Kasarani Campus
Off Thika Road
P. O. Box 49274, 00101
NAIROBI
Westlands Campus
Pamstech House
Woodvale Grove
Tel. 4442212
Fax: 4444175

KIRIRI WOMENS' UNIVERSITY OF SCIENCE AND TECHNOLOGY
UNIVERSITY EXAMINATIONS, 2024/2025 ACADEMIC YEAR
FOURTH YEAR, SECOND SEMESTER EXAMINATION
FOR THE DEGREE OF BACHELOR OF SCIENCE (MATHEMATICS)

KMA 2416: MULTIVARIATE METHODS II

DATE: 4TH DECEMBER 2024
Time: 11:30AM – 1:30PM

INSTRUCTIONS TO CANDIDATES
ANSWER QUESTION ONE (COMPULSORY) AND ANY OTHER TWO QUESTIONS

QUESTION ONE: COMPULSORY (30 MARKS)

a) A random sample of size 10 was obtained from a bivariate normal population with mean vector $\underline{\mu}$ and variance-covariance matrix Σ_0 (known) where

$\Sigma_0 = \begin{bmatrix} 4 & 4.2 \\ 4.2 & 9 \end{bmatrix}$, $\underline{\mu}_0 = (6 \quad 5)'$. Given that $\overline{\underline{X}} = (5.8, \ 5.2)'$, carry out the test at $\alpha = 0.01$ level of significance for

$$H_0 : \underline{\mu} = \underline{\mu}_0$$
$$H_1 : \underline{\mu} \neq \underline{\mu}_0$$

(5 Marks)

b) Consider the following multivariate data from two groups. The sample data consists of measurements on two variables (e.g., height and weight) for two different groups:

Group 1 (Sample Size $n_1 = 5$) Group 2 (Sample Size $n_1 = 5$)

$$\mathbf{X} = \begin{pmatrix} 1.5 & 65 \\ 1.6 & 70 \\ 1.7 & 75 \\ 1.8 & 80 \\ 1.9 & 85 \end{pmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{pmatrix} 1.4 & 60 \\ 1.5 & 65 \\ 1.6 & 70 \\ 1.7 & 75 \\ 1.8 & 80 \end{pmatrix}$$

You are required to conduct a quadratic form analysis to evaluate whether there is a significant difference between the mean vectors of the two groups by performing the following tasks:

i. Compute the sample mean vectors $\overline{\underline{\mathbf{X}}}$ and $\overline{\underline{\mathbf{Y}}}$. **(4 Marks)**

ii. Calculate the sample covariance matrices $\mathbf{S}_X$ and $\mathbf{S}_Y$ **(6 Marks)**

iii. Compute the pooled covariance matrix $\mathbf{S}_X$. **(3 Marks)**

iv. Construct the quadratic form $\mathbf{Q} = \left(\overline{\underline{\mathbf{X}}} - \overline{\underline{\mathbf{Y}}}\right)^T \mathbf{S}_P^{-1} \left(\overline{\underline{\mathbf{X}}} - \overline{\underline{\mathbf{Y}}}\right)$ and interpret the result

(3 Marks)

c) Consider the following dataset consisting of two variables *X* and *Y* collected from a sample of 5 observations to be used in Principal Component Analysis (PCA):

X	2	3	5	8	9
Y	3	4	7	10	12

Using the data provided above:

i. Construct the sample covariance matrix of the dataset. **(4 Marks)**

ii. Calculate the eigenvalues of the covariance matrix. **(5 Marks)**

QUESTION TWO: (20 MARKS)

a) Let $X_1, X_2, \ldots\ldots X_n$ be i.i.d random variables with mean μ and variance σ^2. Given that $Q = (X_1 - X_2)^2 + (X_2 - X_3)^2 + \ldots\ldots\ldots (X_{n-1} - X_n)^2$. Find:

i. Expectation of Q, $E(Q)$ **(6 Marks)**

ii. Hence or otherwise, find the unbiased estimator of σ^2 **(2 Marks)**

b) A researcher collected data on two types of fruits based on their weight (grams) to be used in discriminant analysis. The data for Type A and Type B is given below:

Type A	150	160	155	162	158
Type B	130	135	128	140	138

Using the provided data, perform the following tasks:

i. Obtain the sample mean and variance for each type of fruit. **(4 Marks)**

ii. Construct the linear discriminant function for this problem. **(5 Marks)**

iii. Classify a new fruit with a weight of 145 grams using the discriminant function. **(3 Marks)**

