



# 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](mailto:info@must.ac.ke) Email: [info@must.ac.ke](mailto:info@must.ac.ke)

---

## University Examinations 2023/2024

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

### SMS 3172: EXPLORATORY DATA ANALYSIS

DATE: DECEMBER 2023

TIME: 2 HOURS

---

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

---

#### QUESTION ONE (30 MARKS)

- a) With the help of a diagram briefly explain the central dogma of statistics (3 marks)
- b) Explain the importance of examining all your variable before making inference from data (4 marks)
- c) With examples, explain the types of data in commonly (4) further, discuss dimensionality of data sets giving examples (3 marks)
- d) Explain the importance of the following in exploratory data analysis
  - i. Histograms (2 marks)
  - ii. Box plots (2 marks)
  - iii. Scatter plots (2 marks)
- e) Explain how you can develop the following in excel
  - i. Frequency table (4 marks)
  - ii. Frequency polygon (3 marks)

## QUESTION TWO (20 MARKS)

- a) Distinguish between exploratory data analysis and confirmatory data analysis (4 marks)
- b) Explain the straight line methodology for confirmatory data analysis according to Turkey (1980) (4 marks)  
Further, explain how exploratory data analysis is incorporated (3 marks)
- c) Discuss the four themes of exploratory data analysis (8 marks)

## QUESTION THREE (20 MARKS)

- a) Define the following
- i. Skewness (2 marks)
  - ii. Kurtosis (2 marks)
  - iii. Standard error (2 marks)
- b) Highlight the steps involved in constructing stem and leaf plot given the data (4 marks)

58	72	64	66	67	92	51	55	69	73
64	59	65	55	75	56	89	60	84	68
74	67	55	68	74	43	67	71	72	66
62	63	83	64	51	63	49	78	65	75

Construct a stem and leaf diagram, give a brief interpretation of the data (6 marks)

- c) Discuss the assumption behind parametric tests (8 marks)

## QUESTION FOUR (20 MARKS)

- a) Discuss various data transformation techniques employed to reduce the effects of outliers in the data before analysis (6 marks)
- b) Explain how it is done in excel (4 marks)
- c) Discuss different types of observational studies
- d) Explain the following (6 marks)
- i. Probabilistic sampling

- ii. Non probabilistic sampling

**QUESTION FIVE (20 MARKS)**

- a) Discuss the following
  - i. Sampling distribution of the mean (3 marks)
  - ii. Sampling distribution of the variance (3 marks)
  - iii. Sampling distribution of the proportions (3 marks)
- b) Discuss the basic concepts of assessing normality (5 marks)
- c) Define
  - i. Linear correlation (2 marks)
  - ii. P- value (2 marks)
  - iii. Errors involving correlation (2 marks)