Ramakrishna Mission Vivekananda University

Belur Math, Howrah, West Bengal

School of Mathematical Sciences, Department of Data Science

M.Sc. in Big Data Analytic 2017, Semester Exam	Date: 5 May 2018
Course : **DA310: Multivariate Statistics**	Time: 2 hrs
Instructor : Dr. Sudipta Das	*Max marks: 50*
Student signature and Id:	

1. (a) Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{60}$ be a random sample of size 60 from a four-variate normal distribution having mean μ and covariance Σ. Specify each of the following completely (mention parameters).
   - Distribution of $\bar{\mathbf{X}}$
   - Distribution of $(\mathbf{X}_1 - \mu)'\Sigma^{-1}(\mathbf{X}_1 - \mu)$
   - Distribution of $n(\bar{\mathbf{X}} - \mu)'\Sigma^{-1}(\bar{\mathbf{X}} - \mu)$
   - Approximate distribution of $n(\bar{\mathbf{X}} - \mu)'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mu)$

   $\bar{\mathbf{X}}$ and $\mathbf{S}$ are the sample mean and the sample variance of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{60}$, respectively.

   (b) What is Box-Cox transformation and why is it needed?

   (c) What is be the statistical distance between the first and last observations of the following data

x_1	1	2	3	4	5	6	7	8	9	10
x_2	18.95	19.00	17.95	15.54	14.00	12.95	8.94	7.49	6.00	3.99

[4+2+5=11]

2. The sample mean vector and the sample covariance matrix, as given below, are calculated from pairs of 42 observations.

$$\bar{x} = \begin{bmatrix} 0.564 \\ 0.603 \end{bmatrix}, \text{ and } S = \begin{bmatrix} 0.0144 & 0.0117 \\ 0.0117 & 0.0146 \end{bmatrix}.$$

   Compare the 95% T^2 and 95% Bonferroni simultaneous confidence intervals. [13]

3. Consider the covariance matrix

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

   (a) Calculate the principal components.

   (b) Draw the scree plot.

   (c) When the principal components are preferred to be derived from the correlation matrix instead from the covariance matrix?

[7+2+2=11]

4. Orthogonal factor model with p features and m common factors is described as follows:

$$X = \mu + LF + \epsilon.$$

(a) Prove that

$$\Sigma = LL' + \Psi$$

and state the assumptions needed to prove it. Is it, always, possible to get a consistent solution to the above equation?

(b) The Σ and L matrices are given as

$$\Sigma = \begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix} \text{ and } L = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix},$$

respectively. Find the Ψ matrix.

(c) State two methods, used for estimation of factor loadings.

(d) Why factor rotation is needed?

(e) In which situation the varimax criterion is used?

[(3+3+1)+2+2+2+2=15]

You may need following values:

$t_{41}(0.05) = 1.68, t_{41}(0.0125) = 2.327,$
$F_{3,2}(0.05) = 19.16, F_{2,3}(0.05) = 9.55, F_{2,40}(0.05) = 3.23, F_{40,2}(0.05) = 19.47$

This exam has total 4 questions, for a total of 50 points and 0 bonus points.