



MACHAKOS UNIVERSITY

University Examinations for 2022/2023 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

THIRD YEAR SEMESTER EXAMINATIONS FOR

BACHELOR OF SCIENCE (MATHEMATICS)

SMA 367: EXPLORATORY DATA ANALYSIS

DATE:

TIME:

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

QUESTION ONE (COMPULSORY) (30 MARKS)

- a) Distinguish the following terms as they apply in data exploration and analysis
- Data skewness and data kurtosis
 - Descriptive analysis and predictive analysis (4 marks)
- b) Highlight four benefits of visual representation of research findings (4 marks)
- c) The following data represents the weight (KG) of newly born babies in Katu mission hospital, present it in a stem and leaf plot:
23, 35, 25, 22, 25, 43, 27, 54, 28, 55, 32, 36, 36, 45, 50 (4 marks)
- d) A farmer wished to evaluate the effect of fertilizer PHK on his 10 acre farm. He noted the yield per acre before and after applying the fertilizer as recorded below

Before	111	113	114	116	116	115	115	114	113	113
After	150	150	155	160	165	165	165	160	160	150

The output analysis for the yield is as summarized below.

The sample t-test output, for the test value=65						
	t	df	Sig.(2-tailed)	Mean difference	95% CI of the difference	
					Lower	Upper
Yield	26.833	9	6.711e-10	2.500	40.2955	47.7095

- i. Based on the t-value and the sig.(2-tailed) value make statistical conclusion
 - ii. Interpret the 95% CI lower and upper difference values
 - iii. Highlight the procedure in getting the output using any statistical software

(6 marks)
- e) Discuss three main components of the research findings report (6 marks)
- f) Briefly explain the following terms as they apply in exploratory data analysis
 - i. Exploratory data analysis (EDA)
 - ii. Robust statistical analysis
 - iii. Prescriptive data analysis

(6 marks)

QUESTION TWO (20 MARKS)

- a) Distinguish between the following terms as they apply in exploratory data analysis
 - i. Parametric and Non-parametric test statistics
 - ii. Prior probabilities and posterior probabilities

(6 marks)
- b) Show that the conjugate prior for poison likelihood is gamma distribution. (6 marks)
- c) Discuss four causes of data asymmetry (8 marks)

QUESTION THREE (20 MARKS)

- a) Complete the two- way ANOVA table by inserting the right figures in the spaces labeled in letters r, s, t, q, v, w, x and y (8 marks)

Two- way ANOVA

Source of Variation	Degrees of Freedom	Sum of Squares	Mean sum of Squares	F – value	F – table
Factor A	3	42	v	t	r
Factor B	2	w	16	q	s
AB	x	136	22.67		
Error	y	210			
Total	21	325			

Commend on the significance of the findings represented in the ANOVA table at $\alpha=5\%$

(2 marks)

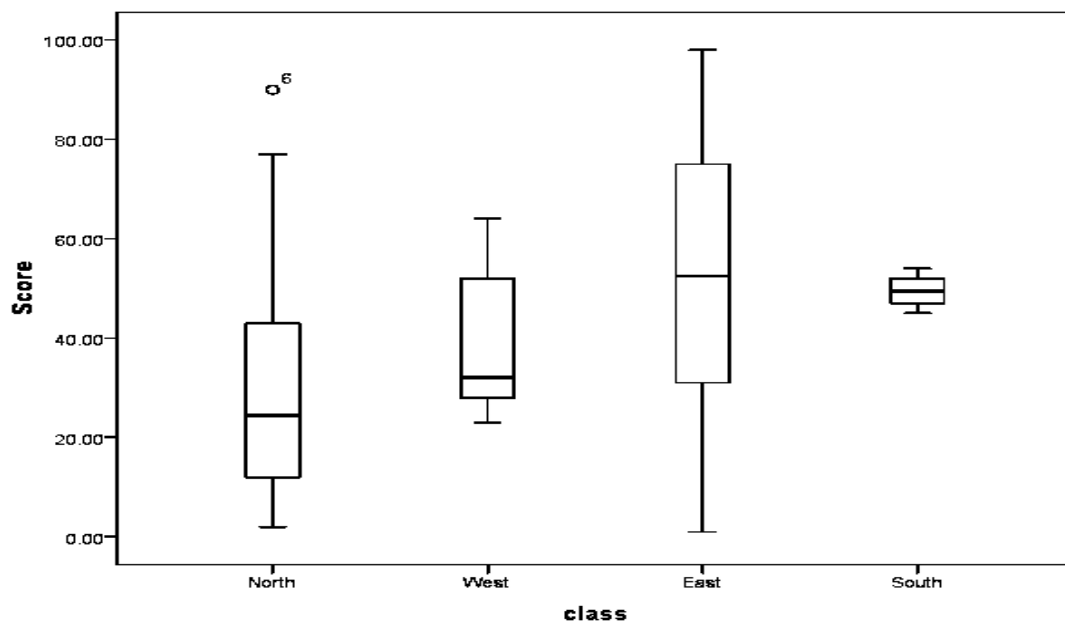
- b) Based on the recent observations 3% of the MU male post graduate students have prostate cancer. The university clinic laboratory can detect 90% of the prostate cancer when present. It also gives 13% false positive results. A male post graduate student in

MU volunteers for a prostate cancer test in the university clinic laboratory, determine the following probabilities;

- (a) That the test result will be positive
- (b) That, given a positive result, he has a prostate cancer
- (c) That, given a negative result, he has no prostate cancer
- (d) That he was misclassified (10 marks)

QUESTION FOUR (20 MARKS)

- a) Show that the conjugate prior for Binomial likelihood is Beta distribution. (6 marks)
- b) The marks distribution for form four classes, Kalaba secondary school were represented in a boxplot as shown below. Interpret the marks distribution for each class (4 marks)



- c) Discuss five areas of computer graphics applications (10 marks)

QUESTION FIVE (20 MARKS)

- a) Highlight four ways of normalizing a skewed data for analysis (4 marks)
- b) Discuss four assumptions of parametric test statistics to apply (8 marks)

- c) The populism of two products, A and B were tested by carrying out an interview among the residents of the two main consumer blocks. The number of consumers per block is summarized in the table below.

Block	Products	
	A	B
Highlands (H)	15	60
Lowlands (L)	55	25

Test whether the popularity levels differed significantly per block at $\alpha = 5\%$

(8 marks)