INDIAN STATISTICAL INSTITUTE

MS QMS

TEST ON MULTIVARIATE ANALYSIS

Date: 21 February 2023 Time: 3 hours Maximum Marks: 50

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

1.

a. What is the objective of carrying out principal components analysis?
b. How the coefficients of the principal components are computed?
c. Suggest the commonly used methods to arrive at the number of principal components.
d. Write the step-by-step procedure for conducting principal component analysis and interpretation of the results.

[10]

2.

a. In factor analysis, assume that a set of manifest variables $x = (x_1, x_2, - - -, x_q)'$ is linked to k unobserved latent variables $f_1, f_2, - - - , f_k$, $k \leq q$ by the model

$$x = \Lambda f + \mu$$

Describe how elements of Λ matrix are computed by clearly stating the assumptions.

b. The correlation matrix of 3 variables is given below. Identify the optimum number of factors using 3 methods.

	x_1	x_2	x_3
x_1	1.00	0.55	0.82
x_2	0.55	1.00	0.43
x3	0.82	0.43	1.00

[15]

3.

a. Explain the procedure to perform hierarchical clustering.
b. The pairwise Euclidean distance matrix of 4 respondents of a market research survey is given below. Group the respondents into the optimum number of clusters using hierarchical cluster analysis. Draw dendrogram. Compute the silhouette distance and confirm whether the number of clusters identified is appropriate.

	1	2	3	4
1	0.00	9.54	2.00	9.64
2	9.54	0.00	9.85	3.00
3	2.00	9.85	0.00	9.54
4	9.64	3.00	9.54	0.00

[15]

4.

a. Explain the k-means cluster analysis procedure.

b. The responses of 4 individuals who participated in a market research survey is given below. Group the 4 participants into 2 clusters using k-means cluster analysis. The initial values of the centroids are given below.

Respondent	Q1	Q2	Q3	Q4
1	6	2	6	1
2	2	7	1	6
3	5	1	7	2
4	1	6	2	7

Centroid	Q1	Q2	Q3	Q4
I	7	1	7	1
II	1	7	1	7

[15]