SCHOOL OF COMPUTER SCIENCE
TAYLOR'S UNIVERSITY

ASSESSMENT BRIEF

MODULE DETAILS	
MODULE NAME	STATISTICAL INFERENCE AND MODELING STATISTICS AND OPERATIONAL RESEARCH
MODULE CODE	ITS66804 STATISTICAL INFERENCE AND MODELING MTH60904 STATISTICS AND OPERATIONAL RESEARCH
LECTURER	HUSNA SARIRAH HUSIN
YEAR/SEMESTER	2025/APRIL

ASSESSMENT DETAILS	
TITLE/NAME	INDIVIDUAL LAB ASSESSMENT
WEIGHT	30%
DATE/DEADLINE	
COURSE OUTCOME(S)	MLO2: Perform statistical inferences and predictions using a suitable statistical tool to solve a given problem.
INSTRUCTIONS	For this assignment, students are required to use any public or built-in dataset available in R. Examples of datasets include *mtcars, iris, airquality*, Boston (from the MASS package), or any other dataset downloaded from public repositories like Kaggle or UCI Machine Learning Repository.
OBJECTIVES	The objective is to analyze the dataset, perform statistical inferences, and build predictive models using R.
DELIVERABLES	1. An R script to answer each question and justifications where necessary. 2. Screenshots of sample queries and their outputs.
EVALUATION CRITERIA	• Correctness: R script syntax and logic are correct. • Completeness: All tasks are completed as per instructions. Screenshots of output must be provided where necessary. • Clarity: The report is well-organized, and explanations are clear. • Efficiency: R scripts are optimized and follow best practices.

Answer ALL questions

Question 1

Perform exploratory data analysis (EDA) to understand the dataset.

a. Load your chosen dataset into R and display the first 6 rows.

(5 marks)

b. Check for missing values in the dataset. If any are found, handle them appropriately.

(5 marks)

c. Generate summary statistics for all numerical variables.

(5 marks)

d. Create at least two visualizations (e.g., scatterplot, histogram, boxplot) to explore relationships or distributions in the data.

(5 marks)

Question 2

Test whether there is a significant difference between two groups in the dataset.

a. Identify two groups in your dataset (e.g., automatic vs manual transmission in *mtcars*).

(5 marks)

b. State the null hypothesis and alternative hypothesis.

(5 marks)

c. Perform an appropriate statistical test (e.g., t-test, ANOVA) to compare the groups.

(5 marks)

d. Interpret the results of the test.

(5 marks)

Question 3

Build a regression model to predict a numerical variable in the dataset.

a. Choose a numerical variable as the response variable (e.g., *mpg* in *mtcars*).

(5 marks)

b. Fit a multiple linear regression model using other variables in the dataset as predictors.

(5 marks)

c. Display the summary of the regression model.

(5 marks)

d. Interpret the coefficients of the model.

(8 marks)

Question 4

Use the regression model to make predictions.

a. Create a new data frame with hypothetical values for the predictors used in the regression model.

(10 marks)

b. Use the regression model to predict the response variable for this new data point.

(10 marks)

Question 5: Evaluate the performance of the regression model.

a. Calculate the R-squared value of the model.

(6 marks)

b. Plot the residuals vs fitted values to check for patterns.

(6 marks)

c. Discuss whether the model assumptions are satisfied.

(5 marks)

--- END OF INDIVIDUAL LAB ASSIGNMENT QUESTIONS ---