

Machakos University

FIRST YEAR EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE, BIOLOGY

SZL 208: COMPUTERS APPLICATIONS

Date.....

Time.....

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

Question one (30 marks)

- a) Briefly define data cleaning (2 marks)
- b) Discuss four advantages of sampling over census (8 marks)
- c) Briefly highlight four ways of dealing with a missing data during research analysis stage (4 marks)
- d) The following data represents, a plant specimen length in centimeters, **2.3, 2.5, 2.5, 2.7, 2.8 3.2, 3.6, 3.6, 4.5, 5.0** present it in a stem and leaf plot (4 marks)
- e) Differentiate the following terms as they apply in scientific research (6 marks)
 - i) Response and independent variables
 - ii) Descriptive and inferential statistics
 - iii) Computer software and Computer system
- f) By citing examples distinguish between experimental research design and non experimental research design (6 marks)

Question Two (20 marks)

- a) Briefly discuss four differences between spreadsheets and databases (8 marks)
- b) Define data integrity. Hence highlight five areas by which the data can be compromised (8 marks)
- c) Differentiate the following functions/terms as they apply in data analysis
 - i) Cases and variables
 - ii) Sorting and splitting the data file (4 marks)

Question Three (20 marks)

- a) Discuss three way of handling an outlier in a data set before data analysis (6 marks)
- b) The lecturer presumed that the average class attendance was 65 students. He did call register for 10 days and observed following: 66, 65, 69, 70, 69, 71, 70, 63, 64 and 68.
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

- State the hypothesis 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)
- c) Discuss any four types of test-statistics and highlight when appropriate to use each (8 marks)

Question Four (20 marks)

- The function $\hat{y} = 5.5 + 2.23x_1 + 0.78x_2 - 1.11x_3$ is model of predicting the expected yield (kg '000') per hectare of a certain crop as influenced by fertilizer $k63(x_1)$ the water content (x_2) and humidity level (x_3). Interpret the values 5.5, 2.23, 0.78 and 1.11 (6 marks)
- Discuss any four advantages of using computers in data analysis (8 marks)
- Outline three reasons for continued scientific research (6 marks)

Question Five (20 marks)

- Highlight three principles of experimental research (3 marks)
- By citing examples distinguish between categorical and continuous variables (6 marks)
- A manufacturer wished to compare the performance of different machine brands, their performance difference is summarized in ANOVA table below.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean sum of Squares	F – value
Between machines	$4 - 1 = 3$	540.69	180.23	$F = \frac{180.23}{7.15} = 25.207$
Error	$16 - 4 = 12$	85.75	7.15	
Total	$16 - 1 = 15$	626.44		

- Based on the table above state the number of machine brands compared (1marks)
- Explain the two sources of variation in column one of the table (2 marks)
- When is it appropriate to use ANOVA and F-test for data analysis (4 marks)
- Was there a significant difference in their performance at $\alpha=5\%$ (4 marks)

