NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY OF SCIENCE AND TECHNOLOGY EDUCATION

DEPARTMENT OF SCIENCE, MATHEMATICS AND TECHNOLOGY EDUCATION

MASTER OF SCIENCE EDUCATION

MULTIVARIATE STATISTICS

PST 6134

Main Examination Paper

NOVEMBER 2024

This examination paper consists of 5 pages

Time Allowed: 3 hours

Total Marks: 100

Special Requirements: Statistical Tables

Examiner's Name: Sesilani Nkomo

External Examiner: Dr Sunzuma G

INSTRUCTIONS

1. This paper consists of **six** questions. Answer **at most five** questions.
2. All answers should be presented in good style.
3. Begin each full question on a new page.

MARK ALLOCATION

QUESTION	MARKS
1.	20
2.	20
3.	20
4.	20
5.	20
6.	20

Q1. Municipal wastewater treatment plants are required by law to monitor their discharges into rivers and streams on a regular basis. Concern about the reliability of data from one of these self-monitoring programs led to a study in which samples of effluent were divided and sent to two laboratories for testing. One-half of each sample was sent to the State Laboratory of Hygiene and one-half was sent to a private commercial laboratory routinely used in the monitoring program. Measurements of biochemical oxygen demand (BOD) and suspended solids (SS) were obtained, for $n = 11$ sample splits, from the two laboratories. The data are displayed below:

EFFLUENT DATA				
Sample j	**Commercial Lab** X_{1j1}**(BOD)**	X_{1j2}**(SS)**	**State Lab of hygiene** X_{2j1}**(BOD)**	X_{2j2}**(SS)**
1	6	27	25	15
2	6	23	28	13
3	18	64	36	22
4	8	44	35	29
5	11	30	15	31
6	34	75	44	64
7	28	26	42	30
8	71	124	54	64
9	43	54	34	56
10	33	30	29	20
11	20	14	39	21

(a) Find the

i) data matrix $\mathbb{d}$, [3]

ii) sample mean $\bar{\mathbb{d}}$ [2]

iii) sample covariance matrix $S_{\mathbb{d}}$ and [4]

iv) $S_{\mathbb{d}}^{-1}$. [5]

b) Test at 95% whether the two laboratories' chemical analyses agree. [5]

© If differences exist what is their nature? i.e the 95% simultaneous confidence intervals for the mean vector $\mathbb{d}$. [6]

Q2. a) Let $A = \begin{pmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{pmatrix}$

i) Find the eigenvalues of A. [5]
ii) Find the corresponding eigenvector for each eigenvalue. [6]
iii) Represent A by its spectral decomposition. [3]

b) State and prove the Cauchy-Schwartz inequality. [6]

Q3. There are two methods of manufacturing washing soap. Two characteristics

$\boldsymbol{X_1}$ = lather
$\boldsymbol{X_2}$ = mildness

are measured. The summary statistics for bars of washing soap produced by methods1 and 2 are:

$n_1 = 6 \qquad \bar{X}_1 = \begin{pmatrix} 8.3 \\ 4.1 \end{pmatrix}, \qquad S_1 = \begin{pmatrix} 2 & 1 \\ 1 & 6 \end{pmatrix}$

$n_2 = 8 \qquad \bar{X}_2 = \begin{pmatrix} 10.2 \\ 3.9 \end{pmatrix}, \qquad S_2 = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$

(a) Assuming equal population variances, test whether the two methods produce soap with same mean characteristics. Use $\alpha = 0.05$. [10]
(b) Given that the population variances are different, test the hypothesis that the mean characteristics are the same for two methods. [10]

Q4. a)The rate at which margarine melts is measured after treatment with three different colouring agents. The agents are extracts from X_1, a tropical grass, $X_{2,}$ a fruit juice and $X_{3,}$ a sulphur compound. The summary statistics for a sample of 30 are:

$\bar{X}' = (4.64 \quad 45.40 \quad 9.97)$,
$S^{-1} = \begin{pmatrix} 0.586 & -0.022 & 0.258 \\ -0.022 & 0.005 & -0.002 \\ 0.258 & -0.002 & 0.402 \end{pmatrix}$.
Test the claim that the mean vector $\mu' = (4 \quad 50 \quad 10)$ at 5% level of significance. [8]

b) A chemist investigating levels of citric acid in oranges from two different plantations, measured the acidity levels of 35 lemons from plantation A and 22 lemons from plantation B, and obtained the following results;

[Negative values reflecting low acidity and conductivity levels, while positive values reflect increasing (high) values]

$$\bar{X}_A = \begin{pmatrix} -0.056 \\ -0.043 \end{pmatrix}, \bar{X}_B = \begin{pmatrix} -0.448 \\ 0.042 \end{pmatrix}, S_{pooled}^{-1} = \begin{pmatrix} 125.158 & -79.423 \\ -79.423 & 118.417 \end{pmatrix}$$

i) Obtain the Fisher's linear discrimination function. [6]

ii) Construct a decision rule for allocating the new observation $X_0 = \begin{pmatrix} X_{01} \\ X_{02} \end{pmatrix}$. [3]

iii) Classify the orange with the following acidity and conductivity $X_0 = \begin{pmatrix} -0.321 \\ -0.144 \end{pmatrix}$, with respect to plantation A or plantation B. Justify your answer. [3]

Q5. a) State the assumptions of MANOVA. [3]

b) Consider the data below, which was recorded on treatments A, B and C.

$Observation$	1	2	3	4	5	6	7
$Treatment$	A	A	A	B	B	C	C
$Response\ X_1$	9	6	9	0	2	3	8
$Response\ X_2$	3	2	7	4	0	8	5

i) Breakdown the observations into $X_{ij} = \bar{X} + (\bar{X}_i - \bar{X}) + (X_{ij} - \bar{X}_i)$ for both variables X_1 and X_2, and the cross products for the variables where X_{ij}, are the observations (i variable 1 and j representing the observation). $\bar{X}$ the overall sample mean, $(\bar{X}_i - \bar{X})$ is the estimated treatment effect and $(X_{ij} - \bar{X}_i)$ the residual. [8]

Assuming normality do the following

ii) Construct the one-way bivariate analysis of variance table. [5]

iii) Evaluate the Wilks's Lambda and test for treatment (depth) differences at the $\alpha = 0.05$ level of significance. Comment on your results. [4]

[Hint: $(\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}})\frac{(\sum n_i-g-1)}{(g-1)} \sim F_{2(g-1),2(\sum n_i-g-1)}$]

Q6. a) The table below shows results of 8 observations recorded on two variables X and Y.

X	80	50	70	75	50	75	62	73
Y	75	52	65	80	49	73	58	74

i) Test the hypotheses $H_0 : \mu = \binom{65}{50}$ against $H_1 : \mu \neq \binom{65}{50}$ (use 0.05 significance level). State any assumptions you make. [9]

ii) Calculate the 90% Boniferroni confidence intervals for μ_1 *and* μ_2. [4]

iii) Find the simultaneous 95% confidence intervals for μ_1 *and* μ_2. [4]

iv) Compare the results in sections (ii) and (iii) and comment. [3]

END OF QUESTION PAPER