



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

SECI2143

PROBABILITY & STATISTICAL DATA ANALYSIS

Assignment 1

Lecturer:

Dr Sharin Hazlin Binti Huspi

Team Members:

Nik Syahdina Zulaikha binti Badrul Hisham (A20EC0108)

Hanis Rafiqah binti Hisham Razuli (A20EC0041)

Semester 2 2020/2021

PSDA

ASSIGNMENT 1

Question 1

- ① 1. a) Malaysiana
b) Malaysiana Super League matches
c) the total number of goals scored
d) the amount of time played before a goal is scored
e) number of goals scored in each match
f) mean number of goals per game was 3.08,
mean number of goals per match as 2.97
2. a) All students at a school in JayBie
b) first year, second year
c) place of birth, gender, age
d) height, weight
e) weight of second pupils, the number of pupils weighed
f) data on first year pupils, the mean of weights recorded

Question 2

(2) 1.

Scales	Frequency				Total
Factors	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. Factor : Price

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	14	$14 \div 20 = 0.7$	14	$14 \div 20 = 0.7$
2	6	0.3	20	$20 \div 20 = 1.0$
3	0	0	20	1.0
4	0	0	20	1.0
Total	20	1.0		

Factor : 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		

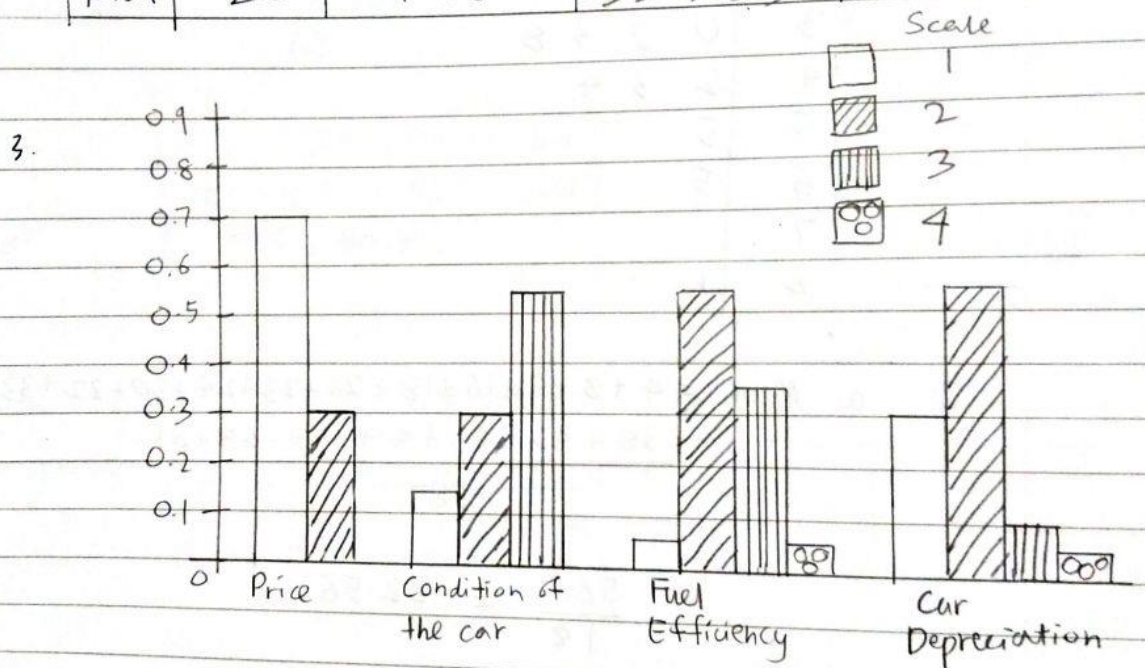
Factor : Fuel Efficiency

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. Factor : 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		



QUESTION 3

① Represent the data in a stem-and-leaf plot.

stem	leaf	
0	4	8
1	6	8
2	1	3
3	0	2
4	2	3
5	5	4
6	5	
7		
8	1	

KEY=16

2. (a) mean

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

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

18

$$= 32.56$$

b. ~~As the~~ Since each value occurs only once in the data set, there is no mode

(c) median

$$\text{median} = \frac{1}{2} [x_9 + x_{10}] \quad x_{\frac{n+1}{2}} \quad (\text{odd})$$

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

$$= \frac{1}{2} [30 + 32] \quad \frac{1}{2} [x_{\frac{n}{2}} + x_{\frac{n}{2}+1}] \quad (\text{even})$$

$$= 31$$

n = number of values in data set

③ Box plot

(a) 1st, 2nd, 3rd Quartiles

$$Q_1 = P_{25} = X_5 = X_5 = 18$$

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

$$Q_2 = P_{50} = \text{median} = 31$$

$$Q_3 = P_{75} = X_5 = X_{14} = 43$$

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

if the index is not a round number, round it up (or down, if it is closer to the lower number) to the nearest whole number

3.		
a.	$Q_1 = 18$	$\min = 4, \max = 81$
	$Q_2 = 31$	
	$Q_3 = 43$	

(b) Interquartile range (IQR)

