LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION – **STATISTICS**

THIRD SEMESTER – **NOVEMBER 2023**

PST 3501 – MULTIVARIATE ANALYSIS

Date: 30-10-2023 Dept. No. Max. : 100 Marks
Time: 01:00 PM - 04:00 PM

PART – A

Answer ALL the questions **(10 x 2 = 20 marks)**

1. Provide any two applications of multivariate analysis
2. Define mean vector and covariance matrix
3. What is use of Hotelling T^2 statistic?
4. State the null and alternative hypotheses of MANOVA
5. Define singular value decomposition
6. Define Expected Cost of Misclassification
7. Provide any two use of principal component bi-plot
8. What is the use of varimax rotation?
9. State the goal of discriminant analysis
10. What is the use of Factor analysis biplot?

PART – B

Answer any FIVE questions **(5 x 8 = 40 marks)**

11. Obtain the characteristic function of Multivariate Normal distribution
12. Let $X^{(1)}$ and $X^{(2)}$ be qx1 and (p-q)x1 partitions of the random vector X where

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim N_p \left[\begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} , \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right]$$

    Show that $X^{(1)} \coprod X^{(2)} \;\; iff \;\; \Sigma_{12} = 0$

13. Discuss Principal Component Analysis in detail
14. Discuss MANOVA for comparing g population mean vectors in detail
15. State and establish Maximization of Quadratic forms for points on a unit sphere
16. Explain Orthogonal Factor Model and illustrate Varimax rotation with an example
17. a) If $X \sim N_p(\mu, \Sigma)$ $then\ show\ that$ $(X-\mu)'\Sigma^{-1}(X-\mu) \sim {\chi^2}_{(p)}$

    b) Discuss the method to detect outliers in multidimensional data using generalized squared distance.
18. Establish Cauchy-Schwarz Inequality for px1 vectors and use the result to establish extended Cauchy-Schwarz Inequality

PART – C

Answer any TWO questions **(2 x 20 = 40 marks)**

19. Determine a) Mean Vector b) Var-Cov matrix c) Correlation matrix and d) Multiple Correlation $R_{1.234}$ based on the data given below (2 + 10 + 5 + 3)

X_1 :	91	95	81	83	76	58	89	79	83	74
X_2 :	55	68	72	91	49	78	54	76	69	73
X_3 :	65	67	89	69	78	87	94	79	96	59
X_4 :	72	64	85	65	75	58	48	72	91	82

20. Determine the three principal component equations based on the var-cov matrix given below and also determine the proportion of variance explained by each principal component (6 +6 + 6 + 2)

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 8 & 2 \\ 0 & 2 & 8 \end{bmatrix}$$

21. Perform Hierarchical Clustering based on a) Single Linkage b) Complete Linkage and c) Average Linkage using the distance matrix given below and obtain the dendrogram based on the three methods (5 +5 +5 +5)

$$\boldsymbol{D} = \begin{bmatrix} 0 & & & & \\ 4 & 0 & & & \\ 7 & 11 & 0 & & \\ 9 & 5 & 9 & 0 & \\ 2 & 8 & 6 & 10 & 0 \end{bmatrix}$$

22. a) Determine Fisher's sample discriminant function for the three population based on the data given below (15 Marks)

$\pi_1 \ (n_1 = 3)$ $\quad \pi_2 \ (n_2 = 3)$ $\quad \pi_3 \ (n_3 = 3)$

$$\boldsymbol{X_1} = \begin{bmatrix} -2 & 5 \\ 0 & 3 \\ -1 & 1 \end{bmatrix} \qquad \boldsymbol{X_2} = \begin{bmatrix} 0 & 6 \\ 2 & 4 \\ 1 & 2 \end{bmatrix} \qquad \boldsymbol{X_3} = \begin{bmatrix} 1 & -2 \\ 0 & 0 \\ -1 & -4 \end{bmatrix}$$

b) Classify a new observation $\boldsymbol{x_0}' = [1 \quad 3]$ into one of the three populations using Fisher classification procedure (5 Marks)

&&&&&&&&&&