DESCRIPTION OF COURSEWORK

Course Code	G0135
Course Name	Probability and Statistics in Real Life
Lecturer	Koh Siew Khew
Academic Session	February 2025
Assessment Title	R Assignment (Team of Two)

A. Course Learning Outcomes (CLO) covered

At the end of this assessment, students are able to:

CLO 1 Apply the concepts and theories in probability and statistics.

CLO 2 Identify the right tools to explore a given data set.

B. University Policy on Academic Misconduct

1. Academic misconduct is a serious offense in Xiamen University Malaysia. It can be defined as any of the following:
   i. **Plagiarism** is submitting or presenting someone else's work, words, ideas, data or information as your own intentionally or unintentionally. This includes incorporating published and unpublished material, whether in manuscript, printed or electronic form into your work without acknowledging the source (the person and the work).
   ii. **Collusion** is two or more people collaborating on a piece of work (in part or whole) which is intended to be wholly individual and passed it off as own individual work.
   iii. **Cheating** is an act of dishonesty or fraud in order to gain an unfair advantage in an assessment. This includes using or attempting to use, or assisting another to use materials that are prohibited or inappropriate, commissioning work from a third party, falsifying data, or breaching any examination rules.
2. All the assessment submitted must be the outcome of the student. Any form of academic misconduct is a serious offense which will be penalised by being given a zero mark for the entire assessment in question or part of the assessment in question. If there is more than one guilty party as in the case of collusion, both you and your collusion partner(s) will be subjected to the same penalty.

C. Questions

This quiz is based on a data file in **R** about cars (Question 1, 2, 3). To retrieve the data, type:

```
> mtcars
```

To learn more about the data set, type:

```
> help(mtcars)
```

QUESTION 1 (using mtcars data) – **10 MARKS**

(a) Name 3 subjects in the data file. [2]

(b) How many of the cars are automatic (transmission)? Justify your answer with an **R** output. [3]

(c) Use the *z*-score method to verify that the variable *Rear Axle Ratio* has no outliers. Provide the relevant **R** output of the summary statistics. [5]

QUESTION 2 (using mtcars data) – **20 MARKS**

(a) Which variable, *gross horsepower* or *weight* (in 1000 lbs), has a stronger linear relationship with *miles/(US) gallon*? Justify your answer with an **R** output. [4]

(b) Produce a scatter plot with **R** of the stronger pair and describe the direction of the relationship. [4]

(c) Treat *miles/(US) gallon* as the response variable and consider the stronger pair. Give a value of the explanatory variable such that an estimation should not be made with it due to the danger of extrapolation. [2]

(d) Treat *miles/(US) gallon* as the response variable and consider the stronger pair. Find the regression line with **R** and interpret the two coefficients in plain English. [5]

(e) Predict the *miles/(US) gallon* for Fiat 128 using the regression line in part (d) and compute the residual by hand. [5]

QUESTION 3 (using mtcars data) – **20 MARKS**

Consider the question "Is there an association between the type of engine of a car and its number of forward gears?"

(a) Produce a frequency contingency table with **R** for the question considered above. How many cells are there in the table? [5]

(b) Let *type of engine* be the response variable. Use **R** to find the conditional proportions and to produce the corresponding stacked bar chart. [5]

(c) Let *number of gears* be the response variable. Use **R** to find the conditional proportions and to produce the corresponding stacked bar chart. [5]

(d) Use one of the charts in parts (b) and (c) to answer the question considered above. Clearly explain your answer. [5]

QUESTION 4 – 20 MARKS

A teacher announces a pop quiz for which Gaston is completely unprepared. The quiz consists of 100 true-false questions. Gaston has no choice but to guess the answer randomly for all 100 questions.

(a) Use **R** to simulate Gaston's answers. Show the **R** code and the **R** output. [5]

(b) The table below shows the 100 correct answers. The answers should be read across rows. What proportion of the questions did Gaston answer correctly? Use **R** to do this. Show the **R** code and the **R** output. [5+5]

Pop Quiz Correct Answers

T	F	T	T	F	F	T	T	T	T	F	T	F	F	T	T	F	T	F
F	F	F	F	F	F	F	T	F	F	T	F	T	F	F	T	F	T	T	F
T	F	F	F	F	F	T	F	T	T	F	T	T	T	F	F	F	F	F	T
T	F	F	T	F	F	T	T	T	T	F	F	F	F	F	F	F	T	F	F
F	F	T	F	F	T	T	F	F	T	F	T	F	T	T	T	T	F	F	F

(c) If we take the average of all the answers to part (b) from our class of 50 students, what should the average be closed to? Why? [5]

QUESTION 5 – 30 MARKS

This question requires you to compare two set of random numbers in terms of their statistics.

(a) Generate two set of data (rand01 and rand02) that consist of 1000 random numbers as follows:

```
set.seed(x)

rand01 = rnorm(1000, mean=0, sd=4)

rand02 = rt(1000, df=3)
```

You may use x as 1, 2 or 3 in this project.

(b) Compare both the data sets in terms of their statistical properties. Your answer must provide some related statistics and diagrams.

Hint: For example, for **S**hape of the data set, statistics such as mean, median, and skewness might be useful. For **O**utlier, histogram and box-whisker plot might be useful. For skewness and kurtosis, you may need to include the *moments* package in your R. Below are some example of statistical measurements:

Center measurement [10]

Methods

1. mean, median, mode.

Spread measurement [10]

Methods

1. range
2. Quartiles (boxplots)
3. Variance
4. Standard deviation

Outlier measurements [5]

1. Boxplot
2. Z-value

Shape measurement [5]

Methods

1. Histogram – check symmetrical/asymmetrical distribution
2. Skewness – compare the mode, median, mean position
3. Use Boxplot to check the skewness.
4. Kurtosis – measure the distribution in terms of height or flatness

D. Instruction to Students

The assignment is due on **7 March 2024, 10:00 AM**.

You are required to submit 2 files:

1. A .txt file that consists of your R commands (just the codes, no output): **ABC1234567_XYZ1234567.txt**
2. Cover Page combines with a .pdf file that consists of your answers to the questions: **ABC1234567_XYZ1234567.pdf**

   Use the XMUM assignment template (word document) made available to you. Start your answer from page 3. You may choose to type your answers. R output must be included. Convert the word document into a pdf file before submitting it on Moodle.