MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

P.O. Box 972-60200 – Meru-Kenya.
Tel: +254(0) 799 529 958, +254(0) 799 529 959, +254 (0)712 524 293
Website: www.must.ac.ke Email: info@mucst.ac.ke

UNIVERSITY EXAMINATIONS 2023/2024

FOURTH YEAR SECOND SEMESTER EXAMINATION FOR DEGREE OF BACHELOR OF SCIENCE IN STATISTICS

SMS 3456: MULTIVARIATE METHODS

DATE: APRIL 2023 **TIME: 2 HOURS**

INSTRUCTIONS: Answer Question ONE and any other TWO questions.

QUESTION ONE (30 MARKS)

a. Briefly explain 5 objectives of Multivariate methods. (5marks)

b. the following data

Variable 1	x_1	5	4	6	2	2	8	3
Variable 2	x_2	5	5.5	4	7	10	5	7.5

a) For this dataset plot 0

i) Dot plot. (2marks)

ii) Scatter plot. (2marks)

b) Interpret the plots. (2marks)

c. A set of paired measurements $(x_{1,}x_2)$ on two variables yields

$\bar{x}_1 = \bar{x}_2$Y2 = 0, $s_{11}4$ and $s_{22=1.}$ Plot an ellipse of unit distance. (5marks)

d. Given the vectors $x^1 = [1,3,2]$and $y^1 = [-2,1,-1]$Determine.

i) The length of x and y . (2marks)

ii) ii. The angle between x and y . (3marks)

e. Suppose $p = 2\ and\ n = 1.$ Consider the random vector $X^l = [(X_1, X_2)]$where X_1 and X_2 have the probability functions.

x_1	-1	0	1
$p_1(x_1)$	0.3	0.3	0.4

x_2	0	1	
$p_2(x_2)$	0.8	0.2	

Determine the mean of vector x (4marks)

f. Let b and d be any two $px1$ vectors. Then show that $\left(\underset{\sim}{b}^1 \underset{\sim}{d}\right)^2 \leq \left(\underset{\sim}{b}^1 \underset{\sim}{b}\right)\left(\underset{\sim}{d}^1 \underset{\sim}{d}\right)$

With equality iff $\underset{\sim}{b} = \underset{\sim}{c}\,\underset{\sim}{d}$ $\left(or\ \underset{\sim}{d} = c\,\underset{\sim}{b}\right)$for some constant c. (5marks)

QUESTION TWO (20 MARKS)

a. Show that $3x_1^2 + 2x_2^2 - 2\sqrt{2}X_1, X_2$ is positive definite. (l0marks)

b. Suppose $\underset{\sim}{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$. Where $X_1 and\ X_2$ has the data given as

$\frac{\alpha_2}{x_1}$	0	1
-1	O .24	0.6
0	0.16	0.14
1	0.40	0.00

Determine

i. The mean of $\underset{\sim}{X}$ (3marks)

ii. The variate of $\underset{\sim}{X}$ (7marks)

QUESTION THREE (20 MARKS)

a. Let X^l = [Xl,X2] be a random vector with mean vector $\underset{\sim}{\mu}^1 = [\mu_1 \mu_2]$and variable covariable matrix

$$\sum_x = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$$

Find the mean vector and covariance matrix for the linear combination

$$Z_1 = X_1 - X_2$$
$$Z_2 = X_1 + X_2$$

in terms of $\underset{\sim}{\mu}_x$ $and\ \sum_x$. (4marks)

b. Given the data matrix

$$X = \begin{bmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 8 \end{bmatrix}$$

i. Plot $n = 3$ data points in $p = 2$ space and locate $\bar{x}$ on the resulting diagrams.

ii. Plot the data as $p = 2$vectors in $n = 3$space. (5marks)

c. Let $x$be distributed as $N_p\left(\mu, \underset{\sim}{\Sigma}\right)$with $\left|\Sigma\right| > 0$. Show that $(\underset{\sim}{x} - \underset{\sim}{\mu})^1 \Sigma^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu})$ is distributed as λ_p^2 when λ_p^2 denotes the chi- square distribution with p degrees of Freedom.

The world's 10 largest companies yield the following data.

COMPANY	x_1= sales	$x_2 = profits$	$x_3 = assets$
Cila group	108.25	17.05	
General Electric	152.36	16.59	
America intel	95.04	10.91	
Bank of America	65.45	14.14	
HSBC Group	62.95	9.52	
Exxomobile	263.99	25.33	
Royal Sunol	265.19	18.54	
BP	285.06	15.73	
LNG Group	92.01	8.16	
Toyota	165.68	11.13	

i. Compute $\bar{x}_1, \bar{x}_2, s_{11}, s_{12}$ (5marks)

ii. Test for normality for this data (6marks)

QUESTION FOUR (20 MARKS)

a) Let $\underset{\sim}{x} = \begin{bmatrix} x_1 \\ \underset{\sim}{x_2} \end{bmatrix}$ be distributed as $N_p\left(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}\right)$ with $\underset{\sim}{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ And $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$

Show that the conditional distribution of $\underset{\sim}{x_1}$ given $\underset{\sim}{X_2} = x_2$ is normal and has

$Mean = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$and covariance $\Sigma_{11}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ (9marks)

b) Let the data matrix for a random sample of size n = 3 from a bivariate normal population be $\underset{\sim}{x} = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}$

i. Evaluate the observed $T^2 \; for \mu_0 = [9,5]^1$ (9marks)

ii. What is the sampling distribution of T^2 (2marks)

QUESTION FIVE (20 MARKS)

a) A summary of a dataset was given as

$$\bar{x} = \begin{bmatrix} 526.29 \\ 54.69 \\ 25.103 \end{bmatrix} \text{ and } S = \begin{bmatrix} 5808.06 & 597.84 & 222.03 \\ 597.84 & 126.05 & 23.39 \\ 222.03 & 23.39 & 23.11 \end{bmatrix}$$

Compute the 95% simultaneous confidence intervals for $\mu_1, \mu_2 \; and \; \mu_3$ (5marks)

a) The table below list observation of three types of calcium measurements in soil and turnip greens

No. of observations	y_1	y_2	x
1	35	3.5	2.80
2	35	4.9	2.70
3	40	30.0	4.38
4	10	2.8	3.21
5	6	2.7	2.73
6	20	2.8	2.81
7	35	4.6	2.88
8	35	10.9	2.90
9	35	8.0	3.25
10	30	1.6	3.20

Using $\mu_0 = (15.0 \;\; 6.0 \;\; 2.85)^1, T_{yx}^2 = 24.559$

i. Determine T_y^2 (10makrs)

ii. Examine what could have changed T_{yx}^2 to T_y^2 (5marks)