



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

Subject : **PROBABILITY & STATISTICAL DATA ANALYSIS**

Section : 05

Assignment :- 01

GROUP NAME: GROUP G

1		Name : NABIL RAYHAN Matric Number : A20EC9107 Phone Number: +60187731706 E-mail :nabilrayhan17@gmail.com
2		Name : NAZMUS SAKIB Matric Number : A20EC4046 Phone Number :+60172081293 E-mail:sakib.n@graduate.utm.my
3		Name : FARHANA TAHSEEN Matric Number : A20EC4015 Phone Number:+60193272280 E-mail :farhanatahseen98@gmail.com
4.		Name :MD.MOHAIMINUL ISLAM FAYSAL Matric Number : A20EC9105 Phone Number:+601127739204 E-mail :zxfaysalgrp@gmail.com
5.		Name : MUHAMMAD ROBY HERFIAN Matric Number : A20EC0321 Phone Number: 628127632798 E-mail :herfian@graduaate.utm.my

ANS TO THE QUESTION NO: 01

1. a.) Football clubs in Malaysia
- b.) the mean number of goals per game.
- c.) the previous season, the number of goals scored in each match is 3.08
In this season, the number of goals scored in each match is 2.9.
- d.) The difference between the mean number of goals per game is -0.11
- e.) the number of goals scored per game that journalist recorded
On the first Saturday of the season.
- f.) Old newspaper articles that showed the mean number of goals scored in each match was 3.08

2. a.) All students at a school in JayPie
- b.) The data recorded for each pupil include place of birth, gender, age, height and weight.
- c.) ~~Each member of the class~~
- d.) a set of bathroom scales will be provided
~~by statistics students~~
- e.) compare the first year student weight data with second year student data

F Student must undergo a medical examination during their first year at the school

E the data weight of second year pupils that collected by statistics students.

Ans. to the ques. NO. 02.

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	7	1	20
Car Depreciation		6	11	2	1	20

2. Frequency distribution for price,

Scale	Frequency	R. Frequency	Com. Frequency	C.R. Freq.
1	14	0.7	14	0.70
2	6	0.3	20	1.0
3	0	0	20	1.0
4	0	0	20	1.0
Total	20	1.0		

Frequency distribution for condition of the car,

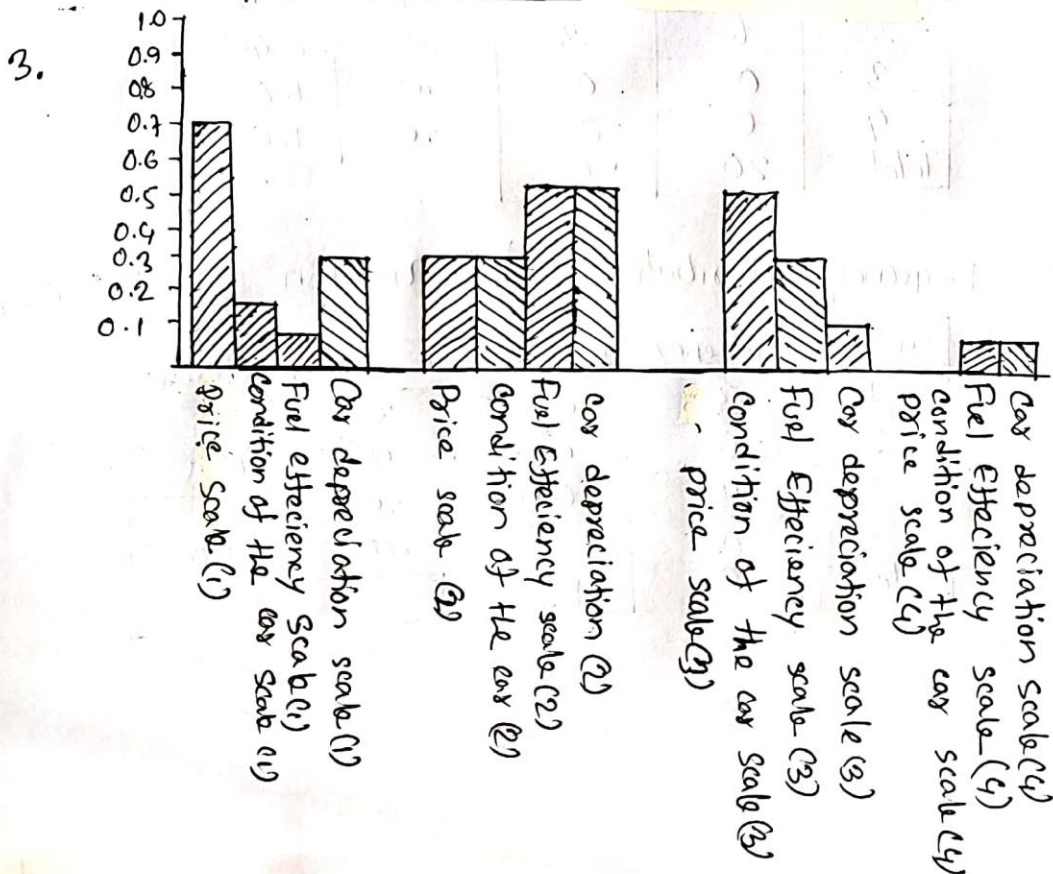
Scale	Frequency	R. Frequency	Com. Frequency	C.R. Freq.
1	3	0.15	3	0.15
2	6	0.30	9	0.45
3	11	0.55	20	1.0
4	0	0	20	1.0
Total	20	1.0		

