MACHAKOS UNIVERSITY

University Examinations for 2022/2023

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

FOURTH YEAR SPECIAL/SUPPLEMENTARY EXAMINATION FOR

BACHELOR OF SCIENCE (STATISTICS AND PROGRAMMING)

SMA 466: MULTIVARIATE STATISTICAL METHODS II

DATE: 7/3/2023 **TIME: 11:00 – 1:00 PM**

INSTRUCTION: *Answer Question* ***ONE*** *which is compulsory and any other* ***TWO*** *Questions*

QUESTION ONE (COMPULSORY) (30 MARKS)

a) Briefly explain the following terms

i) Multivariate Analysis (2 marks)

ii) Principal Component Analysis (2 marks)

b) Give three importance of Multivariate Normal Distribution. (3 marks)

c) If $\boldsymbol{y}$ is a random vector taking on any possible value in a multivariate population, show that the population covariance matrix is defined as

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix}$$ (5 marks)

d) Suppose that $\boldsymbol{X}$ has mean zero and covariance $\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$. Let $Y = X_1 + X_2$. Write Y as a linear transformation, that is, find the transformation matrix $\boldsymbol{A}$. Then compute $Var(Y)$. (5 marks)

e) Suppose $\boldsymbol{y}$ is $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 4 & -3 & 0 \\ -3 & 6 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

Determine whether:

i) y_1 and y_3 are independent (2 marks)

ii) (y_1, y_3) and y_2 are independent (3 marks)

f) Assume $\boldsymbol{y}$ and $\boldsymbol{x}$ are sub vectors, each 2×1, where $\begin{pmatrix} y \\ x \end{pmatrix}$ in $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \left(\begin{array}{c} 2 \\ -1 \\ \hline 3 \\ 1 \end{array}\right), \qquad \Sigma = \left(\begin{array}{cc|cc} 7 & 3 & -3 & 2 \\ 3 & 6 & 0 & 4 \\ \hline -3 & 0 & 5 & -2 \\ 2 & 4 & -2 & 4 \end{array}\right)$$

Find:

(i) $E(y|x)$ (8 marks)

QUESTION TWO (20 MARKS)

a) Suppose y is $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} -2 \\ 3 \\ -1 \\ 5 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 11 & -8 & 3 & 9 \\ -8 & 9 & -3 & 6 \\ 3 & -3 & 2 & 3 \\ 9 & 6 & 3 & 9 \end{pmatrix}$$

Find the distribution of $z = 4y_1 - 2y_2 + y_3 - 3y_4$ (7 marks)

b) Given the following data from a bivariate normal distribution,

$$X = \begin{pmatrix} 24 & 22 & 20 & 17 & 27 & 21 & 22 & 19 & 17 \\ 21 & 24 & 29 & 24 & 22 & 20 & 25 & 26 & 25 \end{pmatrix}$$

test at $\alpha = 0.05$ level of significance the hypothesis: $H_0 : \mu = (24 \quad 19)'$ versus $H_1 : \mu \neq (24 \quad 19)'$ (13 marks)

QUESTION THREE (20 MARKS)

a) Give brief explanation of any three properties of the Multivariate normal distribution. (6 marks)

b) Given a random vector $\mathbf{X} = (X_1, X_2, X_3)'$ with mean vector $\boldsymbol{\mu} = (1,2,3)'$ and covariance matrix

$$\Sigma = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 9 \end{pmatrix}$$

Compute the following:

(i) The regression function of X_1 on X_2 and X_3. (5 marks)

(ii) the multiple correlation coefficient between X_1 and (X_2, X_3). (4 marks)

c) Given that $X \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\mu = \begin{pmatrix} 60 \\ 48 \\ 24 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 9 & 6 & \frac{12}{5} \\ 6 & 36 & -12 \\ \frac{12}{5} & -12 & 16 \end{pmatrix}$

Find the conditional variance of X_1 when given X_2 and X_3 (5 marks)

QUESTION FOUR (20 MARKS)

a) Explain two properties of the canonical correlations. (4 marks)

b) The data below consist of head measurements on first and second sons. Define y_1 and y_2 as the measurements on the first son and x_1 and x_2 for the second son.

First Son		Second Son	
Head Length	Head Breadth	Head Length	Head Breadth
y_1	y_2	x_1	x_2
191	155	179	145
181	148	185	149
176	144	171	142
189	150	190	149
188	152	197	159
179	158	186	148
174	150	185	152
188	151	187	158
195	155	183	158
181	145	182	146

(i) Find the mean vector for all four variables to two decimal place and partition it into $\begin{pmatrix} \bar{\mathbf{y}} \\ \bar{\mathbf{x}} \end{pmatrix}$. (4 marks)

(ii) Given that

$\sum y_1^2 = 339730, \sum y_2^2 = 227584, \sum x_1^2 = 340819, \sum x_2^2 = 227124, \sum y_1 y_2 = 277903,$
$\sum y_1 x_1 = 339983, \sum y_1 x_2 = 277589, \sum y_2 x_1 = 278323, \sum y_2 x_2 = 227195, \sum y_1 x_1 = 278115$

The covariance matrix to two decimal places for all four variables and partition it into $\begin{pmatrix} \mathbf{S}_{\mathbf{yy}} & \mathbf{S}_{\mathbf{yx}} \\ \mathbf{S}_{\mathbf{xy}} & \mathbf{S}_{\mathbf{xx}} \end{pmatrix}$ (12 marks)

QUESTION FIVE (20 MARKS)

a) State the assumptions for using MANOVA (4 marks)

b) Let $X \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a tri-variate normal random vector. Suppose a certain sample gave

$$\boldsymbol{\Sigma} = \begin{pmatrix} 64 & 0 & 8 \\ 0 & 4 & 0 \\ 8 & 0 & 16 \end{pmatrix}$$

Find:

i) The eigenvalues of this matrix give your values in radical form where possible. (7 marks)

ii) The first principal component. (6 marks)

iii) The total variance explained by the component in percentage to two decimal place. (3 marks)