COURSE:

SECI2143-02 PROBABILITY & STATISTICAL DATA ANALYSIS

FACULTY:

FACULTY OF ENGINEERING

SCHOOL:

SCHOOL OF COMPUTING

TITLE:

ASSIGNMENT 1

LECTURER'S NAME:

DR. CHAN WENG HOWE

NAME	MATRIC NO.
NURARISSA DAYANA BINTI MOHD SUKRI	A20EC0120
SAKINAH AL'IZZAH BINTI MOHD ASRI	A20EC0142
QAISARA BINTI ROHZAN	A20EC0133

Question 1

1. (a) the population : All the Malaysian a Super League matches in the season.
   (b) the sample : each match total number of goals scored and the amount of time played before a goal is scored.
   (c) a discrete variable : number of goals scored
   (d) a continuous variable : the amount of time played
   (e) primary data : First Saturday of the season.
   (f) secondary data : old newspapers articles

2. (a) the population : All second year students at a school in JayBie
   (b) the sample : A high proportion of second year students at a school in JayBie.
   (c) a discrete variable : number of pupils weighed.
   (d) a continuous variable : weight
   (e) primary data : the weight of second year students
   (f) secondary data : the data on first year students.

QUESTION 2

1. Summarize Table 1 in a frequency table as below

Factors \ Scales	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

scale

1 → 4

1 = very important

4 = very unimportant

2. Summarize all the factors (price, condition of the car, fuel efficiency and car depreciation) from Table 1 in a frequency distribution. For each of the factor, use the table as below:

Factor 1: Price

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	14	0.70	14	0.7
2	6	0.30	20	1.0
3	0	0.00	20	1.0
4	0	0.00	20	1.0
Total	20	1.00		

Factor 2: 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		

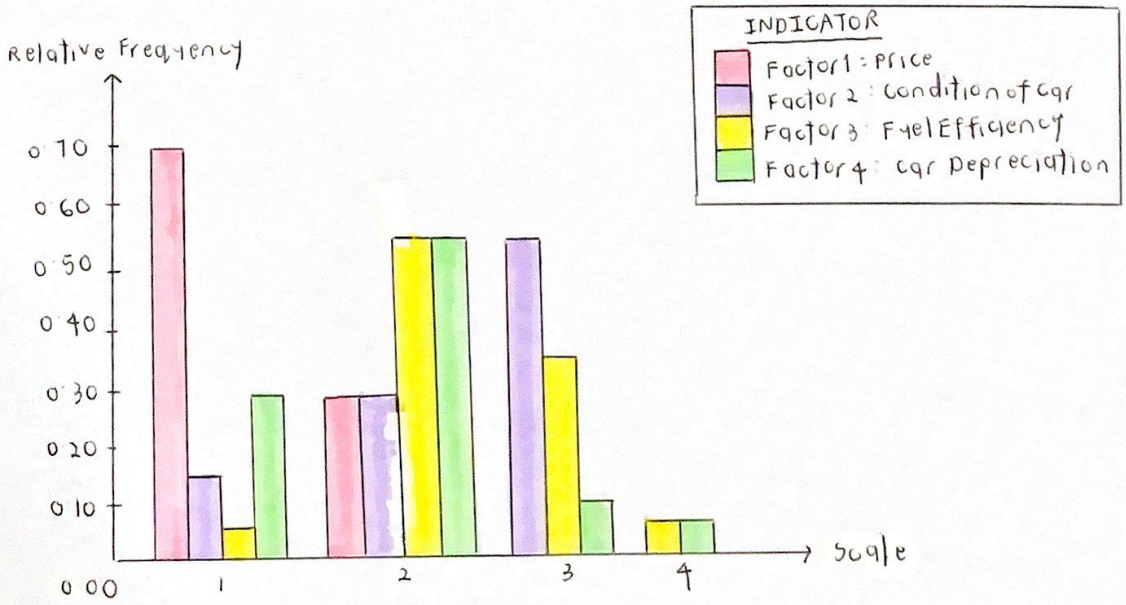
Factor 3: Fuel Efficiency

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		

Factor 4: 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		

3. Draw the results from (b) using a comparative bar chart



QUESTION 3

Data of number of days taken by 18 to finish developing a mobile app

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

1. Represent the data in a 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) Mean

$$\bar{x} = \frac{\Sigma x}{n}$$

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

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

$$\bar{x} = 32.56$$

b) Mode

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

Since each value occurs only once in the data set, there is no mode for this set of data.

