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

Midterm Exam — Date: February 25, 2020
Maximum marks: 40 — Duration: 2 hours

Answer as many questions you can.

The maximum you can score is only 40 marks

1. An airline wants to know the customer preference on travel between two cities of the country. The company has conducted a conjoint analysis. The details of the conjoint analysis with aggregate ranking are given below. Analyze the data, compute part-worth utilities and importance scores. What suggestions can you provide to the airline to optimize customer satisfaction? [15]

Combinations	Ticket Cost	Onboard Food	Transfers	Seat Selection	Aggregate Score
1	Rs. 4000	Need to pay	Direct	Free	8.0
2	Rs. 4000	Need to pay	One-stop	Paid	6.5
3	Rs. 4000	Free	Direct	Paid	7.5
4	Rs. 4000	Free	One-stop	Free	5.5
5	Rs. 7500	Need to pay	Direct	Paid	3.5
6	Rs. 7500	Need to pay	One-stop	Free	2.5
7	Rs. 7500	Free	Direct	Free	7.0
8	Rs. 7500	Free	One-stop	Paid	4.5

2. [15]

   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

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

   Write down the different assumptions made on the error matrix μ to solve the aforementioned equation.

   b. The data on review time, test coverage and review coverage of 7 projects are given below. Kindly classify the projects into 3 clusters. The initial centroid values are also given in the table below:-

Project	Review time	Test coverage	Review coverage
1	30.0	100.0	70.0
2	20.0	70.0	100.0
3	30.0	50.0	50.0
4	30.0	70.0	71.0
5	30.0	70.0	100.0
6	30.0	100.0	70.0
7	20.0	50.0	50.0

Centroid	Review time	Test coverage	Review coverage
1	30.0	100.0	70.0
2	26.7	70.0	90.3
3	25.0	50.0	50.0

3. [15]

a. Describe two methods to identify the optimum value of k (number of clusters) in cluster analysis?

b. Five software development projects have been classified into 2 clusters based on KEDB and skill of the developer as shown below. The centroids of the two clusters also are given below. Compute the silhouette width of each project and interpret the results

Projects	KEDB	Developer Skill	Cluster
1	20	3.17	1
2	21	3.35	2
3	21	3.25	2
4	22	3.49	1
5	22	3.29	1

Centroid	KEDB	Developer Skill
1	22	3.39
2	20	3.26