



GWANDA STATE UNIVERSITY

Faculty of Computational Sciences

DEPARTMENT OF MATHEMATICS AND STATISTICS

Statistical Software

CMS 2205

Examination Paper

April 2025

This examination paper consists of 4 printed pages

Time Allowed: 3 hours

Total Marks: 100

Examiner's Name: Mr. E. Utete

INSTRUCTIONS

Answer **ALL** questions in Section A and **ANY THREE** questions in Section B

ADDITIONAL REQUIREMENTS

Section A use Answer Booklet.

Section B Use a computer with R programming Software.

Save your work on the Desktop in a folder named name_surname_regnumber.

Practical work is done using R Markdown, save all your work in a folder.

Page 1 of 4

Copyright: Gwanda State University, 2025

Section A: Theory Questions (25 Marks)

Answer All Questions Using Answer Booklet

- A1** (a) What is the purpose of the `dplyr` package in R? Name three key functions from this package and explain their use. [5]
- (b) Explain the difference between `ggplot2` and base R graphics. Provide an example of when you might prefer one over the other. [5]
- (c) What is R Markdown, and how does it support reproducible research? [5]
- A2** (a) Compare R, SPSS, and SAS in terms of:
- i. Ease of use
 - ii. Cost
 - iii. Flexibility for advanced statistical analysis
- [7]
- (b) In what scenarios would you recommend using SPSS or SAS over R? Justify your answer. [3]

Section B: Practical Tasks: Answer the whole of this section using R Markdown, type clearly all answers, comments, R codes and outputs. Save your output file as a Microsoft Word Document. (Total:75 Marks)

Answer Any 3 Out of 4 Questions

Each Question is Worth 25 Marks: Use a Computer with R programming Software

- B3** You are provided with a dataset named `students.csv` containing the following variables:
- ID: Student ID
 - Name: Student name
 - Age: Student age
 - Gender: Student gender (Male/Female)
 - Score: Exam score (out of 100)

Tasks:

- (a) Load the dataset into R. [2]

- (b) Display the first 10 rows of the dataset. [2]
- (c) Calculate the mean, median, and standard deviation of the `Score` variable. [4]
- (d) Create a new variable `Grade` based on the following criteria:
 - i. A if `Score >= 80`
 - ii. B if `60 <= Score < 80`
 - iii. C if `Score < 60`
- (e) Filter the dataset to include only female students and display the result. [6]
- (f) Display the cleaned dataset on your R Markdown document. [5]

B4 Using the `students.csv` dataset:

- (a) Create a histogram of the `Score` variable with appropriate labels and title. [6]
- (b) Create a boxplot of `Score` by `Gender` and interpret the result. [6]
- (c) Create a scatterplot of `Age` vs. `Score` and add a trendline. [7]
- (d) Create a bar chart showing the average `Score` by `Gender` and add appropriate labels. [6]

B5 You are provided with a dataset named `sales.csv` containing the following variables:

- **Region:** Sales region (North, South, East, West)
- **Product:** Product type (A, B, C)
- **Revenue:** Sales revenue (in thousands)

Tasks:

- (a) Fit a linear regression model to predict `Revenue` based on `Product`. [8]
- (b) Interpret the coefficients of the model. [5]
- (c) Calculate and interpret the R-squared value. [5]
- (d) Create a residual plot and comment on the model's assumptions. [7]

B6 You are provided with a dataset named `employees.csv` containing the following variables:

- **ID:** Employee ID
- **Name:** Employee name
- **Department:** Department (HR, Finance, IT, Sales)
- **Salary:** Monthly salary (in USD)
- **Experience:** Years of experience

Tasks:

- (a) Write a function in R to calculate the bonus for each employee based on the following criteria:
- i. If `Experience` ≥ 5 , `bonus` = 20% of `Salary`
 - ii. If `Experience` < 5 , `bonus` = 10% of `Salary`
- [8]
- (b) Use the `apply()` function to calculate the total salary (`Salary` + `Bonus`) for each employee. [6]
- (c) Create a new dataset containing only employees from the "Sales" department. [6]
- (d) Write a loop to print the names of all employees with a salary greater than \$5000. [5]

End of Examination Paper

Total: 100 marks.