UNIVERSITY EXAMINATIONS
2022/2023 ACADEMIC YEAR
YEAR FOUR SEMESTER ONE EXAMINATIONS
FOR THE DEGREE OF
BACHELOR OF SCIENCE

COURSE CODE: STA 421
COURSE TITLE: MULTIVARIATE DATA ANALYSIS

DATE: 10/08/2023 **TIME: 8:00AM-10:00AM**

INSTRUCTIONS TO CANDIDATES

ANSWER **QUESTIONS ONE** AND **ANY OTHER TWO**.

TIME: 2 HOURS

QUESTION ONE

[30 MARKS]

(a) (i) What is a mean vector [1 mark]

(ii) Describe how multivariate data are arranged [2 marks]

(b) The data below shows the scores of a sample of 15 students in mathematics, English and Kiswahili CATS in a certain school

$$\boldsymbol{X} = \begin{bmatrix} 4 & 8 & 6 & 8 & 9 \\ 8 & 7 & 4 & 4 & 10 \\ 10 & 9 & 7 & 5 & 9 \end{bmatrix}$$

Obtain

(i) Mean Vector **[3 marks]**

(ii) Variance-Covariance matrix **[5 marks]**

(iii) Correlation matrix **[3 marks]**

(c) Let $\underline{x} = [5,1,3]'$ and $\underline{y} = [-1,3,1]'$ Find

(i) The length of $\underline{x}$ **[1mark]**

(ii) The angle between $\underline{x}$ and $\underline{y}$ **[2marks]**

(iii) The length of the projection of $\underline{x}$ on $\underline{y}$ **[1mark]**

(d) A random sample of 10 was obtained from a bivariate normal population with mean vector $\boldsymbol{\mu}$ and a known variance-covariance matrix $\Sigma_0 = \begin{bmatrix} 4 & 4.2 \\ 4.2 & 9 \end{bmatrix}$ Find the principal component and hence Test at $\alpha = 0.01$ level of significance for $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ where $\mu_0 = 6.5$ and the sample mean vector is $\overline{x} = (5.8, 5.2)'$ **[8 marks]**

(e) If $\underline{a}' = (a_1, a_2, \ldots, a_p)$ is a non-zero vector of order p and $y = \underline{a'X}$. Show that

i) $E(\underline{a'X}) = \underline{a'\mu}$ **[1 marks]**

ii) $Var(\underline{a'X}) = \underline{a'}\Sigma\underline{a}$ **[3 marks]**

QUESTION TWO

[20MARKS]

a) Let $X_1, X_2 \ldots X_n$ be random sample from a joint distribution which has mean vector μ and covariance matrix $\frac{\Sigma}{n}$. Show that sample mean, $\overline{X}$ is unbiased; i.e. $E(\overline{X}) = \mu$ and $\frac{n}{n-1}\Sigma$ is the unbiased estimator of the sample variance S_n **[6 marks]**

b) Let $\Sigma = \begin{bmatrix} 25 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 4 \end{bmatrix}$ **[5 marks]**

Determine the principal component Y_1, Y_2, Y_3. Comment on the eigenvectors (and principal components) associated with eigenvalues that are not distinct?

c) Find all the Eigen-values of A, where, $A = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ **[6 marks]**

d) Show that A is a positive definite matrix **[4 marks]**

QUESTION THREE **[20 MARKS]**

(a) Consider the random variable $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \sim N\left[\begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 & 0 & 1 & 3 \\ 0 & 4 & 1 & 1 \\ 1 & 1 & 3 & 1 \\ 3 & 1 & 1 & 9 \end{bmatrix}\right]$. Find the conditional distribution of $\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ given $X_2 = x_2$ and $X_4 = x_4$. [8 marks]

(b) For a bivariate normal distribution, use the data below to test at $\alpha = 0.05$ level the hypothesis

$$H_0: \boldsymbol{\mu} = (3.4, 6)' \quad Vs$$
$$H_1: \boldsymbol{\mu} = (3.4, 6)'$$

$$\underline{X} = \begin{bmatrix} 3 & 4 & 5 & 6 & 2 \\ 9 & 5 & 7 & 2 & 8 \end{bmatrix}$$ **[7 marks]**

