MACHAKOS UNIVERSITY

University Examinations for 2020/2021

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

SECOND YEAR FIRST SEMESTER EXAMINATION FOR

BACHELOR OF SCIENCE IN PROGRAMMIN & STATISTICS,
SST 200: INTRODUCTION TO COMPUTER INTERACTIVE STATISTICS

DATE: 12/8/2021 **TIME: 8:30 – 10:30 AM**

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

QUESTION ONE (COMPULSORY) (30 MARKS)

a) The following is weight in kilograms for ten randomly sampled Machakos University students in K63 class of 2019; 45, 39, 57, 65, 77, 49, 50, 66, 71, 52. Write the R codes for determining the following measurements; Mean, Median, Variance, Standard deviation, and Range. (7 marks)

b) Highlight four properties of a good estimator (4 marks)

c) Highlight four causes of asymmetry in data distribution (4 marks)

d) The following table shows the number of household members in certain town in 2018.

No of House hold Members	Percentage
1	18
2	32
3	20
4	19
5	7
6 or more	4

(i) Calculate the mean and standard deviation of the number of households. (6 marks)

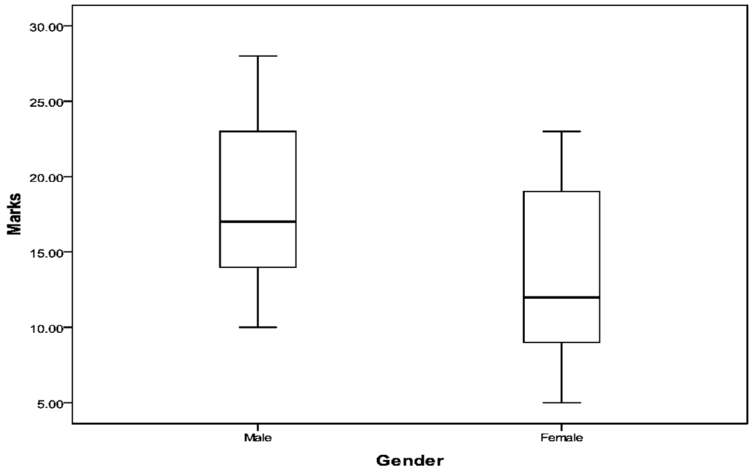
(ii) Assuming the data is based on a single random sample of 445; calculate a 95% confidence interval for the mean household size (5 marks)

e) Briefly explain the meaning of multicollinearity and highlight two ways of handling it during data analysis (4 marks)

QUESTION TWO (20 MARKS)

a) Discuss three ways of handling a missing data during analysis stage (6 marks)

b) The figure below represents the CAT marks for BSc class in statistics per gender. Use it to estimate the following



i) The median mark and the inter-quartile range per gender

ii) The gender with the higher dispersion.

iii) Write the R code for the output (8 marks)

c) A standardized stochastic test was given to 50 ladies and 75 gentlemen. The ladies made an average grade of 76 with a standard deviation of 6, while the gentlemen made an average grade of 82 with a standard deviation of 8. Find a 95% confidence interval for the difference μ_1 and μ_2, where μ_1 is the mean score of all gentlemen and μ_2 is the mean score of all ladies who took this test. (6 marks)

QUESTION THREE (20 MARKS)

a) Highlight any three components that must form part of the statistical research findings report (3 marks)

b) 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.

i. Present it in the stem-and-leaf plot

ii. Write the R code for stem and leaf plot (7 marks)

c) Discuss five areas prone to abuse of statistics (10 marks)

QUESTION FOUR (20 MARKS)

a) (i) Briefly state your understanding for exploratory data analysis. (3 marks)

(ii) Discuss any four ways of carrying out data exploration prior to further analysis (8 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

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
iv) Write the R codes for the above output
v) When is it appropriate to use T-test as the test statistics (9 marks)

QUESTION FIVE (20 MARKS)

a) Discuss three ways of handling an outlier during analysis stage (3 marks)

b) The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2 and 9.6 liters. Find a 95% confidence interval for the mean of all such containers, assuming an approximate normal distribution. (6 marks)

c) The following table represents the output of comparing the maize mean yield for different plots on applying the nitrogen fertilizer at varied levels of growth stages.

Source of Variation	**df**	**Sum of Squares**	**Mean sum of Squares**	**F – value computed**	**F – value Standard**
Between groups	3	*w*	180.23	*y*	*z*
Within groups	*u*	85.75	*x*		
Total	15	626.44			

i) Complete the missing values; *u, w, x, y* and *z* (5 marks)

ii) Based on the ANOVA table values was fertilizer top dressing timing a significant factor at α=5% (2 marks)

iii) When is it appropriate to use ANOVA and F-test for data analysis (2 marks)

iv) Write the R code for the above output (2 marks)