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

Midterm Exam   Date: February 25, 2019
Maximum marks: 50   Duration: 2 hours

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

1. 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 f and error matrix μ to solve the aforementioned equation.   [10]

2.
   a. Explain Mahalanobis distance?
   b. The surface hardness, case depth and dimensional variation are three important performance characteristics of a heat treatment process. The data on the deviation from the mean vector and inverse of the variance-covariance matrix of a pilot sample of 6 parts are given below. Compute the Mahalanobis distance and arrange the parts in the increasing order of distance from the center   [15]

Deviation from the mean vector

Job Id	Surface Hardness	Case Depth	Dimensional Variation
1	-0.417	0.268	0.057
2	0.183	-0.342	-0.009
3	1.983	0.028	0.029
4	-2.917	0.438	-0.121
5	0.883	0.138	-0.026
6	0.283	-0.532	0.071

Inverse of the variance-covariance matrix

	Surface Hardness	Case Depth	Dimensional Variation
Surface Hardness	0.813	0.451	-11.316
Case Depth	0.451	13.477	34.156
Dimensional Variation	-11.316	34.156	524.792

3.

    a. Describe the optimization problem to be solved to get the principal components? Give three different methods to identify the optimum number of principal components?

    b. The correlation matrix of 3 variables from a chemical process is given below. Carryout principal component analysis, identify the variances of principal components, draw the scree plot and shortlist the optimum number of principal components   [20]

	Temperature	Time	Kappa Number
Temperature	1.00	-0.86	0.51
Time	-0.86	1.00	0.43
Kappa Number	0.51	0.43	1.00

4. Answer *any two* of the following:-   [10]

    a. Explain briefly the correspondence analysis?
    b. Explain briefly the multidimensional scaling?
    c. Similarities and differences between factor analysis and principal components analysis