Assignment 1

SECI2143 – 08 PROBABILITY & STATISTICAL DATA ANALYSIS

LECTURER: Dr Sharin Hazlin Binti Huspi

Name & Matrics No:

ANG WAN XIN	A20EC0015
LOW JIA YING	A20EC0070
NUR DINIE SAJEEDA BINTI AZMAN	A20EC0112

Question 1

1. (a) Population = football clubs in Malaysiana
   (b) Sample = All the Malaysiana Super League matches in the season.
   (c) Discrete variable = the total number of goals scored
   (d) Continuous variable = the amount of time played before a goal is scored.
   (e) Primary data = the total number of goals scored for each match in the season.
   (f) Seconday data = the mean number of goals scored in each match in the previous season.

2. (a) Population = All students at a school in JayBie.
   (b) Sample = second year pupils on lunch time
   (c) Discrete variable = age
   (d) Continuous variable = height
   (e) Primary data = weight of second year pupils
   (f) Secondary data = data of pupils collected during their first year

Question 2

1.

Scales	Frequency				Total
Factors	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	0.70	14	0.70
2	6	0.30	20	1.00
3	0	0	20	1.00
4	0	0	20	1.00
Total	20	1.00		

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	20	1.00
Total	20	1.00		

~~Fuel Efficiency~~ 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		

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		

Relative Frequency
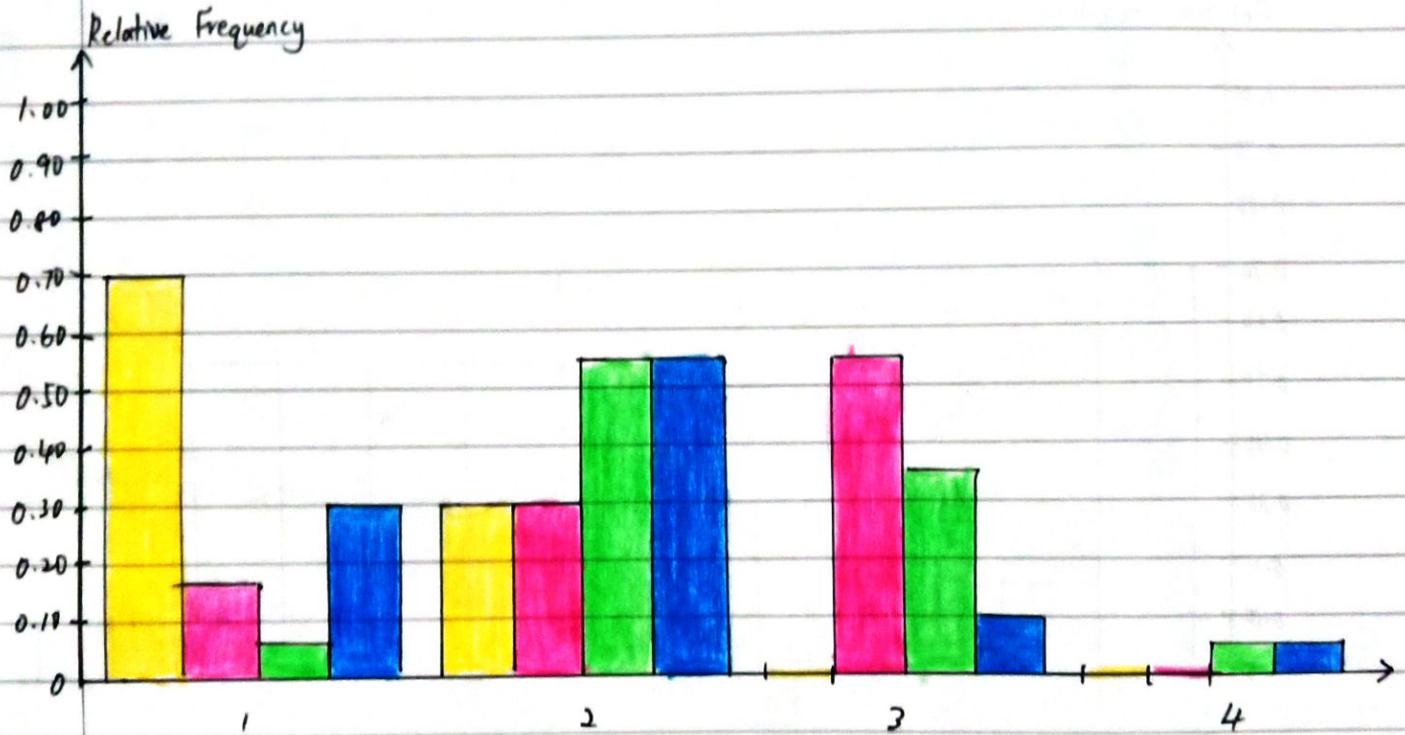
Factors:

