LAIKIPIA UNIVERSITY

UNIVERSITY EXAMINATIONS

SECOND SEMESTER 2023/2024 ACADEMIC YEAR

FOURTH YEAR EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE (ECONOMIC & STATISTICS)

STAT 422: APPLIED MULTIVARIATE ANALYSIS

STREAM: R ***TIME: 2 HRS***

DAY: WEDNESDAY[8.30A.M – 10.30A.M] ***DATE: 10/04/2024***

THIS QUESTION PAPER CONSISTS OF SIX (6) PAGES

PLEASE DO NOT OPEN UNTIL THE INVIGILATOR SAYS SO

INSTRUCTIONS: ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS

QUESTION ONE (30 MARKS)

a) Assume y and X are sub-vectors, each 2×1, where $\begin{pmatrix} y \\ x \end{pmatrix}$ is $N_4(\mu, \Sigma)$ with

$$\mu = \begin{pmatrix} 2 \\ 1 \\ \ldots\ldots \\ 3 \\ 1 \end{pmatrix} \quad and \quad \Sigma = \begin{pmatrix} 7 & 3 & -3 & 2 \\ 3 & 6 & 0 & 4 \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ -3 & 0 & 5 & -2 \\ 2 & 4 & -2 & 4 \end{pmatrix}$$

Required

i) Find (E(y/x) when $X=\begin{pmatrix} 4 \\ 2 \end{pmatrix}$ **(6 Marks)**

ii) Cov(y/x) **(4 Marks)**

b) Explain Five objectives of scientific investigations to which multivariate analysis methods most naturally lend themselves **(10 Marks)**

c) Explain the importance of rotated plots in three dimensions in data analysis. **(3 Marks)**

d) Which variables have significant correlation and which one does not have significant correlation **(7 Marks)**

		Dominant radius	Radius	Dorminat Humerous	Humerous	Dorminat Ulna	Ulna
Dominant radius	Pearson Correlation	1	.838**	.786**	.754*	.717*	.434
	Sig. (2-tailed)		.002	.007	.012	.020	.210
	N	10	10	10	10	10	10
Radius	Pearson Correlation	.838**	1	.832**	.922**	.461	.540

	Sig. (2-tailed)	.002		.003	.000	.180	.107
	N	10	10	10	10	10	10
Dorminat Humerous	Pearson Correlation	.786**	.832**	1	.932**	.252	.272
	Sig. (2-tailed)	.007	.003		.000	.483	.448
							10
Humerous	Pearson Correlation	.754*	.922**	.932**	1	.278	.357
	Sig. (2-tailed)	.012	.000	.000		.437	.311
							10
Dorminant Ulna	Pearson Correlation	.717*	.461	.252	.278	1	.642*
	Sig. (2-tailed)	.020	.180	.483	.437		.045
							10
Ulna	Pearson Correlation	.434	.540	.272	.357	.642*	1
	Sig. (2-tailed)	.210	.107	.448	.311	.045	
							10

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

QUESTION TWO (20 MARKS)

a) State six the properties of a random $p \times 1$ vector y from a multivariate normal random variables distribution $N_p(\mu, \Sigma_p)$ **(12 Marks)**

b) The table below presents the 2005 attendance (millions) at the fifteen most visited national parks and their size(acres).

National park	Size(acres)	Visitors(millions)
Arcadia	47.4	2.05
Bruce	35.8	1.02
Valley	32.9	2.53
Everglades	1508.5	1.23
Canyon	1217	4.4
Teton	310	2.46
Hot	5.6	1.34
Colympic	922.7	3.14
Mount	235.6	1.17
Rocky	265.8	2.8
Shenan	199	1.09
Yellow	2219.8	2.84
Yose	761.3	3.3
Zion	146.6	2.59

Required

i) Calculate the correlation coefficient **(4 Marks)**

ii) Identify the park that is unusual **(2 Marks)**

iii) Drop this unusual point and recalculate the correlation coefficient **(4 Marks)**

iv) Comment on the effect of this one point. (**2 Marks)**

QUESTION THREE (20 MARKS)

a) Let $X_1, \quad X_2, \quad X_3, \quad ,X_{20}$ be a random sample of size n=20 from a $N_6(\mu, \Sigma)$.

Specify each of the following completely

i) The distribution of $(X_1 - \mu)^T \Sigma^{-1}((X_1 - \mu))$ **(2 Marks)**

ii) The distribution of $\bar{X}$ **(2 Marks)**

iii) The distribution of $\sqrt{n}(\bar{X}-\mu)$ **(2 Marks)**

b) State the assumptions of MANOVA **(4 Marks)**

c) Let the vector X be partitioned into X1 and X2 such that the dimension of X1 is q1 and X2 is p-q vector. Find the distribution of X1 given X2=x2. **(10 Marks)**

Solution

$Let\ X=\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ be distributed as $Np(\mu,\Sigma)\ \ and\ \ \mu=\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\Sigma=\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

QUESTION FOUR (20 MARKS)

a) Let X be distributed as $N_3(\mu,\Sigma)$ where $\mu^T=(1,\ -1,\ 2)$ and

$$\Sigma=\begin{pmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{pmatrix}$$

Which of the following variables are independent? Explain,

i) $X_1\ \ and\ \ X_2$ **(2 Marks)**

ii) $X_1\ \ and\ \ X_3$ **(2 Marks)**

iii) $X_2\ \ and\ \ X_3$ **(2 Marks)**

iv) $(X_1\ ,X_3)\ ,X_2$ **(2 Marks**

v) $X_1\ and\ ,X_1+3X_2-2X_3$ **(4 Marks)**

b) Explain what is meant by the principal components analysis. What does the firsts and second principal components represent? **(4 Marks)**

i) Give four examples of real life multivariate data **(4 Marks)**

QUESTION FIVE (20 MARKS)

a) Evaluate T^2 for testing $Ho: \mu^T = (7,11)$, using the data **(12 Marks)**

$$X = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix}$$

b) Specify the distribution of T^2 for the situation in a) **(4 Marks)**

c) Test The hypothesis $Ho: \mu^T = (7,11)$ **(4 Marks)**