MURANG'A UNIVERSITY OF TECHNOLOGY

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF APPLIED SCIENCES

UNIVERSITY ORDINARY EXAMINATION

2017/2018 ACADEMIC YEAR

EXAMINATION FOR MASTER OF SCIENCE IN STATISTICS

AMS 603: MULTIVARIATE METHODS

DURATION: 3 HOURS

DATE: 15TH AUGUST, 2018

TIME: 9.00 – 12.00 NOON

Instructions to Candidates:

1. Answer **Any Four** questions.
2. Mobile phones are not allowed in the examination room.
3. You are not allowed to write on this examination question paper.

QUESTION ONE

a. Let $X \sim N_p(\mu, \sum)$ and let X be partitioned into two vectors as:

$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ with $E(X_i) = \mu_i$, $Var(X_i) = \sum_{ii}$ and $Cov(X_i, X_j) = \sum_{ij}$. Show that

$X_2 | X_1 \sim N\left(\mu_2 + \sum_{21} \sum_{11}^{-1} (X_1 - \mu_1), \sum_{22} - \sum_{21} \sum_{11}^{-1} \sum_{12}\right)$ (13 marks)

b. Let $X \sim N_3(\mu, \sum)$ with:

$\mu = \begin{pmatrix} 44 \\ 20 \\ 16 \end{pmatrix}$, $\sum = \begin{pmatrix} 64 & 8 & -16/5 \\ 8 & 16 & 4 \\ -16/5 & 4 & 4 \end{pmatrix}$ find:

a. The distribution of $Y = \begin{pmatrix} X_1 - X_3 \\ X_1 + 2X_2 + X_3 \end{pmatrix}$ (6 marks)

b. The distribution of X_1 given that $X_2 = 15$ and $X_3 = 18$ (6 marks)

QUESTION TWO

a. Let $X = [X_1, X_2, ..., X_p]$ be a random vector with covariance matrix $\sum$. Define a principal component based on X and state the main difference between principal component analysis and factor analysis. (10 marks)

b. Let $X \sim N(\mu, \sum)$ be a tri-variate normal random vector. Suppose a certain sample gave:

$$S = \begin{pmatrix} 64 & 0 & 8 \\ 0 & 4 & 0 \\ 8 & 0 & 16 \end{pmatrix}$$

Find the eigenvalues of this matrix, the first two principal components and the total variance explained by the components. (15 marks)

QUESTION THREE

a. Given the following data from a bivariate normal distribution,

$$X = \begin{pmatrix} 24 & 22 & 20 & 17 & 27 & 21 & 22 & 19 & 17 \\ 21 & 24 & 29 & 24 & 22 & 20 & 25 & 26 & 25 \end{pmatrix}$$

test at $\alpha = 0.01$ level the hypothesis: $H_0 : \mu = (24\ 19)'$ *versus* $H_1 : \mu \neq (24\ 19)'$

(13 marks)

b. Using the maximum likelihood estimation method, show that the mean vector and the dispersion matrix for a multivariate normal distribution are given by: $\hat{\mu} = \bar{X}$ *and* $\sum = S$ respectively. (12 marks)

QUESTION FOUR

a. Given that $\bar{X}_1 = \begin{pmatrix} 33 & 10 & 10 \end{pmatrix}'$ comes from population 1 and $\bar{X}_2 = \begin{pmatrix} 27 & 19 & 8 \end{pmatrix}'$ comes from population 2 and both populations have the same covariance matrix.

$$S = \begin{pmatrix} 20 & -4 & 15 \\ -4 & 16 & 0 \\ 15 & 0 & 4 \end{pmatrix}$$

Use the discriminant rule to classify: $X = \begin{pmatrix} 34 & 11 & 8 \end{pmatrix}$ (16 marks)

b. The following are factor loadings computed from a study on ratings of some 329 communities according to nine criteria. Using a cut-off of 0.5, determine which are large and small. Interpret the results for the 3 factors. (9 marks)

Variable	1	2	3
Climate	0.286	0.076	0.841
Housing	0.698	0.153	0.084
Health	0.744	-0.410	-0.020
Crime	0.471	0.522	0.135
Transportation	0.681	-0.156	-0.148
Education	0.498	-0.498	-0.253
Arts	0.861	-0.115	0.011
Recreation	0.642	0.322	0.044
Economics	0.298	0.595	-0.533

QUESTION FIVE

a. Given that the covariance matrix for p=3 is defined by:

$$P = \begin{pmatrix} 1 & 0.63 & 0.45 \\ & 1 & 0.35 \\ & & 1 \end{pmatrix}$$

i. Show that the standardized random variables $Z_1, Z_2,$ *and* Z_3 can be generated by one factor model defined by:

$Z_1 = 0.9F_1 + \varepsilon_1,\ Z_2 = 0.9F_1 + \varepsilon_2,\ Z_3 = 0.9F_1 + \varepsilon_3$ where: $Var(F_1) = Cov(\varepsilon, F_1) = 0$

$$\psi = Cov(\varepsilon) = \begin{pmatrix} 0.19 & 0 & 0 \\ 0 & 0.15 & 0 \\ 0 & 0 & 0.75 \end{pmatrix}$$ (12 marks)

ii. Compute the communalities (7 marks)

iii. Compute the structure loadings (6 marks)