



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

**PROBABILITY & STATISTICAL DATA
ANALYSIS**

SEMESTER II 2020/2021

SECI2143-05

ASSIGNMENT 1

NAME OF STUDENTS	MATRIC NO.
BRENDAN DYLAN GAMPA ANAK JOSEPH DUSIT @ DUSIT	A20EC0021
HAFIY HAKIMI BIN SAIFUL REDZUAN	A20EC0040
MOHAMAD HAZIQ ZIKRY BIN MOHAMMAD RAZAK	A20EC0079
AUM JEEVAN A/L AUM NIRANGKAR	A20EC0017

LECTURER: DR. NOR AZIZAH BINTI ALI

SUBMISSION DATE: 17TH APRIL 2021

Question 1

1. a) the population
= Football clubs in Malaysia
- b) the sample
= All the Malaysia Super League matches in the season.
- c) a discrete variable
= home win
- d) a continuous variable
= Amount of time played before a goal is scored
- e) primary data
= 2.97
- f) secondary data
= 3.08

2. a) the population
= All students at a school in JayBre
- b) the sample
= first year
- c) a discrete variable
= gender
- d) a continuous variable
= weight
- e) primary data
= weight of second year pupils
- f) secondary data
= weight of first year pupils

QUESTION 2 [34 marks]

A small, explorative survey was done among 20 respondents in the AERON car park to determine what factors were important to buyers when buying a car. The four most important factors considered by the buyers were price, condition of the car, fuel efficiency, and car depreciation.

Buyers were given a questionnaire that had a 4-point interval scale on which they could rate their preference. One on the scale meant very important and four very unimportant. In between points on the scale were intended to allow for degrees of preference between the polar extremes. Table 1 shows the results as follows:

Table 1: Surveys on Car Buying Factors

Respondents	Factors			
	Price	Condition of the Car	Fuel Efficiency	Car Depreciation
A	1	2	2	2
B	2	2	3	1
C	1	3	3	2
D	2	1	4	2
E	1	2	3	2
F	1	3	3	1
G	2	3	2	3
H	1	3	2	1
I	1	1	2	1
J	2	1	2	2
K	1	2	3	3
L	2	3	3	2
M	1	3	1	2
N	1	3	2	2
O	1	3	2	4
P	1	3	2	2
R	2	2	2	1
S	1	3	2	2
T	1	2	3	2
U	1	3	2	1

1. Summarize Table 1 in a frequency table as below. (10 marks)

Scales Factors	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. Summarize all the factors (price, condition of the car, fuel efficiency, and car depreciation) from Table 1 in a frequency distribution. For each of the factor, use the table as below:
(1 table for a factor = 5 marks, 4 tables = 20 marks)

Table 1: 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	1.0		

Table 2: 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		

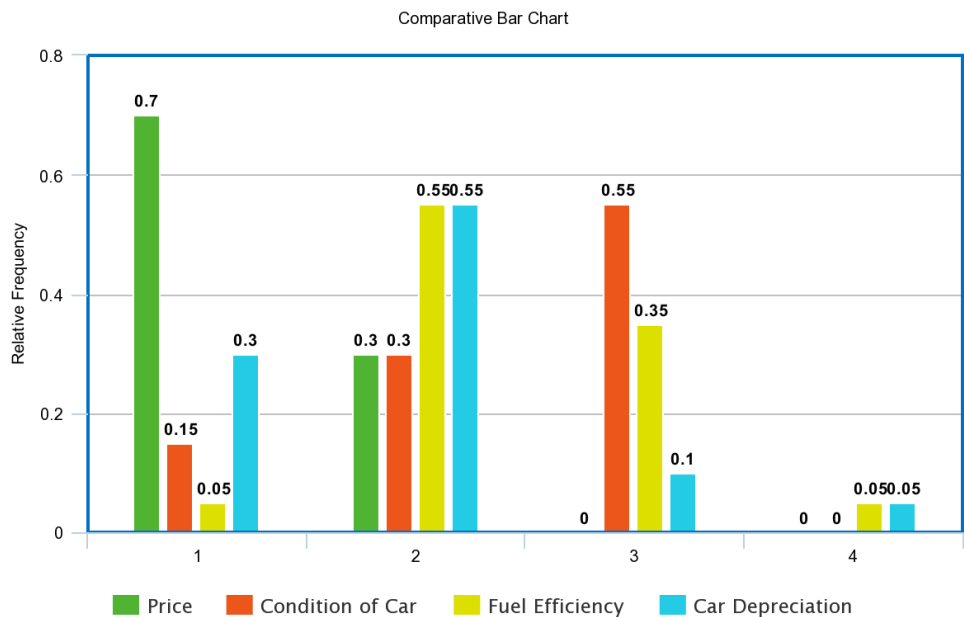
Table 3: Fuel efficiency

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		

Table 4: 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		

3. Draw the results from (b) using a comparative bar chart. (4 marks)



Question 3

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

1. Stem-and-Leaf plot

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 = 0/4
= 4

2. a) mean

$$\bar{x} = \frac{4 + 8 + 9 + 16 + 18 + 21 + 23 + 24 + 30 + 32 + 33 + 38 + 42 + 43 + 44 + 55 + 65 + 81}{18}$$
$$= 32.56$$

