LOYOLA COLLEGE (AUTONOMOUS) CHENNAI – 600 034

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

THIRD SEMESTER – **NOVEMBER 2024**

PST3MC01 MULTIVARIATE ANALYSIS

Date: 07-11-2024 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	**Fill in the blanks**	
a)	Function associates the ……. Value for every outcome of a random experiment.	
b)	The …….. measures the strength and direction of the relationship between two variables while controlling for the effect of one or more additional variables.	
c)	A key assumption of MANOVA is that the dependent variables should be …… distributed within each group.	
d)	Canonical correlation is a statistical technique used to explore the relationship between two sets of …… variables.	
e)	The …… method is a common approach in cluster analysis that groups data points based on the distance between them, often using Euclidean distance as a metric.	
	SECTION A – K2 (CO1)	
	Answer ALL the questions	**(5 x 1 = 5)**
2	**Define the following**	
a)	Probability distribution of a discrete random variable.	
b)	Generalized T^2 statistic	
c)	State any one of the Properties of Wishart distribution.	
d)	Factor rotation	
e)	What is the similarity coefficient when 0-0 matches totally ignored and double weight for unmatched pairs in denominator?	
	SECTION B – K3 (CO2)	
	Answer any THREE of the following	**(3 x 10 = 30)**
3	Compute the correlation matrix from the following covariance matrix: $\sum = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 5 \end{bmatrix}$	
4	Explain Hotelling T^2 and Mahalanobi's D^2 statistics.	
5	Discuss canonical correlation coefficient.	

6	What are the roles of Kaiser Meyer Olkin test and Bartlett's test of Sphericity in Factor analysis.
7	Explain any four distance measures for numeric data.

SECTION C – K4 (CO3)

	Answer any TWO of the following **(2 x 12.5 = 25)**
8	Explain expected cost of misclassification for classifying two populations.
9	Discuss the Partial correlation coefficient and Multiple correlation coefficient.
10	State the properties of Wishart Distribution, Repeated Measures Design and MANOVA.
11	Determine the optimum number of clusters (k) using Elbow method for the following data: X_1: 1 2 4 5, X_2: 1 1 3 4.

SECTION D – K5 (CO4)

	Answer any ONE of the following **(1 x 15 = 15)**
12	Find the mean vector covariance matrix and correlation matrix for two random variables x_1 and x_2, when their joint probability function is represented in the following table:

x1 \x2	0	1
-1	0.24	0.06
0	0.16	0.14
1	0.4	0

13	The maximum likelihood estimators μ and $\sum$ are $\hat{\mu} = \bar{X}$ and $\sum = \frac{1}{n}s$, $$where\ s = \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'$$

SECTION E – K6 (CO5)

	Answer any ONE of the following **(1 x 20 = 20)**
14	Determine the principal component and the proportion of variance explained by each component based on variance covariance matrix given below: $$\sum = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$
15	Construct the Dendogram using Single Linkage Method with Euclidean distance for the following data:

National track Record for Women			
Country	100m	200m	400m
Argentina	11.57	22.94	52.5
China	10.79	22.01	49.81
India	11.56	23.86	55.08
USA	10.49	21.34	48.83
Russia	10.77	21.8	49.11

$$$$$$$$$$$$$$$