BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: IMSC | **SEMESTER: V/ADD**
BRANCH: CQEDS | **SESSION: MO/2025**

SUBJECT: ED303 MULTIVARIATE DATA ANALYSIS

TIME: 3 Hours | **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions, each of 10 marks, and a total of 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. Before attempting the question paper, be sure that you have got the correct question paper.
5. Tables/Data handbook/Graph paper, etc., to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	Find the maximum likelihood estimators of the 2x1 mean vector μ and the variance-covariance matrix Σ based on the random sample $X = \begin{bmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 4 & 7 \end{bmatrix}$.	[5]	1	2
Q.1(b)	Let $X = (X1, X2, X3)^T$ be a 3-dimensional random vector having a multivariate normal distribution with mean vector $(0,0,0)^T$ and covariance matrix $\Sigma = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 4 \end{pmatrix}$. Let $\alpha^T = (2, 0, -1)$ and $\beta^T = (1, 1, 1)$. Find the E(trace($XX^T\alpha\alpha^T$)) and Var(trace($X\alpha^T$)).	[5]	2	4
Q.2(a)	Evaluate the value of the T^2 statistics and test the hypothesis H_0: $\mu = \begin{bmatrix} 7 \\ 11 \end{bmatrix}$ vs H_1: $\mu \neq \begin{bmatrix} 7 \\ 11 \end{bmatrix}$ using the sample data $X = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix}$, at 5% level of significance. Use $F_{2,2} = 19.00$.	[5]	2	3
Q.2(b)	Discuss Bonferroni simultaneous confidence intervals for the mean vector.	[5]	3	3
Q.3(a)	Let $f_1(x) = \frac{1}{2}(1-\mid x \mid)$ for $\mid x \mid \leq 1$ and $f_2(x) = \frac{1}{4}(2-\mid x - 0.5 \mid)$ for $-1.5 \leq x \leq 2.5$ probability density functions for two populations. Suppose we have a new observation, say Z, which actually belongs to one of these two populations, but the exact location is unknown. First, (a) Sketch the two densities, and then (b)Obtain the Classification regions and the rule when both populations are equally probable and the costs of misclassifications are equal.	[6]	3	4
Q.3(b)	For a classification problem, define the following terms (i) Costs of Misclassification (ii) Expected cost of misclassification (iii) Actual error rate (AER) (iv) Apparent error rate (APER)	[4]	3	3

PTO

Q.4(a) Municipal wastewater treatment plants must regularly monitor their effluent. To examine whether two laboratories provide consistent chemical analyses, samples of effluent were split and sent to a commercial laboratory and to the State Laboratory of Hygiene. Measurements of biochemical oxygen demand (BOD) and suspended solids (SS) were obtained for *n* = *5* sample splits. The data are shown below. [5] 4 3

Sample j	Commercial x_{1j1}(BOD)	Commercial x_{1j2}(SS)	State x_{2j1}(BOD)	State x_{2j2}(SS)
1	6	27	25	15
2	6	23	28	13
3	18	64	36	22
4	8	44	35	29
5	11	30	15	31

Do the two laboratories' mean chemical analyses (BOD and SS) agree? If differences exist, what is their nature? $F_{2,3}(0.05)$ =9.55, and $S_d = \begin{pmatrix} 73.50 & -50.25 \\ -50.25 & 254.30 \end{pmatrix}$.

Q.4(b) Explain the mechanism of MANOVA and its importance. [5] 4 3

Q.5(a) Find the principal components and the proportion of the total population variance explained by each when the covariance matrix is [6] 5 4

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & \sigma^2\rho & 0 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ 0 & \sigma^2\rho & \sigma^2 \end{bmatrix}, \quad -\frac{1}{\sqrt{2}} < \rho < \frac{1}{\sqrt{2}}$$

Q.5(b) The eigenvalues and eigenvectors of the correlation matrix ρ in are: [4] 5 3

$\lambda_1 = 1.96$, $e_1' = [0.625, 0.593, 0.507]$

$\lambda_2 = 0.68$, $e_2' = [-0.219, -0.491, 0.843]$

$\lambda_3 = 0.36$, $e_3' = [0.749, -0.638, -0.177]$

Assuming an m = 1 factor model, calculate the loading matrix L and the matrix of specific variances Ψ using the principal component solution method.

::::::20/11/2025::::::E