UTM UNIVERSITI TEKNOLOGI MALAYSIA	DEPARTMENT OF APPLIED COMPUTING		
	SUBJECT: PROBABILITY & STATISTICAL DATA ANALYSIS		
	ASSESSMENT: ASSIGNMENT 2	**CODE:** SECI 2143	**WEEK:** 7

Name : Soh Zen Ren
Wong Hui Shi
Teoh Wei Jian

Matric ID / IC No : A20EC0152
A20EC0169
A20EC0229

Question 1, 2, 3	Wong Hui Shi
Question 4 ,5	Soh Zen Ren
Question 6,7	Teoh Wei Jian

Assignment 2 : Discrete & Continuous Probability Distribution

ANSWERS SHEET

Question 1

1.(a) Population = Football clubs in Malaysiana

(b) Sample = Malaysiana Super League matches

(c) Discrete variable = Total number of goals scored

(d) Continuous variable = Amount of time played before a goal is scored.

(e) Primary Data = Number of goal scored in each match in the season.

(f) Secondary data = Mean number of goals per game in the previous season.

2.(a) Population = All students at a school in Jaybie

(b) Sample = Second year students

(c) Discrete variable = place of birth

(d) Continuous variable = height

(e) Primary data = weight of second year pupils

(f) Secondary data = Data on first year pupils

Question 2

1.

Scales / Factors	Frequency				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. Price :

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	14	0.7	14	0.7
2	6	0.3	20	1.0
3	0	0	20	1.0
4	0	0	20	1.0
Total	20	1.0		

Condition of the car :

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	3	0.15	3	0.15
2	6	0.3	9	0.45
3	11	0.55	20	1.0
4	0	0	20	1.0
Total	20	1.0		

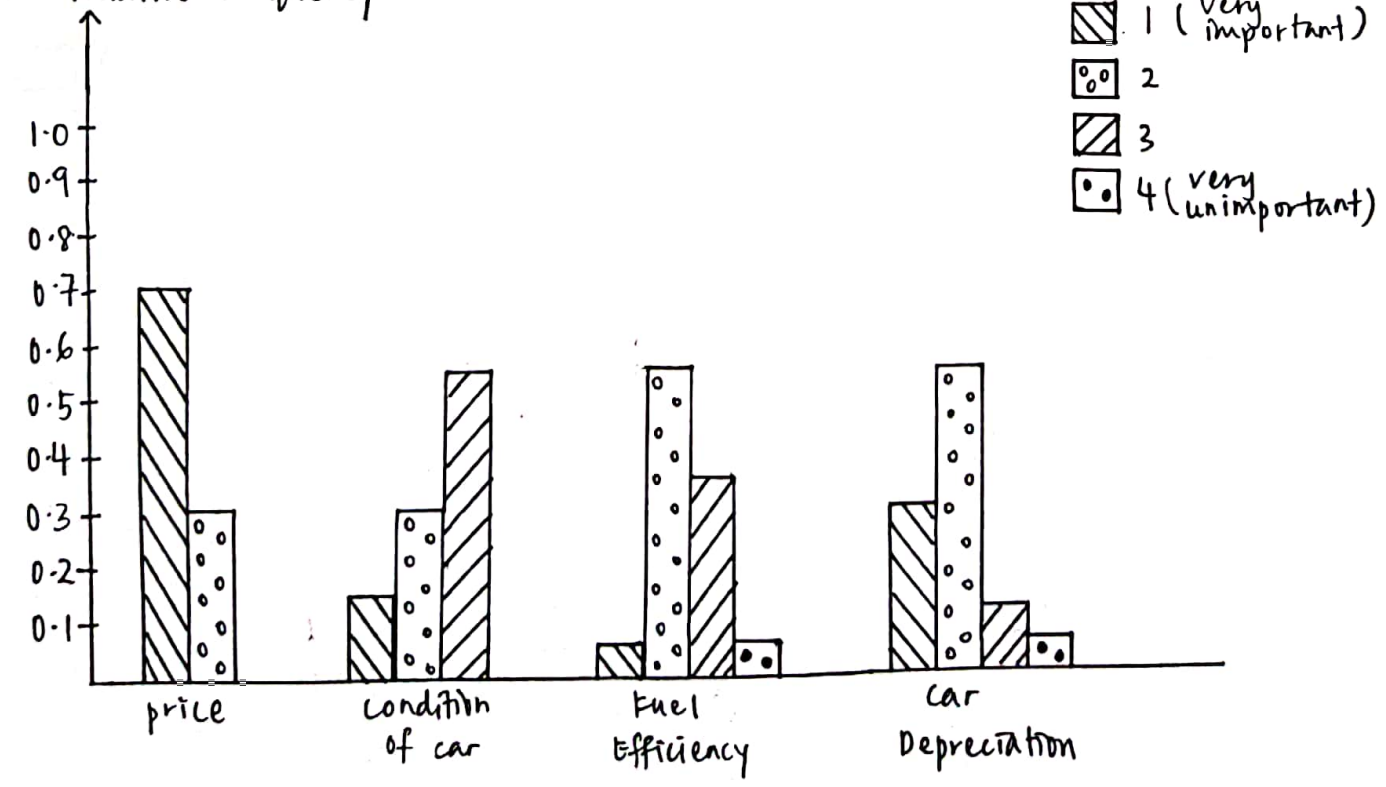
Fuel Efficiency :

