Reg. No.																

Question Paper Code	12906

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2024

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

	Question	*Marks*	*K – Level*	*CO*
1.	Let X be the distribution as $N_3(\mu,\sum)$ when $\mu=(1,-1,2)$ and $\sum\begin{pmatrix}4 & 0 & -1\\ 0 & 5 & 0\\ -1 & 0 & 2\end{pmatrix}$. Check whether X_1 and X_2 are independent?	*2*	*K2*	*CO1*
2.	Determine $\mu_x, \mu_y, \sigma_x, \sigma_y$ and ρ_{xy} for the following $\frac{1}{2\pi}e^{-\frac{1[2x^2+y^2+2xy-22x-14y+65]}{2}}$	*2*	*K2*	*CO1*
3.	Define Multicollinearity and the method used to detect it.	*2*	*K1*	*CO2*
4.	Write the formula to find the residual of Multivariate linear regression?	*2*	*K1*	*CO2*
5.	Define Discriminant Analysis and its application.	*2*	*K1*	*CO3*
6.	What is the test statistic for the variables $p=2$ and population $g\geq 2$ in MANOVA?	*2*	*K2*	*CO3*
7.	What is the covariance structure for the orthogonal factor model?	*2*	*K1*	*CO4*
8.	Define specific variance.	*2*	*K1*	*CO4*
9.	What is hierarchical clustering?	*2*	*K1*	*CO5*
10.	What is Average linkage?	*2*	*K1*	*CO5*

PART - B (5 × 16 = 80 Marks)

Answer ALL Questions

11\. a) i) If X ~ $N_3(\mu,\sum)$, $\mu=\begin{pmatrix}5\\3\\7\end{pmatrix}$ *and* $\sum=\begin{pmatrix}4 & 0 & -1\\ 0 & 5 & 0\\ 1 & 0 & 2\end{pmatrix}$, then find *8 K3 CO1*

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

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

ii) Let $X\sim N_4(\mu,\sum)$, $\mu=\begin{pmatrix}5\\6\\7\\8\end{pmatrix}$ and $\sum=\begin{pmatrix}2 & 0 & 1 & 0\\ 0 & 3 & 2 & 0\\ 1 & 2 & 4 & 0\\ 0 & 0 & 0 & 9\end{pmatrix}$. *8 K3 CO1*

a) Find the distribution of $\begin{pmatrix}x_2\\x_4\end{pmatrix}$.

b) Find the distribution of x_1-x_4.

OR

b) i) Consider a bivariate normal distribution with $\mu_1 = 1,\ \mu_2 = 3, \sigma_{11} = 2, \sigma_{22} = 1$ and $\rho_{12} = -0.8$ — *8 K3 CO1*

a) Write bivariate normal density.

b) Write the squared statistical distance expression.

ii) Find the maximum likelihood estimates of 2 x 1 mean vector μ and 2 x 2 covariance matrix Σ based on the random sample — *8 K3 CO1*

$X = \begin{pmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 7 & 7 \end{pmatrix}$ From a bivariate normal population.

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 15, 87, 92, 95, 96, 98, 99, 99, 100, 302. Perform outlier test and find the outliers if any. — *8 K3 CO2*

OR

b) Given the mean vector and covariance matrix of Y,X_1,X_2 — *16 K3 CO2*

$\mu = \begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 10 & 1 & -1 \\ 1 & 7 & 3 \\ -1 & 3 & 2 \end{pmatrix}.$

Determine the following,

(i) The best linear predictor $\beta_0 + \beta_1 X_1 + \beta_1 X_2$

(ii) Mean square Error

(iii) Multiple correlation coefficient

(iv) Also verify that mean square error equals $\sigma_{yy}(1-\rho_y^2(\mathrm{x}))$.

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

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

OR

b) Calculate the Linear Discriminant score to classify the observation x_0. — *16 K3 CO3*

$\pi_1 : X_1 = \begin{pmatrix} -2 & 5 \\ 0 & 3 \\ -1 & 1 \end{pmatrix}\ n_1 = 3, \quad \overline{X_1} = \begin{pmatrix} -1 \\ 3 \end{pmatrix},\ S_1 = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}$

$\pi_2 : X_2 = \begin{pmatrix} 0 & 6 \\ 2 & 4 \\ 1 & 2 \end{pmatrix}\ n_2 = 3, \quad \overline{X_2} = \begin{pmatrix} 1 \\ 4 \end{pmatrix},\ S_2 = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}$

$$\pi_3: X_3 = \begin{pmatrix} 1 & -2 \\ 0 & 0 \\ -1 & -4 \end{pmatrix} \; n_3 = 3, \quad \overline{X_3} = \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \; S_3 = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}$$

Given $p_1 = p_2 = 0.25$ and $p_3 = 0.50$. $X_0' = (-2, -1)$.

14. a) i) Determine the population Principal Components Y_1 and Y_2 for the covariance matrix $\Sigma = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}$. Also calculate the proportion of the total population variance explained by the first Principle Component. *8 K3 CO4*

ii) Find the Principal Components for the following sample data: *8 K3 CO4*

Features	s_1	s_2	s_3	s_4
X_1	4	8	13	7
X_2	11	4	5	14

OR

b) In a consumer preference study, a random sample of customers were asked to rate several attributes of a new product *16 K3 CO4*

$$R = \begin{pmatrix} 1 & 0.02 & 0.96 & 0.42 & 0.01 \\ 0.02 & 1 & 0.13 & 0.71 & 0.85 \\ 0.96 & 0.13 & 1 & 0.5 & 0.11 \\ 0.42 & 0.71 & 0.5 & 1 & 0.79 \\ 0.01 & 0.85 & 0.11 & 0.79 & 1 \end{pmatrix}$$ is the correlation matrix.

The eigen values more than unity are 2.85 and 1.81.

$F_1 = (0.56 \;\; 0.78 \;\; 0.65 \;\; 0.94 \;\; 0.80)^T$

$F_2 = (0.82 \;\; -0.53 \;\; 0.75 \;\; -0.10 \;\; -0.54)^T$ are the loading factors.

Find the residual matrix.

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) 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