Price

Condition of the car

Fuel efficiency

Car Depreciation
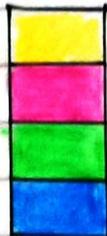
Question 3

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: $8 \mid 1$ represent 81 days

2) a) mean = $4+8+9+16+18+21+23+24+30+32+32+33+42$
$+42+43+44+55+65+81$ over 18

$= 32.56$

b) mode = no mode

c) median = $\dfrac{30+32}{2} = 31$

3) box plot.

a) $Q_1 = \dfrac{25}{100}(18) = 4.5 \approx 5 \rightarrow 18$

$Q_2 = $ median $= 31$

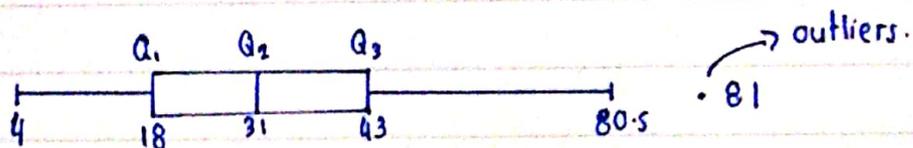$Q_3 = \dfrac{75}{100}(18) = 13.5 \approx 14 \rightarrow 43$

b) $IQR = Q_3 - Q_1$
$= 43 - 18$
$= 25$

c) outliers $= \left[Q_1 - 1.5\,IQR \, , \, Q_3 + 1.5\,IQR \right]$
$= \left[18 - 1.5(25) \, , \, 43 + 1.5(IQR) \right]$
$= \left[-19.5 \, , \, 80.5 \right]$

$Q_1 \quad Q_2 \quad Q_3$ — outliers.
$4 \quad 18 \quad 31 \quad 43 \quad 80.5 \quad \cdot 81$

Question 4

1) $\bar{x} = \dfrac{1(175000) + 5(250000) + 4(500000) + 1(700000)}{11}$

    = RM 375 000

2) median = $\dfrac{11+1}{2}$ = 6 → RM 250 000

3) mean - because mean shows the average values for 11 houses.

Question 5

1.

Age	ratio
Current smoker	nominal
Body mass Index (BMI)	interval
Hypertension	nominal

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

3.

