NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY
FACULTY OF SCIENCE AND TECHNOLOGY EDUCATION
DEPARTMENT OF SCIENCE, MATHEMATICS AND TECHNOLOGY EDUCATION

PST 2137 BIOMATHEMATICS

Special Examination Paper

August 2024

This examination paper consists of 10 pages

Time Allowed: 3 hours
Total Marks: 100
Special Requirements:

- Graph Paper
- Statistics Tables

Examiner's Name: Mr P Moyo
External Examiner: Dr P Manyanga

INSTRUCTIONS

1. Answer **both** questions in **Section A** and any **three (3)** questions from **Section** B.
2. Start each question on a new page.

MARK ALLOCATION

QUESTION	MARKS
1.	20
2.	20
3.	20
4.	20
5.	20
6.	20
TOTAL	100

SECTION A

Answer both questions

1. The following are masses in kgs, of five animal offspring delivered by a female parent on the same day and raised under different conditions.

75 107 270 139 71

a. Find the value of the following parameters for these data.
   i. Mean [2]
   ii. Variance. [6]
   iii. standard deviation. [2]
   iv. coefficient of variance. [2]

b. Write short notes on the following in relation to experimental design; giving relevant examples.
   i. Replication [4]
   ii. Randomisation [4]

2. An educational researcher wanted to determine if there is an association between teachers' gender and attitudes towards within-class ability grouping. After undertaking a survey on a sample of 400 teachers. The following results emerged:

	Attitude Towards within- class ability grouping		
GENDER	Favorable	Neutral	Negative
Female	90	70	46
Male	85	58	59

a. State the **data requirements** to enable a Chi-square test to be carried out? [3]
b. Write down the null and alternate hypothesis for this research? [2]
c. Carry out a test at 1% significance level to ascertain if teachers' attitudes towards within-class ability grouping are **independent** of gender and comment on the results obtained. [15]

SECTION B

Answer any three questions

3. Ten candidates sat examinations in Statistics and Biomathematics. Their corresponding marks were: -

Statistics (x)	18	20	30	50	46	54	60	80	88	82
Biomathematics (y)	42	54	60	64	62	68	80	66	82	100

a. Plot these points on a scatter diagram. [3]

b. Calculate the equation of the regression line of y on x using the method of least squares. [4]
c. Using your regression line, calculate the Biomathematics (y) marks when Statistics (x) marks are 10 and 90 respectively. [2]
d. Hence plot the regression line on your diagram. [1]
e. The results in the table above were run in SPSS for the analysis of variance (ANOVA). Complete the ANOVA table below and test whether at 1% level of significance the slope B is positive. [10]

Model	SS	df	MS	F
Regression	?	1	?	17.885
Residual	?	8	?	
Total	?	9		

4. At an athletics academy, it used to take an average of 90 minutes for athletes to jog a 10km stretch. Recently the academy introduced a special diet to the athletes. The coach at the academy wants to find if the mean time taken by the athletes to run the 10km is different from 90 minutes. A sample of 20 athletes showed that it took, on average, 85 minutes for them to run the 10km. It is known that the jogging times for all athletes are normally distributed with a population standard deviation of 7 minutes.
    a. Find the p-value for the test that the mean jogging time for the 10km stretch on the new diet is different from 90 minutes. (17)
    b. What will your conclusion be if $\alpha = 0.01$? (3)

5. Vision, or more especially visual acuity, depends on a number of factors. A study was undertaken in Australia to determine the effect of one of these factors: racial variation. Visual acuity of recognition as assessed in clinical practice has a defined normal value of 20/20 (or zero in log scale). The following summarised data on monocular visual acuity (expressed in log scale) were obtained from two groups:

	Australian males of European origin		Australian males of Aboriginal origin
n_1	89	n_2	107
$\bar{x}_1$	-0.20	$\bar{x}_2$	-0.26
s_1	0.18	s_2	0.13

a. Conduct an independent two-sample t-test to determine whether or not the mean visual acuity indices differ between the European and Aboriginal groups at the $\alpha = 0.05$ significance level. [17]
b. State 3 requirements for an **independent two sample t-Test.** [3]

6. The table below shows data for four groups, as collected in the study referred to in question 5 to determine the effect of racial variation on visual acuity.

