



ALUPE UNIVERSITY

**OFFICE OF THE DEPUTY VICE CHANCELLOR
ACADEMICS, RESEARCH AND STUDENTS' AFFAIRS**

UNIVERSITY EXAMINATIONS

2024/2025 ACADEMIC YEAR

FIRST YEAR SECOND SEMESTER REGULAR MAIN EXAMINATION

**FOR THE DEGREE OF BACHELOR OF SCIENCE
COMPUTER SCIENCE**

COURSE CODE: STA 122

COURSE TITLE: PROBABILITY AND STATISTICS I

DATE: 25th APRIL 2025

TIME: 2:00-5:00PM

INSTRUCTION TO CANDIDATES

- SEE INSIDE

THIS PAPER CONSISTS OF 3 PRINTED PAGES

PLEASE TURN OVER

STREAM: CS

DURATION: 3 Hours

INSTRUCTION TO CANDIDATES

- Answer ALL questions from section A and any TWO from section B*
- Do not write on the question paper.*

SECTION A COMPULSARY (34 MARKS)

QUESTION ONE (18 MARKS)

- a) The number of patients seen by a clinician per day in 20 consecutive days was recorded as follows: 22 27 39 53 29 58 44 21 45 13 18 12 27 29 34 15 10 34 40 20
- (i) Arrange the data in ascending order and compute the range. **(2 marks)**
 - (ii) Compute the lower quartile, middle quartile, and upper quartile of these data **(3 marks)**
 - (iii) Determine the Inter-Quartile Range (IQR) **(1 mark)**
 - (iv) Represent the data in a box-and-whisker diagram and use it to comment on the shape of the distribution **(3 marks)**
 - (v) Construct an ungrouped frequency distribution table to summarize these data **(3 marks)**
 - (vi) State two disadvantages of using ungrouped frequency tables to summarize data **(2 marks)**
 - (vii) Compute the mean and variance of these data **(4 marks)**

QUESTION TWO (16 MARKS)

- a) Distinguish between primary data and secondary data and give examples in each case. **(4 marks)**
- b) With the help of suitable examples, distinguish between the following terms as used in biostatistics:
- (i) Nominal and Ordinal data **(2 marks)**
 - (ii) Discrete and continuous variables **(2 marks)**
 - (iii) Quantitative data and Qualitative data **(2 marks)**
- c) Outline the 6 step-procedure followed when conducting a statistical hypothesis test **(6 marks)**

SECTION B, ANSWER ANY TWO QUESTIONS (26 MARKS)

QUESTION THREE (13 MARKS)

- a) Discuss any three methods of collecting primary data in your area of practice **(3 marks)**
- b) Describe what you understand by the term 'sampling' as used in biostatistics. Explain why it is necessary. **(3 marks)**
- c) Discuss the following methods of sampling as used in statistics. Where possible give a scenario when the said method would be appropriate for sample election.
- (i) Stratified random sampling **(2 marks)**
 - (ii) Judgement sampling **(2 marks)**
 - (iii) Multi-stage sampling **(3 marks)**

QUESTION FOUR (13 MARKS)

- a) Define the term correlation analysis and state two types of correlation **(3 marks)**
- b) If 4% of pesticides manufactured by a firm are ineffective, find the probability that in a sample of 200 pesticides;
- i. At most 2 are ineffective **(2 marks)**
 - ii. At least 3 are ineffective **(2 marks)**
- c) CAT and Examination marks for 10 students are indicated in the table below.

| | | | | | | | | | | | |
|------|----|----|----|----|----|----|----|----|----|----|----|
| CAT | 15 | 16 | 17 | 14 | 18 | 19 | 15 | 13 | 12 | 20 | 21 |
| Exam | 34 | 37 | 44 | 39 | 48 | 52 | 61 | 28 | 30 | 50 | 55 |

Compute the Spearman's Rank correlation coefficient.

State whether there is a statistical relationship between CAT marks and Examination Marks. **(6 marks)**

QUESTION FIVE (13 MARKS)

- a) During a study, data on the following variables was collected for each patient in a hospital ward (see table). For each of the following variables, state whether it is qualitative or quantitative; hence further classify into either nominal, ordinal, discrete or continuous. **(10 marks)**

| Variable | Type | Subtype |
|-----------------------------------|------|---------|
| Exact age of a patient (in years) | | |
| Weight (in grams) | | |
| Height (in meters) | | |
| Systolic blood pressure | | |
| Blood type | | |

- b) If the mean and variance of a binomial distribution are 12 and 4 respectively. Determine the number of trials (n) and probability of success (p). **(3 marks)**

QUESTION SIX (13 MARKS)

- a) Recently, research efforts have focused on the problem of predicting a Manufacturer's Market share by using information on the quality of its product. Suppose that the following data are available on Market share in percentage (Y) and the product quality

(X) measured on a scale of 0-100 determined by an objective evaluation procedure:

| | | | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|
| X | 27 | 39 | 73 | 66 | 33 | 43 | 47 | 55 | 60 | 68 | 70 | 75 | 82 |
| Y | 2 | 3 | 10 | 9 | 4 | 6 | 5 | 8 | 7 | 9 | 10 | 13 | 12 |

- (i) Draw a scatter diagram for these data and comment on the possible relationship between X and Y **(2 marks)**
- (ii) Find the equation of the linear regression model X and Y **(4 marks)**
- (iii) Predict the market share of a product whose rating is 76 **(1 mark)**
- (iv) Compute the product Moment Correlation coefficient between X and Y and interpret it. **(2 marks)**
- b) It is expected that 10% of production from a continuous process will be defective. Find the probability that in a sample of 10 units chosen at random;
- (i) Exactly two will be defective **(2 marks)**
- (ii) At least two will be defective **(2 marks)**