M. Sc. Semester-II Examination 2017
Statistics
Course: MSC-22
(Applied Multivariate Analysis)

Time : Three Hours **Full Marks : 40**

Questions are of value as indicated in the margin

Answer **any four** questions

1. Let $\underline{X}' = [X_1 X_2 .... X_p]$ have common covariance matrix Σ, with eigenvalue – eigenvector pairs $(\lambda_1, \underline{e}_1)...,(\lambda_p, \underline{e}_p)$ where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$. Let $Y_1 = \underline{e}'_1\underline{X}, Y_2 = \underline{e}'_2\underline{X}, ..., Yp = \underline{e}\ p'\underline{X}$ be the principal components. Then show that

   (i) $\sum_1^p \text{var}(X_i) = \sum_1^p \text{var}(Y_i) = \sum_{i=1}^p \lambda_i$

   (ii) $\rho_{y_i, X_k} = \frac{e_{ki}\sqrt{\lambda_i}}{\sigma_{kk}}$ (symbols having their usual meanings) 5+5=10

2. (a) Discuss the concept of factor analysis.

   (b) Propose a factor analytic model clearly stating the mathematical format and other requisites. 5+5=10

3. (a) Suppose $p \leq q$ and let the random vectors $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ have

   $\text{cov}(\underline{X}^{(1)}) = \sum_{11}^{p\times p}, \text{cov}(\underline{X}^{(2)}) = \sum_{22}^{q\times q}$ and $\text{cov}(\underline{X}^{(1)}, \underline{X}^{(2)}) = \sum_{12}^{p\times q}$ where Σ has full rank. For coefficient vectors $\underline{a}^{p\times 1}$ and $\underline{b}^{q\times 1}$, form the linear combinations $\underline{U} = \underline{a}'\underline{X}^{(1)}$ and $V = \underline{b}'\underline{X}^{(2)}$. Then show tha t

   $$\max_{a,b} corr(\underline{u}, \underline{v}) = \rho_1^*$$

   attained by the linear combinations $U_1 = \underline{e}'_1\Sigma_{11}^{-1/2}\underline{X}^{(1)}$ and $V_1 = \underline{f}_1'\ \Sigma_{22}^{-1}\underline{X}^{(2)}$.

   (b) In factor analytic model, What would you choose for primary analysis-correlation matrix or covariance matrix? Why? 7+3=10

4. (a) Discuss a procedure of testing the equality of mean vectors of K multinormal populations, where they share the common covariance matrix.

   (b) Discuss different proximity measures for binary data and comment on their performances. 7+3=10

5. (a) Derive a discriminant rule based on the ML method for two populations minimizing the expected cost of misclassification. Consider prior probability $\pi_1 = \frac{1}{3}$ and the expected cost of misclassification $c\left(2/1\right) = 2c\left(1/2\right)$.

   (b) Define Fisher's linear discriminant rule. For two population case under multinormal set up with common covariance matrix deduce its expression. 6+4=10

6. Write short notes on

   (a) Principal component Method (b) Single Linkage Method 5+5=10