



MACHAKOS UNIVERSITY

University Examinations 2019/2020 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

.....YEAR SPECIAL /SUPPLEMENTARY EXAMINATION FOR

BACHELOR OF SCIENCE IN STATISTICS AND PROGRAMMING

SMA 360: MULTIVARIATE STATISTICAL METHOD

DATE:

TIME:

INSTRUCTIONS TO CANDIDATES

Answer ALL the questions in Section A and ANY TWO Questions in Section B

SECTION A

QUESTION ONE (30 MARKS) (COMPULSORY)

- a) Briefly state assumptions for MANOVA. (4 marks)
- b) Observations on three responses collected for three treatments based on tri-variate normal

random vector $\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ are as given below

Treatment 1:

x_1	4	6
x_2	2	2
x_3	0	2

Treatment 2:

x_1	10	4	10	8
x_2	3	3	0	2
x_3	6	3	4	3

Find the:

- i. matrix of sum of squares due to treatment (4 marks)
- ii. matrix of residual sum of squares (4 marks)

- iii. Wilk's lambda statistics and use it to test the hypothesis that there is no treatment effects ($\alpha = 0.05$) (4 marks)
- c) Given that $\vec{X} \sim N_3(\mu, \Sigma)$ with
- $$\mu = \begin{pmatrix} 27 \\ 25 \\ 30 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 36 & 12 & -2 \\ 12 & 64 & 0 \\ -2 & 0 & 49 \end{pmatrix}$$
- Find;
- The distribution of $Y = \begin{pmatrix} -2X_2 + X_3 \\ X_1 + X_2 - X_3 \end{pmatrix}$ (3 marks)
 - The correlation matrix for the data (2 marks)
 - The distribution of $X_1/X_3 = 25$ where $X_1 = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ (4 marks)
 - The regression function of X_3 on X_1 and X_2 (2 marks)
 - The partial correlation coefficient between X_1 and X_2 for fixed values of X_3 (2 marks)
 - The conditional variance of X_1 when the effects of X_2 and X_3 have been eliminated. (2 marks)

SECTION B

QUESTION TWO (20 MARKS)

- Discuss the objectives of scientific investigation based on multivariate statistical data analysis. (5 marks)
- Let $\mathbf{X} \sim N(\mu, \Sigma)$ be tri-variate normal random vector with sample variance-covariance matrix $S = \begin{pmatrix} 1 & 0.67 & -0.1 \\ 0.67 & 1 & -0.29 \\ -0.1 & -0.29 & 1 \end{pmatrix}$. Find;
 - The first principal component and the percentage of total variance explained by the component. (11 marks)
 - The correlation between the first principle component and the first variable. (4 marks)

QUESTION THREE (20 MARKS)

- Suppose that $\mathbf{X} \sim N_p(\mu, \Sigma)$ with $\Sigma > 0$ and

$$f(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

Derive the Maximum Likelihood estimates for the mean vector and dispersion matrix, $(\boldsymbol{\mu}, \Sigma)$. (10

marks)

- b) Given that $\bar{\mathbf{x}}_1 = (33 \ 12 \ 10)'$ comes from population 1 and $\bar{\mathbf{x}}_2 = (27 \ 19 \ 8)'$ comes from population 2 and both populations have the same sample covariance matrix

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

Use an appropriate discriminant rule to classify a new observation $\mathbf{x} = \begin{pmatrix} 34 \\ 11 \\ 8 \end{pmatrix}$ (10

marks)

QUESTION FOUR (20 MARKS)

- a) Explain hypothesis testing under MANOVA. (5 marks)

- b) Consider the data matrix $\mathbf{X} = \begin{bmatrix} 3 & 4 & 5 & 6 & 2 \\ 9 & 5 & 7 & 2 & 8 \end{bmatrix}$. Test at $\alpha = 0.05$ the hypothesis

$$H_0: \boldsymbol{\mu} = (3.4 \ 6)' \quad (10 \text{ marks})$$

$$H_1: \boldsymbol{\mu} \neq (3.4 \ 6)'$$

- c) Briefly explain the main idea behind discriminant analysis stating the two types of errors. (5 marks)

QUESTION FIVE (20 MARKS)

- a) Describe briefly Principle component analysis. (5 marks)

- b) Explain the difference between Canonical correlation analysis and Cluster analysis as used in multivariate analysis. (5 marks)

- c) Let $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ be tri-variate normal random vector. Suppose a certain sample gave

$$S = \begin{pmatrix} 4 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 25 \end{pmatrix}$$

Find the;

- i. Eigenvalues of this matrix. (5 marks)

- ii. First principal component and the proportion of variance explained by the first two principal components. (5 marks)