	Australian males of European origin		Australian males of Aboriginal origin
n_1	89	n_2	107
$\bar{x}_1$	-0.20	$\bar{x}_2$	-0.26
s_1	0.18	s_2	0.13

	Australian females of European origin		Australian females of Aboriginal origin
n_3	63	n_4	54
$\bar{x}_3$	-0.13	$\bar{x}_4$	-0.24
s_3	0.17	s_4	0.18

a. State 3 assumptions of ANOVA. [3]

b. Carry out a One-Way Analysis of Variance at 5% significance level to test whether the mean visual indices for these four racial groups are the same? Assume that all the assumptions required to apply the One-Way ANOVA procedure hold true?

Follow the steps suggested below.

i. State the null and alternative hypothesis. [2]

ii. State the rejection criterion. [3]

iii. Complete the One-Way ANOVA table. [10]

iv. Make a decision and conclude. [2]

END OF QUESTION PAPER

LIST OF SELECTED FORMULAE

For ungrouped data:

	Population	Sample
mean	$\bar{x} = \frac{\sum x}{N}$	$\bar{x} = \frac{\sum x}{n}$
Geometric mean	$\bar{x}_g = \sqrt[N]{x_1 x_2 x_3 \dots x_N}$	$\bar{x}_g = \sqrt[n]{x_1 x_2 x_3 \dots x_n}$
Standard deviation	$\sigma = \sqrt{\frac{1}{N}\left(\sum x^2 - \frac{(\sum x)^2}{N}\right)}$	$s = \sqrt{\frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)}$

For grouped data:

	Population	Sample
mean	$\bar{x} = \frac{\sum xf}{\sum f}$	$\bar{x} = \frac{\sum xf}{\sum f}$
Weighted mean	$\bar{x}_w = \frac{\sum xw}{\sum w}$	$\bar{x}_w = \frac{\sum xw}{\sum w}$
Geometric mean	$\bar{x}_g = \sqrt[\sum f]{x_1^{f_1} x_2^{f_2} x_3^{f_3} \dots x_N^{f_N}}$	$\bar{x}_g = \sqrt[\sum f]{x_1^{f_1} x_2^{f_2} x_3^{f_3} \dots x_N^{f_n}}$
Standard deviation	$\sigma = \sqrt{\frac{1}{N}\left(\sum x^2 f - \frac{(\sum xf)^2}{N}\right)}$	$s = \sqrt{\frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)}$

Pearson coefficient of Skewness

$skewness = \frac{3(mean - median)}{\text{standard deviation}}$ or $skewness = \frac{3(mean - mode)}{\text{standard deviation}}$

Discrete Random Variables

Expectation and Variance	For the binomial Distribution $X \sim Bin(n;p)$	For the Poisson Distribution $X \sim P(\lambda)$
$E(X) = \sum xP(X = x)$ $Var(X) = E(X^2) - [E(X)]^2$ where: $E(X^2) = \sum x^2 P(X = x)$	$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where: $x = 0, 1, 2, 3 \dots, n$ $E(X) = np$ $var(X) = np(1-p)$	$P(X = x) = \frac{(e^{-\lambda})\lambda^x}{x!}$ $x = 0, 1, 2, 3 \dots$ $E(X) = \lambda$ $var(X) = \lambda$

Continuous Random Variables and the Normal Distribution

z value for an x value: $X \sim N(\mu; \sigma^2)$	Value of x when μ, σ, and z-score are known:
$z = \frac{x-\mu}{\sigma}$	$x = \mu + \sigma z$,

Hypothesis Tests about the Single Mean and Proportion

Test statistic Z for μ when σ is known	Test statistic T for μ when σ is unknown	Test statistic Z for p when the sample is large
$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$	$T = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$, where: s =Sample standard deviation $\bar{x}$ =Mean of the sample	$Z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$ where: $\hat{p}$ =Sample proportion i.e. $\hat{p}=\frac{x}{n}$

Hypothesis Testing: The Paired t test

Test statistic for a test of hypothesis about μ_d using the t distribution:
$T = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$ where: Sample mean for paired differences: $\bar{d} = \frac{\sum d}{n}$ Sample standard deviation for paired differences: $s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}}$

