MACHAKOS UNIVERSITY

University Examinations 2021/2022 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

FOURTH YEAR SECOND SEMESTER EXAMINATION FOR

BACHELOR OF SCIENCE (STATISTICS AND PROGRAMMING)

BACHELOR OF SCIENCE (MATHEMATICS)

BACHELOR OF EDUCATION (SCIENCE)

SMA 466: MULTIVARIATE STATISTICS II

DATE: 17/12/2021 **TIME: 2:00 – 4:00 PM**

INSTRUCTION:

Answer Question One and Any Other Two Questions .

QUESTION ONE (30 MARKS)

a) Given five measurements on variables X_1, X_2 and X_3.

X_1	9	2	6	5	8
X_2	12	8	6	4	10
X_3	3	4	0	2	1

Obtain $\overline{\boldsymbol{X}}$, $\boldsymbol{S}_n$, and $\boldsymbol{R}$ (6 marks)

b) Let $\overline{\boldsymbol{X}}' = [X_1, X_2]$ be a random vector with mean vector $\boldsymbol{\mu}' = [\mu_1, \mu_2]$ and covariance matrix

$$\boldsymbol{\Sigma}_X = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

i. Find the means and covariance matrix for the linear combinations

$$Z_1 = X_1 - X_2$$

$$Z_2 = X_1 + X_2$$

ii. Interpret the covariance matrix $\boldsymbol{Z}' = [Z_1, Z_2]$ when $\sigma_{11} = \sigma_{22}$. (6 marks)

c) i Evaluate T^2 for testing H_0: $\boldsymbol{\mu}' = [7,11]$, (8 marks)

using the data $\boldsymbol{X} = \begin{bmatrix} 2 & 8 & 6 & 8 \\ 12 & 9 & 9 & 10 \end{bmatrix}$

iii. Specify the distribution of T^2 for the situation in (a).

iv. Using i) and ii), test H_0 at the $\alpha = 0.05$ level. What conclusion do you reach?

d) Consider the data matrix $\boldsymbol{X} = \begin{bmatrix} -1 & 2 & 5 \\ 3 & 4 & 2 \\ -2 & 2 & 3 \end{bmatrix}$. Determine $\mathbf{S}$ and calculate the generalized sample variance $|\mathbf{S}|$. Interpret $|\mathbf{S}|$ geometrically. (6 marks)

e) Are the following distance functions valid for the distance from the origin? Explain.

i. $x_1{}^2 + 4x_2{}^2 + x_1x_2 = (\text{distance})^2$

ii. $x_1{}^2 - 2x_2{}^2 = (\text{distance})^2$ (4 marks)

QUESTION TWO (20 MARKS)

a) A likelihood argument provides additional support for pooling the two independent sample covariance matrices to estimate a common covariance matrix in the case of normal populations. Give the likelihood function $L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma)$ for two independent sample sizes n_1 and n_2 from $N_p(\boldsymbol{\mu}_1, \Sigma)$ and $N_p(\boldsymbol{\mu}_2, \Sigma)$ populations, respectively. Show that this likelihood is maximized by the choices $\hat{\boldsymbol{\mu}}_1 = \overline{\boldsymbol{x}}_1$, $\hat{\boldsymbol{\mu}}_2 = \overline{\boldsymbol{x}}_2$ and (8 marks)

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n_1 + n_2}[(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2] = \left(\frac{n_1 + n_2 - 2}{n_1 + n_2}\right)\boldsymbol{S}_{\text{pooled}}$$

Hint: Use the exponential inequality: For $b > 0$ and a positive definite matrix B,

$$\frac{1}{|\Sigma|^b}\exp\left\{-\frac{1}{2}\text{trace}(\Sigma^{-1}B)\right\} \leq \frac{1}{|\text{B}|^b}(2b)^{bp}\exp\{-bp\}$$

for all positive definite Σ and equality is achieved when $\Sigma = \frac{B}{2b}$.

b) Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{20}$ be independent with sample size $n = 20$ from $N_6(\boldsymbol{\mu}, \ \Sigma)$ population. Specify each of the following completely (12 marks)

i. The distribution of $(\boldsymbol{X}_1 - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu})$.

ii. The distribution of $\overline{\boldsymbol{X}}$ and $\sqrt{n}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})$.

iii. The distribution of $(n-1)\boldsymbol{S}$.

iv. Find the marginal distributions for each of the random vectors

$$\boldsymbol{V}_1 = \frac{1}{4}\boldsymbol{X}_1 - \frac{1}{4}\boldsymbol{X}_2 + \frac{1}{4}\boldsymbol{X}_3 - \frac{1}{4}\boldsymbol{X}_4$$

and

$$\boldsymbol{V}_2 = \frac{1}{4}\boldsymbol{X}_1 + \frac{1}{4}\boldsymbol{X}_2 - \frac{1}{4}\boldsymbol{X}_3 - \frac{1}{4}\boldsymbol{X}_4$$

v. Find the joint density of the random vectors $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ defined in (d).

