



# ASSIGNMENT 1

---



**UTM**  
UNIVERSITI TEKNOLOGI MALAYSIA

**SCHOOL OF COMPUTING**  
Faculty of Engineering

SCHOOL OF COMPUTING  
FACULTY OF ENGINEERING  
UNIVERSITI TEKNOLOGI MALAYSIA  
81310, UTM JOHOR BAHRU, JOHOR  
+607-5538828 (ACADEMIC OFFICE)  
+60 – 5538822 (FAX)



## PREPARED BY:

- |    |                                |           |
|----|--------------------------------|-----------|
| 1. | GUI YU XUAN                    | A20EC0039 |
| 2. | NURZARIFAH BINTI AZIZAN        | A20EC0127 |
| 3. | ROHAIZAAZIRA BINTI MOHD ZAWAWI | A20EC0138 |

|                     |                                                                |
|---------------------|----------------------------------------------------------------|
| SECTION/COURSE CODE | : [SECI 2143] [01] – [PROBABILITY & STATISTICAL DATA ANALYSIS] |
| LECTURER            | : [DR. SHARIN HAZLIN HUSPI]                                    |
| SESSION/SEMESTER    | : [2020/2021] - [02]                                           |

Case Study Report submitted to School of Computing, Univerisiti Teknologi Malaysia in partial fulfilment to complete the SECI 2143 Assignment for Semester 2 session 2020/2021

## Question 1.

1

1. The population: Football clubs in Malaysia  
The sample: Malaysians Super League Match  
Discrete variable: Total number of goals scored  
continuous variable: Amount of time played before a goal is scored  
Secondary data: Record total numbers of goal scored  
Primary data: old newspaper article, showed that in previous season, the mean number of goals per game is 3.08.
2. The population: second year pupils.  
The sample: second year who had been weighed  
Discrete variable: Gender  
Continuous variable: height  
Primary data: collect data on weight of second year's pupils.  
Secondary data: ~~A summary of data collected is available on request.~~  
A summary data collected on first year pupils.

## Question 2.

1.

| Scales<br>Factor | Frequency |    |    |   | Total |
|------------------|-----------|----|----|---|-------|
|                  | 1         | 2  | 3  | 4 |       |
| Price            | 14        | 6  | 0  | 0 | 20    |
| Condition        | 3         | 6  | 11 | 0 | 20    |
| Fuel Efficiency  | 1         | 11 | 7  | 1 | 20    |
| Car Depreciation | 6         | 11 | 2  | 1 | 20    |

20

10

2.

| Scale | Frequency | Relative frequency | Cumulative frequency | Cumulative Relative frequency |
|-------|-----------|--------------------|----------------------|-------------------------------|
| 1     | 14        | $14/20 = 0.70$     | 14                   | 0.70                          |
| 2     | 6         | $6/20 = 0.30$      | 20                   | 1.00                          |
| 3     | 0         | 0.00               | 20                   | 1.00                          |
| 4     | 0         | 0.00               | 20                   | 1.00                          |
| Total | 20        | 1.0                |                      |                               |

Table of Price Factor

| Scale | Frequency | Relative frequency | Cumulative frequency | Cumulative Relative frequency |
|-------|-----------|--------------------|----------------------|-------------------------------|
| 1     | 3         | $3/20 = 0.15$      | 3                    | 0.15                          |
| 2     | 6         | $6/20 = 0.30$      | 9                    | 0.45                          |
| 3     | 11        | $11/20 = 0.55$     | 20                   | 1.00                          |
| 4     | 0         | 0.00               | 20                   | 1.00                          |
| Total | 20        | 1.0                |                      |                               |

Table of Condition of the car Factor.

10

| Scale | Frequency | Relative frequency     | Cumulative frequency | Cumulative Relative frequency |
|-------|-----------|------------------------|----------------------|-------------------------------|
| 1     | 1         | $\frac{1}{20} = 0.05$  | 1                    | 0.05                          |
| 2     | 11        | $\frac{11}{20} = 0.55$ | 12                   | 0.60                          |
| 3     | 7         | $\frac{7}{20} = 0.35$  | 19                   | 0.95                          |
| 4     | 1         | $\frac{1}{20} = 0.05$  | 20                   | 1.00                          |
| Total | 20        | 1.00                   |                      |                               |

