



PROBABILITY & STATISTICAL DATA ANALYSIS

SECI2143-02

ASSIGNMENT 1

LECTURER'S NAME: DR. CHAN WENG HOWE

Prepared by: -

ADRINA ASYIQIN BINTI MD ADHA	A20EC0174
MADINA SURAYA BINTI ZHARIN	A20EC0203
NAYLI NABIHAH BINTI JASNI	A20EC0105
NUR SYAMALIA FAIQAH BINTI MOHD KAMAL	A20EC0118

Question 1

- 1
- a) Population of all Malaysia Super League Matches
- b) Sample of the Malaysia Super League Matches
- c) Number of goals scored
- d) The amount of time played before a goal is scored
- e) The record of each matches
- f) Old newspapers articles showed that in the previous season

- 2
- a) Population of students at a school in Jaybre
- b) Sample of second year pupils at the school
- c) gender
- d) weight
- e) The report about the new weight of second year pupils
- f) The report about data of the first year students

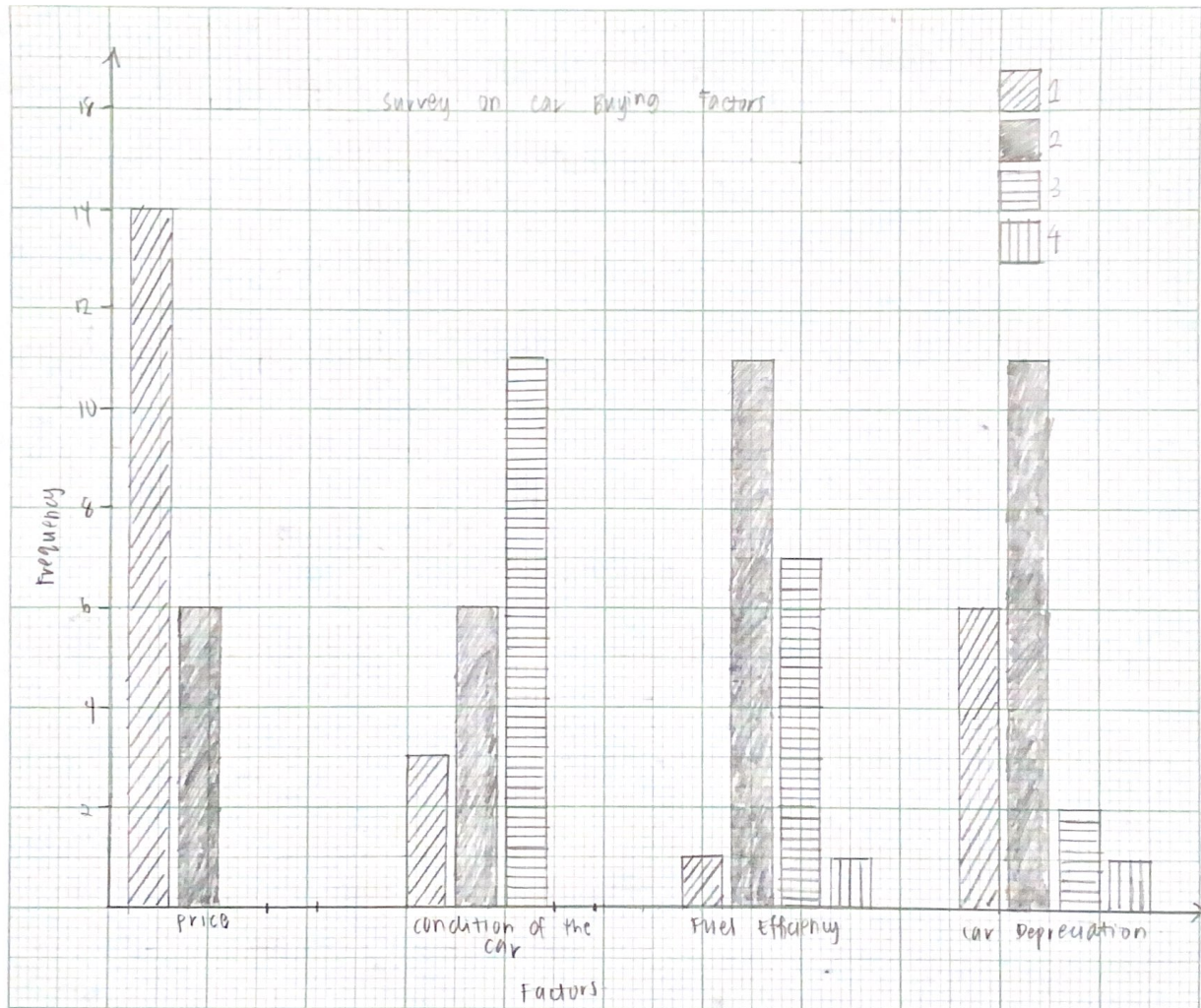
Question 2

1.

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	6 7	1	20
car Depreciation		6	11	2	1	20

2.

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	24	0.3	24	0.3
2	36 34	0.425	58	0.725
3	20	0.25	78	0.975
4	2	0.025	80	1.00
Total	80	1.00		



Question 3

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 $1|6 = 16$

2. a) Mean

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

$$= 32.56$$

b) Mode

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

Based on the arranged data set, no mode is exist in this set of data.

c) Median

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

$$\frac{18}{2} = 9$$

$$\frac{30 + 32}{2} = 31 \text{ (median)}$$

3.

$\begin{array}{ccccccc} & & & Q_1 & & & \\ & & & \downarrow & & & \\ 4, & 8, & 9, & 16, & 18, & 21, & 23, & 24, \\ Q_2 & & & & Q_3 & & & \\ \downarrow & & & & \downarrow & & & \\ 30, & 32, & 33, & 38, & 42, & 43, & 44, \\ & & & & & & \\ 55, & 65, & 81 \end{array}$

