



MACHAKOS UNIVERSITY

University Examinations 2022/2023 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

THIRD YEAR SECOND SEMESTER EXAMINATION FOR

BACHELOR OF EDUCATION

SZL 300: BIOSTATISTICS

DATE: 18/4/2023

TIME: 8:30 – 10:30 AM

Instructions to the Candidate:

1. Answer **Question 1** and any other **two** questions.
2. You must have a scientific calculator for this paper.

QUESTION ONE (30 MARKS)

- (a) Differentiate each of the following terms as they apply in biostatistics
- i. Sample statistic and Population parameter
 - ii. Sampling error and Non-Sampling error
 - iii. Parametric test statistics and Non-parametric test statistics (6 marks)
- (b) The following is a record of the students who attended the lectures in the first day of the semester across different disciplines; 11, 17, 16, 3, 21, 22, 27, 19, 17, 6. By constructing a frequency table and a class interval of; 1-5, 6-10... determine the
- i. Mean
 - ii. Median
 - iii. Standard deviation
 - iv. Karl Pearson Coefficient of skewness. (12 marks)
- (c) A researcher presumed that the average microorganism abundance per square metre was 65 in an experimental plot unit. He did abundance count for 10 randomly selected plots and observed the following: 66, 65, 69, 70, 69, 71, 70, 63, 64 and 68 microorganisms per plot. The t-test output was as given

The sample t-test output, for the test value=65						
	t	df	Sig. (2-tailed)	Mean difference	95% CI of the difference	
					Lower	Upper
Attendance	2.825	9	0.020	2.500	0.4979	4.5021

- Construct the hypothesis statement for the above scenario
- Based on the t-value and the sig.(2-tailed) value make statistical conclusion
- Interpret the 95% CI lower and upper difference values (6 marks)

(d) Determine the regression equation of Y on X given that

$$\sum xy = 125, \sum x^2 = 2500, \bar{x} = 50 \text{ and } \bar{y} = 5. \text{ And hence calculate the value of Y when}$$

$$X=20 \quad (6 \text{ marks})$$

QUESTION TWO (20 MARKS)

- Highlight four advantages of using mean as a measure of central tendency in statistics (4 marks)
- Differentiate between the terms; *discrete variable and continuous variable* as used in statistics, giving an example for each. (4 marks)
- The data below is a summary of the calves weight before and after administering some fattening drug.

Calves	A	B	C	D	E	F	G	H
Weight Before	85	124	172	123	111	139	99	77
Weight after	105	160	175	143	156	127	95	100

Test the hypothesis that on average the fattening drug did not result to any improvement on the calves' weight. (8 marks)

- A forester took the height of two types of Eucalyptus tree and at one point recorded the data below.

Type of Eucalyptus	n	Height (ft)	
		Mean	Standard deviation
Red gum	15	12	3
Cider gum	30	16	2.5

Which of the two Eucalyptus varieties had higher height distribution variability? (4 marks)

QUESTION THREE (20 MARKS)

- The population standard deviation for the amount of aspirin in a batch of analgesic tablets is known to be 7 mg of aspirin. If after randomly selecting five tablets and analyzing it was found to contain 245 mg of aspirin, what is the 95% confidence interval for the population's mean? (5 marks)

- b) The ANOVA table output below shows the data results after analysing the effect of different fertilizers on the yield tried in different plots.

ANOVA Table					
Source of Variance	df	Sum sq	Mean sq	F value	Pr(>F)
Fertilizer	2	9.10	x	1.21	0.326
Residue	15	y	3.77		

Using the output values;

- Determine the values of x and y
- Make a statistical conclusion of the effect of different fertilizer on the yields

(5 marks)

- c) A random sample of 400 medical practitioners in East Africa was selected for in service clinical psychology training. As part of the training each trainee was asked to choose a country he/she preferred for an attachment from a list of three countries. The results are shown in the table below.

Age group	Country		
	Kenya	Tanzania	Rwanda
Under 30	120	30	50
31 – 50	10	75	15
51 and above	10	30	60

Test the hypothesis that the populations were homogeneous with respect to the country they preferred for an attachment despite their age difference.

(10 marks)

QUESTION FOUR (20 MARKS)

- a) A researcher in crop production is interested in studying the relationship between green grams yield and spacing. A random sample of 5 plots and their corresponding spacing in inches was recorded as shown in the table below:

Plot No.	1	2	3	4	5
Yield	70	15	52	85	25
Crop spacing	18	06	12	16	10

- Compute the *Karl Pearson's coefficient of correlation* between the yield and the crop spacing and interpret it. (6 marks)
- Compute the coefficient of determination and Interpret its value (2 marks)

- b) A meteorological department in an equatorial climate zone recorded the rainfall intensity at three points A, B and C for five days as summarized below:

A: 23000, 41000, 54000, 66000, 78000.

B: 20000, 30000, 34000, 40000, 44000.

C: 45000, 55000, 60000, 70000, 72000

Use Kruskal-Wallis test to determine if the rainfall intensity differed significantly among the three points. (12 marks)

QUESTION FIVE (20 MARKS)

- a) Highlight three characteristics of normally distributed data (3 marks)
- b) The normal heart rate for adults is assumed to be normally distributed. A random sample of 36 adults was found to have the heart beat rate of 70 beats per minute with a standard deviation of 13 rates. If an adult is randomly picked from the population, determine,
- The probability that his heart rate ranges between 65 to 80 per minute (6 marks)
 - The number of adults who may be subjected to further check-ups if it constituted of the adults whose heart rate is below 50 beats per minute (4 marks)

- c) The data below shows the mycelia coverage area in a rotting tree trunk.

Days	1	2	3	4	5
Radius of colonized trunk	2	5	3	8	7

- Determine the regression equations of the radius of the colonized trunk on days
- Assuming a linear relationship, predict the radius of the colonized trunk on the 10th day. (7 marks)