KABARAK UNIVERSITY

UNIVERSITY EXAMINATIONS

2008/2009 ACADEMIC YEAR

FOR THE DEGREE OF BACHELOR OF ECONOMICS & MATHEMATICS

COURSE CODE: MATH 425

COURSE TITLE: **MULTIVARIATE ANALYSIS**

STREAM: Y4S2

DAY: **MONDAY**

TIME: **2.00 – 4.00 P.M.**

DATE: **15/12/2008**

INSTRUCTIONS TO CANDIDATES:

1. Answer Question **ONE** and any other **TWO** Questions
2. Observe further instructions on your answer booklet.

PLEASE TURN OVER

QUESTION ONE (30 MARKS)

(a) What is the purpose of studying "canonical correlation analysis" as a multivariate method? **(2 mks)**

(b) Consider three variables X_1, X_2 and X_3 whose covariance matrix is given by

$$\sum = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

i) Find the principal components of $\mathbf{X}^T = [X_1, X_2, X_3]$. **(11 mks)**

ii) What proportion of the total variation in the original data is accounted for by the first principal component? **(1 mk)**

iii) Is it important to extract principal components from such a variance-covariance matrix? Why? **(2 mks)**

(c) Briefly explain the difference between the Wishart and Hotellings T^2 distributions. **(2 mks)**

(d) The scores on two CATs for each of the 8(eight) students were as follows

Student	**CAT 1**	**CAT2**
1	2	9
2	16	20
3	8	0
4	18	11
5	10	13
6	4	7
7	12	16

Compute;

i) sample covariance matrix (show all your working steps) **(5 mks)**

ii) sample correlation matrix. **(3 mks)**

(e) Briefly explain the use of each of the following multivariate techniques

i) Discrimination and classification **(2 mks)**

ii) Multivariate analysis of variance **(2 mks)**

QUESTION TWO (20 MARKS)

(a) State and explain THREE uses of principal component analysis. **(9 mks)**

(b) The following data represent measurements on two-year-old boys from highland areas in Kenya. X_1 = Height (cm), X_2 = chest circumference (cm), X_3 = MUAC (the mid-upper-arm circumference). Suppose $\underline{\mu}_0^T = [90, 58, 16]$ are the height, chest and MUAV means of similar children form lowlands in Kenya. Test the hypothesis that the highland children have the same means. (Use $\alpha = 0.01$)

Height	Chest Circumference	MUAC
80	58.4	14
75	59.2	15
78	60.3	15
75	57.4	13
79	59.5	14
78	58.1	14.5
75	58.0	12.5
64	55.5	11
80	59.2	12.5

(11 mks)

QUESTION THREE (20 MARKS)

(a) Suppose that **X**: 4 x 1 and is normally distributed with mean vector $\underline{\mu}$ and covariance $\sum$, where $\underline{\mu}^T = (2, 5, 3, 6)$ and

$$\sum = \begin{pmatrix} 11 & 5 & 2 & 3 \\ 5 & 4 & 1 & 0 \\ 2 & 1 & 3 & 0 \\ 3 & 0 & 0 & 2 \end{pmatrix}$$

Determine

(i) the joint distribution of $Y_1 = X_1 - X_2 + X_3 - 3X_4$ and $Y_2 = X_2 - X_3 + 4X_4$

(i) the marginal distribution of $\underline{X}_1 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

(ii) the conditional distribution of $\underline{X}_2 = \begin{pmatrix} x_3 \\ x_4 \end{pmatrix}$ given $\underline{X}_1 = \begin{pmatrix} 8 \\ 6 \end{pmatrix}$

QUESTION FOUR (20 MARKS)

In a certain experiment, the correlation matrix was forms to be

$$R = \begin{pmatrix} 1.0 & 0.4 & 0.5 & 0.6 \\ 0.4 & 1.0 & 0.3 & 0.4 \\ 0.5 & 0.3 & 1.0 & 0.2 \\ 0.6 & 0.4 & 0.2 & 1.0 \end{pmatrix}$$

Determine;

(a) the canonical pairs of variates **(5 mks)**

(b) the canonical pairs of variates (u_1, v_1) and (u_2, v_2) **(15 mks)**

QUESTION FIVE (20 MARKS)

Suppose we have the variables X_1 and X_2 whose correlation matrix is given by

$$R = \begin{matrix} 1 & r \\ r & 1 \end{matrix} \quad \text{where } r > 0$$

(a) Find the principal components **(15 mks)**

(b) What percentage of the total variation is explained by the first principal component? **(2 mks)**

(c) Suppose that $r < 0$ or that $r = 0$. What effect does this have on the eigenvalues and hence on the principal components? **(3 mks)**