Department Of Mathematics

Indian Institute of Technology Delhi

MAL 466 / MTL766 – Multivariate Statistical Analysis

Minor Examination

Time: 1hr 30 Mins | **Date 25-03-2017** | **Total Marks: 40**

Q1. Let X be a $n \times p$ data matrix. Suppose $\underset{\sim}{\bar{x}}$ ($p \times 1$) and S are the sample mean vector and sample covariance matrix, respectively, computed from X.

(a) Prove that S is positive definite if the columns of the mean corrected matrix $X - \underset{\sim}{1}\,\underset{\sim}{\bar{x}}^{T}$ are linearly independent.

$n \times p \quad - n\times 1 \times 1 \times p$

(b) Justify that $(\underset{\sim}{x} - \underset{\sim}{\bar{x}})^{T} S^{-1} (\underset{\sim}{x} - \underset{\sim}{\bar{x}})$ is a valid statistical distance measure of an observation vector $\underset{\sim}{x}$ from the data mean vector $\underset{\sim}{\bar{x}}$, whose value does not depend on the scales of measurement of components of $\underset{\sim}{x}$.

[4 + 4 = 8]

$a^T X \qquad x^T a$

Q2. Use the geometric interpretation in n-space of the generalised sample variance computed

from $n \times p$ data matrix X to justify that it gives a joint measure of variation of p-component variables of the measurement $\underset{\sim}{x}$.

What is a major weakness of this measure of joint variation?

[7]

Q3. Let $\underset{\sim}{X}^{T} = [X_1, X_2, X_3]$ be a multi-normal random vector having a $N_3(\mu, \Sigma)$ distribution. A random sample of size 10 on $\underset{\sim}{X}$ yielded the sample mean vector

$\underset{\sim}{\bar{x}}^{T} = [0.8, 0.5, 0.4]$ and sample covariance matrix $S = \begin{bmatrix} 0.85 & 0.63 & 0.17 \\ 0.63 & 0.57 & 0.13 \\ 0.17 & 0.13 & 0.17 \end{bmatrix}$

Use the above information to compute

(a) Maximum likelihood estimate of regression of $[X_1, X_2]$ on X_3.

(b) An unbiased estimate of covariance matrix Cov $\begin{bmatrix} X_1 - X_2 \\ X_1 - X_3 \end{bmatrix}$ $\qquad \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$

$E\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$s = \frac{1}{n-1} \Sigma (x_{ij} - \mu_j)$

μ_{11}

Q4. Let a population distribution be $N_p(\underset{\sim}{\mu}, \Sigma)$. Based on a random sample of size n from this population, develop a large sample approximate likelihood ratio test for the hypothesis $H_0 : \Sigma = \Sigma_0$ against $H_1 : \Sigma \neq \Sigma_0$, Σ_0 being a given positive definite matrix.

[7]

Q5. 42 observations on variables

$$X_1 = (\text{radiations with doors closed})^{1/4}$$

$$X_2 = (\text{radiations with doors opened})^{1/4}$$

taken on each of 42 randomly selected microwave ovens used by households. The following gives summary statistics of the data.

$$\bar{\underset{\sim}{x}} = \begin{bmatrix} 0.564 \\ 0.603 \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} 0.0144 & 0.0117 \\ 0.0117 & 0.0146 \end{bmatrix}$$

(a) Stating the assumptions made, test at 5% level of significance the hypothesis that $[\mu_1, \mu_2] = [0.57, 0.59]$ against any other values, where $\mu_1 = E(X_1)$ and $\mu_2 = E(X_2)$.

(b) Find the simultaneous 95% T^2- confidence intervals for μ_1, μ_2, and $\mu_1 - \mu_2$.

(c) Use the confidence intervals found in part (b) to test at 5% level of significance the hypothesis $\mu_1 = \mu_2 = 0.58$. Comment on the validity of the test for this hypothesis.

(d) If the sample size 42 is regarded as large in comparison with variable size $p = 2$, will the assumptions and conclusions of the test in part (a) change? If so how?

[4 + 3 + 2 + 2 = 11]