Also draw the dendrograms.