(c) Let $\underline{x}$ be a random vector having the covariance matrix

$$\Sigma = \begin{Bmatrix} 4 & 1 & 0 \\ 1 & 16 & -5 \\ 0 & -5 & 36 \end{Bmatrix}$$

Obtain

(i) Square root of $\Sigma = \left(V^{\frac{1}{2}}\right)$ [2 marks]

(ii) Inverse of the square root $\Sigma = \left(V^{\frac{1}{2}}\right)^{-1}$ **[1 mark]**

(iii) Correlation matrix ρ defined by

$$\rho = \left(V^{\frac{1}{2}}\right)^{-1} \Sigma \left(V^{\frac{1}{2}}\right)^{-1}$$ **[2marks]**

QUESTION FOUR **[20 MARKS]**

a.) Consider the following $n = 7$ observations on $p = 2$ variables

x_1	3	4	2	6	8	2	5
x_2	5	5.5	4	7	10	5	7.5

(i) Compute the sample means $\bar{x}_1$ and $\bar{x}_2$ and the sample variances S_{11} and S_{22} (4marks)

(ii) Compute the sample covariance S_{12} and the sample correlation coefficient r_{12} and interpret these quantities (5marks)

(iii) Display the sample mean array $\bar{x}$, the sample correlation array R and the sample variance-covariance S_{12} (3marks)

b.) A researcher considered five companies, x_1, x_2, x_3, x_4, x_5 of Uchumi, Tuskys, Union Carbide, Eveready and Total for weekly rates of returns respectively. The means and correlation matrix, **R** are given below:

$$\overline{X} = \begin{bmatrix} 0.0054 \\ 0.0048 \\ 0.0057 \\ 0.0063 \\ 0.0037 \end{bmatrix} \text{ and } R = \begin{bmatrix} 1.000 & 0.577 & 0.509 & 0.387 & 0.462 \\ 0.577 & 1.000 & 0.599 & 0.389 & 0.322 \\ 0.509 & 0.599 & 1.000 & 0.436 & 0.426 \\ 0.387 & 0.389 & 0.436 & 1.000 & 0.523 \\ 0.462 & 0.322 & 0.426 & 0.523 & 1.000 \end{bmatrix}$$

The eigenvalues and corresponding normalized eigenvectors of R were determined by a computer and are given below:

$\hat{\lambda}_1 = 2.857, \hat{\gamma}_1 = [0.464, 0.457, 0.470, 0.421, 0.421]$

$\hat{\lambda}_1 = 0.809, \hat{\gamma}_2 = [0.240, 0.509, 0.260, -0.526, -0.582]$

$\hat{\lambda}_3 = 0.540, \hat{\gamma}_3 = [-0.612, 0.178, 0.335, -0.541, -0.435]$

$\hat{\lambda}_4 = 0.452, \hat{\gamma}_4 = [0.387, 0.206, -0.662, 0.472, -0.382]$

$\hat{\lambda}_5 = 0.343, \hat{\gamma}_5 = [-0.451, 0.676, -0.400, -0.1761, 0.385]$

i) Write down principal components that accounts for the communality of at least 73% of variations.

[5 marks]

ii) Interpret the results of the above PCA as fully as possible in terms of data characterization **[3 marks]**

QUESTION FIVE **[20 MARKS]**

(a) Find the maximum likelihood estimators of the mean vector $\underline{\mu}$ and covariance matrix Σ based on the data matrix (6marks)

$$x = \begin{bmatrix} 42 & 4 \\ 52 & 5 \\ 48 & 4 \\ 58 & 3 \end{bmatrix}$$

(b) Given the data matrix $x = \begin{bmatrix} 1 & 2 & 5 \\ 4 & 1 & 6 \\ 4 & 0 & 4 \end{bmatrix}$

Define $X_c = X - 1\,\overline{x}\,'$ as the mean corrected data matrix.

(i) Obtain the mean corrected data matrix (4marks)

(ii) Obtain the sample covariance matrix (4marks)

(iii) The generalized variance and hence verify that columns of mean corrected data matrix are linearly dependent. (3marks)

(iv) Specify a vector $a' = [a_1\ a_2 a_3]$ that establishes the linear dependence (3marks)