c) Median

(30 + 32) ÷ 2 = 31

∴ Median = 31

3. a) $Q1 = 25\%$

$N = 18$

$P = 25$

$i = \frac{25}{100}(18)$

$= 4.5$

$i \approx$ 5th number

$\therefore Q1 = 18$

$Q2 = 50\%$

$P = 50$

$i = \frac{50}{100} \times (18)$

$i = 9$

$Q2 = (Y[9] + Y[10]) \div 2$

$= (30 + 32) \div 2$

$\therefore Q2 = 31$

$Q3 = 75\%$

$P = 75$

$i = \frac{75}{100} \times 18$

$i = 13.5$

$i \approx$ 14th number

$\therefore Q3 = 43$

b) Interquartile Range (IQR)

$IQR = Q3 - Q1$

$= (43) - (18)$

$IQR = 25$

c) Mild and Extreme outliers

$mild = 1.5(IQR)$

$= 1.5 \times 25$

$= 37.5$

$extreme = 3(IQR)$

$= 3 \times 25$

$= 75$

$lower\ limit = Q_1 - 1.5 \times IQR$

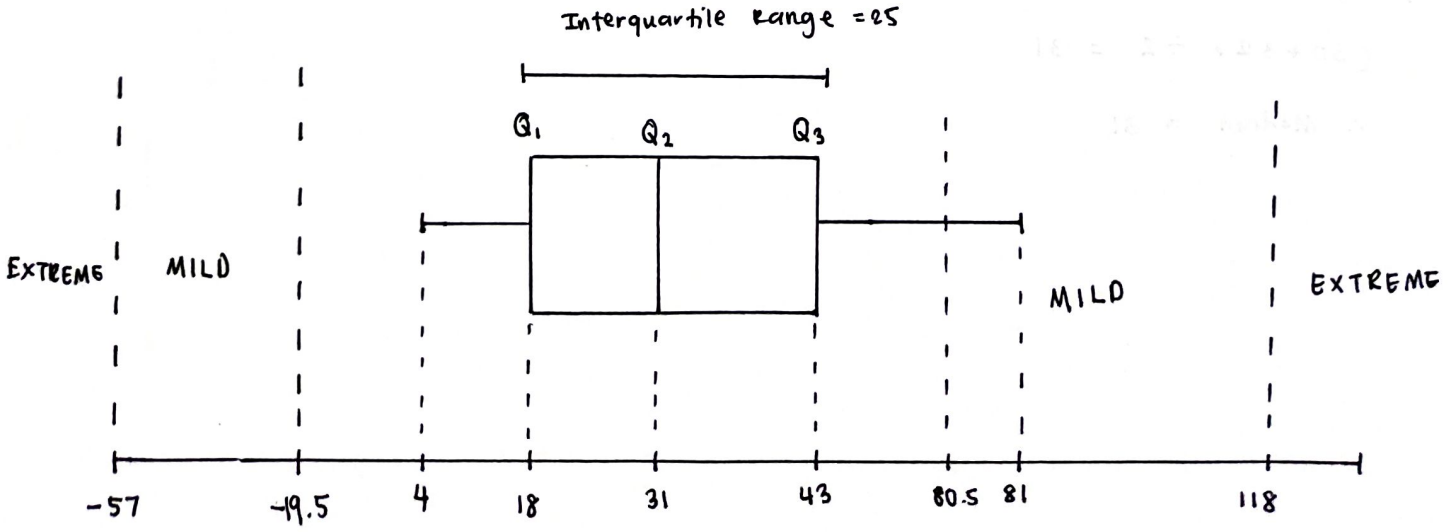
$= 18 - 1.5 \times 25$

$= -19.5$

$Upper\ limit = Q_3 + 1.5 \times IQR$

$= 43 + 1.5 \times 25$

$= 80.5$



Question 1

The value of 11 houses in a new development area in Secydai are shown in Table 2.

