Ramakrishna Mission Vivekananda University

Belur Math, Howrah, West Bengal

School of Mathematical Sciences, Department of Computer Science

M.Sc. in Big Data Analytic 2018-20, End Semester Exam — Date: 05 May 2019

Course : **DA310: Multivariate Statistics** — Time: 3 hrs

Instructor : Dr. Sudipta Das — *Max marks: 100*

Student signature and Id:

---

1. (a) You are given the following $n = 3$ observations on $p = 2$ variables:

$$\text{variable 1: } x_{11} = 2, x_{21} = 3, x_{31} = 4$$

$$\text{variable 2: } x_{12} = 1, x_{22} = 2, x_{32} = 4$$

   i. Plot the pairs of observations in the two-dimensional "variable space".

   ii. Plot the data as two points in the "item space".

[2+2=4]

   (b) Let

$$A = \begin{bmatrix} 9 & -2 \\ -2 & 9 \end{bmatrix},$$

   Write the spectral decomposition of A. Hence, find A^{-1}. [4+2=6]

2. Let $\mathbf{X} = [X_1 \ X_2 \ X_3]'$ be distributed as $N_3(\mu, \Sigma)$, where $\mu' = [1, -1, 2]$ and

$$\Sigma = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix},$$

   (a) Find ρ_{13}

   (b) Find the correlation between X_1 and $\frac{1}{2}X_2 + \frac{1}{2}X_3$.

   (c) What is the conditional distribution of X_1, given that $X_3 = x_3$.

[3+3+4=10]

3. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{100}$ be a random sample of size 100 from a four-variate normal distribution having mean μ and covariance Σ. Specify each of the following completely (mention the parameters).

   (a) Distribution of $\bar{\mathbf{X}}$

   (b) Distribution of $(\mathbf{X}_1 - \mu)'\Sigma^{-1}(\mathbf{X}_1 - \mu)$

   (c) Distribution of $n(\bar{\mathbf{X}} - \mu)'\Sigma^{-1}(\bar{\mathbf{X}} - \mu)$

   (d) Approximate distribution of $n(\bar{\mathbf{X}} - \mu)'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mu)$

   $\bar{\mathbf{X}}$ and $\mathbf{S}$ denote the sample mean and the sample variance of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{100}$, respectively. $[2 \times 4 = 8]$

4. Let $X_1, X_2, \ldots, X_5$ be a random sample from $N_2(\mu, \Sigma)$, where

$$\bar{\mathbf{x}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{\Sigma} = \begin{bmatrix} 3 & 0.1 \\ 0.1 & 1 \end{bmatrix}.$$

Test the hypothesis $H_0 : \mu = (0,0)'$ Vs $H_1 : \mu \neq (0,0)'$ at 5% level of significance. [12]

5. Perspiration from 20 healthy females was analyzed. Three components, X_1 = sweat rate, X_2 = sodium content and X_3 = potassium content, were measured. The sample mean and inverse of the sample covariance matrix are given as follows

$$\bar{\mathbf{x}} = \begin{bmatrix} 4.640 \\ 45.400 \\ 9.965 \end{bmatrix} \text{ and } \mathbf{S}^{-1} = \begin{bmatrix} 0.586 & -0.022 & 0.258 \\ -0.022 & 0.006 & -0.002 \\ 0.258 & -0.002 & 0.402 \end{bmatrix}.$$

Determine the axes of the 90% confidence ellipsoid for μ. Determine also the lengths of these axes. [12]

6. Construct a 2×2 non-trivial covariance matrix for which the variation explained by the first principal component remains same irrespective of considering the covariance or corresponding correlation matrix. [6]

7. Find the principal components and the proportion of the total population variance explained by each when the correlation matrix is

$$\rho = \begin{bmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{bmatrix}.$$

[10]

8. Consider an $m = 1$ factor model for the population with covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & .4 & .9 \\ .4 & 1 & .7 \\ .9 & .7 & 1 \end{bmatrix}.$$

Show that there is a unique choice of $\mathbf{L}$ and $\mathbf{\Psi}$ with $\mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi}$, but that $\psi_3 < 0$, so the choice is not admissible. [10]

9. Write short notes on the following

   (a) Ridge Regression

   (b) CART

[6+6=12]

10. Let

$$Y_{n\times 1} = Z_{n\times(p+1)}\beta_{(p+1)\times 1} + \epsilon_{n\times 1},$$

where $E(\epsilon) = 0_{n\times 1}$ and $Cov(\epsilon) = \sigma^2 I$. Show that the least square estimate of β is

$$\hat{\beta} = (Z'Z)^{-1}Z'Y.$$

Hence, prove that $Cov(\hat{\beta}) = \sigma^2(Z'Z)^{-1}$ and $Cov(\hat{\epsilon}) = \sigma^2[I - H]$, where $H = Z(Z'Z)^{-1}Z'$ and $\hat{\epsilon} = Y - Z\hat{\beta}$. [5+5+5=15]

You may need following values:

$t_{41}(0.05) = 1.683, t_{41}(0.025) = 2.020, t_{41}(0.0125) = 2.327,$
$\chi^2_2(0.05) = 5.99$
$F_{3,17}(.05) = 0.115, F_{3,17}(.1) = 0.193, F_{3,17}(.9) = 2.437, F_{3,17}(.95) = 3.197$

This exam has total 10 questions, for a total of 105 points and 0 bonus points.