Frequency distribution for Fuel Efficiency,

Scale	Frequency	R. Frequen	Cum. Frequen	C.R. Frequen
1	3	0.05	1	0.05
2	11	0.55	12	0.60
3	7	0.35	19	0.95
4	1	0.05	20	1.0
Total	20	1.0		

Frequency distribution for Car Depreciation,

Scale	Frequency	R. Frequency	Cum. Frequen	C.R. Frequen
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	1.0
Total	20	1.0		



Answer to the question no. 3

①

The data has been represented in a Stem-and-Leaf plot below —

Stem	Leaf
0	8 4 9
1	8 6
2	1 3 4
3	0 2 8 3
4	3 4 2
5	5
6	5
7	
8	1

Key : 0/8 means 8 days



Stem : tens ; leaf : ones

(2)

(a)

$$\text{Mean} = \frac{30+43+32+21+65+18+4+18+16+38+2+44+33+23+24+18+12+5}{18}$$

$$= 32.56.$$

(b)

Since each value in the given data set appears only once, the data set has no mode.

(c)

The data has been arranged in order below —

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

$\therefore$ Median = Mean of two middle numbers

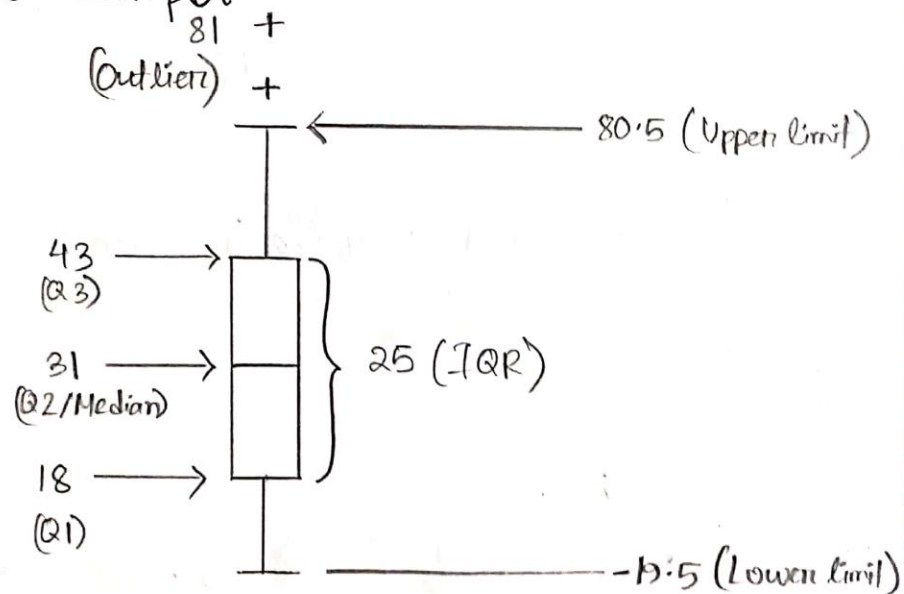
$$= \frac{30+32}{2}$$

$$= 31.$$

Ans: Mean = 32.56, no mode, median = 31.

iii. (3)

The data has been represented below in a modified box plot —



(a)

We know,

$$i = \frac{P}{100} (N)$$

For the 1st quartile,

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

$$= 4.5$$

$$\therefore \text{1st quartile, } Q_1 = Y[5] \\ = 18$$

For the 2nd quartile (median),

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

$$= 9$$

$$\therefore \text{2nd quartile (median), } Q_2 = \frac{(Y[9] + Y[9+1])}{2} \\ = \frac{30 + 32}{2} \\ = 31$$

For the 3rd quartile,

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

$$= 13.5$$

$$\therefore \text{3rd quartile, } Q_3 = 4[14] \\ = 43.$$

$$\text{Ans: } Q_1 = 18, Q_2 = 31, Q_3 = 43.$$

(b)

$$\begin{aligned} \text{Interquartile Range (IQR)} &= Q_3 - Q_1 \\ &= 43 - 18 \\ &= 25. \end{aligned}$$

$$\text{Ans: } \text{IQR} = 25.$$

(c)

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

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

From the data set, we can see that,

since $81 > \text{Upper limit}$,

81 is a mild outlier in the data set.

