MACHAKOS UNIVERSITY

University Examinations for 2022/2023 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

THIRD YEAR FIRST/SECOND SEMESTER EXAMINATION FOR

BACHELOR OF SCIENCE (MATHEMATICS AND COMPUTER SCIENCE)

BACHELOR OF SCIENCE (STATISTICS AND PROGRAMMING)

BACHELOR OF SCIENCE (ECONOMICS AND STATISTICS)

BACHELOR OF SCIENCE (ACTUARIAL SCIENCE)

BACHELOR OF SCIENCE (MATHEMATICS)

BACHELOR OF EDUCATION (SCIENCE)

SMA 360 MULTIVARIATE STATISTICAL METHODS I

DATE: **TIME:**

INSTRUCTIONS TO CANDIDATES:

Answer Question One and Any Other Two Questions

QUESTION ONE (COMPULSORY) (30 MARKS)

a) Define the p-dimensional multivariate normal density. (2 marks)

b) State four properties of the multivariate normal distribution. (4 marks)

c) Suppose $\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and given that $\boldsymbol{Y_1} = \boldsymbol{X_1} - \boldsymbol{X_2}$ and $\boldsymbol{Y_2} = \boldsymbol{X_1} - \boldsymbol{X_3}$**.** Determine

i. $E(\boldsymbol{Y})$ (2 marks)

ii. $Var(\boldsymbol{Y})$ (4 marks)

d) Given two random variables X and Y with joint and marginal probabilities as below, determine the variance covariance matrix.

	0	1	P(x)
-1	0.24	0.06	0.3
0	0.16	0.14	0.3
1	0.40	0.00	0.4
P(y)	0.8	0.2	1

(8 marks)

e) Given the covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}$**,** determine the correlation matrix ρ.

(5 marks)

f) Let **X** have the covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Determine

i) Σ^{-1} (1 mark)

ii) The eigenvalues and eigenvectors of Σ. (2 marks)

iii) The eigenvalues and eigenvectors of Σ^{-1}. (2 marks)

QUESTION TWO (20 MARKS)

a) Consider the data matrix $\boldsymbol{X} = \begin{bmatrix} 1 & 4 & 3 \\ 6 & 2 & 6 \\ 8 & 3 & 3 \end{bmatrix}$ and the linear combinations $\boldsymbol{b}'\boldsymbol{X} = [1 \quad 1 \quad 1]\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ and $\boldsymbol{c}'\boldsymbol{X} = [1 \quad 2 \quad -3]\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$. Evaluate the sample means, variances and covariance of $\boldsymbol{c}'\boldsymbol{X}$ and $\boldsymbol{b}'\boldsymbol{X}$. (6 marks)

b) Let $\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\mu' = [1 \quad -1 \quad 2]$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$. Determine whether the following random variables are independent. Explain.

i. X_1 and X_2

ii. X_1 and X_3

iii. X_2 and X_3

iv. (X_1, X_3) and X_2

v. X_1 and $3X_1 - 2X_2 + X_3$ (6 marks)

c) Energy consumption in 2001, by state, from major sources $x_1 = petroleum$, $x_2 = natural\ gas$, $x_3 = hydroelectric\ power$, $x_4 = nuclear\ electric\ power$ is recorded in quadrillions of BTUs. The resulting mean and covariance matrix are

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

i) Using the summary statistics, determine the sample mean and variance of a state's total energy consumption for these major sources. (4 marks)

ii) Determine the sample mean and variance of the excess of petroleum consumption over natural gas consumption. Determine the sample covariance of this variable with the total variable in part i) above. (4 marks)

QUESTION THREE (20 MARKS)

a) Let **X** be $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}' = [2 \quad -3 \quad 1]$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}$. Determine the distribution of $3X_1 - 2X_2 + X_3$. (5 marks)

b) Determine the maximum likelihood estimates of the 2 x 1 mean vector $\boldsymbol{\mu}$ and the 2 x 2 covariance matrix $\boldsymbol{\Sigma}$ based on the random sample

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

from a bivariate normal population. (5 marks)

c) Given that $f(x_1, x_2)$ is the bivariate normal density, show that $f(x_1|x_2)$ is $N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}{}^2}{\sigma_{22}}\right)$. (10 marks)

QUESTION FOUR (20 MARKS)

a) Consider a bivariate normal distribution with $\mu_1 = 0$, $\mu_2 = 1$, $\sigma_{11} = 2$, $\sigma_{22} = 1$ and $\rho_{12} = 0.5$.

i. Write out the bivariate normal density. (8 marks)

ii. Write out the squared statistical distance expression as a function of x_1 and x_2. (2 marks)

b) Let $\boldsymbol{X_1}, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X_4}, \boldsymbol{X_5}$be independent and identically distributed random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ **.**

i. Determine the mean vector and covariance matrices for each of the two linear combinations of random vectors

$\frac{1}{5}\boldsymbol{X_1} + \frac{1}{5}\boldsymbol{X}_2 + \frac{1}{5}\boldsymbol{X}_3 + \frac{1}{5}\boldsymbol{X_4} + \frac{1}{5}\boldsymbol{X_5}$ and $\boldsymbol{X_1} - \boldsymbol{X}_2 + \boldsymbol{X}_3 - \boldsymbol{X_4} + \boldsymbol{X_5}$

in terms of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. (8 marks)

ii. Determine the covariance between the two linear combinations of random vectors. (2 marks)

QUESTION FIVE (20 MARKS)

a) Suppose **y** and **x** are subvectors such that y is 2x1 and **x** is 3x1, with **μ** and $\boldsymbol{\Sigma}$ partitioned accordingly:

$$\mu = \begin{pmatrix} 3 \\ -2 \\ 4 \\ -3 \\ 5 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} 14 & -8 & 15 & 0 & 3 \\ -8 & 18 & 8 & 6 & -2 \\ 15 & 8 & 50 & 8 & 5 \\ 0 & 6 & 8 & 4 & 0 \\ 3 & -2 & 5 & 0 & 1 \end{pmatrix}$$

Assume that $\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix}$ is distributed as $N_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Determine

i. $E(\boldsymbol{y}|\boldsymbol{x})$ (5 marks)

ii. $cov(\boldsymbol{y}|\boldsymbol{x})$ (5 marks)

b) Let $\boldsymbol{X_1}, \boldsymbol{X}_2, \ldots, \boldsymbol{X_n}$ be a random sample from a joint distribution that has mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$**.** Prove that $\overline{\boldsymbol{X}}$ is an unbiased estimator of $\boldsymbol{\mu}$ and its covariance matrix is $\frac{1}{n}$ $\boldsymbol{\Sigma}$**.** (10 marks)