$Q_1 (25\%)$

$$18 \times \frac{25}{100} = 4.5$$

$$\approx 5$$

$$Q_1 = 18$$

$Q_2 (50\%)$

$$18 \times \frac{50}{100} = 9$$

$$Q_2 = 30$$

$Q_3 (75\%)$

$$18 \times \frac{75}{100} = 13.5$$

$$\approx 14$$

$$Q_3 = 43$$

$$IQR = Q_3 - Q_1$$

$$= 43 - 18$$

$$= 25$$

Outliers

$$\text{Mild} = 1.5 \times IQR$$

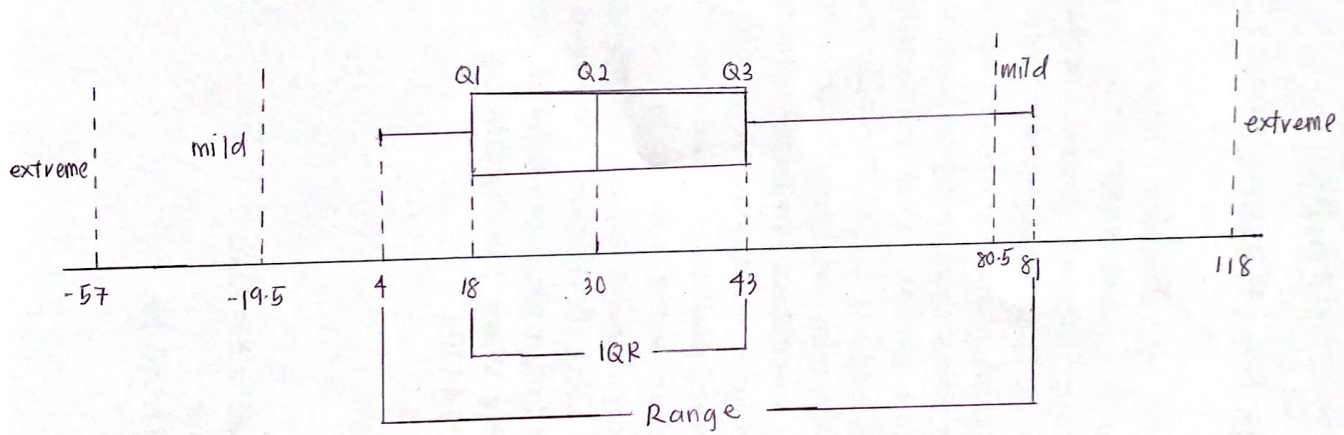
$$= 1.5 \times 25$$

$$= 37.5$$

$$\text{Extreme} = 3 \times IQR$$

$$= 3 \times 25$$

$$= 75$$



Question 4

$$1. \quad \bar{x} = \frac{(1 \times 175k) + (5 \times 250k) + 4(500k) + 1(700k)}{11}$$

$$= \frac{175k + 1250k + 2000k + 700k}{11}$$

$$= \frac{4125k}{11}$$

$$= 375k \approx \text{RM } 375,000.00$$

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

m-c →

	f	< cf
175k	1	1
250k	5	6
500k	4	10
700k	1	11

median class = $\left(\frac{11}{2}\right)$: 5-5th observation.