Ans: Mild outlier = 81.

Ans to the question no: 04

Table. 2

Value per house (x_i)	Numbers of houses (f_i)	$f_i x_i$	commulative frequency
RM 175,000	1	(175000×1) $= 175,000$	1
RM 250,000	5	1250000	$1+5$ $= 6$
RM 500,000	4	2000000	10
RM 700,000	1	700,000	11
	$n = 11$	$\sum f_i x_i =$ 4125000	

04

①

Mean value,

We know,

$$\text{Mean} = \frac{\sum f_i x_i}{n}$$

$$= \frac{4125000}{11}$$

$$= 375,000.$$

04
02

Data set $\rightarrow$ 175000, 250000, 250000, 250000, 250000,
250000, 500000, 500000, 500000, 500000, 500000, 700000

$$\text{Median value} = (N+1) \div 2$$

$$= (11+1) \div 2$$

$$= 6^{\text{th}} \text{ number value.}$$

$$\text{So, Median is} = 250,000.$$

04
03

$$\text{Mean} = 375000$$

$$\text{Median} = 250,000$$

$$\text{So, Mean} > \text{Median}$$

The median is usually preferred to other in these situations because the value of the mean can be ~~distorted~~ distorted by the outliers. So, Median is best represents to the central tendency. Also Median value is represent the middle value of the averaged data set.

Question No: 5

1. • Age - Ratio
- Current Smoker - Nominal
- Body Mass Index (BMI) - Interval
- Hypertension - ~~Ratio~~ Nominal

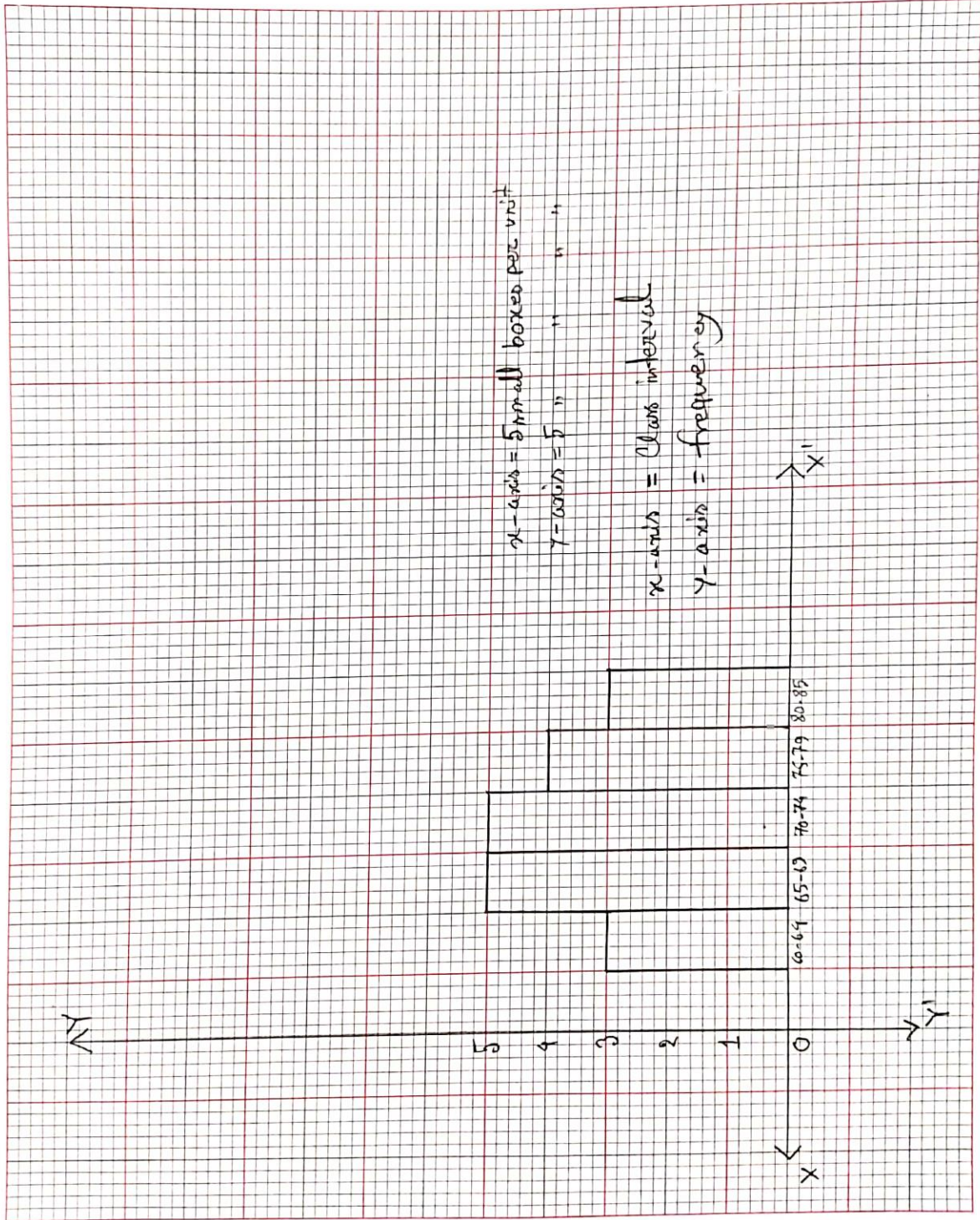
2.

