INDIAN STATISTICAL INSTITUTE

MS in QMS

TEST ON MULTIVARIATE DATA ANALYSIS (Back paper)

Date: 05 June 2026 Time: 3 hours Maximum Marks: 50

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

1. Explain how multicollinearity in regression models leads to inflated variances of the estimated coefficients. Explain principal component Regression? What are the advantages and disadvantages of principal component regression? Show that principal components are uncorrelated?

[10]

2.
   a. Explain Factor analysis. What are the assumptions used to estimate the factors? How factor analysis is different from multiple linear regression?
   b. What is KMO statistic? How it is useful in factor analysis? How KMO statistic is computed?
   c. Explain Bartlett's test of Sphericity? Write down the null and alternative hypothesis in Bartlett's test of sphericity? Explain the formula used to compute the test statistic and p value in Bartlett's test of sphericity?

[12]

3.
   a. State and prove the theorem that the first principal component maximizes variance among all possible linear combinations of the original variables. Why is this property fundamental to the construction of PCA?
   b. State and prove the theorem of rotational indeterminacy in factor analysis. Why does the factor solution remain unchanged under orthogonal rotation?

[10]

4.
   a. State and prove the theorem that the cluster mean as centroid minimizes the sum of squared distances from centroid within a cluster. Why is this property fundamental to the optimality of the k-means clustering algorithm?
   b. State and prove the convergence theorem for k-means clustering. Explain why the iterative process of reassigning points and updating centroids is guaranteed to converge to a local optimum, and discuss the implications of this result.

[10]

5. A popular ice cream parlor is looking to create a new dessert by combining their ice cream with a sauce/syrup and topping it off with delicious decorations. To determine customer preferences, the parlor has conducted a conjoint analysis and collected the data, which is presented below

Case No	Ice Cream	Sauce/Syrup	Toppings	Aggregate Score
1	Vanilla	Honey	Dry Fruits	6.50
2	Vanilla	Caramel	Cherries	6.00
3	Vanilla	Chocolate	Roasted Cashew nuts	5.80
4	Vanilla	Pineapple	Lychees	5.50
5	Chocolate	Honey	Cherries	6.00
6	Chocolate	Caramel	Dry Fruits	4.50
7	Chocolate	Chocolate	Lychees	2.50
8	Chocolate	Pineapple	Roasted Cashew nuts	6.80
9	Strawberry	Honey	Roasted Cashew nuts	9.30
10	Strawberry	Caramel	Lychees	6.00
11	Strawberry	Chocolate	Dry Fruits	7.00
12	Strawberry	Pineapple	Cherries	9.50
13	Mango	Honey	Lychees	6.00
14	Mango	Caramel	Roasted Cashew nuts	7.30
15	Mango	Chocolate	Cherries	6.50
16	Mango	Pineapple	Dry Fruits	8.00

a. Please analyze the data and calculate the part worth utilities and importance scores.
b. Based on your analysis, which ice cream dessert is the most preferred and which is the second most preferred?
c. Kindly estimate the expected scores for the best and second-best ice cream dessert?

[13]