Table 2

Value per House	Number of Houses (n)
RM 175,000	1
RM 250,000	5
RM 500,000	4
RM 700,000	1

$n = 1 + 5 + 4 + 1$

$n = 11$ = Total number of houses

Based on the Table:

i) Find the mean value of these houses in RM

$$\sum f_i X_i = 175000(1) + 250000(5) + 500000(4) + 700000(1)$$
$$= 4125000$$

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i X_i}{n} = \frac{4125000}{11} = \text{RM } 375,000$$

ii) Find the median value of these houses in RM

(6th) median

Data Arranged in order: 175000, 250000, 250000, 250000, 250000, 250000, 500000, 500000, 500000, 500000, 700000

There are 11 numbers in the order:
So the number in the middle would be the median $(11+1) \div 2 = 12 \div 2 = 6$th number

So the median is RM 250 000

iii) State which measures the central tendency, the mean or the median, "best" represents the values of the 11 houses. Justify your answer.

The mean because it measures all values in a data set and works as an average

Question 5.

1. Age – ratio
   current smoker – Nominal
   Body Mass Index (BMI) – interval
   Hypertension – Nominal

2.

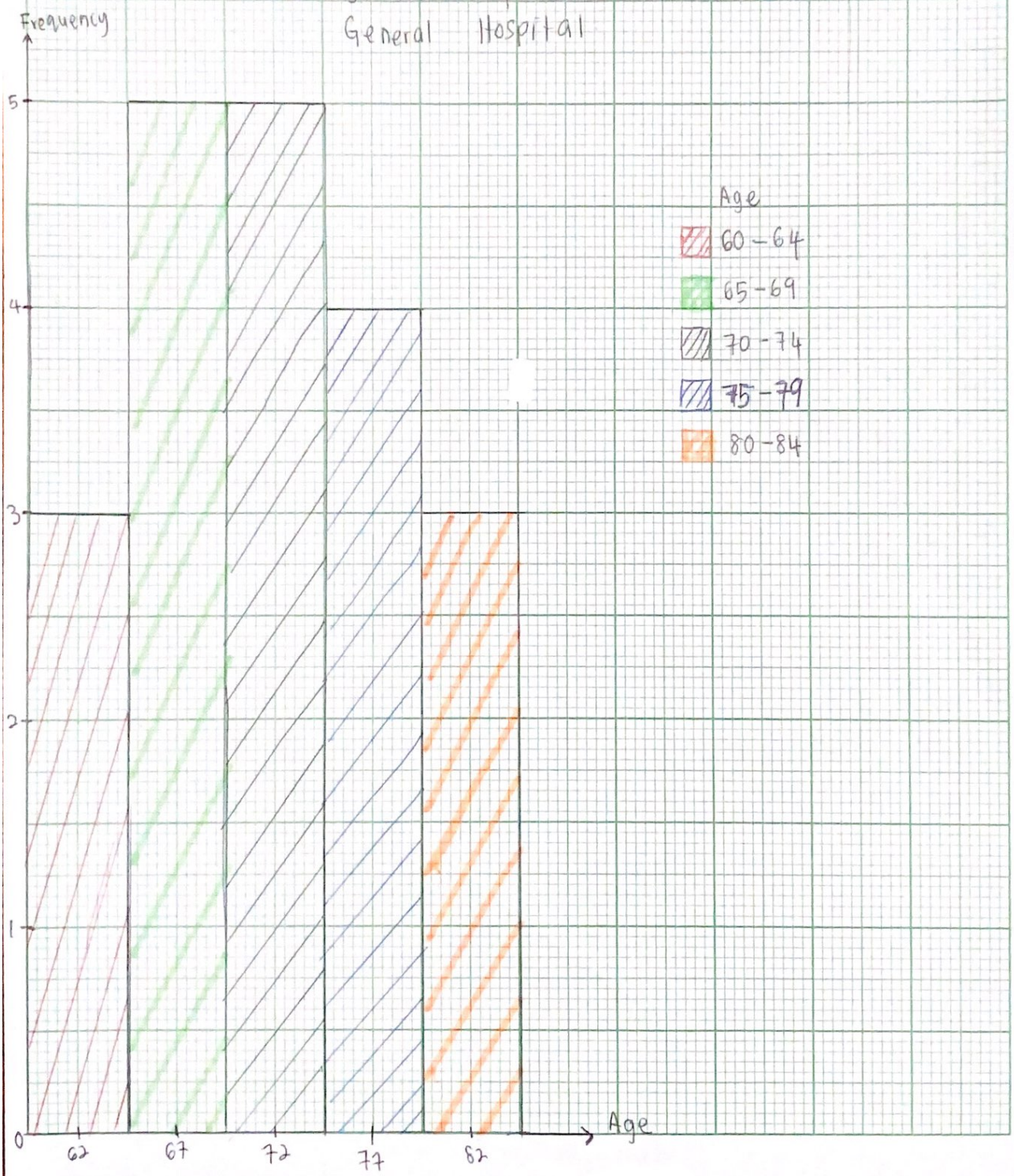
class interval	category	Frequency
0<BMI <18.5	underweight	0
18.5 <= BMI <25.0	Normal	12
25.0 <= BMI <30.0	overweight	7
30.0 <= BMI	Obesity	1