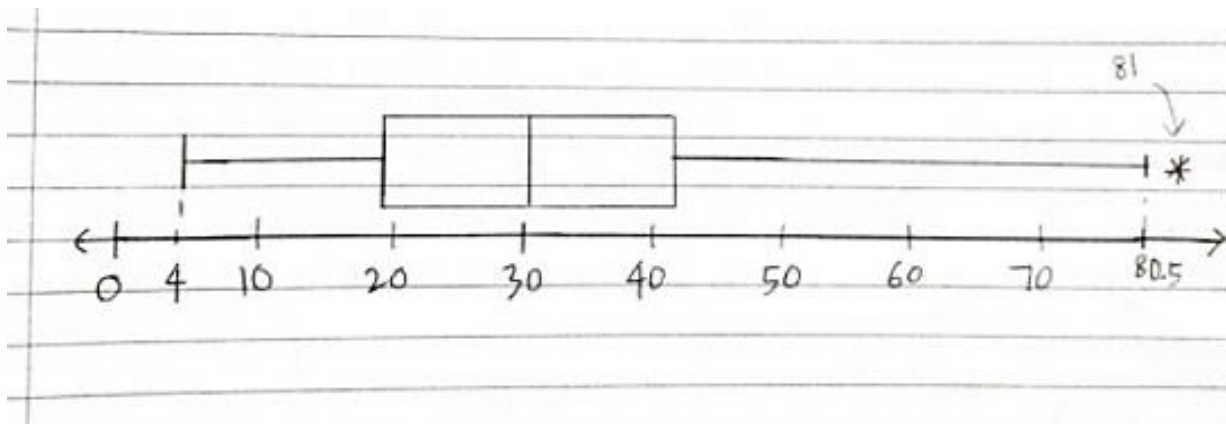
$$\begin{aligned} \text{IQR} &= Q_3 - Q_1 \\ &= 43 - 18 \\ &= 25 \end{aligned}$$

(c) mild and/or extreme outliers in the dataset

$$\begin{aligned} \text{Lower fence} &= Q_1 - 1.5(Q_3 - Q_1) \\ &= 18 - 1.5(25) \\ &= -19.5 \end{aligned}$$

$$\begin{aligned} \text{Upper fence} &= Q_3 + 1.5(Q_3 - Q_1) \\ &= 43 + 1.5(25) \\ &= 80.5 \end{aligned}$$

c)



Question 4

$$\textcircled{4} 1. \text{ mean} = \frac{175\,000(1) + 250\,000(5) + 500\,000(4) + 700\,000(1)}{11}$$

$$= \frac{4\,125\,000}{11} = 375\,000$$

$$\text{mean} = \text{RM } 375\,000.00$$

(b) median (km)

$$\text{median} = X_{\frac{11+1}{2}} = X_{\frac{12}{2}} = X_6$$

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

3. mean, represent average

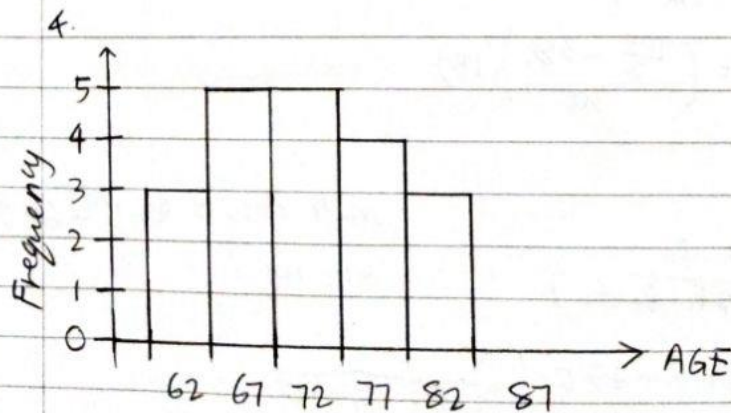
Question 5

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	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	-



Question 6

time	frequency	x	F	fx
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
	100			6650

← mode class
← median class

(a) median, mean, mode

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

$$\begin{aligned} \text{median} &= L + \left[\frac{\frac{N}{2} - F_b}{f_{\text{med}}} \right] (C) \\ &= 60.5 + \left[\frac{\frac{100}{2} - 46}{25} \right] (15) \\ &= 62.9 \end{aligned}$$

$$\text{Mean} = \frac{3(23) + 13(38) + 30(53) + 25(68) + 14(83) + 8(98) + 4(113) + 2(128) + 1(143)}{100}$$

$$= \frac{6650}{100} = 66.5$$

$$\text{mode class} = 46-60$$

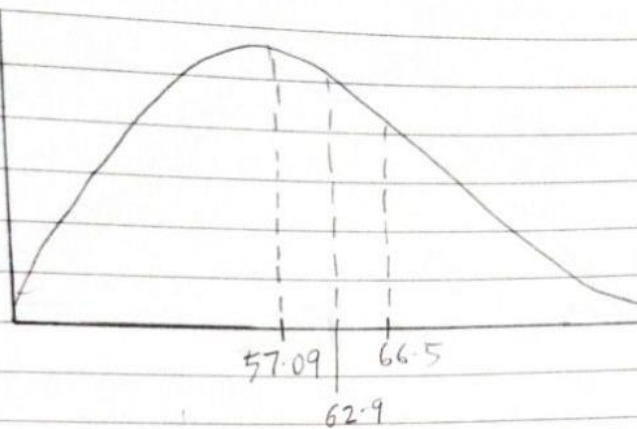
$$\text{mode} = l + c \left[\frac{f - f_b}{2f - f_b - f_a} \right]$$

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

$$= 57.09$$

Question 6

2.



b) right skewed
leptokurtic

c) The average time is 66.5 minutes

The distribution is positively skewed as
the data falls toward the right side of the scale