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: www.must.ac.ke Email: info@mucst.ac.ke

UNIVERSITY EXAMINATIONS 2023/2024

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

CDS 3200: STATISTICS FOR DATA SCIENCE

DATE: JANUARY 2025 **TIME: 2 HOURS**

INSTRUCTIONS: Answer Question ONE and any other TWO questions.

QUESTION ONE (30 MARKS)

a) Explain how the following can be computed in Excel
   i. Descriptive statistics (2 Marks)
   ii. Frequency table (4 Marks)
   iii. Frequency polygonic (3 Marks)

b) Describe data type and data dimensionality (6 Marks)

c) Define the following
   i. Parameter (2 Marks)
   ii. Statistic (2 Marks)
   iii. Statistical significance (2 Marks)

d) Highlight techniques used to identify outliers in data before analysis. (4 Marks)

e) i. Discuss the chebysher's theorem. (2 Marks)

ii. 1Q scores have a mean of 100 and standard deviation of 15 what can we conclude from Chebysher's theorem. (3 Marks)

QUESTION TWO (20 MARKS)

a) given the mean height of men as 68.34 and standard deviation as 3.02 ln, the weight as 170.551b and standards deviation as 20.12 lb

   i. compare variation among heights to variation among weight. (3 Marks)

   ii. compare the height of 785 ln to the weight of 231.5 lb. (3 Marks)

b) with examples discuss different types of observational studies. (8 Marks)

c) briefly discuss probabilistic and non-probabilistic sampling method. (6 Marks)

QUESTION THREE (20 MARKS)

a) Define

   i. skewness (2 Marks)

   ii. Kurtosis (2 Marks)

b) Discuss the following probability distributions applicability. Demonstrate how they can be computed in Excel.

   i. Binominal distribution (3 Marks)

   ii. Poisson distribution (3 Marks)

   iii. Uniform distribution (3 Marks)

   iv. Normal distribution (4 Marks)

c) With diagram explain the central dogma of statistics. (3 Marks)

QUESTION FOUR (20 MARKS)

a) Discuss the following

   i. Sampling distribution of the mean (3 Marks)

   ii. Sampling distribution of variance (2 Marks)

   iii. Sampling distribution of proportion (4 Marks)

b) Discuss the concept of assessing normality. (5 Marks)

c) Define the following.

   i. Parametric tests (2 Marks)

   ii. Non-parametric (2 Marks)

d) Discuss the underlying assumptions under parametric tests (2 Marks)

QUESTION FIVE (20 MARKS)

a) Define
   i. Linear correlation (2 Marks)
   ii. P-value (2 Marks)

b) Explain
   i. How scatler plots are computed in Excel (2 Marks)
   ii. Computing correlation coefficient r in excel (4 Marks)
   iii. Interpretation of r. (3 Marks)

c) Given the data

Cost of p	0.15	0.35	1.15	1.25	1.63	2.25
Fare	0.15	0.29	1.02	1.15	1.22	1.01

   i. Compute the correlation coefficient (4 Marks)

d) Discuss
   i. One failed test
   ii. Two failed test