



MACHAKOS UNIVERSITY

University Examinations 2019/2020 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

SECOND YEAR SPECIAL /SUPPLEMENTARY EXAMINATION FOR
BACHELOR OF SCIENCE IN MATHEMATICS AND COMPUTER SCIENCE

BACHELOR OF SCIENCE IN STATISTICS AND PROGRAMMING

SST 200: INTRODUCTION TO COMPUTER INTERACTIVE STATISTICS

DATE: 20/01/2021

TIME: 2.00-4.00 PM

INSTRUCTIONS

Answer question one and any other two Questions Time 2 hours

QUESTION ONE (30 MARKS)

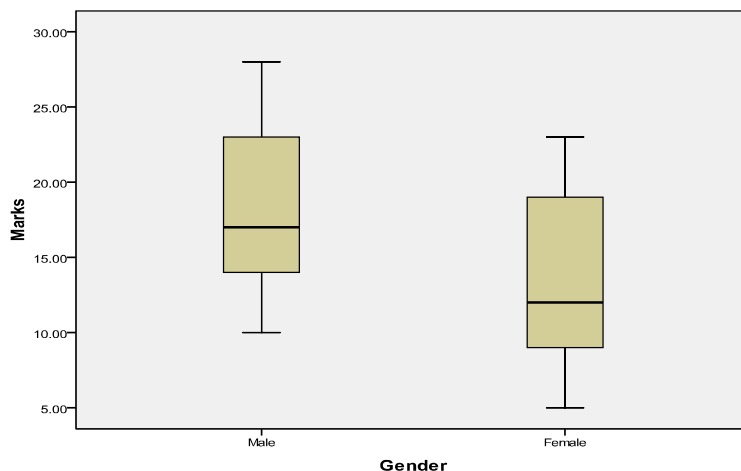
- a) Differentiate between each of the following terms:
- i. Sample statistic and Population parameter
 - ii. Point and interval estimation
 - iii. Sampling error and Sampling variability (6 marks)
- b) Highlight four properties of a good estimator (4 marks)
- c) Highlight four causes of asymmetry in 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. (5 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)
- f)
- i. The mean and standard deviation for the high jump of a random sample of 36 university students are calculated to be 2.6m and 0.3m, respectively. Find a 95% confidence interval for the mean of the entire class.
 - ii. How large a sample is required in (i) if we want to be 95% confident of μ is off by less than 0.05? (6 marks)

QUESTION TWO

- 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. (5 marks)
- c) Highlight three limitations of sampling (3 marks)
- d) 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 96% 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

- a) 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 (5 marks)
- b) Discuss five areas prone to abuse of statistics (10 marks)
- c) The mean lifetime of a sample of 100 light tubes produced by a company is found to be 1,580 hours with standard deviation of 90 hours. Test the hypothesis that the mean lifetime of the tubes produced by the company is 1600 hrs. (5 marks)

QUESTION FOUR

- a) i What is the meaning of exploratory data analysis?
- ii Discuss four ways of carrying out data exploration prior to further analysis (9 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 (6 marks)
- c) In a random sample of $n = 500$ college students owning mobile phones in Machakos town, it was found that $x = 340$ owned smart phones. Find a 95% confidence interval for the actual proportion of students in this town with smart phones. (5 marks)

QUESTION FIVE

- a) Discuss two ways of handling an outlier during analysis stage (2 marks)
- b) Differentiate between each of the following terms as they apply in hypothesis testing:
- i. Type I and type II error
- ii. One tail and two tails

- iii. Null hypothesis and Alternative hypothesis (6 marks)
- c) 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)
- d) In a batch chemical process, two catalysts are being compared for their effect on the output of the process reaction. A sample of 12 batches is prepared using catalyst 1 and a sample of 10 batches was obtained using catalyst 2. The 12 batches for which catalyst 1 was used gave an average yield of 85 with a sample standard deviation of 4, while the average for the second sample gave an average of 81 and a sample standard deviation of 5. Find a 90% confidence interval for the difference between the population means, assuming the populations are approximately normally distributed with equal variances (6 marks)