NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY OF APPLIED SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

STATISTICAL PROGRAMMING

SIA2105

Examination Paper

First Semester, 2024

This examination paper consists of 3 pages

Time Allowed:	3 hours
Total Marks:	100
Examiner's Name:	Mr. W. Marabada
External Examiner:	Dr L. C. Sakala

INSTRUCTIONS

1. Answer any four (4) questions
2. Each question carries 25 marks

MARK ALLOCATION

QUESTION	MARKS
1.	**25**
2.	**25**
3.	**25**
4.	**25**
5.	**25**

QUESTION ONE

a) Explain why might an organization elect to use proprietary software instead of open-source software. **[7]**

b) Discuss any 4 functions that can be used for debugging in R. **[8]**

c) Describe the key stages involved in data wrangling, highlighting its importance to the data analysis process. **[10]**

QUESTION TWO

a) Write an R function that takes an array of numbers x and returns the smallest number in the array. [5]

b) Write an R function to check whether the given number is even or odd. **[7]**

c) Given Java and R programming languages, which one would you prefer for data science analytics and why? **[13]**

QUESTION THREE

a) Explain the term cluster analysis. **[2]**

b) Show how variables are declared and explain why it is necessary to declare variables in programming. **[5]**

c) Write a function that recursively computes the n'th Fibonacci number. **[8]**

d) Identify and explain 5 data types available in R. **[10]**

QUESTION FOUR

a. Explain how a confusion matrix is used to evaluate the effectiveness of the model. **[5]**

b. What are the problems that statistical programmers are likely to face when handling large volumes of data? **[10]**

c. Describe the 5 essential components available in statistical programming. **[10]**

QUESTION FIVE

a. Write a code snippet in R that will list all the data sets available in all R packages. **[5]**

b. State with reasons why reserved words cannot be used as constants, variables, or any other identifier names in a program. Give any three reserved words available in R Language. **[10]**

c. Analyse the potential challenges and limitations of using R as a statistical programming language over other programming languages. **[10]**

END OF QUESTION PAPER

