M.Sc. Examination, 2018
Semester-II
Statistics
Course : MSC-22
(Applied Multivariate Analysis)

Time : 3 Hours **Full Marks : 40**

Questions are of value as indicated in the margin

Answer **any four** questions

1. (a) Assume that you wish to analyse a data matrix consisting of p orthogonal, standardized and centered columns with n observations. What would be the percentage of the variance explained by the first principal component? What percentage of variance would be explained by the first q principal components? 5

   (b) Apply principal component analysis (PCA) to $\sum = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $0 < \rho < 1$ Give the percentage of variance to be explained by the first PC. Also write down the first PC. 5

2. (a) Suppose that the data matrix consists of two columns, $\underset{\sim}{x}_1$ and $\underset{\sim}{x}_2$ and that $\underset{\sim}{x}_2 = 2\underset{\sim}{x}_1$. What do the eigen values and given vectors of the empirical correlation matrix R look like? How many eigen values are non zero? What percentage of variance is explained by the first factor? 7

   (b) What is a screw plot. State its uses. 3

3. (a) Does factor model solution always exist? Give reasons. 4

   (b) Describe why factor rotation is necessary. 3

   (c) Differentiate between discriminant and cluster analysis. 3

4. Let U_1 and U_2 be two independent uniform random variables on (0,1). Suppose that $\underset{\sim}{X} = (X_1, X_2, X_3, X_4)'$ where $X_1 = U_1, X_2 = U_2, X_3 = U_1 + U_2$ and $X_4 = U_1 - U_2$. Compute the correlation matrix of R of $\underset{\sim}{X}$. How many PCs (principal component) are of interest here? Show that $\underset{\sim}{\alpha}_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 1, 0\right)'$ and $\underset{\sim}{\alpha}_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 1, 0\right)'$ are eigenvectors corresponding to the nontrivial eigenvalues. Interpret the first two PCs obtained. 4+1+4+1=10

5. (a) Describe the steps of K-Means method of clustering. What is a dendrogram? Can you use dendrogram to show the clustering steps in K-Means method. 5

   (b) Derive the Maximum Likelihood (ML) discriminant rule based on observations of a one-dimensional variable with exponential distribution. 5

P.T.O.

6. (a) Show that discriminant rule based on minimizing the expected cost of misclassification with prior probability $\pi_1 = \frac{1}{3}$ and the cost of misclassification $C(2/1) = 2C(1/2)$ is identical with ML(Maximum likelihood) method of discriminant rule for 2 populations. 7

(b) Hence or otherwise find the optimum classification rule for two univariate normal populations with means μ_1, μ_2 and variances σ_1^2, σ_2^2 respectively based on one observation. 3

7. (a) What is canonical correlation and canonical variables? 4

(b) Let $\rho_{12} = \begin{pmatrix} \rho & \rho \\ \rho & \rho \end{pmatrix}$ and $\rho_{11} = \rho_{22} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ corresponding to the equal correlation structure where $\underset{\sim}{X}^{(1)}$ and $\underset{\sim}{X}^{(2)}$ each has two components. Determine the canonical variates corresponding to the non-zero canonical-correlation. 6