Class Interval	Category	Tally	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

Number of total Patient: 20.

3.

Class Interval	Class Boundaries	Class Midpoint	Tally	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	61



ANS TO THE QUESTION NO: 06

Ans to the question No: 6

Time	frequency (fi)	Midpoint (xi)	commulative frequency	fi xi	
16-30	3	23	3	69	
31-45	13	38	16	494	
46-60	30	53	46	1590	
61-75	25	68	71	1700	
76-90	14	83	85	1162	
91-105	8	98	93	784	
106-120	4	113	97	452	
121-135	2	128	99	256	
136-150	1	143	100	143	
	N=100			$\sum f_i x_i = 6650$	

$$\begin{aligned}\textcircled{1} \text{ Mean} &= \frac{\sum f_i x_i}{n} \\ &= \frac{6650}{100} \\ &= 66.5.\end{aligned}$$

06
Q1 Median,

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

$$= 61-75.$$

$$L = 61$$

$$N = 50$$

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

$$cf_p = 46$$

$$w = 14$$

$$\text{Median} = L + \frac{\frac{N}{2} - cf_p}{f_{\text{(med)}}} \times w$$

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

$$= 63.24.$$

06 01

Mode,-

$$\text{Modal class} = \text{Dec 46} - 60$$

Hence,

$$l = 46$$

$$h = 14$$

$$f_1 = 30$$

$$f_0 = 23$$

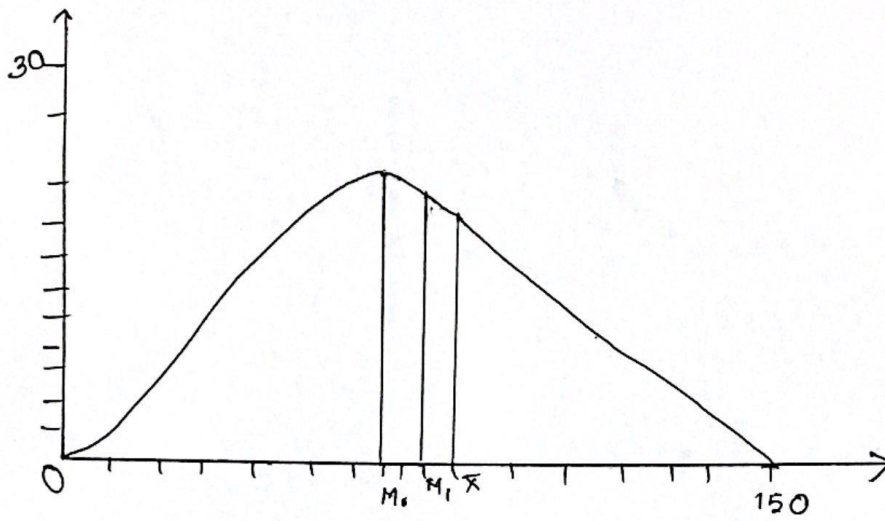
$$f_2 = 25$$

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

$$= 46 + 14 \times \frac{30 - 23}{(2 \times 30) - 23 - 25}$$

$$= \text{Dec 46} \cdot 56 \cdot 818 \cdot$$

06
02
a)



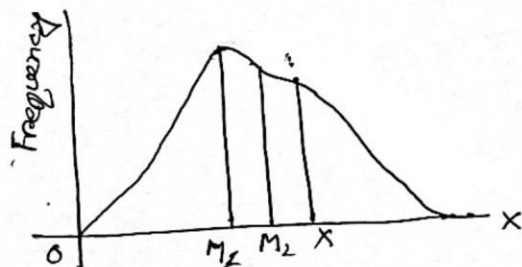
06

02

Q In the shape of distribution,

we can tell that

It is positively skewed distribution.



Positively skewed distribution

02 C

from, Q, we see that it's positively skewed distribution, that's mean the mode $<$ the mean and median.

Mode $<$ Median $<$ Mean.

$K > 0$. So, it's leptokurtic.

It's tails are longer and fatter and often it's central peak is higher and sharper.

