LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

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

THIRD SEMESTER – NOVEMBER 2016

ST 3815 - MULTIVARIATE ANALYSIS

Date: 01-11-2016 Dept. No. Max. : 100 Marks
Time: 09:00-12:00

SECTION – A

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

1. Let X,Y and Z have trivariate normal distribution with null mean vector and covariance matrix $\begin{bmatrix} 2 & 5 & 0 \\ 5 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$. Find the distribution of X+Y.
2. Given the 3 component vector $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_3\left[\begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -1 \\ 0 & 4 & 1 \\ -1 & 1 & 6 \end{pmatrix}\right]$, find the mean of the conditional distribution of X_1 given $X_2 = 0$ *and* $X_3 = 0$.
3. Test $\mu = (0 \;\; 0)'$ at level 0.05, in a bivariate normal population with $\sigma_{11} = \sigma_{22} = 5$ and $\sigma_{12} = -2$, using the sample mean vector $\bar{x} = (7 \;\; -3)'$ based on sample size 10.
4. Explain use of the partial and multiple correlation coefficients.
5. Describe a) Common factor b) Communality.
6. What is the difference between classification problem into two classes and testing problem?
7. Distinguish between principal component analysis and factor analysis.
8. Let Y_i be the ith principal component of the system involving variables $X_1, X_2, ..., X_p$ and e_i be the eigen vector corresponding to the ith largest eigen root λ_i. Find the correlation between Y_i and the variable X_k .
9. Explain MANOVA.
10. Write a short note on data mining.

PART– B

Answer anyFIVE questions **(5X8=40 marks)**

11. Find the multiple correlation coefficient between X_1 and $X_2, X_3, \ldots, X_p$. Prove that the conditional variance of X_1 given the rest of the variables cannot be greater than unconditional variance of X_1.
12. Derive the characteristic function of multivariate normal distribution
13. Obtain the rule to assign an observation of unknown origin to one of two p-variate normal populations having the same dispersion matrix.

14. Show that the sample generalized variance is zero if and only if the rows of the matrix of deviation are linearly dependent.
15. Using the Likelihood ratio procedure, develop the linear discriminant function and its variance.
16. Giving suitable examples explain how factor scores are used in data analysis
17. Let $(X_i, Y_i)'$ i = 1,2,3 be independently distributed each according to bivariate normal with mean vector and covariance matrix as given below. Find the joint distribution of the six variables. Also find the joint distribution of $\overline{X}$ *and* $\overline{Y}$.

    Mean Vector: $(\mu, \tau)'$, covariance matrix: $\begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}$.
18. Outline single linkage and complete linkage procedures with an example

PART– C

Answer anyTWO questions **(2 X 20 =40marks)**

19. Derive the distribution function of the generalized T^2 – Statistic
20. a) Write short notes on repeated measurement design.

    .

    b) If $X \sim N_p(\mu, \Sigma)$ then prove that $Z = DX \sim N_p(D\mu, D\Sigma D')$ where D is qxp matrix rank q≤p. (10+10)
21. Consider the two data sets $X_1 = \begin{pmatrix} 3 & 7 \\ 2 & 4 \\ 4 & 7 \end{pmatrix}$ *and* $X_2 = \begin{pmatrix} 6 & 9 \\ 5 & 7 \\ 4 & 8 \end{pmatrix}$ from populations Π_1 *and* Π_2 respectively, for which $\overline{x}_1 = (3 \;\; 6)'$, $\overline{x}_2 = (5 \;\; 8)'$ *and* $S_{pooled} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$.

    a) Calculate the linear discriminant function.

    b) Classify the observation $x_0' = (2 \; 7)$ to population π_1 or population π_2 using the decision rule with equal priors and equal costs. (14+6)
22. a)Outline the procedure to extract principal components from a given covariance matrix.

    b) Prove that under some assumptions (to be stated), variance covariance matrix can be written as $\Sigma = LL' + \Psi$ in the factor analysis model. Also discuss the effect of an orthogonal transformation. (8+12)

****