NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY OF SCIENCE AND TECHNOLOGY EDUCATION

DEPARTMENT OF SCIENCE, MATHEMATICS AND TECHNOLOGY EDUCATION

PST 2137 BIOMATHEMATICS

Main Examination Paper **First Semester**

November 2024

This examination paper consists of 10 pages

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

- Graph Paper
- Statistics Tables

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

INSTRUCTIONS

1. Answer **both** questions in section A and **any three** questions in 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	120

SECTION A

Answer both questions

---

1. The following data shows the scores obtained by 5 students in test 1 and test 2.

Student	A	B	C	D	E
Test 1 score	8	7	9	5	1
Test 2 score	4	7	6	8	10

a.
   i. Plot the data on a scatter plot. [3]
   ii. Comment on the relationship in performance in the two tests. [3]

b.
   i. Calculate the Spearman-Brown rank order correlation coefficient between test 1 and test 2. [12]
   ii. Comment on the relationship of students' performance in test 1 and test 2. [2]

2. A researcher wishes to determine if indiscipline is related to socioeconomic status. He takes a random sample of 100 students and compares the socio-economic statuses of those who have been associated with indiscipline and those who were not. The table below shows the summary of the data.

	Low income	Middle income	High income
Indiscipline	15	21	23
No indiscipline	14	18	19

   a. Using Chi-square, test at 2.5% significance level, to determine if socio-economic status and indiscipline are independent. [16]
   b. State 3 requirements for a Chi-square test. [4]

SECTION B

Answer any three questions

3. A group of 10 students were given an examination. The table below shows the scores obtained.

Student	A	B	C	D	E	F	G	H	I	J
Score	15	10	9	1	11	10	8	10	7	4

   a. Using the ungrouped data, calculate
      i. Arithmetic mean. [2]
      ii. The mode. [2]
      iii. The median. [2]
      iv. Inclusive range. [2]
      v. Standard deviation. [2]

   b. Distinguish between the following.
      i. Positive and negative skewness. [5]
      ii. Type I and type II error. [5]

4. A researcher conducted a study to compare the effectiveness of two different teaching methods, traditional and online, in improving students' test scores. The researcher randomly assigned 50 students to traditional teaching method and another 50 to the online teaching method. After implementing the teaching methods, the test scores were collected for both groups. The researcher performed a two-sample t-test to determine if there is a significant difference in test core between the traditional teaching method and the online teaching method. The obtained p-value from the t-test is 0.032.
   a. Describe the general steps involved in hypothesis testing. [5]
   b. Carry out a hypothesis testing procedure at 5% significance level to determine if there is a significant difference between the mean test scores between traditional teaching method and online teaching method. [10]
   c. State and briefly explain the assumptions of a two-sample t-test. [5]

5. Twenty-two second year students were randomly assigned to four groups to experiment with four different methods of teaching Biomathematics. At the end of the semester, the same test was given to all 22 students. The table below gives the scores of students in the four groups.

Method I	Method II	Method III	Method IV
19	14	11	24
21	16	14	19
26	14	21	21

24	13	13	26
18	17	16	20
	13	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 Biomathematics score of all second-year students taught by each of these four methods is 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. [2]
   (iii)Complete the One-Way ANOVA table. [11]
   (iv) Make a decision and conclude. [2]

6. Ten candidates sat examinations in Biomathematics and Biology. Their corresponding marks were: -

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

(i) Calculate the equation of the regression line of y on x using the method of least squares. [4]

(ii) Using your regression equation, calculate the biology (y) marks when Biomathematics (x) marks are 10 and 90 respectively. [2]

(iii) Hence plot the regression line on your diagram. [4]

(iv) The results in the table above were run in SPSS for the analysis of variance (ANOVA). Complete the ANOVA table below and test at 1% level of significance whether the slope B is positive. [10]

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

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	$\bar{x}_g = \sqrt[N]{x_1 x_2 x_3 \ldots x_N}$	$\bar{x}_g = \sqrt[n]{x_1 x_2 x_3 \ldots x_n}$
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	$\bar{x_g} = \sqrt[\sum f]{x_1^{f_1} x_2^{f_2} x_3^{f_3} \ldots x_N^{f_N}}$	$\bar{x_g} = \sqrt[\sum f]{x_1^{f_1} x_2^{f_2} x_3^{f_3} \ldots x_N^{f_n}}$
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: $x = 0,1,2,3\ldots\ldots, n$ $E(X) = np$ $var(X) = np(1-p)$	$P(X = x) = \frac{(e^{-\lambda})\lambda^x}{x!}$ $x = 0,1,2,3\ldots\ldots$ $E(X) = \lambda$ $var(X) = \lambda$

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	$SS_{Total} = \sum y^2 - \frac{(\sum y)^2}{n}$	$n-1$		