



MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

P.O. Box 972-60200 – Meru-Kenya

Tel: +254(0) 799 529 958, +254(0) 799 529 959, + 254 (0) 712 524 293,

Website: info@must.ac.ke Email: info@must.ac.ke

University Examinations 2023/2024

SECOND YEAR FIRST SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR
OF SCIENCE DATA SCIENCE

CDS 3200: STATISTICS FOR DATA SCIENCE

DATE: DECEMBER 2023

TIME: 2 HOURS

INSTRUCTIONS: Answer question *one* and any other *two* questions

QUESTION ONE (30 MARKS)

- a) With examples define types of data and variables
(5 marks)
- b) Define the following
- i. Parameter (2 marks)
 - ii. Statistic (2 marks)
 - iii. Statistical significance (2 marks)
- c) Briefly explain how the following can be computed in excel
- i. Descriptive statistics (2 marks)
 - ii. Frequency table (4 marks)
 - iii. Frequency polygon (3 marks)
- d) i. Discuss the chebyshev's theorem (2 marks)
- ii. 10 scores have a mean of 100 and a standard deviation of 15. What can we conclude from chebyshev's theorem (2 marks)

- e) Given the mean height of men as 68.34 in and standard deviation as 3.02 in the weight as 170.55lb and standard deviation as 20.12lb
- i. Compare variation among heights do variation among weights
(2 marks)
 - ii. Compare the height of 78.5 in to the weight of 231.5lb
(2 marks)
 - iii. Discuss how outliers are determined in a data set
(2 marks)

QUESTION TWO (20 MARKS)

- a) With examples, discuss different types of observational studies
(8 marks)
- b) Briefly discuss probability and non-probabilistic sampling methods commonly used in statistics
(8 marks)
- c) Define the following
- i. Skewness (2 marks)
 - ii. Kurtosis (2 marks)

QUESTION THREE (20 MARKS)

- a) Discuss the following probability distributions applicability. Demonstrate how they can be computed in excel
- i. Binomial distribution (3 marks)
 - ii. Poisson distribution (3 marks)
 - iii. Uniform distribution (3 marks)
 - iv. Normal distribution (4 marks)
- b) Briefly discuss the following
- i. Sampling distribution of the mean (3 marks)

- ii. Sampling distribution of variance (2 marks)
- iii. Sampling distribution of proportions(2 marks)

QUESTION FOUR (20 MARKS)

- a) Discuss the basic concepts of assessing normality
(8 marks)
- b) With examples, discuss errors in hypothesis testing
(4 marks)
 - i. Linear correlation coefficient (2 marks)
 - ii. P- value (2 marks)
 - iii. Errors involving correlation (4 marks)

QUESTION FIVE (20 MARKS)

- a) Explain
 - i. How scatter plots are done in excel (2 marks)
 - ii. Computing correlation coefficient r in excel
(4 marks)
 - iii. Interpretation of r (2 marks)

b) Given the data

Cost of pizza	0.15	0.35	1.15	1.25	1.63	2.25
Subway fare	0.15	0.29	1.02	1.15	1.22	1.01

- i. Compute the correlation coefficient (4 marks)
- ii. Use the P-value method to test the claim that there is a linear correlation between the data given (4 marks)
- c) Discuss
 - i. One tailed test (2 marks)
 - ii. Two tailed test (2 marks)