$$= 250k \approx \text{RM } 250,000.00$$

3. - The median, best represents the value.

- median has better measure of central tendency.

Question 5

1.

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

2.

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

3.

Class Interval	Class Boundary	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)

frequency



7

6

5

4

3

2

1

0

62

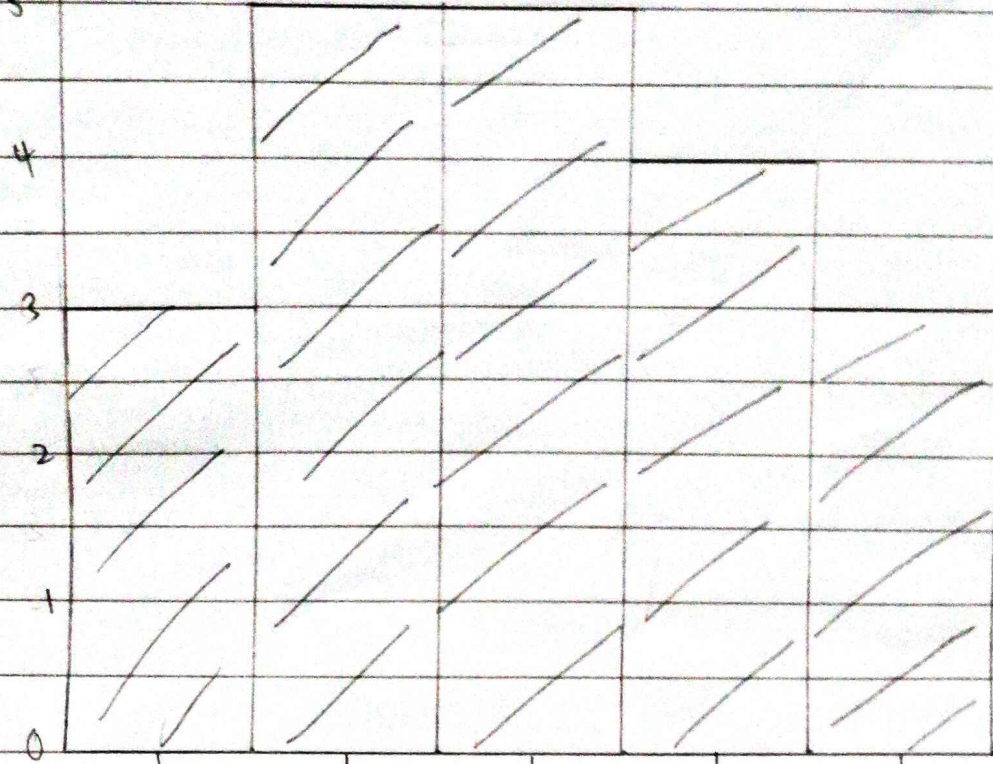
67

72

74

82

> Age



Time (minutes)	Frequencies, f_i	midpoint, X_i	class interval	Cumulative frequency	$f_i X_i$
16-30	3	23	15.5 - 30.5	3	69
31-45	13	38	30.5 - 45.5	16	494
46-60	30	53	45.5 - 60.5	46	1590
61-75	25	68	60.5 - 75.5	71	1700
76-90	14	83	75.5 - 90.5	85	1162
91-105	8	98	90.5 - 105.5	93	784
106-120	4	113	105.5 - 120.5	97	452
121-135	2	128	120.5 - 135.5	99	256
136-150	1	143	135.5 - 150.5	100	143

