MACHAKOS UNIVERSITY

University Examinations 2022/2023 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

SECOND YEAR SECOND SEMESTER EXAMINATION FOR

BACHELOR OF SCIENCE (BIOLOGY)

SZL 208: COMPUTER APPLICATIONS IN BIOLOGY

DATE: 19/4/2023 **TIME: 2:00 – 4:00 PM**

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 (COMPULSORY) (30 MARKS)

a) Differentiate the following terms as they apply in data analysis
   i. Sample statistic and Population parameter
   ii. A variable and a case
   iii. Discrete variable and continuous variable
   iv. Data mining and data screening (8 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. Construct the hypothesis statements 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) Highlight four ways by which the data integrity can be compromised. (4 marks)

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

e) Discuss three ways of handling an outlier at data analysis stage (6 marks)

QUESTION TWO (20 MARKS)

a) Discuss any four types of parametric test-statistics and state when appropriate to use each (4 marks)

b) Explain three challenges of data cleaning (6 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 below

	Country		
Religion	Kenya	Tanzania	Rwanda
Christianity	120	30	50
Muslim	10	75	15
Others	10	30	60

Test the hypothesis that the populations were homogeneous with respect to the country they preferred for an attachment despite their religion affiliation (10 marks)

QUESTION THREE (20 MARKS)

a) Distinguish each of the following terms as they apply in data analysis using computers

i. Computer system and computer software
ii. Spread sheet and database
iii. Probability sampling and non-probability sampling
iv. Sampling errors and non-sampling errors (8 marks)

b) A manufacturer wished to establish whether the airbags influenced the car sales. The data below shows the number and the type of a car sold and the airbags which were placed on both the driver's and the passenger seat, driver's seat only or none.

Data						
	Compact	Large	Midsize car	Small car	Sporty	Van
Driver & Passenger	2	4	7	0	3	0
Driver	9	7	11	5	8	3
None	5	0	4	16	3	6

Data output		
$\chi - Squared$	df	P-value
33.001	10	0.00027

Based on the data output and with justification give a statistical conclusion on whether there was a correlation between the types of car sold and the type of Air bags it had at $\alpha = 0.05$. (6 marks)

c) Discuss three benefits of data coding for analysis (6 marks)

QUESTION FOUR (20 MARKS)

a) Briefly explain two types of publications (4 marks)

b) Explain four advantages of using computer in research (4 marks)

c) Outline the main components of a research findings report (6 marks)

d) The following was a height in feet for sample of students from BSc in statistics class; 3.3, 4.5, 5.5, 5.7, 4.8 6.2, 5.6, 4.6, 4.5, 5.0. Present it in the stem-and-leaf plot (6 marks)

QUESTION FIVE (15 MARKS)

a) Differentiate the following terms as they apply in research
   i. Null hypothesis and alternative hypothesis
   ii. Type I error and Type II error in hypothesis testing
   iii. Confidence level and significance level (6 marks)

b) A contracting farming agency collected the data below according to the listed variables in order to predict the Q plant seeds rate of germination:

X1	23	34	56	78	15	21	24
X2	34	56	78	79	90	65	14
X3	3	4	5	6	4	3	4
Y	4	5	12	23	4	5	8

Y= number of seeds that germinated
X1= Soil moisture content (%)
X2= Soil temperature $\left({}^{0}\mathrm{F}\right)$
X3 = PH level

The agency used ANOVA and regression for analysis, 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	*x*	37.47033	*z*	0.637562
Residual	3	175.0176	*y*		
Total	6	287.4286			

i. State the statistical model in this analysis. (2 marks)
ii. Explain the difference between R Square and Adjusted R Square (4 marks)
iii. Complete the values x, y and z in the ANOVA (4 marks)
iv. Interpret the finding in this analysis (4 marks)