Table for Fuel Efficiency factor

| Scale | frequency | Relative frequency     | Cumulative frequency | Cumulative Relative frequency |
|-------|-----------|------------------------|----------------------|-------------------------------|
| 1     | 6         | $\frac{6}{20} = 0.30$  | 6                    | 0.30                          |
| 2     | 11        | $\frac{11}{20} = 0.55$ | 17                   | 0.85                          |
| 3     | 2         | $\frac{2}{20} = 0.10$  | 19                   | 0.95                          |
| 4     | 1         | $\frac{1}{20} = 0.05$  | 20                   | 1.00                          |
| Total | 20        | 1.00                   |                      |                               |

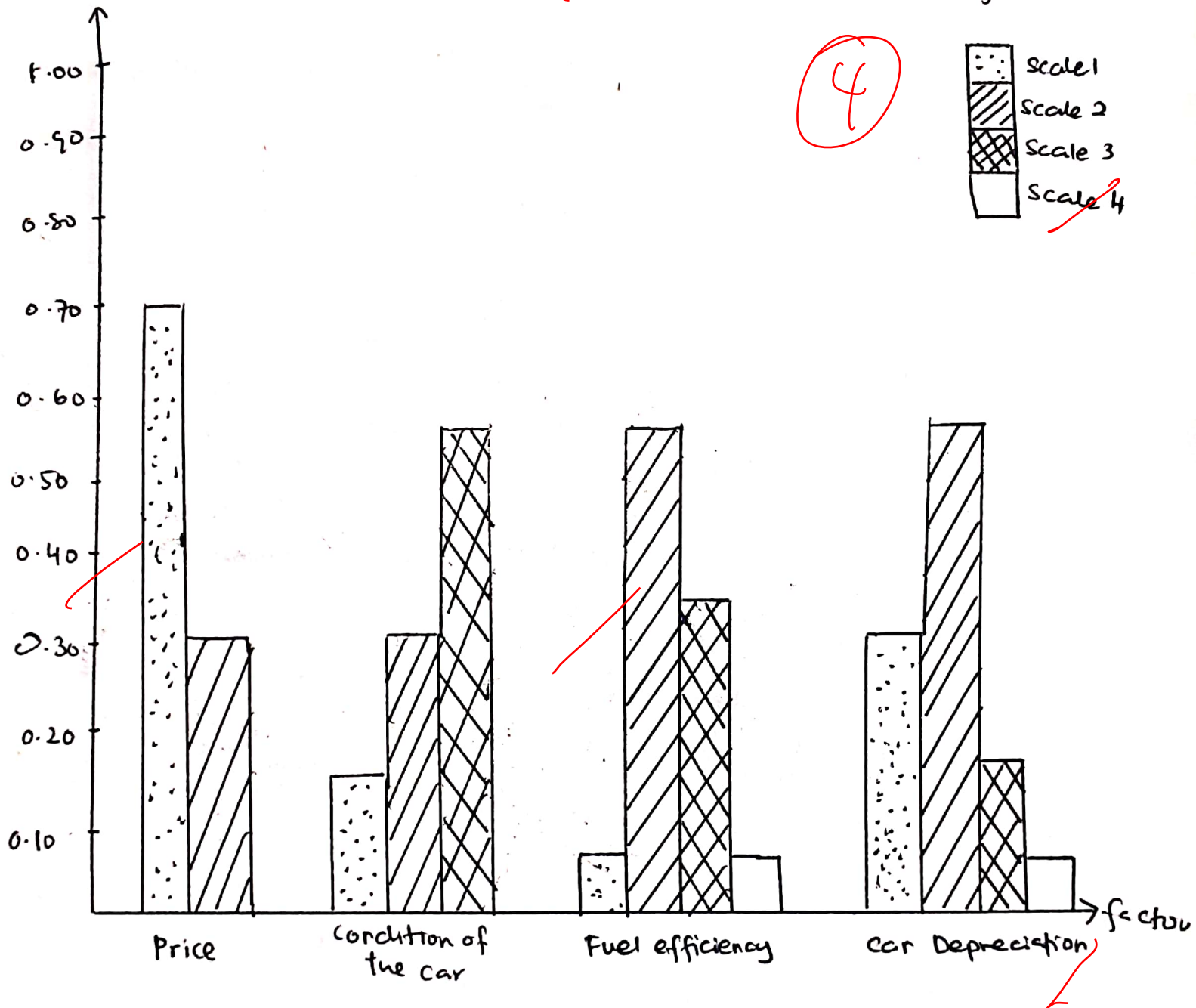
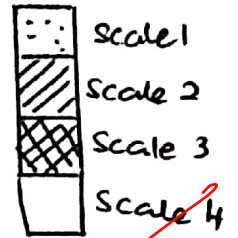
Table for Car Depreciation factor.

