LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

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

THIRD SEMESTER – **APRIL 2023**

PST 3501 – MULTIVARIATE ANALYSIS

Date: 02-05-2023 Dept. No. Max. : 100 Marks
Time: 09:00 AM - 12:00 NOON

Section - A

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

1. Provide the density for Bi-variate Normal distribution.
2. Show by an example that the variance covariance matrix is positive definite.
3. Provide the formula to calculate multiple correlation coefficient from the values of sample correlation matrix.
4. What are Factor loadings and communality?
5. Define first principal component and kth principal component.
6. What is difference between orthogonal rotation and Non-orthogonal factor rotation?
7. What is DB scan? And state any two advantage of DB scan method?
8. What is difference between Agglomerative Hierarchical clustering and Divisive Hierarchical clustering methods?
9. State the Null and alternative hypothesis for MANOVA in comparing 'g' populations means and provide the expression to determine within group sum of squares and between group sum of squares.
10. Define expected cost of misclassification (ECM) with an example.

Section - B

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

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

    Then show that $X^{(1)} \sim N_q(\mu^{(1)}, \Sigma_{11})$ *and* $X^{(2)} \sim N_{P-q}(\mu^{(2)}, \Sigma_{22})$.

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. a) Let X be a p-variate random vector then prove that $X \sim N_p(\mu,\Sigma)$ if and only if every linear combination of $X_1, X_2, \ldots, X_p$ is normally distributed. (4)

    b) If $X \sim N_p(\mu,\Sigma)$ and D is of order qxp ($q \le p$) with rank q, then $DX \sim N_p(D\mu, D\Sigma D')$ (4)

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

    Then show that conditional distribution of

$$X^{(1)} \mid X^{(2)} = x^{(2)} \sim N_q\left(\mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}\left(x^{(2)} - \mu^{(2)}\right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

$$and\ X^{(2)} \mid X^{(1)} = x^{(1)} \sim N_{P-q}(\mu^{(2)}, \Sigma_{22})$$

15. Determine the three first order partial correlation coefficients and all possible multiple correlation coefficient based on the correlation matrix given below

$$\hat{\rho} = \begin{bmatrix} 1 & 0.63 & 0.45 \\ 0.63 & 1 & 0.35 \\ 0.45 & 0.35 & 1 \end{bmatrix}$$

16. State and establish Maximization of Quadratic forms for points on a unit sphere.
17. Discuss MANOVA for comparing 'g' population mean vectors in detail.

18. a) Discuss the steps involved in constructing principal component biplot and Factor Analysis biplot and state their uses. (5)
    b) Explain the Steps involved in K-Means clustering algorithm. (3)

Section - 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 :	83	71	65	54	69	37	21	79	53	55
X_2 :	59	48	72	70	46	38	60	59	65	71
X_3 :	67	52	86	65	72	85	92	77	94	60
X_4 :	73	67	83	63	73	56	45	70	89	84

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)

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

$$\boldsymbol{X_1} = \begin{bmatrix} -3 & 7 \\ 0 & 4 \\ -2 & 0 \end{bmatrix} \qquad \boldsymbol{X_2} = \begin{bmatrix} 1 & 8 \\ 3 & 5 \\ 2 & 3 \end{bmatrix} \qquad \boldsymbol{X_3} = \begin{bmatrix} 2 & -3 \\ 0 & 1 \\ -1 & -3 \end{bmatrix}$$

b) Classify a new observation $\boldsymbol{x_0}' = [2 \quad 4]$ into one of the three population using Fisher classification procedure. (5 Marks)

$$$$$$$