QUESTION THREE: (20 MARKS)

a) In an experiment involving two correlated variables, the following statistics were obtained $\underline{\overline{X}} = \begin{pmatrix} \overline{X}_1 \\ \overline{X}_2 \end{pmatrix} = \begin{pmatrix} 10.0 \\ 10.0 \end{pmatrix}$ and $S = \begin{bmatrix} 0.7986 & 0.6793 \\ 0.6793 & 0.7343 \end{bmatrix}$

Determine:

i. The principal components for this problem **(6 Marks)**

ii. The percentage variance explained by each component. **(3 Marks)**

iii. Correlations of the principal components with the original variables **(5 Marks)**

b) Let $\underline{X}$ be a p-variate random vector with mean vector $\underline{\mu}$ and variance-covariance matrix $\Sigma = ((\sigma_{ij})i, j = 1, 2, \ldots\ldots, p)$. Let A be a symmetric matrix such that the quadratic form on $\underline{X}$ is given by $Q = \underline{X}' A \underline{X}$. Show that $E\left(\underline{X}' A \underline{X}\right) = Trace(A\Sigma) + \underline{\mu}' A \underline{\mu}$ **(6 Marks)**

QUESTION FOUR: (20 MARKS)

a) Let $\underline{X}_1, \underline{X}_2, \underline{X}_3, \ldots\ldots, \underline{X}_n$ be *i.i.d* $N_p(\underline{\mu}, \Sigma)$, where Σ is unknown.

Given $\underline{\overline{X}} = (160.53, 23.81)'$ and $S = \begin{bmatrix} 32.95 & 1.78 \\ 1.78 & 1.77 \end{bmatrix}$, $n = 15$. Carry out a $\alpha = 0.05$ level test for $H_0 : \underline{\mu} = \underline{\mu}_0 = (164.51 \;\; 25.49)'$ against $H_1 : \underline{\mu} \neq \underline{\mu}_0$ **(7 Marks)**

b) Explain how the following tests are used to decide on the number of principal components to be retained

i. Caeser's Test

ii. Catell's Scree Test

iii. Bartlett's Test **(6 Marks)**

c) Two bivariate normal populations are mixed together. It was later decided that the two populations be separated. The parameters of the two distributions are

$P_1 : \underline{X} \sim N_2(\underline{\mu}_1, \Sigma)$ and $\underline{\mu}_1 = (6.2\,,\ 3.8)'$

$P_2 : \underline{X} \sim N_2(\underline{\mu}_2, \Sigma)$ and $\underline{\mu}_2 = (5.8\,,\ 3.5)'$ and $\Sigma = \begin{bmatrix} 25 & 16 \\ 16 & 16 \end{bmatrix}$

Using this information,

i. Construct the optimal linear discriminant rule **(5 Marks)**

ii. Hence or otherwise, classify a new observation $\underline{X} = (6.0\,,\ 3.4)'$ **(2 Marks)**

QUESTION FIVE: (20 MARKS)

a) Given $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, and $f(\underline{X}) = C\exp\left\{-\frac{1}{2}Q\right\}$ where Q has the usual meaning. Show that $E(Q) = p$ **(4 Marks)**

b) The data below refers to nutritional contents of three diets. Variables measured are

$\underline{\mathbf{Y}} = [\mathbf{Y}_1\,(Ascorbic), \mathbf{Y}_2\,(Riboflavin)]'$

Diet A ($n_1 = 2$)	Diet B ($n_2 = 3$)	Diet C ($n_3 = 3$)
$\begin{bmatrix} 0.25 \\ 1.50 \end{bmatrix}, \begin{bmatrix} 0.59 \\ 1.78 \end{bmatrix}$	$\begin{bmatrix} 1.85 \\ 2.90 \end{bmatrix}, \begin{bmatrix} 3.50 \\ 4.00 \end{bmatrix}, \begin{bmatrix} 1.80 \\ 3.15 \end{bmatrix}$	$\begin{bmatrix} 0.74 \\ 0.95 \end{bmatrix}, \begin{bmatrix} 1.25 \\ 1.80 \end{bmatrix}, \begin{bmatrix} 0.95 \\ 1.55 \end{bmatrix}$

i. Write down appropriate statistical model for analyzing this data **(3 Marks)**

ii. Find the between groups (B) and within groups (W), SS and CP Matrices **(6 Marks)**

iii. Form a MANOVA Table **(4 Marks)**

iv. Test for equality of diet content at 0.1 level of significance **(3 Marks)**