

# Summative Assessment RM&DA

Your Name

2025-06-08

## Table of contents

|  |          |
|--|----------|
| Short Brief . . . . .                      | 1        |
| <b>Section 1: Data Wrangling</b>           | <b>2</b> |
| Step 1 - Load the database . . . . .       | 2        |
| Step 2 - Data wrangling . . . . .          | 2        |
| Step 3 - Data Summarisation . . . . .      | 3        |
| <b>Create scatter plot of mpg vs wt</b>    | <b>3</b> |
| <b>Create boxplot of mpg by cyl</b>        | <b>4</b> |
| <b>Create bar chart of cylinder counts</b> | <b>4</b> |
| <b>Perform t-test</b>                      | <b>4</b> |
| <b>Calculate correlation</b>               | <b>5</b> |
| <b>Perform chi-squared test</b>            | <b>5</b> |
| <b>Fit linear model</b>                    | <b>5</b> |
| <b>Fit logistic model</b>                  | <b>5</b> |

### Short Brief

- This is an open-book exam. You may use any resources available to you (notes, books, online documentation).
- You are expected to work individually. Collaboration is not permitted.

- All code and answers must be clearly presented in this document.
- Ensure this document renders successfully to both an HTML and a PDF file.
- Show all your work, including the R code used to arrive at your answers.

## Section 1: Data Wrangling

This section uses the `mtcars` dataset, which contains data about various car models.

### Step 1 - Load the database

Load the dataset

```
# Place your code here
```

Inspect the dataset for basic characteristics such as type of variables, dimensions of the dataset, etc.

```
# Place your code here
```

### Step 2 - Data wrangling

- Create a new variable based on an existing one

The `am` column in the `mtcars` dataset represents the transmission type (0 = automatic, 1 = manual). Create a new factor variable called `transmission_type` based on the `am` column to achieve the result below:

|                   | am | transmission_type |
|-------------------|----|-------------------|
| Mazda RX4         | 1  | Manual            |
| Mazda RX4 Wag     | 1  | Manual            |
| Datsun 710        | 1  | Manual            |
| Hornet 4 Drive    | 0  | Automatic         |
| Hornet Sportabout | 0  | Automatic         |
| Valiant           | 0  | Automatic         |

- Create `transmission_type`

```
# Place your code here
```

### Step 3 - Data Summarisation

Calculate the mean mpg (miles per gallon) for cars with 4 cylinders (cyl) and for cars with 6 cylinders.

- Calculate mean mpg for 4 and 6 cylinder cars

```
# A tibble: 2 x 2
  cyl  mean_mpg
  <dbl>    <dbl>
1     4      26.7
2     6      19.7
```

```
# Place your code here
```

Question 4 ([Insert points here])

Filter the data to include only observations where the car has a manual transmission and has more than 4 gears. How many observations meet these criteria?

```
manual_and_more_than_4_gears <- mtcars %>% filter(am == 1 & gear > 4)
nrow(manual_and_more_than_4_gears)
```

Your answer here

Section 2: Data Visualization This section continues to use the mtcars dataset.

Question 5 ([Insert Points Here])

Create a scatter plot showing the relationship between mpg and wt (weight). Add appropriate labels for the axes and a title for the plot.

### Create scatter plot of mpg vs wt

```
ggplot(mtcars, aes(x = wt, y = mpg)) + geom_point() + labs(title = "Relationship between MPG and Weight", x = "Weight (1000 lbs)", y = "MPG")
```

*Your written answer for Question 5 goes here.*

Question 6 ([Insert Points Here])

Generate a boxplot to visualize the distribution of mpg across different cyl (number of cylinders) values.

## Create boxplot of mpg by cyl

```
ggplot(mtcars, aes(x = factor(cyl), y = mpg)) + geom_boxplot() + labs(title = "MPG Distribution by Number of Cylinders", x = "Number of Cylinders", y = "MPG")
```

*Your written answer for Question 6 goes here.*

Question 7 ([Insert Points Here])

Create a bar chart showing the count of cars for each number of cylinders (cyl).

## Create bar chart of cylinder counts

```
ggplot(mtcars, aes(x = factor(cyl))) + geom_bar() + labs(title = "Car Count by Number of Cylinders", x = "Number of Cylinders", y = "Count")
```

*Your written answer for Question 7 goes here.*

Section 3: Basic Statistics This section uses the mtcars dataset.

Question 8 ([Insert Points Here])

Perform an independent samples t-test to compare the mean mpg for cars with automatic and manual transmissions. State the null and alternative hypotheses, report the p-value, and interpret the results.

## Perform t-test

```
t.test(mpg ~ am, data = mtcars)
```

*Your written answer for Question 8 goes here.*

Question 9 ([Insert Points Here])

Calculate the Pearson correlation coefficient between mpg and hp (horsepower). Interpret the strength and direction of the correlation.

## Calculate correlation

```
cor(mtcarsmpg, mtcarshp)
```

*Your written answer for Question 9 goes here.*

Question 10 ([Insert Points Here])

Conduct a chi-squared test to examine the relationship between cyl (number of cylinders) and am (transmission type). State the null and alternative hypotheses, report the chi-squared statistic and p-value, and interpret the results.

## Perform chi-squared test

```
table_cyl_am <- table(mtcarscyl, mtcarsam) chisq.test(table_cyl_am)
```

*Your written answer for Question 10 goes here.*

Question 11 ([Insert Points Here])

Fit a linear model to predict mpg using wt (weight) and hp (horsepower). Provide the R code, summarize the model, and interpret the coefficients for wt and hp.

## Fit linear model

```
linear_model_mpg <- lm(mpg ~ wt + hp, data = mtcars) summary(linear_model_mpg)
```

*Your written answer for Question 11 goes here.*

Question 12 ([Insert Points Here])

Fit a logistic regression model to predict am (transmission type) using mpg and wt (weight). Provide the R code, summarize the model, and interpret the odds ratio for mpg.

## Fit logistic model

```
logistic_model_am <- glm(am ~ mpg + wt, data = mtcars, family = "binomial") summary(logistic_model_am) exp(coef(logistic_model_am))
```

*Your written answer for Question 12 goes here.*

Section 4: Basic Introduction to Multivariate Analysis This section uses the iris dataset.

Question 13 ([Insert Points Here])

Briefly explain the purpose of Principal Component Analysis (PCA). In the context of the 'iris'