KABARAK UNIVERSITY

EXAMINATIONS

2008/2009 ACADEMIC YEAR

FOR THE DEGREE OF BACHELOR OF SCIENCE IN ECONOMICS AND MATHEMATICS

COURSE CODE: **MATH 425**

COURSE TITLE: **MULTIVARIATE ANALYSIS**

STREAM: **Y4S1**

DAY: **TUESDAY**

TIME: **9.00 – 11.00 A.M.**

DATE: **24/03/2009**

INSTRUCTIONS:

Answer questions **ONE** and any other **TWO** questions.

PLEASE TURN OVER

QUESTION ONE (20 MARKS)

a) Suppose that $\underline{X}$ is d-variate normally distributed (μ, Σ) $, \Sigma > 0$ let $y = a' \underline{X}$ where a' a non-zero vector is. Find the pdf of y using the mgf technique

(6 marks)

b) Describe how the following work in multivariate analysis;

- **i)** Discriminant analysis **(4 marks)**
- **ii)** Principal component analysis **(4 marks)**
- **iii)** Factor analysis **(4 marks)**

c) The data matrix for a sample of size n = 3 for a bivariate normal distribution is given by

$$X = \begin{pmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{pmatrix}$$

Calculate the value of T^2 *for* $\mu_0 = \begin{pmatrix} 9 \\ 5 \end{pmatrix}$ **(8 marks)**

Use the data to test the hypothesis

$$H_O : \underline{\mu} = \underline{\mu}_O$$

$$H_O : \underline{\mu} \neq \underline{\mu}_O$$

Use $\alpha = 0.1$ **(4 marks)**

QUESTION TWO (20 MARKS)

The director of a management school wants to do discriminant analysis concerning the effects of two factors, namely, the yearly spending (in Kshs) on infrastructures of the school(X_1) and yearly spending on interface events of the school (X_2) on the grading of the school by an inspection team. He has collected the data for the last 12 years and submitted to the inspection team as shown below. Based on the data, the committee has awarded one of the following grades for each year as shown in the same table.

Year	**Grade**	**Infrastructure(** X_1)	**interface events** (X_2)
1	Below	3	4
2	Below	4	5
3	Above	10	7
4	Below	5	4
5	Below	6	6
6	Above	1	4
7	Below	7	4
8	Above	12	5
9	Below	8	7
10	Below	9	5
11	Above	13	6
12	Above	14	8

a) Design the discriminant function, Y= a X_1 + b X_2 **(10 marks)**

b) Compute the discriminant ratio k and identity the variable which is more important in the relation to the other variable **(10 marks)**

QUESTION THREE (20 MARKS)

A marketing manager of a two-wheeler company designed questionnaires to study the customer's feedback about its two-wheeler and in turn he is keen in identifying the factors of his study. He has identified six variables which are listed below

1 Fuel efficiency X_1

2 Life of two-wheeler X_2

3 Handling convenience X_3

4 Quality of original spares X_4

5 Breakdown rate X_2

6 Price X_6

So the company administered a questionnaire among 10 customers to obtain their opinion on the above variables. The range of the scores for each of the above variables is assumed to be between 0 and 10. The score 0 means low ratings and 10 mean high rating. The survey data are summarized below. Perform factor analysis using centroid method to identify three factors which can represent the variables of the study.

Survey data

Respondents	X_1	X_2	X_3	X_4	X_5	X_6
1	6	8	9	9	1	2
2	4	4	6	8	2	1
3	4	1	6	5	1	2
4	1	2	6	3	2	3
5	4	3	5	5	2	3
6	4	4	6	8	2	3
7	3	3	0	9	1	3
8	7	7	6	9	9	2
9	5	3	1	8	1	2
10	5	4	2	3	1	0

QUESTION FOUR (20 MARKS)

a) Suppose that $\underline{X}$ is d-variate normally distributed (μ, Σ) , show that the random vector $\underline{Z} = \Sigma^{\frac{1}{2}} (X - \mu)$ is distributed normally with($\underline{0}$, I_d) **(10 marks)**

b) Suppose that $\underline{X}$ is d-variate normally distributed (μ, Σ) where

$$\underline{\mu} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & -1 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Find the

i) pdf of $Y = 2X_1 + X_2 - X_3$ **(4 marks)**

ii) Joint pdf of $Y_1 = X_1 - X_2$ *and* $Y_2 = X_2 - X_3$ **(6 marks)**

QUESTION FIVE (20 MARKS)

Bivariate data were collected in a fully randomized experiment. The variates observed were X_1, total seed yield per plant in grammes, and X_2, the weight of 100 randomly sampled seeds per plant in milligrammes. The four treatments were equally spaced levels of a nitrogenous fertilizer.

Nitrogen Level	Replicates				Treatment Total
1	$\begin{bmatrix} 21 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 25 \\ 8 \end{bmatrix}$	$\begin{bmatrix} 20 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 24 \\ 10 \end{bmatrix}$	$\begin{bmatrix} 90 \\ 42 \end{bmatrix}$
2	$\begin{bmatrix} 31 \\ 9 \end{bmatrix}$	$\begin{bmatrix} 23 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 24 \\ 13 \end{bmatrix}$	$\begin{bmatrix} 28 \\ 10 \end{bmatrix}$	$\begin{bmatrix} 106 \\ 44 \end{bmatrix}$
3	$\begin{bmatrix} 34 \\ 10 \end{bmatrix}$	$\begin{bmatrix} 29 \\ 14 \end{bmatrix}$	$\begin{bmatrix} 35 \\ 11 \end{bmatrix}$	$\begin{bmatrix} 32 \\ 13 \end{bmatrix}$	$\begin{bmatrix} 130 \\ 48 \end{bmatrix}$
4	$\begin{bmatrix} 33 \\ 14 \end{bmatrix}$	$\begin{bmatrix} 38 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 34 \\ 13 \end{bmatrix}$	$\begin{bmatrix} 35 \\ 13 \end{bmatrix}$	$\begin{bmatrix} 140 \\ 52 \end{bmatrix}$

i) Carry out Canonical Variate Analysis (**5 marks**)

ii) Estimate the linear relationship applying to the components of the treatment means (**5 marks**)

iii) Use CVA to find the linear compound of the response which has the best linear relationship with the nitrogen levels and estimate the relationship (**5 marks)**

iv) The second root of $|H - \theta R| = 0$ in part (iii) above is zero. Show that the corresponding canonical variate expresses an exact linear relationship between the estimates of the treatments obtained using the fitted values of the linear relationship with nitrogen levels derived in part (iii) **(5 marks)**