MACHAKOS UNIVERSITY

University Examinations 2021/2022 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

THIRD YEAR FIRST SEMESTER EXAMINATION FOR

BACHELOR OF EDUCATION (SCIENCE)

SZL 300: BIOSTATISTICS

DATE: 29/8/2022 **TIME: 11.00-1.00 PM**

INSTRUCTION:

Answer Question One and Any Other Two Questions

You must have a scientific calculator for this paper.

QUESTION ONE (COMPULSORY) (30 MARKS)

a) Differentiate 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) 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

i. State the hypothesis for the above scenario

ii. Based on the t-value and the sig.(2-tailed) value make statistical conclusion

iii. Interpret the 95% CI lower and upper difference values (6 marks)

c) Determine the regression equation of Y on X given that

$\sum xy = 130$, $\sum x^2 = 2400$, $\bar{x} = 60$ and $\bar{y} = 4$. And hence calculate the value of Y when X=20 (6 marks)

d) Explain any **three** measurement scales for statistical data as outlined by Stevens (1946), giving two examples in each case. (6 marks)

e) Briefly explain three desirable properties of a good estimator that should be pursued when estimating a population parameter using a sample statistics. (6 marks)

QUESTION TWO (20 MARKS)

a) Define the term *variable* as used in statistics, giving an example. (2 marks)

b) The data below is a summary of the calves weight before and after a 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)

c) A random sample of 400 persons was selected from each of three age groups and each person was asked to specify which of the three presidential candidates she/ he preferred. The results are shown in table 1:

Table 1

Age group	Presidential Candidate		
	A	B	C
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 presidential candidate they preferred despite their age difference. (10 marks)

QUESTION THREE (20 MARKS)

a) 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 table below shows the frequency distribution of net monthly salaries of 60 randomly sampled employees in one of the Kenyan public University.

Monthly Net Salary Ksh '000'	Number of Employees
65-84	9
85-104	*"a"*
105-124	17
125-144	10
145-164	*"b"*
165-184	4
185-205	5

Given that the monthly average net salary is Ksh 122,500, determine

i. The values of "*a*" and "*b*" (6 marks)

ii. The Median (3 marks)

iii. The Standard deviation (3 marks)

iv. The Karl Pearson's coefficient of Skewness (3 marks)

QUESTION FOUR (20 MARKS)

a) You are subcontracted by a city consultant to establish whether the salaries of ABC company employees significantly differ per department. The following is a representation of staff for three departments randomly sampled:
Sales department: 23000, 41000, 54000, 66000, 78000.
Maintenance department: 20000, 30000, 34000, 40000, 44000.
Process and Packaging department: 45000, 55000, 60000, 70000, 72000
Use Kruskal Wallis test to refute or affirm the claim of salary difference per department (12 marks)

b) A researcher in crop production is interested in studying the relationship between green grams yield and spacing. A random sample of 10 plots and their corresponding spacing in inches was recorded as shown in the table below:

Plot No.	1	2	3	4	5	6	7	8	9	10
Yield	54	75	36	70	15	52	32	15	85	25
Crop spacing	12	11	08	18	06	12	10	09	16	10

i. Compute the *Pearson's product moment* co-efficient of correlation between the yield and the crop spacing. (7 marks)

ii. Interpret the co-efficient of correlation obtained in (a)(i) above. (1 mark)

QUESTION FIVE (20 MARKS)

a) Highlight two properties of a normal distribution data. (2 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,

i. The probability that his heart rate ranges between 65 to 80 per minute (6 marks)

ii. 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) A contracting farming agency collected the data below according to the listed variables in order to predict the yield of biashara green grams in future:

X1	23	34	56	78	15	21	24
X2	34	56	78	79	90	65	14
X3	23	12	12	13	14	16	23
Y	4	5	12	23	4	5	8

Y= Yield per acre in kg '000'

X1= Amount of PHk fertilizer applied per plant in grams

X2= Amount water applied per plant in litters for the season

X3 = Spacing in inches between two consecutives plants

The agency used ANOVA and regression, the following is part of the results:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.625373
R Square	0.391092
Adjusted R Square	-0.21782
Standard Error	7.63801
Observations	7

ANOVA

	df	*SS*	*MS*	*F*	*Significance*
Regression	3	112.411	37.47033	0.642284	0.637562
Residual	3	175.0176	58.33919		
Total	6	287.4286			

i. State the statistical model in this analysis. (2 marks)

ii. Interpret the finding in this analysis (at 5%significance level) (3 marks)

iii. Explain the difference between R Square and Adjusted R squared (3 marks)