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	1·0
Total	20	1·0		

Car Depreciation :

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	6	0·3	6	0·3
2	11	0·55	17	0·85
3	2	0·1	19	0·95
4	1	0·05	20	1·0
Total	20	1·0		

3. Relative Frequency



Question 3

1. Stem-and-Leaf plot

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: 0|4 = 4

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

$= 32.5$

(b) Mode ⇒ no mode, since all data have only 1 frequency.

(c) Median, $\frac{18}{2} = 9$

∴ Since it is an integer, we need to divide value number 9th and 10th with 2.

value number 9th ⇒ 30

value number 10th ⇒ 32

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

3.

(a) 1st, 2nd, 3rd quartile

1st quartile, $\frac{1}{4} \times 18 = 4.5 \approx 5$, ∴ data 5th is 18 #

2nd quartile, $\frac{1}{2} \times 18 = 9$ $9^{th} = 30$, $10^{th} = 32$, $\frac{30+32}{2} = 31$ #

3rd quartile, $\frac{3}{4} \times 18 = 13.5 \approx 14$, ∴ data 14th is 43 #

(b) IQR = 3rd quartile − 1st quartile = 43 − 18

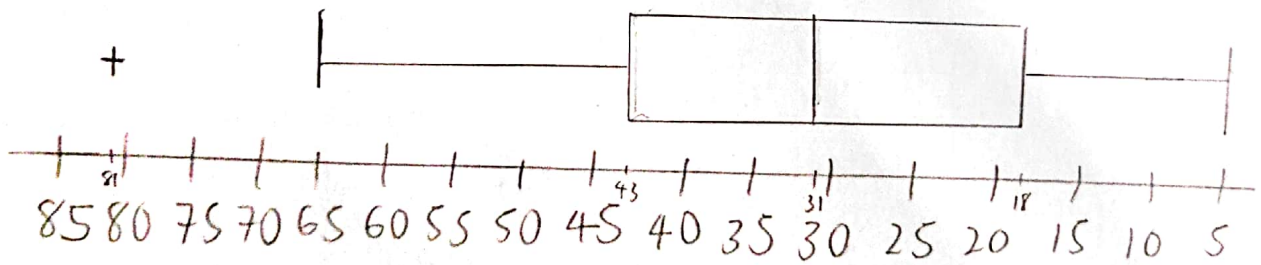
= 25 #

(c) outliers, $Q_3 + 1.5 \times 25$

$= 43 + 1.5 \times 25 = 80.5$

$Q_1 - 1.5 \times 25$

$= 18 - 1.5 \times 25 = -19.5$



Question 4

1. Mean, $\overline{x}$

$$\overline{x} = \frac{(175\,000)1 + (250\,000)5 + (500\,000)4 + (700\,000)1}{11}$$

$$= 375{,}000$$

2. Median (RM175,000 ×1, RM250 000 ×5, RM500 000 ×4, RM700 000 ×1)

$\frac{11}{2} = 5.5 \approx 6$ ①

∴ The house at number 6th in ascending order is the house that cost 250,000.

