



# 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](http://www.must.ac.ke) Email: [info@mucst.ac.ke](mailto:info@mucst.ac.ke)

---

## UNIVERSITY EXAMINATIONS 2024/2025

SECOND YEAR, FIRST SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE IN COMMUNITY HEALTH AND DEVELOPMENT

### HPP 3206: BIOSTATISTICS AND COMPUTING

DATE: JANUARY 2025

TIME: 2 HOURS

---

**INSTRUCTIONS: Answer Question ONE and any other TWO questions.**

---

#### QUESTION ONE (30 MARKS)

- a) Define the term Biostatistics (1 Mark)
- b) Differentiate between the following terms
- i. Primary and secondary data sources (2 Marks)
  - ii. Descriptive and inferential statistics (2 Marks)
  - iii. Null and alternative hypothesis (2 Marks)
- c) A random variable follows the normal distribution with a mean = 74.6 and a standard deviation 18.2 what is the probability that the variable will take in a value between 59.5 and 86.3 (4 Marks)
- d) A study was conducted on a group of 20 patients to record the number of days each patient took to recover from a mild respiratory infection after receiving treatment. The days taken for recovery were as follows:

Days to Recover (X): 3,3 5, 5, 6, 7, 8, 8, 8, 8,8, 8, 11, 11, 11, 12, 12, 12, 13,13.

- i. Construct a ungrouped frequency distribution table of the recovery days (3 Marks)
- ii. Calculate the mean number of days for recovery. (2 Marks)
- iii. Identify the mode of the recovery days (1 Mark)
- iv. Calculate the standard deviation of the recovery days (3 Marks)



MUST is ISO 9001:2015 and



ISO/IEC 27001:2013 CERTIFIED

- e) Imagine you are tasked with presenting data on the incidence of diabetes in different age groups across three regions. What types of charts or visualizations would you use to clearly communicate this information to both clinical professionals and a general audience, and why? Describe how each visualization type could enhance the understanding of trends, comparisons, and potential correlations in the data. (6 Marks)
- f) A health researcher is studying the average systolic blood pressure (SBP) of adults in a specific community. She randomly samples 50 adults and finds a sample mean SBP of 130 mmHg with a standard deviation of 15 mmHg.
- Construct a 95% confidence interval for the true mean systolic blood pressure of all adults in this community. (3 Marks)
  - Interpret the confidence interval in the context of this study (1 Mark)

### QUESTION TWO (20 MARKS)

- a) Describe two probability and two non-probability sampling methods applicable in your study area (8 Marks)
- b) In a study on the effectiveness of a new vaccination program to prevent a particular disease, 500 individuals were followed over a year. Out of 250 vaccinated individuals, 20 developed the disease. Among 250 unvaccinated individuals, 60 developed the disease.
- Create a 2x2 contingency table, giving row and column totals (2 Marks)
  - Use multiple bar charts to represent the data (4 Marks)
  - Is there an association between vaccination programs and disease development status? Carry out the necessary test, Use ch-square test: (6 Marks)

### QUESTION THREE (20 MARKS)

- a) The data below shows the average weight loss (Y in pounds) for people who went to a certain gym for physical exercise for one month for various numbers x (hours) per week

X (hours)	5	7	9	12	18	25
Y (Pounds)	3.4	5.3	7.4	10.4	14.5	19.9

- Draw a scatter plot of this data and comment on the relationship between x and y.

- (3 Marks)
- ii. Determine the correlation coefficient ( $r$ ) and interpret (7 Marks)
- iii. Compute the coefficient of determination and interpret (3 Marks)
- iv. Write the equation of the relationship between  $x$  and  $y$  in the form  $y = a + bx$ . (5 Marks)
- v. Predict the weight loss for someone who would exercise in that gym for a month at a rate of 50 hours per week and comment on your answer. (2 Marks)

**QUESTION FOUR (20 MARKS)**

- a) Explain two reasons for sampling as opposed to census (2 Marks)
- b) The following data shows blood lead levels measured in children or vulnerable populations to assess lead poisoning exposure in communities living near car battery manufacturing company.

14.3 14.4 14.5 14.7 14.9 15.0 15.4 15.5 15.8 16.0 16.1 16.2 17.0 17.6 17.9 18.5 18.9 19.9 20.7  
 20.8 11.9 21.2 13.8 14.6 16.2 17.4 12.9 14.9 15.7 16.3 13.6 15.1 16.8 15.7 14.5 14.9 16.1 13.6  
 14.2 18.4

- i. Construct a grouped frequency distribution table with 5 classes (5 Marks)
- ii. Present the data using a graph/chart of your choice. (3 Marks)
- iii. Compute two measures of location from the data (6 Marks)
- iv. What is the coefficient of variation. (4 Marks)

**QUESTION FIVE (20 MARKS)**

- a) Differentiate between discrete and continuous random variables (2 Marks)
- b) Explain any 3 data collection methods applicable in your field of study (6 Marks)
- c) List the steps in test of hypothesis (5 Marks)
- d) In a local community, the average daily intake of fruits and vegetables recommended by health authorities is 5 servings. A community health researcher believes that the residents of this community consume less than the recommended amount on average. To investigate, the researcher collects a random sample of 25 residents and finds that their average daily intake is 4.2 servings, with a standard deviation of 1.1 servings. Test  $H_0: \mu = 5$  (7 Marks)  
 Use  $t = 2.69$