3.

class interval	class Boundaries	class midpoint	Frequency	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	

Nama: ..... Tarikh: .....

Age of 20 patients of the General Hospital



QUESTION 6

1. Find Median, Mean, Mode.

Time cm)	class Boundaries	Midpoint, X	frequency, f	fX	cumulative frequency, cf
16 - 30	15.5 - 30.5	23	3	69	3
31 - 45	30.5 - 45.5	38	13	494	16
46 - 60	45.5 - 60.5	53	30	1590	46
61 - 75	60.5 - 75.5	68	25	1700	71
76 - 90	75.5 - 90.5	83	14	1162	85
91 - 105	90.5 - 105.5	98	8	784	93
106 - 120	105.5 - 120.5	113	4	452	97
121 - 135	120.5 - 135.5	128	2	256	99
136 - 150	135.5 - 150.5	143	1	143	100
Total			100	6650	

Mean

$\Sigma fx = 6650$

$n = 100$

$$\bar{x} = \frac{\Sigma fx}{n}$$

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

$$\bar{x} = 66.5$$

Mode

$l = 45.5$

$h = 15$

$f_1 = 30$

$f_0 = 13$

$f_2 = 25$

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

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

$$\text{mode} = 57.09$$

Median

$N/2 = 50$, median class = 61 - 75

$L = 60.5$

$N = 100$

$cf_p = 46$

$f_{med} = 25$

$W = 15$

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

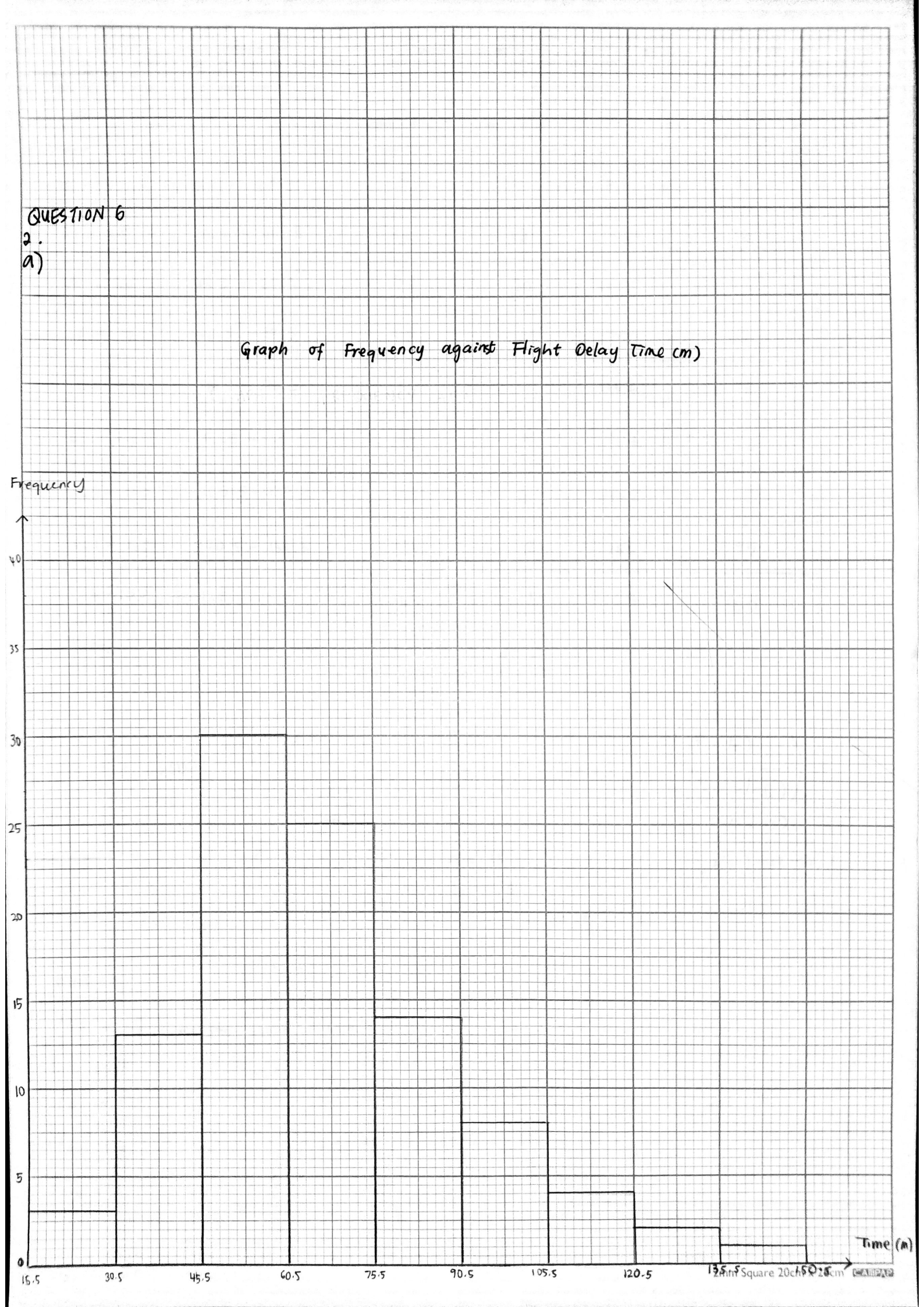
$$= 60.5 + \left(\frac{50 - 46}{25} \right)(15)$$

