INDIAN STATISTICAL INSTITUTE

MS in QMS

TEST ON MULTIVARIATE DATA ANALYSIS

Date: 21 February, 2025   Time: 2 hours   Maximum Marks: 50

Answer as many questions as you can. The maximum you can score is 50

1. Explain the concept of the multivariate normal distribution and its properties?

   [5]

2. Explain how to determine if a given set of data follows a multivariate normal distribution. What methods and tests can be used to assess this?

   [5]

3. Describe the role of the determinant and the inverse of the covariance matrix in the multivariate normal distribution. Why are these important?

   [5]

4. Explain the concept of factor analysis and its main objectives. How does it differ from other dimensionality reduction techniques like principal component analysis (PCA)?

   [5]

5. How do you determine the number of factors to retain in factor analysis? Discuss the criteria and methods used for deciding the appropriate number of factors.

   [5]

6. Consider a dataset with three variables x_1, x_2, and x_3. The correlation matrix for these variables is given as follows:

1	0.8	0.6
0.8	1	0.7
0.6	0.7	1

   Perform the factor analysis using the principal component method.

   [10]

7. Discuss the various types of clustering algorithms, such as hierarchical clustering and k-means clustering. Highlight their key differences.

[5]

8. How do you determine the optimum number of clusters in cluster analysis? Discuss the criteria and methods used for deciding the appropriate number of clusters.

[5]

9. Consider the following dataset

Cases	x_1	x_2
1	2	3
2	3	3
3	8	7
4	9	6

Perform k means cluster analysis to cluster the cases into 2 clusters. Consider the initial centroids as

Cluster	x_1	x_2
I	2	3
II	8	7

[10]