(University of Choice)
MASINDE MULIRO UNIVERSITY OF SCIENCE AND TECHNOLOGY
(MMUST)
MAIN CAMPUS

UNIVERSITY EXAMINATIONS
MAIN EXAM

2021/2022 ACADEMIC YEAR

THIRD YEAR SECOND SEMESTER EXAMINATION
FOR THE DEGREE OF BACHELOR OF SCIENCE IN EPIDEMIOLOGY AND
BIOSTATISTICS (BSc EPIMED)

COURSE CODE: HEM 323
COURSE TITLE: EXPLORATORY DATA ANALYSIS

DATE: 19/04/2022   **TIME: 3.00-5.00 PM**

INSTRUCTIONS TO CANDIDATES:
Answer all Questions from section A and any other two questions from section B
TIME: 2 Hours

MMUST observes ZERO tolerance to examination cheating
Paper Consists of 3 Printed Pages. Please Turn Over

1

SECTION A (40 MKS) COMPULSORY

Question one

1. Distinguish between the following terms;   (6 marks)
   a. Exploratory data analysis and classical data analysis
   b. Measures of central tendency and measures of dispersion
   c. Mode and median

2. In a common test for which the pass mark is 30, the distribution of marks of passing candidates classified by sex were given below;

Marks	30-34	35-39	40-44	45-49	50-54	55-59
Number of students	5	10	15	30	5	5

   Determine;
   a. Standard deviation   (8 marks)
   b. Coefficient of variation   (2 marks)

3. The number of birds visiting a bird table was counted each minute for 15 minutes. The distribution was as follows;
   9,4,7,5,3,8,6,5,8,10,5,6,5,8,10
   Find;
   a. Mean   (3 marks)
   b. Mode   (1 marks)
   c. Median   (1 marks)
   d. Comment on the skewness of this distribution.   (7 marks)

4. The table below shows Total fertility Rate (TFR) from 1999 census of Kenya by age group as at 24th August 1999;

Age group	15-19	209-24	25-29	30-34	35-39	40-44	45-49
TFR	3	5	4	7	6	4	2

   a. Define a Histogram, hence represent this information on a Histogram   (6 marks)
   b. Define a Frequency Polygon, hence represent this information on a Frequency Polygon
                                                            (6 marks)

SECTION B (ANSWER ANY TWO QUESTIONS)

QUESTION ONE (15 MKS)

The masses of fish caught by fishermen in a day are as shown in the table below

Mass (Kg)	0 – 4	5 – 9	10 14	15 – 19	20 – 24	25 – 29	30 – 34	35 – 39
No. of fish	2	6	20	12	10	5	6	2

  a. Compute the first three moments about the mean   (9 marks)
  b. Comment on the Kurtosis of the data above   (6 marks)

QUESTION TWO (15 MKS)
  a. State 3 advantages and 3 disadvantages of arithmetic mean as a measure of central tendency.

(6 marks)

  b. Below is the frequency distribution which resulted when the weight (in Kg) of 50 calves in a dairy

   farm were measured;

Weight	170	172.5	175	177.5	180	182.5	185	187.5	190
Frequency	1	2	4	6	8	9	7	6	7

Determine;
  i.   Mean using assumed mean   (3 marks)
  ii. The median   (3 marks)
  iii. The upper quartile   (3 marks)

QUESTION THREE (15 MKS)
  a. State and describe the major tasks in data mining.   (6 marks)
  b. Determine the mean deviation of the following data;   (9 marks)

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	10	25	30	20	15