$$\text{median} = 62.9$$

QUESTION 6

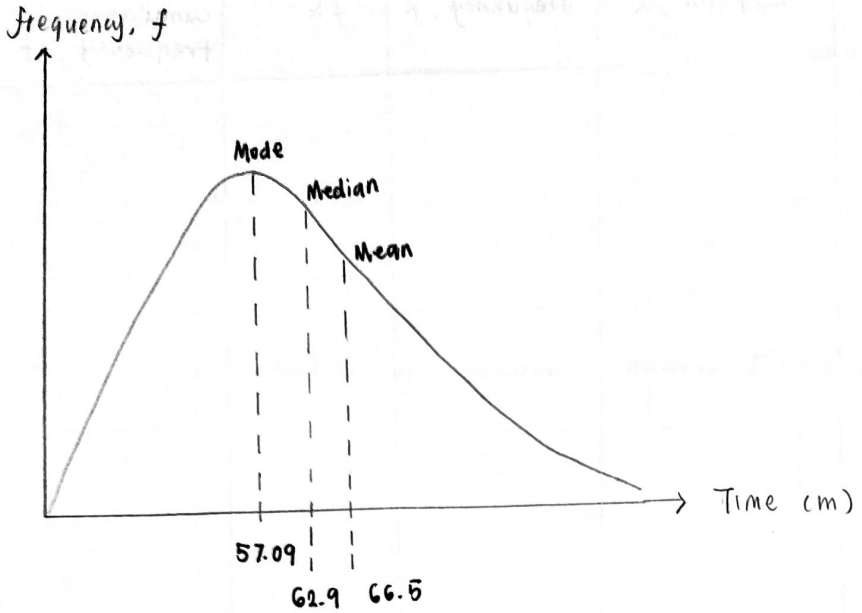
2.

a)



QUESTION 6

2. a. Distribution Graph of frequency against flight Delay Time



b. The shape of the distribution are positive skew / right skewed as the mean has the largest value, while mode is smallest. The shape ~~is~~ also has high kurtosis which is leptokurtic because the peak is higher and sharper.

c. The shape of the graph tells us that the data are moderately skewed and positively skewed. It also shows that the data are heavy-tailed or contains outliers. Outliers ~~stretr~~ stretch the horizontal axis of the graph which makes it appears narrow in vertical, thereby leptokurtic distribution.