



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

SESSION 2020/2021 SEMESTER 2

SECI2143 - PROBABILITY & STATISTICAL DATA ANALYSIS

Dr. Nor Azizah Ali

Section 05

Assignment 1

No	NAME	MATRIC NUMBER
1	Omar Mokhtar Bin Yohan	A20EC0128
2	Muhammad Sulaiman Daud Syu'aib Bin Yaacob	A20EC0100
3	Ibtesham Ahmed Promit	A20EC4027
4	Md Mirza Shihab	A16CS4016

Question 1

1. (a) population = football clubs in Malaysia.
(b) sample = Malaysian Super League matches in the season
(c) discrete variable = number of goals scored
(d) continuous variable = amount of time played before a goal is scored
(e) primary data = mean number of goals per match was 2.97
(f) secondary data = mean number of goals per match was 3.08

2. (a) population = All students at a school in Johore

(b) sample = first year students

(c) discrete variable = age

(d) continuous variable = height

(e) primary data = mean of the height of second-year pupils

(f) secondary data = data on the first year pupils which include place of birth, gender, age, height and weight

Question 2

1.	Scales	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

price

2.	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	0.0	20	1.0
	4	0	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	0.30	9	0.45
3	11	0.55	20	1.00
4	0	0.00	20	1.00
Total	20	1.00		

Fuel efficiency

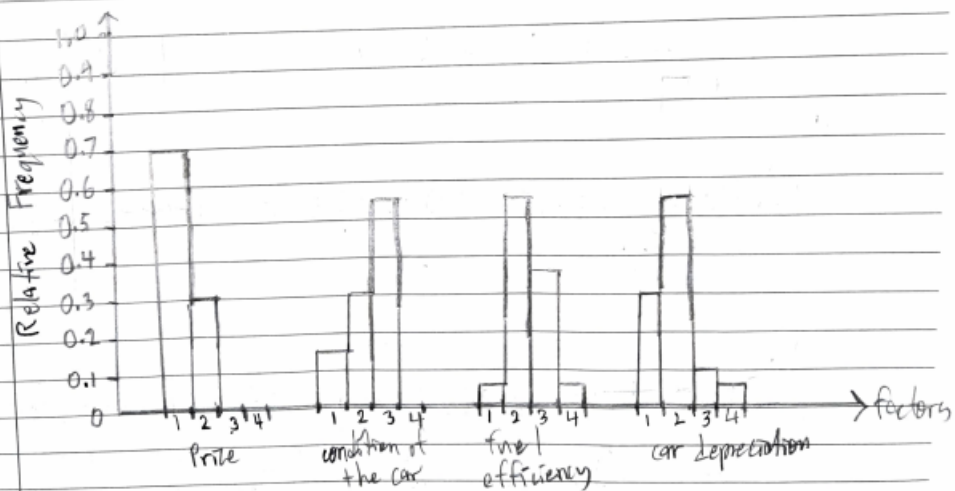
Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	1	0.05	1	0.05
2	11	0.55	12	0.60
3	7	0.35	19	0.95
4	1	0.05	20	1.00
Total	20	1.00		

Question 2

2. car depreciation

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Freq
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.



Question 3

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

1.	stem	Leaf
	0	4 8 9
	1	6 8
	2	1 3 4 key = 1 6 = 16
	3	0 2 3 8
	4	2 3 4
	5	5
	6	5
	8	1

2. a) $\text{mean} = \frac{586}{18} = 32.56$

b) There is no mode for this data

c) $\text{Median} = \frac{18+1}{2} = 9.5 = \frac{30+32}{2} = 31$

3. a) $Q_1 = \frac{25}{100}(18) = 4.5 = 5 = 18$
 $Q_2 = \frac{50}{100}(18) = 9 = \frac{30+32}{2} = 31$
 $Q_3 = \frac{75}{100}(18) = 13.5 = 14 = \frac{43+44}{2} = 43.5 = 43$

b) $IQR = 43 - 18 = 25$

c) $\text{Outliers (L)} = 18 - 1.5 \times 25 = -19.5$
 $\text{Outliers (U)} = 43 + 1.5 \times 25 = 80.5$

= Thus, the outliers is mild because $81 > 80.5$.

Question 4

1.
$$\text{mean} = \frac{(1 \times 175000) + (5 \times 250000) + (4 \times 500000) + (1 \times 700000)}{11}$$
$$= \text{RM } 375000$$

2.
$$\text{median} = \left(\frac{n+1}{2} \right)^{\text{th}} \text{ observation}$$
$$= \left(\frac{11+1}{2} \right)^{\text{th}} \text{ observation}$$
$$= 6^{\text{th}} \text{ observation}$$
$$\text{median} = \text{RM } 250000$$

3.

mean because the values are more balance between the highest price and the lowest price.

Ques-5

1.

Age	Ratio
Current smoker	Normal
Body Mass Index BMI	Interval
Hypertension	Nominal

2.