# Factor Affect when Buying A car

Relative Frequency

key:

4



### Question 3

4 8 9 16 18 21 23 24 30 32 33 38 42 43 44 55 65 81

1.

| Stem | Leaf    |
|------|---------|
| 0    | 4 8 9   |
| 1    | 6 8     |
| 2    | 1 3 4   |
| 3    | 0 2 3 8 |
| 4    | 2 3 4   |
| 5    | 5       |
| 6    | 5       |
| 7    |         |
| 8    | 1       |

$7\frac{1}{2}$

key = 116 means 16

2. a) mean =  $\frac{\sum x}{n}$

$$= \frac{4+8+9+16+18+21+23+24+30+32+33+38+42+43+44+55+65+81}{18}$$

$$= \frac{586}{18}$$

$$= 32.56$$

b) mode = no mode

c) median,  $i = \frac{n}{2}$

$$= \frac{18}{2}$$

$$= 9$$

$$\frac{Y[9] + Y[10]}{2} = \frac{30 + 32}{2}$$

$$= 31$$

$$3. a) 1st \text{ quartile}, i = \frac{25}{100} \times 18$$

$$= 4.5, k = 5$$

$$Y[5] = 18$$

$$\therefore Q_1 = 18$$

$$2nd \text{ quartile}, i = \frac{50}{100} \times 18$$

$$= 9, k = 9$$

$$Y[9] + Y[10] \div 2 = \frac{30 + 32}{2}$$

$$= 31$$

$$\therefore Q_2 = 31$$

$$3rd \text{ quartile}, i = \frac{75}{100} \times 18$$

$$= 13.5, k = 14$$

$$Y[14] = 43$$

$$\therefore Q_3 = 43$$

$$b) IQR = Q_3 - Q_1$$

$$= 43 - 18$$

$$= 25$$

c) lower limit

$$= Q_1 - 1.5(IQR)$$

$$= 18 - 1.5(25)$$

$$= -19.5$$

upper limit

$$= Q_3 + 1.5(IQR)$$

$$= 43 + 1.5(25)$$

$$= 80.5$$

$\therefore$  The mild outliers is 81

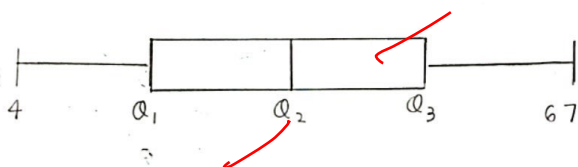
$$\begin{aligned} & \text{lower limit} \\ &= Q_1 - 3(IQR) \\ &= 18 - 3(25) \\ &= -57 \end{aligned}$$

$\therefore$  There is no extreme outliers

2

$$\begin{aligned} & \text{upper limit} \\ &= Q_3 + 3(IQR) \\ &= 43 + 3(25) \\ &= 118 \end{aligned}$$

modified boxplot



+  
81

#### Question 4

5

1. mean =  $\frac{\sum fx}{n}$

$$= \frac{175000(1) + 250000(5) + 500000(4) + 700000(1)}{1 + 5 + 4 + 1}$$

$$= \frac{4125000}{11}$$

$$= \text{RM } 375\,000$$

2. median

| Value per House | Number of Houses | Cumulative frequency |
|-----------------|------------------|----------------------|
| RM 175,000      | 1                | 1                    |
| RM 250,000      | 5                | 6                    |
| RM 500,000      | 4                | 10                   |
| RM 700,000      | 1                | 11                   |

$$i = \frac{11}{2}$$

$$= 5.5, k = 6$$

$$Y_{[6]} = 250\,000$$

$$\text{median} = \text{RM } 250\,000$$

3. Mean is the best to represent the values of 11 houses.  
This is because mean shows the average value of 11 houses.

# Question 5

18

1)

|                       |          |
|-----------------------|----------|
| Age                   | Ratio    |
| Current Smoker        | Nominal  |
| Body Mass Index (BMI) | Interval |
| Hypertension          | Nominal  |

