



SECI 2143-01
PROBABILITY & STATISTICAL DATA ANALYSIS

ASSIGNMENT 1

LECTURER:

Dr. Sharin Hazlin Huspi

SUBMITTED BY:

Amir Iskandar Bin Norkhairulazaddin (A20EC0011)

Iman Ehsan Bin Hassan (A20EC0048)

Navinthra Rao A/L Venkatakumar (A20EC0104)

Semester 2 2020/2021

Question 1

- 1)
 - a) All of the Malaysian Super League matches
 - b) All of the Malaysian Super League matches in the season.
 - c) total number of goals scored
 - d) Amount of time played before a goal is scored
 - e) Result of the matches
 - f) mean number of goals scored in each match

- 2)
 - a) All students at a school in JayBie
 - b) First year student at a school in JayBie
 - c) Age (in years)
 - d) Height
 - e) Number of pupils weighed
 - f) Place of birth

Question 2

1-

Factors	Scales				Total
	1	2	3	4	
Price	14	6	0	0	20
Condition of the car	3	6	11	0	20
Fuel Efficiency	1	11	7	1	20
Car Depreciation	6	11	2	1	20

2.i) Price

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	14	0.70	14	0.70
2	6	0.30	20	1.00
3	0	0.00	20	1.00
4	0	0.00	20	1.00
Total	20	1.00		

ii) Condition of the car

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	3	0.15	3	0.15
2	6	0.30	9	0.45
3	11	0.55	20	1.00
4	0	0.00	20	1.00
Total	20	1.00		

iii) Fuel Efficiency

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	1	0.05	1	0.05
2	11	0.55	12	0.60
3	7	0.35	19	0.95
4	1	0.05	20	1.00
Total	20	1.00		

