



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

SECI2143-01
PROBABILITY & STATISTICAL DATA
ANALYSIS

ASSIGNMENT 1

NAME : INDIRA A/P THANGARAJ (A20EC0049)
MUHAMMAD AIMAN BIN ABDUL RAZAK
(A20EC0082)
LECTURER : DR. SHARIN HAZLIN BINTI HUSPI

QUESTION 1

Question 1

- 1) (a) Football clubs in Malaysia. → [redacted]
(b) Super League Matches in this season. [redacted]
(c) Total number of goals scored.
(d) Amount of time played before a goal is scored.
(e) Number of goals scored in each match on the first Saturday of the Season.
(f) mean number of goals per game for previous season.
- 2) (a) All the students at a school in JayBie
(b) Second and first year pupil
(c) Gender of the first year pupil
(d) Weight of a pupil
(e) Weight of second year pupil
(f) Weight of first year pupil.

QUESTION 2

Question 2

①

Scale	Frequency				Total
	1	2	3	4	
Factors					
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	0.00	20	1.00
4	0	0.00	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	0.00	20	1.00
Total	20	1.00		

Fuel Efficiency :

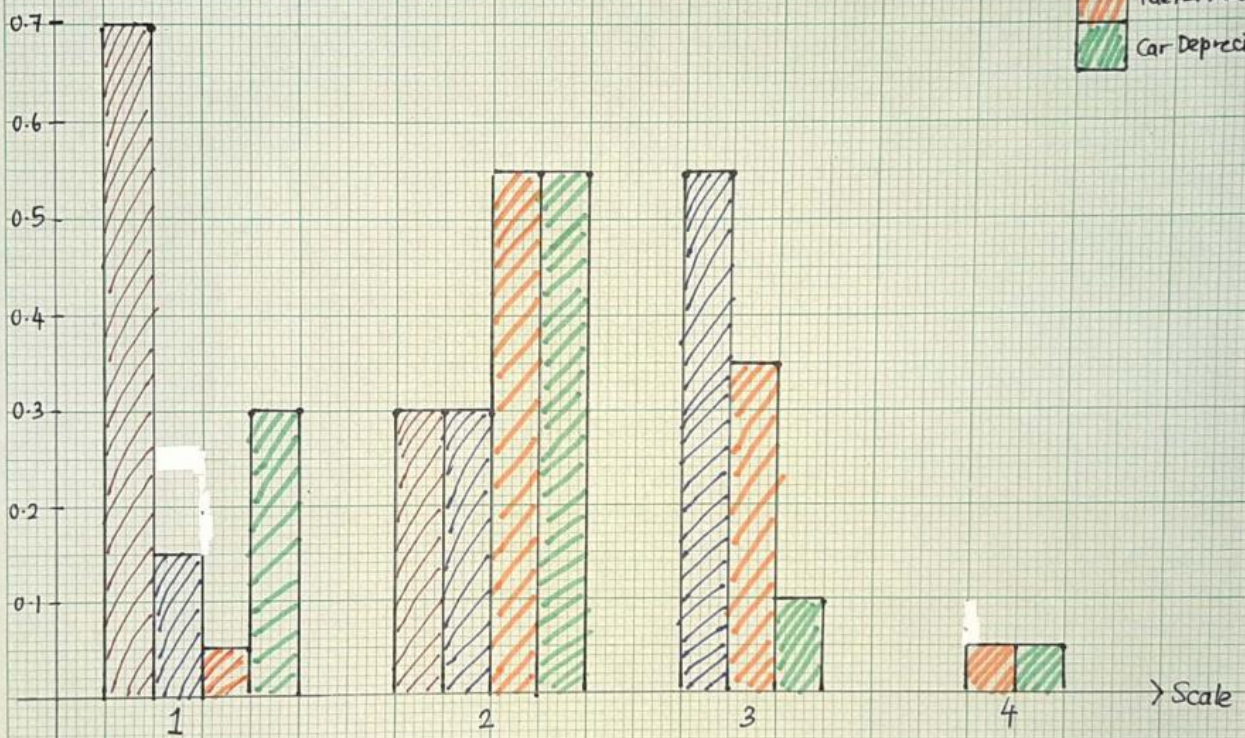
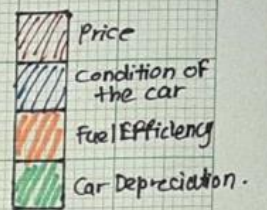
Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	1	0.05	1	0.05
2	11	0.55	12	0.60
3	7	0.35	19	0.95
4	1	0.05	20	1.00
Total	20	1.00		

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		

Important factors when buying
a car.

Relative
Frequency
↑



QUESTION 3

Question 3.