2

2)

| Class Interval         | Category    | Frequency |
|------------------------|-------------|-----------|
| $0 < BMI < 18.5$       | Underweight | 0         |
| $18.5 \leq BMI < 25.0$ | Normal      | 12        |
| $25.0 \leq BMI < 30.0$ | Overweight  | 7         |
| $30.0 \leq BMI$        | Obesity     | 1         |

2

3)

| Class Interval | Class boundaries | Class midpoint | Frequency | Cumulative Frequency |
|----------------|------------------|----------------|-----------|----------------------|
| 60-64          | 59.5 - 64.5      | 62             | 3         | 3                    |
| 65-69          | 64.5 - 69.5      | 67             | 5         | 8                    |
| 70-74          | 69.5 - 74.5      | 72             | 5         | 13                   |
| 75-79          | 74.5 - 79.5      | 77             | 4         | 17                   |
| 80-84          | 79.5 - 84.5      | 82             | 3         | 20                   |
| Total          |                  |                | 20        |                      |

11

For class 60-64,

$$\text{Lower boundary} = 60 - \left(\frac{65-64}{2}\right) = 59.5$$

$$\text{Upper boundary} = 64 + \left(\frac{65-64}{2}\right) = 64.5$$

$$\text{Class midpoint} = \frac{64.5 + 59.5}{2} = 62$$



3

Question 6

5.5

| TIME (MINUTES) | Class Boundaries | Class Midpoint | Frequency | Cumulative Frequency |
|----------------|------------------|----------------|-----------|----------------------|
| 16-30          | 15.5 - 30.5      | 23             | 3         | 3                    |
| 31-45          | 30.5 - 45.5      | 38             | 13        | 16                   |
| 46-60          | 45.5 - 60.5      | 53             | 30        | 46                   |
| 61-75          | 60.5 - 75.5      | 68             | 25        | 71                   |
| 76-90          | 75.5 - 90.5      | 83             | 14        | 85                   |
| 91-105         | 90.5 - 105.5     | 98             | 8         | 93                   |
| 106-120        | 105.5 - 120.5    | 113            | 4         | 97                   |
| 121-135        | 120.5 - 135.5    | 128            | 2         | 99                   |
| 136-150        | 135.5 - 150.5    | 143            | 1         | 100                  |

For class 16-30,

$$\text{lower boundary} = 16 - \left( \frac{31-30}{2} \right) = 15.5$$

$$\text{class midpoint} = \frac{30+16}{2} = 23$$

$$\text{upper boundary} = 30 + \left( \frac{31-30}{2} \right) = 30.5$$

$$i_n = \frac{50}{100} (100) = 50$$

$$\text{median class} = 61-75 \mid 60.5 - 75.5$$

$$L = 60.5 \text{ (lower class limit)}$$

$$N = 100 \text{ (total number of observations)}$$

$$C_{fp} = 46 \text{ (cumulative frequency of the class preceding the median class)}$$

$$f_{med} = 25 \text{ (frequency of median class)}$$

$$w = 15 \text{ (median class size)}$$

$$\text{median} = L + \frac{\frac{N}{2} - C_{fp}}{f_{med}} (w)$$

$$= 60.5 + \frac{\frac{100}{2} - 46}{25} (15)$$

$$= 62.8 \text{ minutes}$$

$$\text{mean} = \frac{\sum fx}{n}$$

$$= \frac{23(3) + 38(13) + 53(30) + 68(25) + 83(14) + 98(8) + 113(4) + 128(2) + 143(1)}{100}$$

$$= \frac{6650}{100}$$

$$= 66.5 \text{ minutes}$$

$$\text{mode class} = 46 - 60 \quad | \quad 45.5 - 60.5$$

$$L = 45.5$$

$$h = 15$$

$$f_0 = 30$$

$$f_1 = 13$$

$$f_2 = 25$$

9.5

$$\begin{aligned} \text{mode} &= L + h \times \left[ \frac{(f_1 - f_0)}{(2f_1 - f_0 - f_2)} \right] \\ &= 45.5 + 15 \times \left[ \frac{(13 - 30)}{(2(13) - 30 - 25)} \right] \\ &= 57.09 \text{ minutes} \end{aligned}$$

2) a)



mode < median < mean

b) The distribution is positively skewed.

c) Most of the flight delay at 57.09 minutes which is less than the average time taken for flight to delay.