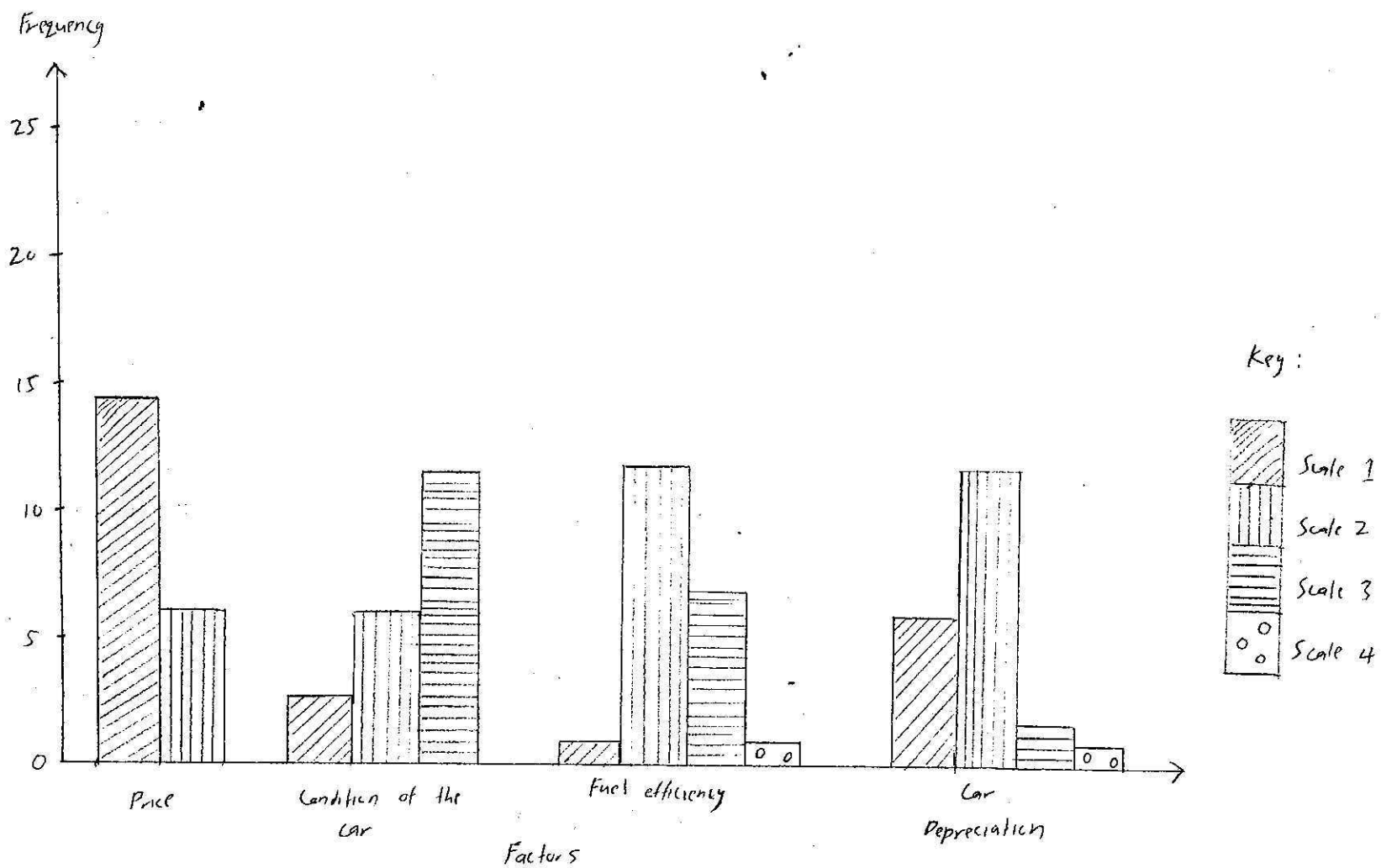
iv) Car Depreciation

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	6	0.30	6	0.30
2	11	0.55	17	0.85
3	2	0.10	19	0.95
4	1	0.05	20	1.00
Total	20	1.00		

Question 2

3)

Important factor when buying a car



Question 3

30 43 32 21 65 8 4 18 16 38 9 44 33 23 24 81 42 58
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
8	1