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 : 1 | 4 = 14

2) (a) $\bar{x} = \frac{\sum x}{n}$

$$= \frac{(4+8+9+16+18+21+23+24+30+32+33+38+42+43+44+55+65+81)}{18}$$
$$= \frac{586}{18}$$

$\therefore$ The mean is 32.56.

$$= 32.56 \#$$

(b) Mode = No mode.

(c) Median

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

Question 3

$$\begin{aligned} 3) (a) \quad i &= \frac{25}{100} \times 18 \\ &= 4.5 \approx 5^{\text{th}} \text{ element} \end{aligned}$$

$$Q_1 = 18$$

$$\begin{aligned} i &= \frac{50}{100} \times 18 \\ &= 9 \end{aligned}$$

$$\begin{aligned} Q_2 &= (Y[9] + Y[10]) \div 2 \\ &= \frac{30 + 32}{2} \\ &= 31, \end{aligned}$$

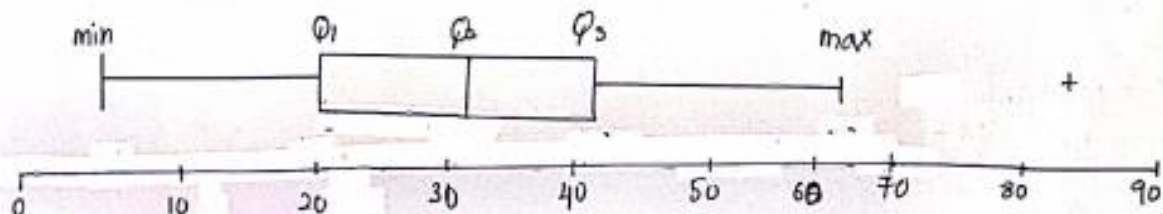
$$\begin{aligned} i &= \frac{75}{100} \times 18 \\ &= 13.5 \approx 14^{\text{th}} \text{ element} \\ Q_3 &= 43. \end{aligned}$$

$$\begin{aligned} (b) \quad IQR &= Q_3 - Q_1 \\ &= 43 - 18 \\ &= 25 \end{aligned}$$

$$\begin{aligned} (c) \quad \text{Lower limit} &= Q_1 - 1.5 IQR \\ &= 18 - (1.5 \times 25) \\ &= -19.5 \end{aligned}$$

$$\begin{aligned} \text{Upper limit} &= Q_3 + 1.5 IQR \\ &= 43 + (1.5 \times 25) \\ &= 80.5 \end{aligned}$$

$\therefore 81$ is the outlier since it is more than 80.5 .



QUESTION 4

Question 4

$$\begin{aligned} 1. \text{ Mean} &= \frac{\sum_{i=1}^n f_i x_i}{n} = \frac{(175000)(1) + (250000)(5) + (500000)(4) + (700000)(1)}{11} \\ &= \frac{175000 + 1250000 + 2000000 + 700000}{11} \\ &= \frac{4125000}{11} = \text{RM } 375000 \end{aligned}$$

$$\begin{aligned} 2. \text{ Median} &= \frac{1}{2} (11) \\ &= 5.5 \approx 6 \\ &= Y_{[6]} = \text{RM } 250000 \end{aligned}$$

3. Mean because it represents the average value of the 11 houses.

QUESTION 5

Question 5

1)

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

2)

CLASS INTERVAL	CATEGORY	FREQUENCY
$0 < \text{BMI} < 18.5$	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 Boundaries	class midpoint	Frequency	Cumulative Frequency
60 - 64	$L: \frac{59+60}{2} = 59.5$ $U: \frac{64+65}{2} = 64.5$	$= \frac{60+64}{2} = 62$	3	3
65 - 69	$L: \frac{64+65}{2} = 64.5$ $U: \frac{69+70}{2} = 69.5$	$= \frac{65+69}{2} = 67$	5	8
70 - 74	$L: \frac{69+70}{2} = 69.5$ $U: \frac{74+75}{2} = 74.5$	$= \frac{70+74}{2} = 72$	5	13
75 - 79	$L: \frac{74+75}{2} = 74.5$ $U: \frac{79+80}{2} = 79.5$	$= \frac{75+79}{2} = 77$	4	17
80 - 84	$L: \frac{79+80}{2} = 79.5$ $U: \frac{84+85}{2} = 84.5$	$= \frac{80+84}{2} = 82$	3	20
Total			20	