Hypothesis Testing: Two independent Samples t-test

Test statistics for a test of hypothesis about $\mu_1 - \mu_2$ for two independent samples from two populations with equal but unknown standard deviations:	Test statistics for a test of hypothesis about $\mu_1 - \mu_2$ for two independent samples from two populations with unequal and unknown standard deviations:
$T = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ where s_p is the pooled standard deviation: $s_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$ $\bar{x}_1$ = Mean of first sample $\bar{x}_2$ = Mean of second sample n_1 = Sample size of 1st sample n_2 = Sample size of 2nd sample s_1 = Standard deviation of first sample s_2 = Standard deviation of second sample	$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$ where Degrees of freedom: $df = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$

Hypothesis Testing: Difference Between Two Population Proportions for Large and Independent Samples

Test statistic Z for a test of hypothesis about $p_1 - p_2$
$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where: $\bar{p}$ = pooled Sample proportion: $\bar{p} = \frac{x_1+x_2}{n_1+n_2}$ $\hat{p}_1$= sample proportion of the first sample i.e. $p_1 = \frac{x_1}{n_1}$ $\hat{p}_2$ = sample proportion of the second sample i.e. $p_2 = \frac{x_2}{n_2}$

Analysis of Variance Table for One-way Anova

Source of variation	Sum of squares (SS)	df	Mean square (MS)	F-ratio
Between Groups	$SS_{between} = \sum \frac{(T_j)^2}{n_j} - \frac{(T)^2}{n}$	$k-1$	$MS_{between} = \frac{SS_{between}}{k-1}$	$F = \frac{MS_{between}}{MS_{within}}$
Within Groups	$SS_{within} = \sum x_{ij}^2 - \sum \frac{(T_j)^2}{n_j}$	$n-k$	$MS_{within} = \frac{SS_{within}}{n-k}$	
Total	$SS_{Total} = \sum x_{ij}^2 - \frac{(T)^2}{n}$	$n-1$		

Chi-square statistics

$$\chi^2 = \sum \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

where

O_{ij} = observed frequency of the cell in i^{th} row and j^{th} column.

E_{ij} = expected frequency of the cell in i^{th} row and j^{th} column.

Expected frequency in any cell = $\frac{\text{(Row total for that cell)} \times \text{(Column total for that cell)}}{\text{Grand Total}}$

Degrees of freedom for a test of independence or homogeneity:

$$df = (R-1)(C-1)$$

where R and C are the total number of rows and columns respectively, in the contingency table

Correlation and Regression analysis

Pearson product-moment coefficient	Spearman's coefficient	Regression Line of y on x
$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$ where: $S_{xy} = \sum xy - \frac{\sum x \sum y}{n}$ $S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$ $S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$	$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$ where: n is the number of pairs of values d = rank$x$ – rank$y$	$y = a + bx$ where: $b = \frac{S_{xy}}{S_{xx}}$ $S_{xy} = \sum xy - \frac{\sum x \sum y}{n}$ $S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$ $a = \frac{1}{n}(\sum y - b\sum x)$

Test statistics for a test hypothesis about B	Test statistics for a hypothesis about ρ
$T = \frac{b-B}{s_b}$ where: $s_b = \sqrt{\frac{MSE}{S_{xx}}}$ $S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$ $t - critical = t_{n-k}(\alpha)$ for one tailed test $t - critical = \pm t_{n-k}(\frac{\alpha}{2})$ for two tailed test	$t = r\sqrt{\frac{n-2}{1-r^2}}$ where: n =number of all sample values r = linear correlation coefficient $t - critical = t_{n-k}(\alpha)$ for one tailed test $t - critical = \pm t_{n-k}(\frac{\alpha}{2})$ for two tailed test

Analysis of Variance Table for Regression Analysis

Source of variation	Sum of squares (SS)	df	Mean square (MS)	F-ratio
Regression	$SS_{regression} = bS_{xy}$	$k - 1$	$MS_{regression} = \frac{SS_{regression}}{k-1}$	$F = \frac{MS_{regression}}{MS_{error}}$
Error	$SS_{error} = S_{yy} - bS_{xy}$	$n - k$	$MS_{error} = \frac{SS_{error}}{n-k}$	
Total	$SS_{Total} = \sum y^2 - \frac{(\sum y)^2}{n}$	$n - 1$		