Key: 8 | 1 means 81

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

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

$$= 32.56$$

b) Mode = no mode in this data

c) Median = $\frac{19}{2}$

$$= 9^{th} \text{ and } 10^{th} \text{ data}$$

$$= \frac{30+32}{2} = 31$$

3) a) 1st Quartile = $0.25 \times 18 = 4.5 \sim 5^{th} \text{ data}$

$$= \frac{4^{th} + 5^{th}}{2} = 18$$

$$= \frac{16+18}{2}$$

$$= 17$$

3rd Quartile = $0.75 \times 18 = 13.5 \sim 14^{th} \text{ data}$

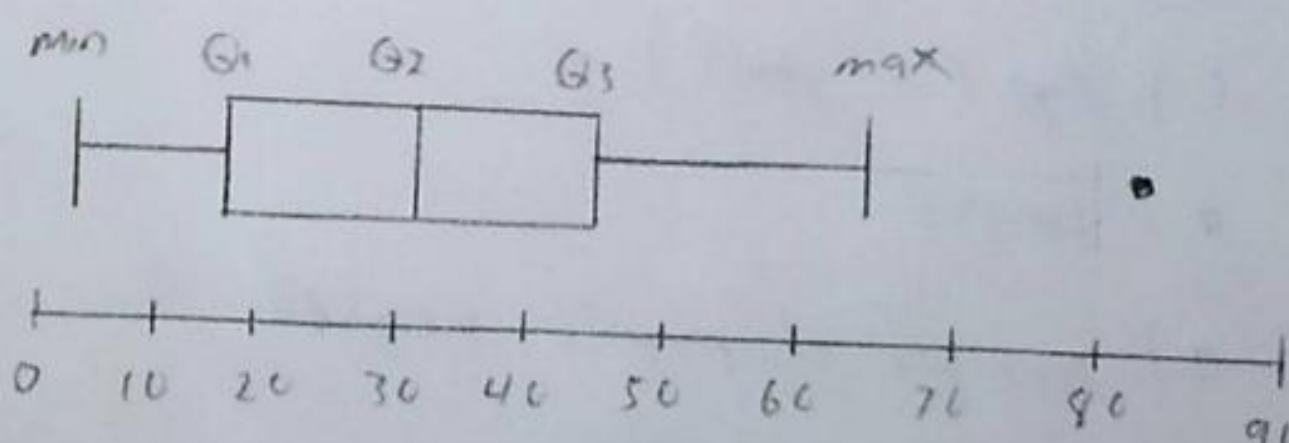
$$= 43$$

b) ~~IGR = $Q_3 - 1.5 IQR$~~

$$IQR = Q_3 - Q_1$$

$$= 43 - 18$$

$$= 25$$



$$\text{Median} = \frac{19}{2}$$

$$= 9.5$$

$$= \frac{9^{th} + 10^{th}}{2}$$

$$= \frac{30+32}{2}$$

$$= 31$$

c) Lower Limit = $Q_1 - 1.5 IQR$

$$= 18 - 1.5(25)$$

$$= -\frac{39}{2}$$

$$= -19.5$$

Upper Limit = $Q_3 + 1.5 IQR$

$$\text{min} = 4$$

$$= 43 + 1.5(25) \text{ max} = 65$$

$$= 80.5$$

$\therefore 81$ is a mild outlier

Q4 1. Mean = $\frac{175000 + 5(250000) + 4(500000) + 700000}{11}$

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

$$= \text{RM } 375000$$

2. Median = $11 \div 2$

$$= 5.5$$

$$= 6^{\text{th}} = \text{RM } 250000$$

3) The mean is the measure of central tendency that is best to represent the value of 11 houses because it gives the average value close or same as the mode which is RM250 000. Thus, mean is the measure of central tendency that is best to measure the value of 11 houses.

Question 5

Question 5

1.

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

2.

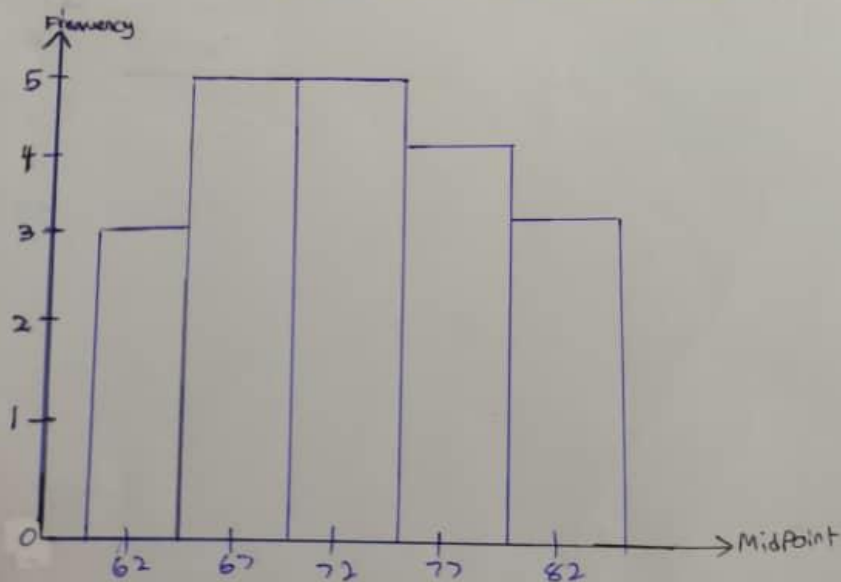
Class Interval	Category	Frequency
$0 < BMI \leq 18.5$	Underweight	0
$18.5 < BMI \leq 25.0$	Normal	12
$25.0 < BMI \leq 30.0$	Overweight	7
$30.0 < BMI$	Obesity	1

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	

4. Histogram (Graph)

Graph of Frequency against MidPoint



Question 6

1)

Midpoint	Cumulative frequency
23	3
38	16
53	46
68	71
83	85
98	93
113	97
128	99
143	100

Median class = 61-75

$$\begin{aligned}
 \text{Median} &= l + \left(\frac{\frac{n}{2} - cf}{f} \right) h \\
 &= 60.5 + \left(\frac{\frac{100}{2} - 46}{25} \right) 15 \\
 &= 62.9
 \end{aligned}$$

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

$$\begin{aligned}
 &= \frac{6650}{100} \\
 &= 66.5
 \end{aligned}$$

$$\begin{aligned}
 \text{Mode} &= l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) h \\
 &= 45.5 + \left(\frac{30 - 13}{2(30) - 13 - 25} \right) 15 \\
 &= 57.09
 \end{aligned}$$

2) b) The distribution is skewed to the right and the kurtosis is leptokurtic because the distribution is peaked.

c) That means all of the data falls to the right side so the mean is greater than the median.

Question 6

2) a)

