Ramakrishna Mission Vivekananda University

Belur Math, Howrah, West Bengal

School of Mathematical Sciences, Department of Computer Science

M.Sc. in Big Data Analytic 2018, Mid Semester Exam — Date: 07 March 2019

Course : **DA310: Multivariate Statistics** — Time: 2 hrs

Instructor : Dr. Sudipta Das — *Max marks: 50*

Student signature and Id:

---

1. (a) Define the distance from the point $P = (x_1, x_2)$ to the origin $O = (0, 0)$ as

$$d(O, P) = (|x_1|^q + |x_2|^q)^{\frac{1}{q}}.$$

Plot the the locus of points whose distance from the origin is 1, for five different values of $q = .5, 1, 2, 4, \infty$.

(b) The sample covariance matrix of a bi-variate n samples with zero sample mean is given as follows.

$$\mathbf{S} = \begin{bmatrix} 8 & -6 \\ -6 & 9 \end{bmatrix}.$$

Find the angle between the two components of the bi-variate data in the n dimensional vector space.

[5+5=10]

2. Energy consumption in 2001, by state, from the major sources x_1 = petroleum, x_2 = natural gas, x_3 = hydroelectric power and x_4 = nuclear electric power is recorded in some unit. The resulting sample mean and covariance matrix are

$$\bar{\mathbf{x}} = \begin{bmatrix} 0.766 \\ 0.508 \\ 0.438 \\ 0.161 \end{bmatrix} \text{ and } \mathbf{S} = \begin{bmatrix} 0.856 & 0.635 & 0.173 & 0.096 \\ 0.635 & 0.568 & 0.128 & 0.067 \\ 0.173 & 0.127 & 0.171 & 0.039 \\ 0.096 & 0.067 & 0.039 & 0.043 \end{bmatrix}.$$

(a) Using the summary statistics, determine the sample mean and variance of a state's total energy consumption for these major sources.

(b) Determine the sample mean and variance of the excess of petroleum consumption over the natural gas consumption.

[4+6=10]

3. (a) Find the maximum likelihood estimates of the 2×1 mean vector μ and the 2×2 covariance matrix Σ based on the random sample

$$\mathbf{X} = \begin{bmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 4 & 7 \end{bmatrix}$$

from a bivariate normal populations.

(b) Let $\mathbf{X} \sim N_3(\mu, \Sigma)$, where $\mu' = [1, -1, 2]$ and $\mathbf{\Sigma} = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$. Are the following random variables independent? Justify

i. (X_1, X_3) and X_2

ii. X_1 and $X_1 + 3X_2 - 2X_3$

[6+4=10]

4. Prove that *Hotteling's* T^2 statistic is unchanged (invariant) under changes in the unit of measurements for $\mathbf{X}_{p\times 1}$ of the form

$$\mathbf{Y} = \mathbf{C}_{p\times p}\mathbf{X} + \mathbf{d}_{p\times 1},$$

where $\mathbf{C}$ is non-singular. [8]

5. 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. [12]

You may need following values:

$t_{41}(0.05) = 1.683, t_{41}(0.025) = 2.020, t_{41}(0.0125) = 2.327,$
$F_{2,40}(0.025) = 4.051, F_{40,2}(0.025) = 39.473, F_{2,40}(0.05) = 3.232, F_{40,2}(0.05) = 19.471$

This exam has total 5 questions, for a total of 50 points and 0 bonus points.

Best of luck!