KABARAK UNIVERSITY

UNIVERSITY EXAMINATIONS

2009/2010 ACADEMIC YEAR

FOR THE DEGREE OF BACHELOR OF SCIENCE IN

ECONOMICS AND MATHEMATICS

COURSE CODE: **MATH 425**

COURSE TITLE: **MULTIVARIATE ANALYSIS**

STREAM: **Y4S2**

DAY: **TUESDAY**

TIME: **9.00 – 11.00 A.M.**

DATE: **23/03/2010**

INSTRUCTIONS:

1. Answer question **ONE** and any other **TWO** questions
2. Begin each question on a separate page
3. Show your workings clearly and neatly.

PLEASE TURN OVER

QUESTION ONE (30 MARKS)

a) Given $\mathbf{x} = (x_1, x_2, x_3)$ where $f(x) = e^{-x2}(x_1 + x_3)$, $0 < x_1 < 1$, $0 < x_3 < 1$
Determine the mean vector of $\mathbf{x}$ (5 marks)

b) If $\mathbf{a} = (a_1, a_2, a_5),\ldots.a_p)$ is a non-zerovector of order p and $\mathbf{y} = \mathbf{a'x}$. Show that

i) $E(\mathbf{a'x}) = \mathbf{a'}\,\mu$ (4 marks)

ii) $Var(\mathbf{a'x}) = \mathbf{a'}\,\Sigma\,\mathbf{a}$ (4 marks)

c) Let $\mathbf{x} = (x_1,x_2,x_3)'$ be a random vector with dispersion matrix with dispersion

$$\Sigma = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{pmatrix}$$

i) Find the correlation matrix ρ and $\mathbf{x}$ (4 marks)

ii) Show that Σ is a positive definite matrix (5 marks)

iii) Find the variance of $z = x_1 - 2x_2 - x_3$ (4 marks)

iv) Find the dispersion matrix of $\mathbf{y} = (y_1, y_2 , y_3)$ where $y_1 = x_1 + x_2$, $y_2 = x_2$ and $y_3 = x_1 - 2x_2 - x_3$ and Covariance of $z_1 = x_1 + x_2$ and $z_2 = x_1 - 2x_2 - x_3$ (8 marks)

QUESTION TWO (20 MARKS)

d) Let $\mathbf{x} = (x_1,x_2,x_3)'$ be a random vector with a mean vector $\mu = (\mu_1\ \mu_2\ \mu_3)$ and dispersion matrix Σ where μ and Σ are unknown, Given the values of a data matri x for a random sample of size n = 5 from $\mathbf{x}$

$$X = \begin{pmatrix} 5 & 6 & 9 \\ 3 & 8 & 7 \\ 5 & 7 & 8 \\ 4 & 5 & 6 \\ 3 & 4 & 5 \end{pmatrix}$$

Compute

i) The sample mean vector (4 marks)

ii) Sample covariance matrix S and the unbiased estimate of S_n (8 marks)

iii) Sample correlation matrix (8 marks)

QUESTION THREE (20 MARKS)

a) Consider $n_1 = n_2 = 50$ observations from two species I_1 and I_2 with measurements x_1 (length) and x_2 (width) in cm.For each observation the ML estimates are i.e the sample mean and variance for each group are;

$\bar{x}_1 = (5.006,\ 3.428)' \qquad \bar{x}_2 = (5.936, 2.770)'$

$$S_1 = \begin{pmatrix} 0.1218 & 0.0972 \\ 0.0972 & 0.1408 \end{pmatrix} \qquad S_2 = \begin{pmatrix} 0.2611 & 0.0835 \\ 0.0835 & 0.0965 \end{pmatrix}$$

i) Find the discriminant rule for allocating a new observation $x = (x_1, x_2)$ to Π_1 or Π_2 (7 marks)

ii) Estimate the probability of misallocation (5 marks)

b) Consider ML discriminant function, proof the probability of misallocation (8 marks)

QUESTION FOUR (2O 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 FIVE (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