

Indian Statistical Institute, Bangalore
M.S. (QMS) First Year
First Semester – Multivariate Data Analysis

Mid Term Exam Duration: 2 Hrs Date: February 20, 2017 Max Marks: 50

Answer as many questions you can.

1. Let a random vector $x = (x_1, x_2, \dots, x_p)'$ is multivariate normally distributed with mean vector μ and variance-covariance matrix Σ .

a. Write the probability density function of x ? Give the maximum likelihood estimator of μ and Σ ?

b. The yield and viscosity are two important output characteristics of pulp manufacturing process. The target values of yield and viscosity are given by $\mu_0 = (36 \ 52)$. A sample of 10 data points on yield and viscosity are collected from the process. The mean vector \bar{x} and the inverse of corrected sample variance-covariance matrix S_u^{-1} are given below. Test whether on an average the process is on the target (F table value = 4.46).

$$\bar{x} = (36.49 \ 52.25)$$

$$S_u^{-1} = \begin{bmatrix} 10.54 & -1.23 \\ -1.23 & 0.34 \end{bmatrix}$$

[12]

2.

a. Define Mahalanobis distance?

b. The data on design review defect density (DR_DD), code review defect density (CR_DD) and system & integration testing defect density (SIT_DD) of nineteen projects are collected from an account. The sample mean vector \bar{x} and the inverse of sample variance-covariance matrix S^{-1} are given below. The data of 3 projects are also given in the table below. Compute the Mahalanobis distance of the projects and arrange them in the increasing order of distance from centre

$$\bar{x} = (0.37 \ 0.91 \ 1.23)$$

$$S^{-1} = \begin{bmatrix} 142.63 & -61.90 & -6.88 \\ -61.90 & 71.43 & -18.05 \\ -6.88 & -18.05 & 16.06 \end{bmatrix}$$

Project Id	DR_DD	CR_DD	SIT_DD
Mondelez 1	0.51	1.13	1.73
Mondelez 2	0.25	0.9	0.8
Mondelez 3	0.47	1.03	1.55

[15]

3.

- a. What is the objective of carrying out principal components analysis? 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. Carry out principal component analysis, identify the variances of principal components and shortlist the optimum number of principal components

	Temperature	Time	Viscosity
Temperature	1.00	-0.96	0.22
Time	-0.96	1.00	-0.24
Viscosity	0.22	-0.24	1.00

[15]

4.

- a. Define single, complete and average linkages?
- b. The data on five variables collected from five entities is given below. Group them using hierarchical clustering method. Identify the optimum number of clusters?

Id	X1	X2	X3	X4	X5
1	2	6	1	5	7
2	6	7	3	6	2
3	7	6	2	5	2
4	1	6	2	6	7
5	4	7	2	2	1

[15]