Indian Statistical Institute, Bangalore
MS (QMS) First Year
First Semester - Multivariate Data Analysis

Final Exam — Date: March 15, 2022
Maximum marks: 50 — Duration: 2 hours

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

1. [6 + 12 = 18]

a. Explain the multivariate analysis of variance (MANOVA)? Describe the important statistics used in MANOVA?
b. The plating thickness and surface finish are two important characteristics of the chromium plating process of piston rods. The mean vector and inverse of the variance covariance matrix of the characteristics are given below:

Mean Vector	
Plating Thickness	Surface Finish
85.519	0.0195

Covariance matrix inverse	Plating Thickness	Surface Finish
Plating Thickness	25.31	3121.63
Surface Finish	3121.63	421652.94

The data of 3 chromium-plated piston rings are given below. Compute Mahalanobis distance and give your comments on the 3 piston rods

Piston Rod Id	Plating Thickness	Surface Finish
1	86.65	0.011
2	85.02	0.025
3	83.85	0.029

2. [10 + 2 + 6 = 18]

a. Explain the procedure for determining the loadings in principal component analysis and factor analysis? Provide the similarities and differences between principal component and factor analysis?

b. Describe the role of the Bartlett test of sphericity in factor analysis?

c. The correlation matrix of sprint productivity, sprint velocity and focus factor of agile software development processes of healthcare vertical is subjected to eigenvalue analysis. The eigenvalues are given below. Identify the optimum number of factors using variance, cumulative % of variance and scree plot methods. The KMO statistics is computed and is found to be 0.64, is it a good idea to go ahead with factor analysis? Justify?

SL No	Eigenvalue
1	1.73
2	1.27
3	0.004

3. [5 + 5 + 8 = 18]

a. Explain the k means clustering procedures?

b. Explain the different methods for identifying the optimum number of clusters in cluster analysis

c. Five customers who participated in a customer satisfaction survey has been clustered into 2 clusters as follows:

Customer	Cluster
1	I
2	II
3	II
4	II
5	I

The pairwise dissimilarity between the customers is given below. Compute silhouette width and construct silhouette plot? Is the clustering appropriate? Give your comments.

Customer Id	1	2	3	4	5
1	0.00	7.94	6.00	7.75	1.73
2	7.94	0.00	9.95	2.24	8.25
3	6.00	9.95	0.00	9.49	6.24
4	7.75	2.24	9.49	0.00	8.43
5	1.73	8.25	6.24	8.43	0.00