LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

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

THIRD SEMESTER – NOVEMBER 2011

ST 3811/3808 - MULTIVARIATE ANALYSIS

Date : 31-10-2011 Dept. No. Max. : 100 Marks

Time : 9:00 - 12:00

SECTION – A

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

1. Briefly explain the 'Data Exploration' stage of Multivariate Analysis.
2. State the relationship among Var-Cov matrix, Correlation matrix and Standard Deviation matrix.
3. Give the motivation for 'statistical distance'.
4. State the kernel estimate of a multivariate p.d.f.
5. State the m.g.f. of multivariate normal distribution.
6. State the T^2 statistic for testing hypothesis about the mean vector of a multivariate normal population.
7. Explain use of 'Dendogram' in cluster analysis.
8. Mention any two criteria for obtaining 'good' classification procedures.
9. State the test for significance of correlation coefficient in a bivariate normal population.
10. Define Multiple Correlation Coefficient.

SECTION – B

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

11. Briefly explain the terms 'Sorting / Grouping' and 'Prediction'. Give real-life examples of these two objectives which are addressed by multivariate methods.
12. Describe the scatter plot enhancement using *'lowess curve'*.
13. If $X \sim N_p(\mu, \Sigma)$ and C is a non-singular matrix of order p x p, show that $CX \sim N_p(C\mu, C\Sigma C^T)$. Hence, deduce the distribution of DX where D is a q x p matrix with rank q ($\leq$ p).
14. If $X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix} \sim N_p(\mu,\Sigma)$ and μ and Σ are correspondingly partitioned as $\begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}$ and $\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, show that $X^{(1)}$ and $X^{(2)}$ are independent if and only if each covariance of a variable from $X^{(1)}$ and a variable from $X^{(2)}$ is zero.
15. Show that the criteria of minimizing 'Total Probability of Misclassification' and maximizing 'Posterior Probabilities' lead to the same procedure as 'Minimum ECM' rule with equal misclassification costs.
16. Develop the MANOVA for comparing mean vectors of a number of normal populations and explain the test procedure for the same.
17. Explain the Weighted Least Squares Method and the Approximate-Simple Method of finding the Factor Scores.
18. A consumer-preference study on a food product was carried out and the ratings given by consumers on five attributes were measured. A factor analysis was performed from the correlation matrix and some partial results of the same are given below. Fill up the missing entries:

Variable	Factor loadings	Rotated Factor	Communalities	Specific

	F_1	F_2	Loadings F_1	F_2	h_i^2	variances
1. Taste	0.56	0.82	0.02	____	____	_____
2. Money value	0.78	– 0.53	____	– 0.01	____	_____
3. Flavour	0.65	0.75	0.13	____	____	_____
4. Suitability	0.94	– 0.11	0.84	____	____	_____
5. Energy	0.80	– 0.54	____	– 0.02	____	_____

SECTION – C

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

19. (a) Develop the multivariate normal density function .
(b) Show that the sample mean vector and the sample var-cov matrix, based on a random sample from a multivariate normal distribution, are independent.
(10 + 10)

20. (a) Discuss 'Hierarchical Agglomerative and Divisive methods' of clustering items. Present the algorithm of agglomerative methods. Present a figurative display of the measure of between-cluster distances in the different linkage methods.
(b) Explaining the notations, enlist any four similarity measures for pairs of items when variables are binary and state their rationale. (12 + 8)

21. Stating the motivation and formal definition, derive the Fisher's (multiple) discriminant functions.

22. (a) Define 'Principal Components' and bring out their relationship with eigen values and eigen vectors of the var-cov matrix of the underlying random vector.
(b) Explain the 'Orthogonal Factor Model' and develop the notions of 'communality' and 'specific variance'. Briefly sketch the 'Principal Component Method' of estimating the parameters of the model. (8 + 12)

* * * * * * *