LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

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

THIRD SEMESTER – **NOVEMBER 2019**

16/18PST3MC01 – MULTIVARIATE ANALYSIS

Date: 29-10-2019 Dept. No. Max. : 100 Marks

Time: 09:00-12:00

Section - A

Answer ALL the Questions **(10 x 2 = 20 marks)**

1. Briefly explain the ‘sorting’ / ‘grouping’ objective of multivariate analysis.
2. If Var-Cov(**X**) = **Σ**, obtain Var-Cov(**DX**).
3. Describe Bubble Plots.
4. Give an example where marginal distributions are normal but joint distribution is non-normal.
5. Define partial correlation coefficient and briefly explain the test for its significance in the case of joint normality.
6. Explain MANOVA.
7. Mention two situations in which standardized variables need to be used as inputs to principal component analysis.
8. Explain any one method of scaling the coefficient vector of a discriminant function and its use.
9. Demonstrate with an example how a factor model reduces the parameter space dimension.
10. Suggest any one similarity coefficient (explaining the notations) for clustering of objects based on binary variables..

Section - B

Answer any FIVE Questions **(5 x 8 = 40 marks)**

11. Present the notion of ‘angle’ between vectors in multidimensional space. Discuss its relevance in measuring the correlation coefficient between two random variables based on a sample of bivariate observations.
12. Show that the var-cov matrix of a random vector **X** is positive definite unless there is an affine dependence among the components of **X** in which case it is non-negative definite.
13. Derive the moment generating function of multivariate normal distribution.
14. Develop the notion of multiple correlation coefficient and derive an expression for it in the case of joint normality.
15. The var-cov matrix of a random vector **X** = $(X_1, X_2, X_3, X_4)^T$ is estimated to be

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.30 & 0.24 & 0.12 & 0.24 \\ 0.24 & 0.81 & 0.18 & 0.54 \\ 0.12 & 0.18 & 0.12 & 0.21 \\ 0.24 & 0.54 & 0.21 & 0.66 \end{bmatrix}$$

    The Eigen vector associated with the largest eigen value of Σ is $(0.31, 0.69, 0.21, 0.)^T$.

    Express the 1st Principal Component as a linear combination of X_i’s and also find what % of total variation is explained by the 1st PC. Evaluate the loadings of each of the X_i’s on the 1st PC by computing the correlations.

16. Define the T^2 Statistic for testing the hypothesis H_0: $\boldsymbol{\mu} = \boldsymbol{\mu_0}$ based on a random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with unknown $\boldsymbol{\Sigma}$ and explain the test procedure. Show that this test is obtained from the Likelihood Ratio Criterion.
17. Explain the three usual criteria for good classification procedures with an illustration for each.
18. Explain hierarchical clustering of objects and present figuratively any two linkage schemes for agglomerative clustering method.

Section - C

Answer any TWO Questions **(2 x 20 = 40 marks)**

19. (a) Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\mathbf{X} = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix}$ where $X^{(1)}$ is qx1 and $X^{(1)}$ is (p–q)x1, and $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ accordingly partitioned. 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. Show that, under these conditions, the marginal distributions of $X^{(1)}$ and of $X^{(2)}$ are also normal.

    (b) Establish that even without independence, the marginal distributions of $X^{(1)}$ and of $X^{(2)}$ are normal. Hence, obtain the conditional distribution of $X^{(1)}$ given $X^{(2)}$. **(10 +10)**

20. (a) Explain the orthogonal factor model and the notions of communalities and specific variances. Bring out the ambiguity in the factor model and give the motivation for factor rotation.

    (b) Explain the weighted least squares approach of obtaining the 'factor scores'. **(12 + 8)**

21. (a) Discuss the motivation given by Fisher for discrimination of two populations and derive the Fisher's Linear Discriminant Function.

    (b) Derive an expression for 'Expected Cost of Misclassification' and the 'Minimum ECM Rule' with reference to two-population classification. **(10 +10)**

22. (a) Consider the following data on five binary variables measured on five individuals:

Individual \ Variable	X_1	X_2	X_3	X_4	X_5
1	0	0	0	1	1
2	1	1	1	0	1
3	0	1	0	1	1
4	0	0	1	0	1
5	1	1	1	0	0

Carry out the clustering of the variables treating positive correlation as similarity coefficient and form two clusters.

(b) Explain non-hierarchical clustering and K-Means method. **(14+ 6)**

* * * * * *