LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION – STATISTICS

THIRD SEMESTER – NOVEMBER 2007

ST 3808 / 3801 - MULTIVARIATE ANALYSIS

BB 7

Date : 24/10/2007 Dept. No. Max. : 100 Marks
Time : 9:00 - 12:00

PART – 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

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix},$$

find the distribution of X+Y.
2. Write the statistic used to test the hypothesis H: $\rho_{12.3} = 0$ in a bivariate normal distribution.
3. Mention any two properties of multivariate normal distribution.
4. Write down the characteristic function of a multivariate normal distribution.
5. Explain the use of partial and multiple correlation coefficients.
6. Define Hotelling's T^2 – statistics. How is it related to Mahlanobis' D^2?
7. Give an example in the bivariate situation that, the marginal distributions are normal but the bivariate distribution is not.
8. Outline the use of discriminant analysis.
9. What are canonical correlation coefficients and canonical variables?
10. Write down any four similarity measures used in cluster analysis.

PART B

Answer any FIVE questions. **(5 X 8 = 40)**

11. Obtain the maximum likelihood estimator Σ of p-variate normal distribution.
12. Let Y ~ N_p (**0** , Σ). Show that $Y'\Sigma^{-1}Y$ has $\chi 2$ 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. Outline single linkage and complete linkage clustering procedures with an example.
15. Let X ~ N_p (μ, Σ). If $X^{(1)}$ and $X^{(2)}$ are two subvectors of X, obtain the conditional distribution of $X^{(1)}$ given $X^{(2)}$.
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 six variables. Also find the joint distribution of $\bar{x}$ and $\bar{y}$.

Mean vector: (μ, τ)', covariance matrix: $\begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}$

18. Write short notes on step-wise regression.

PART C

Answer any two questions. **(2 X 20 = 40)**

19. a) 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 of rank $q \leq p$.

b). Consider a multivariate normal distribution of X with

$$\mu = \begin{pmatrix} 8 \\ -2 \\ 0 \\ 3 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 7 & 5 & 1 & 4 \\ 5 & 4 & 8 & -6 \\ 1 & 8 & 3 & 7 \\ 4 & -6 & 7 & 2 \end{pmatrix}.$$

Find i) the conditional distribution of $(x_2, x_4) / (x_1, x_3)$.

ii) $\sigma_{33.42}$ (10 + 10)

20. a) What are principal components?. Outline the procedure to extract principal components from a given correlation matrix.

b) What is the difference between classification problem into two classes and testing problem. (14 + 6)

21. a) Derive the distribution function of the generalized T^2 – statistic.

b) Test at level 0.05 ,whether $\mu = (0 \ \ 0)'$ in a bivariate normal population with $\sigma_{11} = \sigma_{22} = 5$ and $\sigma_{12} = -2$, by using the sample mean vector $\bar{x} = (7 \ \ -3)'$ based on a sample size 10. (15 + 5)

22. a) Explain the method of extracting canonical correlations and their variables from a dispersion matrix.

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

****