Class Interval	Class Boundaries	Class Midpoint	Frequency	Cumulative frequency
60-64	59.5 - 64.5	$\frac{64-60}{2} = 62$	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.
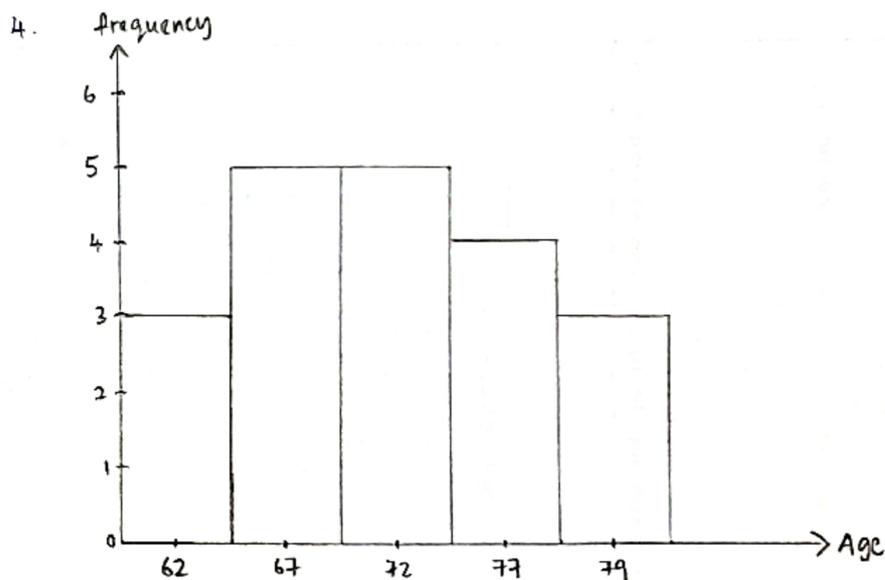
Question 6

1.

Time (minutes)	Frequencies	Midpoint	Cumulative frequency
16 - 30	3	$\frac{30+16}{2} = 23$	3
31 - 45	13	38	16
46 - 60	30	53	46
61 - 75	25	68	71
76 - 90	14	83	85
91 - 105	8	98	93
106 - 120	4	103	97
121 - 135	2	108	99
136 - 150	1	143	100
Total	100	—	.

(a) median $= L + \frac{\frac{N}{2} - CF_p}{f_{med}}(W)$ $\frac{N}{2} = \frac{100}{2} = 50$

$$= 60.5 + \frac{50 - 46}{25}(30 - 16)$$

$$= 62.74 \text{ minutes}$$

(b) mean $= \frac{\Sigma fx}{\Sigma x}$

$$= \frac{3(23) + 13(38) + 30(53) + 25(68) + 14(83) + 8(98) + 4(103) + 2(128) + 1(143)}{100}$$

$$= \frac{6610}{100}$$

$$= 66.1 \text{ minutes}$$

(c) mode $= l + h\left(\frac{f_1 - f_0}{2(f_1) - f_0 - f_2}\right)$

$$= 45.5 + 14\left(\frac{30 - 13}{2(30) - 13 - 25}\right)$$

$$\approx 56.32 \text{ minutes}$$

2. (a)
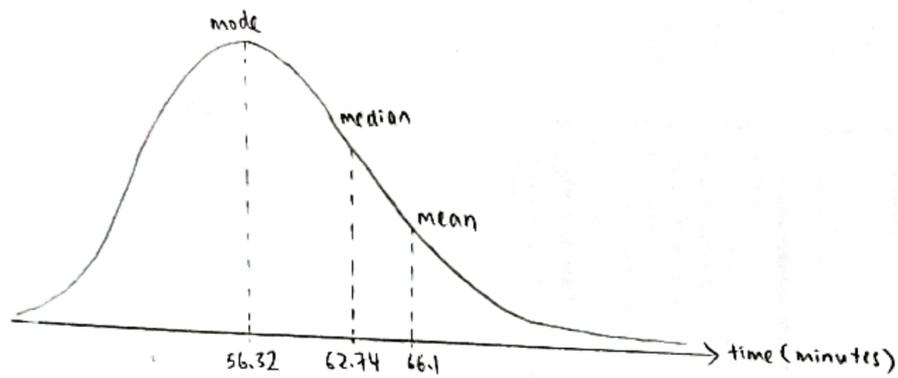
(b) - shape of distribution: positively skewed and is leptokurtic.
   - positively skewed because mode < median < mean.
   - leptokurtic because k > 0.

(c) - most of the flight delayed about 46 to 60 minutes. later than its
     scheduled time.

   - the mean of time delayed for all flights is about 66.1 minutes.

   - most of the flight delayed later than its scheduled time.