 LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

**M.Sc.** DEGREE EXAMINATION - **STATISTICS**

THIRD SEMESTER – **APRIL 2012**

# ST 3811/3808 - MULTIVARIATE ANALYSIS

Date : 21-04-2012 Dept. No. Max. : 100 Marks

Time : 1:00 - 4:00

SECTION – A

Answer ALL the following questions: (10 x 2 = 20 marks)

1. Define var-cov matrix, correlation matrix and standard deviation matrix. State the relationship among them.
2. State the general expressions for the mean vector and var-cov matrix of linear transformations of a random vector.
3. State the general form of ‘statistical distance’.
4. Explain ‘Bubble Plots’.
5. Define ‘Partial Correlation Coefficient’.
6. State the one-way MANOVA model.
7. State the test for significance of correlation coefficient in a bivariate normal population.
8. Give one reason why in K-means algorithm, the number of clusters ‘K’ is kept an open question.
9. State the postulates on the ‘common factors’ and the ‘specific factors’ in the orthogonal factor model.
10. Explain a situation where the ‘challenge’ of ‘Classification’ arises.

SECTION – B

Answer any FIVE questions: (5 x 8 = 40 marks)

1. Briefly explain the terms ‘Data Reduction / Structural Simplification’ and ‘Sorting / Grouping’. Give real-life examples of these two objectives which are addressed by multivariate methods.
2. Explain probability plots in general and how it is used for investigation of multivariate normality assumption.
3. Derive the moment generating function of multivariate normal distribution.
4. If X =  ~ Np (μ,Σ) and μ and Σ are accordingly partitioned as  and where  ≠ 0, derive the conditional distribution of X(1) given X(2).
5. Derive Fisher’s linear discriminant function for discriminating two populations.
6. Mention the three linkage methods for hierarchical clustering and present a figurative display of the measure of between-cluster distances in each method.
7. Develop the Hotelling’s T2 test through the likelihood ratio criterion. (P.T.O)
8. Give the motivation and the formal definition of Principal Components. State the ‘Maximization Lemma’ (without proof) and hence, obtain the PCs for a random vector

SECTION – C

Answer any TWO questions: (2 x 20 = 40 marks)

1. (a) If and Sn­ are the sample mean vector and var-cov matrix from a sample of size ‘n’ from a multivariate population with mean vector **μ** and var-cov matrix **Σ** , show that  is an unbiased estimator of **μ** but Sn is a biased estimator of **Σ**.

(b) Derive the MLEs of the parameters of multivariate normal distribution.

(10+10)

1. (a) Consider the partitions in Q.No. (14). Let =  where= ith row of  . Show that, for every vector α,

(i) Var[ Xi – X(2) ] ≤ Var [ Xi – α′ X(2) ]

(ii) Corr ( Xi, X(2) ) ≥ Corr ( Xi­ , α′ X(2) ).

Hence, obtain an expression for the multiple correlation coefficient between Xi and X(2).

(b) Find the mean vector and the var-cov matrix for the bivariate normal distribution whose p.d.f. is

*f(x,y)* = exp  (12 + 8)

1. (a) Exhibit the ‘ambiguity’ in the factor model. Bring out the need for factor rotation and explain the ‘Varimax’ criterion for rotation.

(b) Explain the ‘Ordinary Least Squares Method’ of estimating the Factor Scores.

(12 + 8)

1. (a) Derive an expression for ‘Expected Cost of Misclassification (ECM)’ for classification involving two populations and obtain the optimum allocation regions for the ‘Minimum ECM Rule’.

(b) Consider the following table on three binary variables measured on five subjects with a view to carry out clustering of the five subjects:

|  |  |
| --- | --- |
| variable  Individual | X1 X2 X5 |
| 1  2  3  4  5 | 0 0 1  1 1 1  0 0 1  0 1 1  1 1 0 |

Obtain the matrices of matches and mismatches, compute the similarity measure

sij =  (under usual notations) and carry out the clustering. (10 + 10)

\* \* \* \* \* \*