3. Mean is the "best" represents the values of 11 houses because mean calculates the average of 11 houses but median just only choose the middle value among the 11 houses. Besides, there is no outlier inside the data so it is possible to include all the values.

outlier limit, IQR (500000 − 250000 = 250000)

$Q_1 - 1.5 \times 250{,}000$

$= 250{,}000 - 1.5 \times 250000 = -125\,000$

$Q_3 + 1.5 \times 250{,}000$

$= 500{,}000 + 1.5 \times 250000 = 875{,}000$

Question 5

1) Age = Ratio
Current Smoker = Nominal
Body Mass Index = Ratio
Hypertension = Nominal

2)

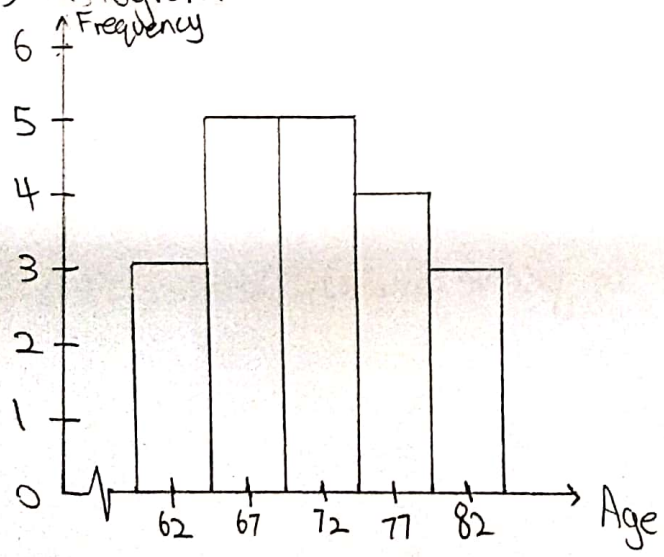
CLASS INTERVAL	CATEGORY	FREQUENCY
0 < BMI <18.5	Underweight	0
18.5<= BMI <25.0	Normal	12
25.0<= BMI <30.0	Overweight	7
30.0<= BMI	Obesity	1

3)

CLASS INTERVAL	CLASS BOUNDARIES	CLASS MIDPOINT	FREQUENCY	C.F.
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		360	20	61

C.F. = Cumulative Frequency

4) Histogram



Question 6

1)

TIME MINUTES	MIDPOINT	FREQUENCIES	C.F.
16-30	23	3	3
31-45	38	13	16
46-60	53	30	46
61-75	68	25	71
76-90	83	14	85
91-105	98	8	93
106-120	113	4	97
121-135	128	2	99
136-150	143	1	100

C.F. = Cumulative Frequency

$100 \div 2 = 50^{th}$

$$\text{median} = L + \frac{\frac{N}{2} - cf_p}{f_{med}}(W)$$

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

$$= 62.9$$

$$\text{mode} = l + h \times \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right]$$

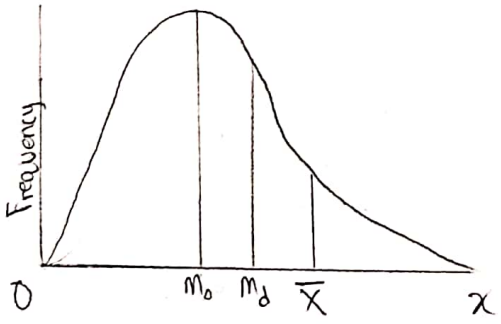
$$= 45.5 + 15 \times \left[\frac{30 - 13}{2(30) - 13 - 25}\right]$$

$$= 57.09$$

$$\text{mean} = \frac{(23\times3)+(38\times13)+(53\times30)+(68\times25)+(83\times14)+(98\times8)+(113\times4)+(128\times2)+(143\times1)}{100}$$

$$= 66.5$$

2a)



b) The distribution is asymmetrical and points in the positive direction. Skewness > 0

c) Mean > Median > Mode