MACHAKOS UNIVERSITY

University Examinations for 2020/2021

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

THIRD YEAR FIRST SEMESTER EXAMINATION FOR

BACHELOR OF EDUCATION (SCIENCE)

BACHELOR OF SCIENCE (BIOLOGY)

SZL 300: BIOSTATISTICS

DATE: 10/8/2021 **TIME: 8:30 – 10:30 AM**

INSTRUCTION: *Answer Question* ***ONE*** *which is compulsory and any other* ***TWO*** *Questions*

QUESTION ONE (30 MARKS)

a) Discus three advantages of sampling over census (6 marks)

b) Differentiate the following terms as they apply in scientific research

i) Response and independent variables

ii) Descriptive and inferential statistics

iii) Null hypothesis and Alternative hypothesis (6 marks)

c) Given below are the lifespan of two insects in hours but from two different mountain zones I and II

Zone	*n*	*Average lifespan in Hours*	*Standard deviation*
I	150	2,500	400
II	100	2,000	200

Determine; i) The zone with a higher dispersion

ii) The combined standard deviation (6 marks)

d) Distinguish in words the following terms as they apply in statistic

i. Mean and Mode

ii. Variance and Range

iii. Data skewness and data kurtosis (6 marks)

e) The data given below represents the age in years of employees of an organisation.

28, 30, 33, 37, 37, 38, 42, 43, 43, 44, 45, 48, 48, 51, 55

Use the data to construct a box and whisker plot. (6 marks)

QUESTION TWO (20 MARKS)

a) The data below is a summary of yields (bags) per five acres before and after applying nitrogen top dressing fertilizer at different seasons around Machakos town.

5 acre plots	A	B	C	D	E	F	G	H
Yield Before	85	124	172	123	111	139	99	77
Yield after top dressing	105	160	175	143	156	127	95	100

Test the hypothesis that on average the top dressing did not result to any improvement on the yield per five acre plot. (8 marks)

b) Below is the daily milk production from a certain dairy farm in Machakos County during a particular week in the year 2020.

Milk Produced in litres	11-20	21-30	31-40	41-50	51-60
Frequency	3	6	11	3	2

Determine;

i. Average milk production per day
ii. Median
iii. Mode and
iv. Standard deviation. (12 marks)

QUESTION THREE (20 MARKS)

a) Highlight two properties of a good measure of data variation (4 marks)

b) The data below relates to the age of 10 employees of XYZ co. and sick off days in a particular month.

Age (years)	20	30	32	35	40	46	52	55	58	62
Sick off (days)	11	12	10	13	14	16	15	17	18	19

Calculate the Karl Pearson's coefficient of correlation and interpret its value. (8 marks)

Calculate the coefficient of determination and interpret its value. (2 marks)

c) Discuss any three types of test-statistics and highlight when appropriate to use each

(6 marks)

QUESTION FOUR (20 MARKS)

a) Two different types of drugs A and B were tried on certain patients for increasing weight, 5 persons were given drug A and 7 persons were given drug B. The increase in weight (in Kgs) is given below:

Drug A	8	12	13	9	3	-	-
Drug B	10	8	12	15	6	8	11

Determine whether the two drugs differ significantly with regard to their effect in increasing weight at $\alpha = 5\%$ (10 marks)

b) From the data given below

i. Use the least square method to determine the equation connecting Y on X

ii. Determine the sum of the residues

X:	1	2	3	4	5
Y:	2	5	3	8	7

(10 marks)

QUESTION FIVE (20 MARKS)

a) Highlight five advantages for continued scientific research (5 marks)

b) A manufacturer wished to compare the performance of different milking machine brands, their performance difference is summarized in ANOVA table below.

Source of Variation	**df**	**Sum of Squares**	**Mean sum of Squares**	**F – value**
Between machines	3	*x*	180.23	*z*
Error	12	85.75	*y*	
Total	15	626.44		

i) Complete the values *x, y* and *z* (3 marks)

ii) Explain the two sources of variation in column one of the table (4 marks)

iii) When is it appropriate to use ANOVA and F-test for data analysis (4 marks)

iv) Was there a significant difference in their performance at α=5% (4 marks)