LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

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

THIRD SEMESTER – **NOVEMBER 2023**

PST3MC01 – MULTIVARIATE ANALYSIS

Date: 30-10-2023 Dept. No. Max. : 100 Marks

Time: 01:00 PM - 04:00 PM

SECTION A – K1 (CO1)

	Answer ALL the questions	**(5 x 1 = 5)**
1	**Define the following**	
a)	Bivariate Normal Distribution	
b)	Orthogonal Factor Model	
c)	Variance captured by kth principal component	
d)	Positive definite matrix with an example	
e)	Hierarchical cluster analysis	

SECTION A – K2 (CO1)

	Answer ALL the questions	**(5 x 1 = 5)**
2	**Fill in the blanks**	
a)	Variance-covariance matrix is a _______________ definite matrix	
b)	An observation which is neither a core point nor a neighbour point is refered to as _____________	
c)	Varimax rotation is an ______________ rotation	
d)	The proportion of variance captured by the underlying factors on a specific observed variable is referred to as _____________	
e)	_____________ is a diagrammatic representation of cluster formation in hierarchical clustering	

SECTION B – K3 (CO2)

	Answer any THREE of the following	**(3 x 10 = 30)**
3	Obtain Bivariate Normal distribution from multivariate normal density by substituting p=2	
4	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 $X^{(1)} \mid X^{(2)} = x^{(2)} \sim N_q\left(\mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}\left(x^{(2)} - \mu^{(2)}\right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$ $and\ X^{(2)} \mid X^{(1)} = x^{(1)} \sim N_{P-q}(\mu^{(2)}, \Sigma_{22})$	
5	If $X^{(1)} \sim N_{p_1}(\mu^{(1)}, \Sigma_{11})\ and\ X^{(2)} \sim N_{p_2}(\mu^{(2)}, \Sigma_{22})$ $X^{(1)} \coprod X^{(2)}$ then $\begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim N_{p_1+p_2} \left[\begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \right]$	
6	Obtain the Moment generating function of Multivariate Normal distribution	
7	Discuss Hotelling T^2 Statistic and Compute Hotelling T^2 Statistic for testing $H_0 : \mu = \begin{pmatrix} 3 \\ 8 \end{pmatrix}$	

	with n=3, p=2 and $X = \begin{bmatrix} 5 & 3 \\ 8 & 2 \\ 6 & 1 \end{bmatrix}$
	SECTION C – K4 (CO3)
	Answer any TWO of the following **(2 x 12.5 = 25)**
8	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. (5 Marks) b) If $X \sim N_p(\mu,\Sigma)$ and D is of order qxp ($q \leq p$) with rank q, then $DX \sim N_p(D\mu, D\Sigma$ (7.5 Marks)
9	Discuss MANOVA for comparing g population mean vectors in detail
10	State and establish maximization of quadratic forms for points on a unit sphere
11	a) Discuss varimax rotation and state its uses b) Discuss the method to detect outliers in multidimensional data using generalized squared distance.
	SECTION D – K5 (CO4)
	Answer any ONE of the following **(1 x 15 = 15)**
12	Define Hierarchical clustering and discuss Single Linkage, Complete Linkage, Average Linkage, Wards method of Hierarchical clustering and also discuss two methods to determine the optimal number of cluster.
13	a) Explain expected cost of misclassification for classifying two populations (4 Marks) b) Discuss the minimum ECM for two normal population with $\Sigma_1 = \Sigma_2 = \Sigma$ (4 Marks) c) Discuss Fisher's Method of discriminating among several populations (7 Marks)
	SECTION E – K6 (CO5)
	Answer any ONE of the following **(1 x 20 = 20)**
14	Determine the principal components based on the var-cov matrix given below and also determine the proportion of variance explained by each principal component (6 +6 + 6 + 2) $$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 8 & 2 \\ 0 & 2 & 8 \end{bmatrix}$$
15	Perform Hierarchical Clustering using the following linkage methods a) Single Linkage (6 Marks) b) Complete Linkage (6 Marks) c) Average Linkage (8 Marks) based on the distance matrix given below and obtain the corresponding dendrogram of each method. $$\boldsymbol{D} = \begin{bmatrix} 0 & & & & \\ 4 & 0 & & & \\ 7 & 11 & 0 & & \\ 9 & 5 & 9 & 0 & \\ 2 & 8 & 6 & 10 & 0 \end{bmatrix}$$

&&&&&&&&&&