

**GPHY251**

**CHUKA**

**UNIVERSITY**



**UNIVERSITY EXAMINATIONS**

**SECOND YEAR EXAMINATION FOR THE AWARD OF DEGREE  
OF BACHELOR OF DEGREE IN PHYSICS**

**GPHY251: STATISTICAL DATA ANALYSIS**

**STREAMS: GPHY251**

**TIME :2 HOURS**

**DAY/DATE: DAY**

**DATE**

**TIME**

**INSTRUCTIONS: Answer question ONE and any other TWO questions**

**QUESTION ONE (30marks)**

- a. Give reasons why statistical literacy is important. [3marks]
- b. Distinguish between a sample and a population [2marks]
- c. What is random sampling? [1marks]
- d. Explain what is meant by an experimental design [3 marks]
- e. What is a normal distribution? [2marks]
- f. Explain the meaning of an outlier in data analysis [1marks]
- g. In a survey of 100 people who had recently purchased motorcycles, data on the following variables was recorded: Gender of purchaser, Brand of motorcycle purchased, Number of previous motorcycles owned by purchaser, Telephone area code of purchaser, Weight of motorcycle as equipped at purchase
  - i. Which of these variables are categorical? [2marks]
  - ii. Which of these variables are discrete numerical? [2marks]
  - iii. Which type of graphical display would be an appropriate choice for summarizing the gender data [1 mark]
  - iv. Which type of graphical display would be an appropriate choice for summarizing the weight data. [1mark]

h. Given the following sorted data

1.2, 1.5, 1.9, 2.4, 2.4, 2.5, 2.6, 3.0, 3.5, 3.8

Determine;

- |                         |          |
|-------------------------|----------|
| i. Median               | [1 mark] |
| ii. Mean                | [2marks] |
| iii. Mode               | [1mark]  |
| iv. Interquartile range | [2marks] |

k. List two graphical methods of analysing bivariate data. [2marks]

l. Explain the importance of inferential statistics [2marks]

m. Define the term hypothesis. [1mark]

### QUESTION TWO (20 marks)

- |                                                                                                             |          |
|-------------------------------------------------------------------------------------------------------------|----------|
| a. Explain the difference between data and a variable                                                       | [2mark]  |
| b. Distinguish between qualitative and quantitative data.                                                   | [2marks] |
| c. Discuss the difference between univariate, bivariate and multivariate data, giving relevant examples.    | [6marks] |
| d. Giving examples, distinguish between discrete and continuous variable.                                   | [4marks] |
| e. Explain the meaning of dependent, independent and control variables as used in statistical data analysis | [6marks] |

### QUESTION THREE (20marks)

- |                                                                     |           |
|---------------------------------------------------------------------|-----------|
| a. Discuss any two types of experimental designs.                   | [4marks]  |
| a. Explain two aspects to be considered when creating a test group. | [ 4marks] |
| b. Distinguish between random and non- random sampling              | [2marks]  |
| c. Discuss two random sampling methods.                             | [4marks]  |
| d. Give two applications of systematic sampling                     | [2 marks] |
| e. State the central limit theorem                                  | [2marks]  |
| f. list any two statistical checks for a normal distribution        | [2marks]  |

**QUESTION FOUR (20marks)**

a. The table 1 below summarizes the resulting data from a study of students scores in a statistical test.

Table 1

| Class interval     | 0 to <10 | 10 to <20 | 20to < 30 | 30 to < 40 | 40 to <50 | 50 to <60 | 60 to <70 |
|--------------------|----------|-----------|-----------|------------|-----------|-----------|-----------|
| Frequency          | 5        | 10        | 20        | 30         | 20        | 10        | 5         |
| Relative frequency |          |           |           |            |           |           |           |

- Complete the table. [3marks]
- Construct a corresponding relative frequency histogram for the scores. [5marks]
- Describe the shape of the histogram? [2marks]

b. Table 2 comprises of two sets of data of a sample.

Table 2

| Set 1      | $X - \bar{X}$ | $(X - \bar{X})^2$ | set 2 | $X - \bar{X}$ | $(X - \bar{X})^2$ |
|------------|---------------|-------------------|-------|---------------|-------------------|
| 24         |               |                   | 28    |               |                   |
| 23         |               |                   | 25    |               |                   |
| 22         |               |                   | 22    |               |                   |
| 22         |               |                   | 22    |               |                   |
| 21         |               |                   | 10    |               |                   |
| 20         |               |                   | 5     |               |                   |
| $\Sigma =$ |               |                   |       |               |                   |

- Fill in the table. [4marks]
- For both sets calculate;
  - Mean ( $\bar{X}$ ) [ 2 marks]
  - Variance [2 marks]
  - Standard deviation [2marks]

### QUESTION FIVE (20Marks)

- a. Explain what is meant by correlation in statistical data analysis. [2marks]
- b. The dose  $x$  (in gms) and concentration in urine,  $y$  (in mg/gm) of a fluid after intravenous administration are measured for 12 people:
- $x$ : 46 53 37 42 34 29 60 44 41 48 33 40
- $y$ : 12 14 11 13 10 8 17 12 10 15 9 13
- i. Prepare a scatter plot for these data. [6marks]
- ii. What does the scatter plot show about the relationship between  $x$  and  $y$ . [2marks]
- iii. Compute the value of correlation coefficient between  $x$  and  $y$ . Is the value of  $r$  consistent with your answer in ( i ) above? [6marks]
- c. Discuss any two data requirements for regression analysis. [4marks]