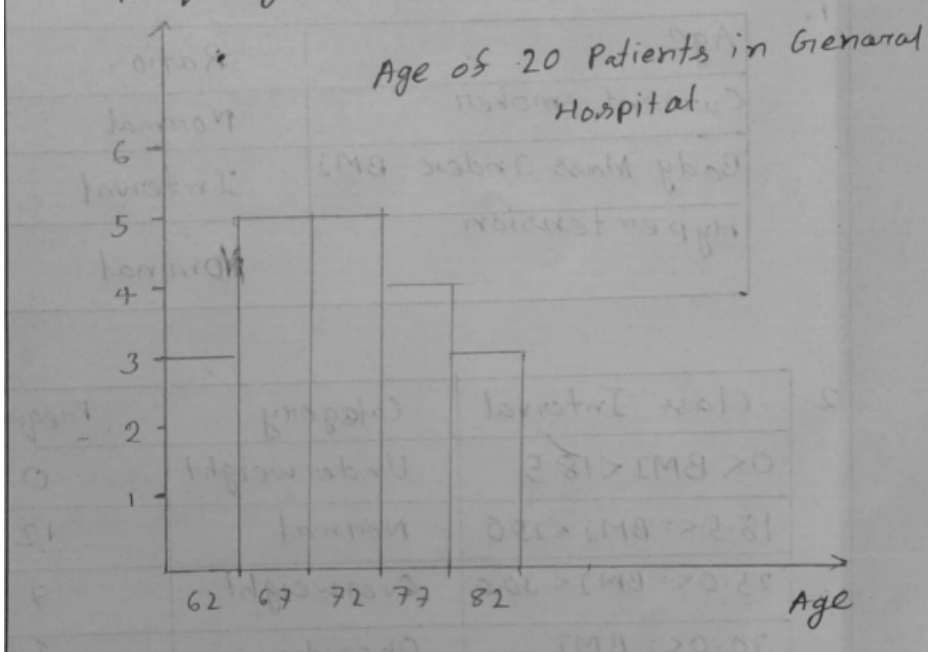
Class Interval	Category	Frequency
$0 < \text{BMI} < 18.5$	Underweight	0
$18.5 \leq \text{BMI} < 25.0$	Normal	12
$25.0 \leq \text{BMI} < 30.0$	Overweight	7
$30.0 \leq \text{BMI}$	Obesity	1

3.

Class Interval	Class Boundaries	Class Midpoint	Frequency	Cumulative Frequency
60-64	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
			20	

$$R = x - x_1 = 84 - 62 = 22$$

4. Frequency



Ques-6

1. $N \div 2 = 100/2 = 50$

$\therefore$ Median class = 61-75

$$L = 61 \quad \text{median} = L + \frac{\frac{N}{2} - cf_p}{f_{\text{med}}} (w)$$

$N = 100$

$cf_p = 46$

$f_{\text{med}} = 25$

$w = 14$

$$= 61 + \frac{50 - 46}{25} (14)$$

$$= 63.24$$

Class interval	Midpoint x_i	Frequency f_i	$f_i x_i$
16-30	23	3	69
31-45	38	13	494
46-60	53	36	1540
61-75	68	25	1700
76-90	83	14	1162
91-105	98	8	784
106-120	113	4	452
121-135	128	2	256
136-150	143	1	143
Total		100	6650

$$\text{mean} = \frac{\sum_{i=1}^n f_i x_i}{n} = \frac{6650}{100} = 66.5$$

$$l = 46$$

$$h = 14$$

$$f_1 = 30$$

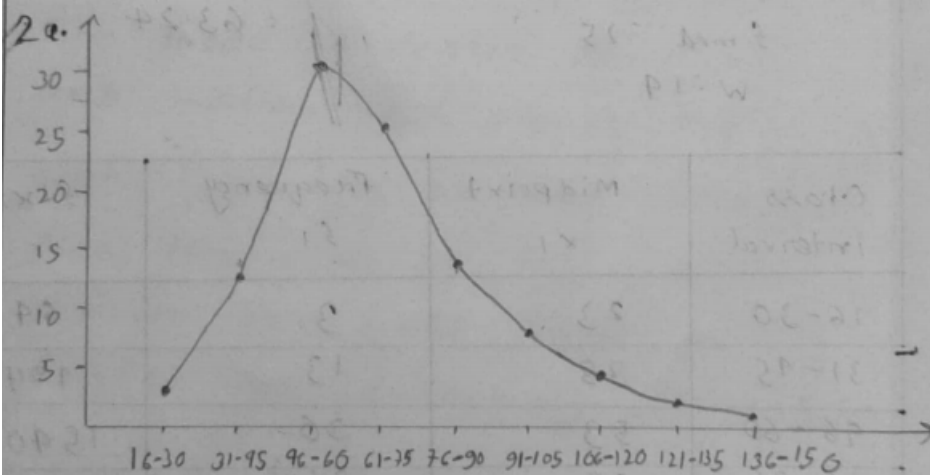
$$f_0 = 13$$

$$f_2 = 25$$

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

$$= 46 + 14 \times \frac{(30 - 13)}{(2 \times 30) - 13 - 25}$$

$$= 56.82$$



2b. From the shape of the distribution we can say, it is a positively skewed distribution to the right.

2c. As it is a positively skewed distribution The mode is lesser than the mean and median. This proves that, the highest point is on the lower side. Also, the mean is higher than both median and mode. Thus, only a few flights are extremely delayed.