modal class ←
median class ←

1. a. Median

$$\text{median class} = \frac{N}{2} = \frac{100}{2} = 50$$

$$\text{median} = L + \frac{\frac{N}{2} - cf_p}{f_{\text{med}}} (W) = 60.5 + \frac{50 - 46}{25} (15) = 62.9^*$$

b. Mean

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^n f_i X_i}{n} = \frac{69 + 494 + 1590 + 1700 + 1162 + 784 + 452 + 256 + 143}{100} \\ &= \frac{6650}{100} = 66.5^* \end{aligned}$$

c. Mode

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

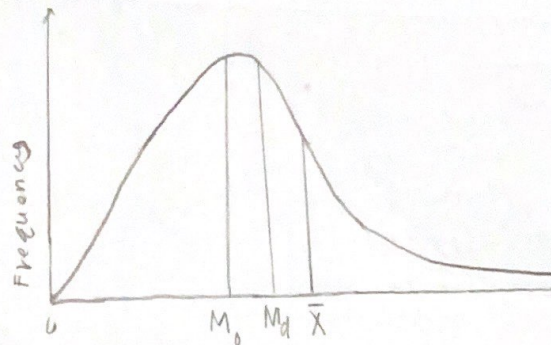
modal class = 46-60

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

$$= 45.5 + 15 \left(\frac{17}{22} \right)$$

$$= 57.091^*$$

2. a. mode < median < mean.



$$\text{Kurtosis} = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{(N-1)s^4}$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

midpoint, x_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^4$
23	1892.25	3580610.0625
38	812.25	659750.0625
53	182.25	33215.0625
68	2.25	5.0625
83	272.25	74120.0625
98	992.25	984560.0625
113	2162.25	4675325.0625
128	3782.25	14305415.0625
143	5852.25	34248830.0625
	$\Sigma 15950.25$	$\Sigma 58561830.56$

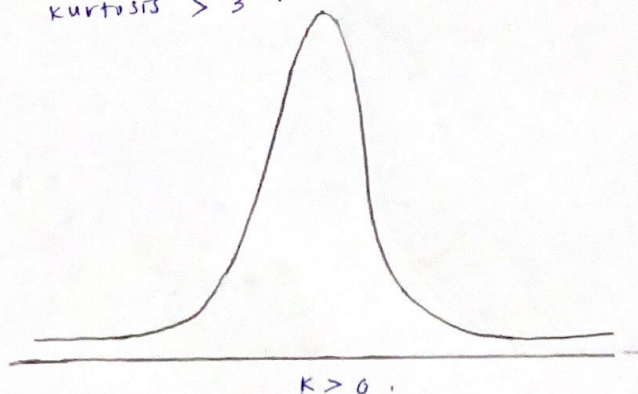
$$\bar{x} = 66.5$$

$$s^2 = \frac{15950.25}{99} = 161.1136$$

$$s = 12.6931$$

$$\text{Kurtosis} = \frac{58561830.56}{(100-1)(12.6931)^4} = 22.7885$$

$$\text{kurtosis} > 3$$



~~b - the distribution is asymmetrical in the positive direction. The shape ^{are} also leptokurtic its tail and are longer and fatter, the central peak is higher and sharper.~~

b.	Skewness	Kurtosis.
	- asymmetrical distribution	- leptokurtic
	- positive skewed.	- longer and fatter tails
	- skewed to the right.	- central peak higher and sharper.

c - Most of the flight delay in a long period of time.