L - Lower boundary

U - Upper boundary



QUESTION 6

Question 6

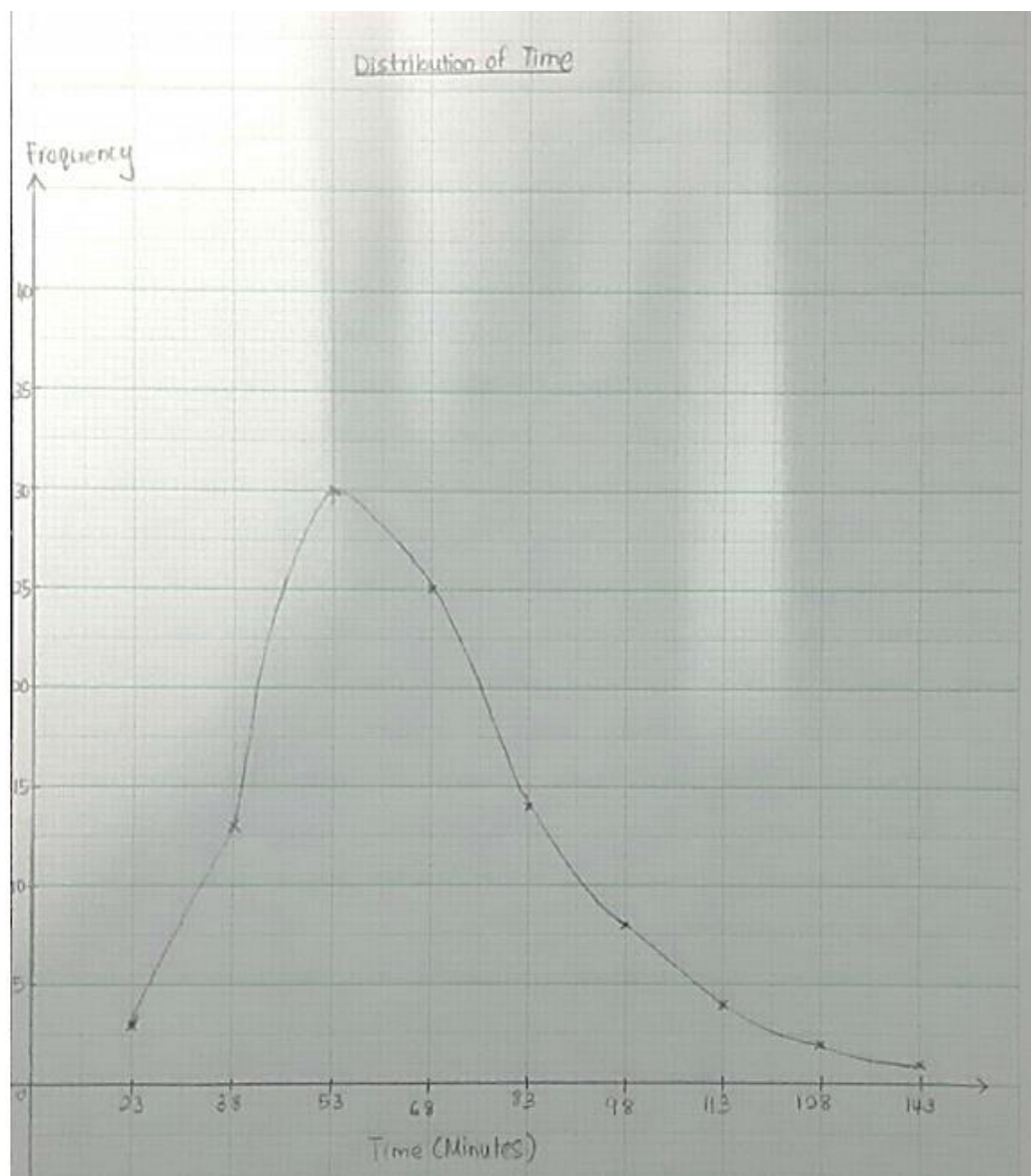
1. Time (minutes)	Midpoint (x)	Frequency (f)	fx	Cumulative Frequency
16-30	23	3	69	3
31-45	38	13	494	16
46-60	53	30	1590	46
61-75	68	25	1700	71
76-90	83	14	1162	85
91-105	98	8	784	93
106-120	113	4	452	97
121-135	128	2	256	99
136-150	143	1	143	100
		100	6650	

$$\text{Mean} = \frac{\sum_{i=1}^n fx_i}{n} = \frac{6650}{100} = 66.5$$

$$\begin{aligned} \text{Median} &= L + \frac{\frac{N}{2} - c_{fe}}{f_{med}} \quad \text{Median class} = 61-75 \\ &= 60.5 + \frac{50 - 46}{25} (15) \\ &= 62.9 \end{aligned}$$

$$\begin{aligned} \text{Mode} &= l + h \times \frac{(f_1 - f_0)}{(2f_1 - f_0 - f_2)} \\ &= 45.5 + 15 \times \frac{(30 - 13)}{(2(30) - 13 - 12.5)} \\ &= 45.5 + 15 \times (17/22) \\ &= 57.09 \end{aligned}$$

Mean > Median > Mode



b) The shape of the distribution is positively skewed and leptokurtic.

c) Based on the distribution graph, we can see that most of the flight delay time is short, with the ~~most~~ highest frequency being 46-60 minutes.