QUESTION THREE (20 MARKS)

a) Consider the covariance matrix (12 marks)

$$\boldsymbol{\Sigma} = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}$$

i. Determine the population principal components Y_1 and Y_2.

ii. Calculate the proportion of total population variance explained by the first principal component.

iii. Obtain the correlation matrix ρ.

iv. Determine the principal components Y_1 and Y_2 from ρ and compute the proportion of total population variance explained by Y_1. Compute the correlations ρ_{Y_1,Z_1} and ρ_{Y_1,Z_2}

b) Observation on two responses are collected for two treatments. The observation vectors $[x_1, x_2]'$ are (8 marks)

Treatment 1:	$[3,3]'$	$[1,6]'$	$[2,3]'$	
Treatment 2:	$[2,3]'$	$[5,1]'$	$[3,1]'$	$[2,3]'$

a) Calculate $\boldsymbol{S}_{\mathbf{pooled}}$.

b) Test $H_0: \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$ employing a two-sample approach with $\alpha = 0.01$.

c) Construct 95% simultaneous confidence intervals for the differences $\boldsymbol{\mu}_{1i}, \boldsymbol{\mu}_{2i},\ i = 1,2$.

QUESTION FOUR (20 MARKS)

a) Suppose that $\overline{\boldsymbol{X}} = (X_1, X_2, X_3)'$ is a random sample of size 3 from a population that has a trivariate normal distribution $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu} = (\mu_1, \mu_2, u_3)'$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & 0 \\ \rho^2 & 0 & 1 \end{pmatrix}$$

Let Y_1, Y_2, Y_3 be random variables defined as $Y_1 = X_1 + X_2$, $Y_2 = X_1 + X_2$ and $Y_3 = X_1 + X_2$.

i. Find the joint probability distribution of $Y_1, Y_2,$ and Y_3.

ii. State the conditional distribution of $Z_1 = Y_2$ given $\boldsymbol{Z}_2 = (Y_1, Y_3)'$. Hence find the conditional mean and conditional variance of Z_1 given $\boldsymbol{Z}_2$.

iii. Obtain the multiple regression function ofZ_1 on $\boldsymbol{Z}_2$ and hence find the corresponding regression coefficients.

iv. Compute the partial correlation coefficient of Y_1 and Y_3 when Y_2 is fixed. (14 marks)

b) Samples of size $n_1 = 45$ and $n_2 = 55$ were taken of Machakos homeowners with and without air conditioning. Two measurements of electrical usage (in Kilowatt hours) were considered. The first is a measure of total on-peak consumption (X_1) during July, and the second is a measure of total off-peak consumption (X_2) during July. The resulting summary statistics are

$$\boldsymbol{X}_1 = \begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix}, \qquad \boldsymbol{S}_1 = \begin{bmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{bmatrix}, n_1 = 45$$

$$\boldsymbol{X}_2 = \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix}, \qquad \boldsymbol{S}_1 = \begin{bmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{bmatrix}, n_2 = 55$$

Use Box's M-test to test the hypothesis $H_0: \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. Here, $\boldsymbol{\Sigma}_1$ is the covariance matrix for the two measures of usage for the population of Machakos homeowners with air conditions and is the electrical usage covariance matrix for the population of Machakos homeowners without air conditioning. Set $\alpha = 0.05$.

QUESTION FIVE (20 MARKS)

a) Consider the sets of points (x_1, x_2) whose "distance" from the origin are given by

$$c^2 = 4{x_1}^2 + 3{x_2}^2 + 2\sqrt{2x_1}x_2$$

for $c^2 = 1$ and for $c^2 = 4$. Determine the major and minor axes of the ellipse of constant distances and their associated length. What will happen as c^2 increases? (6 marks)

b) You are given the random vector $\mathbf{X}' = [X_1, X_2, X_3, X_4]$ with mean vector $\boldsymbol{\mu}' = [4,3,2,1]$ and variance-covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \begin{bmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 9 & -2 \\ 2 & 0 & -2 & 4 \end{bmatrix}$$

Partition $\mathbf{X}$ as

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^{(1)} \\ \cdots\cdots \\ \boldsymbol{X}^{(2)} \end{bmatrix}$$

Let $\mathbf{A} = (1, 2)$ and $\mathbf{B} = \begin{bmatrix} 1 & -2 \\ 2 & -1 \end{bmatrix}$ and consider the linear combinations $\mathbf{A}\boldsymbol{X}^{(1)}$ and $\mathbf{B}\boldsymbol{X}^{(2)}$. Find

i. $E\left(\boldsymbol{X}^{(1)}\right),\ E\left(\boldsymbol{X}^{(2)}\right), E\left(\mathbf{A}\boldsymbol{X}^{(1)}\right), E\left(\mathbf{B}\boldsymbol{X}^{(2)}\right),$

ii. $\mathrm{Cov}\left(\boldsymbol{X}^{(1)}\right),\ \mathrm{Cov}\left(\boldsymbol{X}^{(2)}\right), \mathrm{Cov}\left(\mathbf{A}\boldsymbol{X}^{(1)}\right), \mathrm{Cov}\left(\mathbf{B}\boldsymbol{X}^{(2)}\right),$

iii. $\mathrm{Cov}\left(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}\right), \mathrm{Cov}\left(\mathbf{A}\boldsymbol{X}^{(1)}, \mathbf{B}\boldsymbol{X}^{(2)}\right).$ (14 marks)