b) Mode

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

= There is no mode because there is only one frequency for each data

c) Median

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

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

3. a) 1st quartile

$$i = \frac{25}{100}(18) = 4.5, k=5$$

$$\rightarrow Y[5] = 18$$

2nd Quartile (median)

$$i = \frac{50}{100}(18) = 9, k=9$$

$$\frac{Y[9] + Y[10]}{2} = \frac{30 + 32}{2} = 31$$

3rd Quartile

$$i = \frac{75}{100}(18) = 13.5, k=14$$

$$\rightarrow Y[14] = 43$$

b) Interquartile Range (IQR)

$$= Q3 - Q1$$

$$= 43 - 18$$

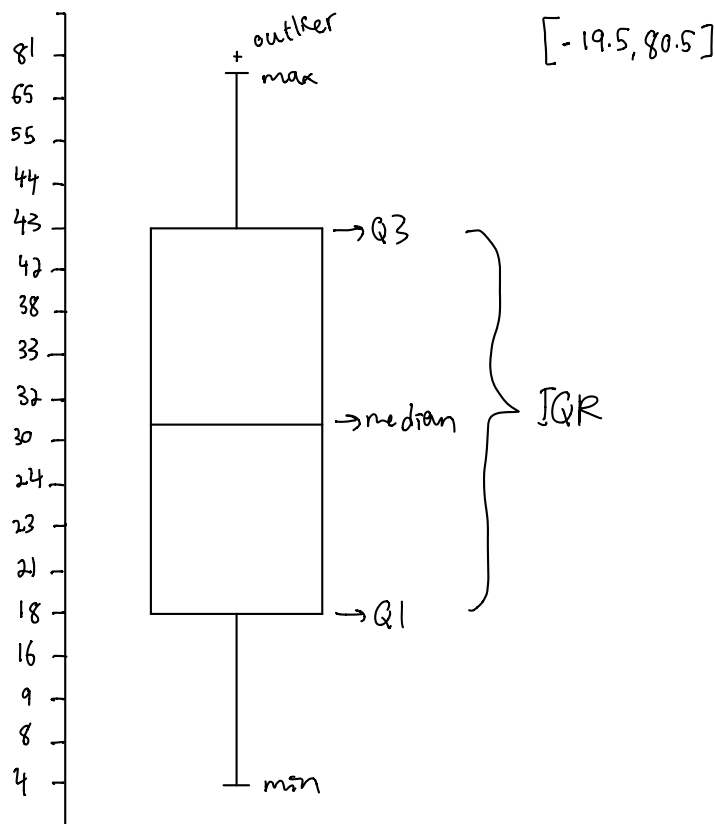
$$= 25$$

c) mild and/or extreme outliers

$$\begin{array}{ll} \text{lower limit: } 18 - (1.5 \times 25) & \text{upper limit: } 43 + (1.5 \times 25) \\ : -19.5 & : 80.5 \end{array} \therefore [-19.5, 80.5]$$

$$= -19.5, 4, 8, 9, 16, 18, 21, 23, 24, 30, 32, 33, 38, 42, 43, 44, 55, 65, 80.5, \textcircled{81}$$

= 81 is mild outlier



Question 4

1. mean in RM

$$\bar{x} = \frac{\text{RM } 175\,000 (1) + \text{RM } 250\,000 (5) + \text{RM } 500\,000 (4) + \text{RM } 700\,000 (1)}{1 + 5 + 4 + 1}$$

$$= \frac{\text{RM } 4\,125\,000}{11}$$

$$= \text{RM } 375\,000$$

2. median in RM

$$\text{median} = \frac{11+1}{2} = 6^{\text{th}} \text{ number}$$

$\therefore$ from the cumulative frequency, the 6th number is RM 250 000

Thus, median value is RM 250 000

3. The median "best" represents the values of 11 houses because its value is closer to most of the values in the table compared to mean.

QUESTION 5

given that $BMI = \frac{\text{weight (kg)}}{(\text{height})^2 (m)}$

1.)

Age	Ratio
Current smoker	Nominal
Body Mass Index (BMI)	Interval
Hypertension	Nominal

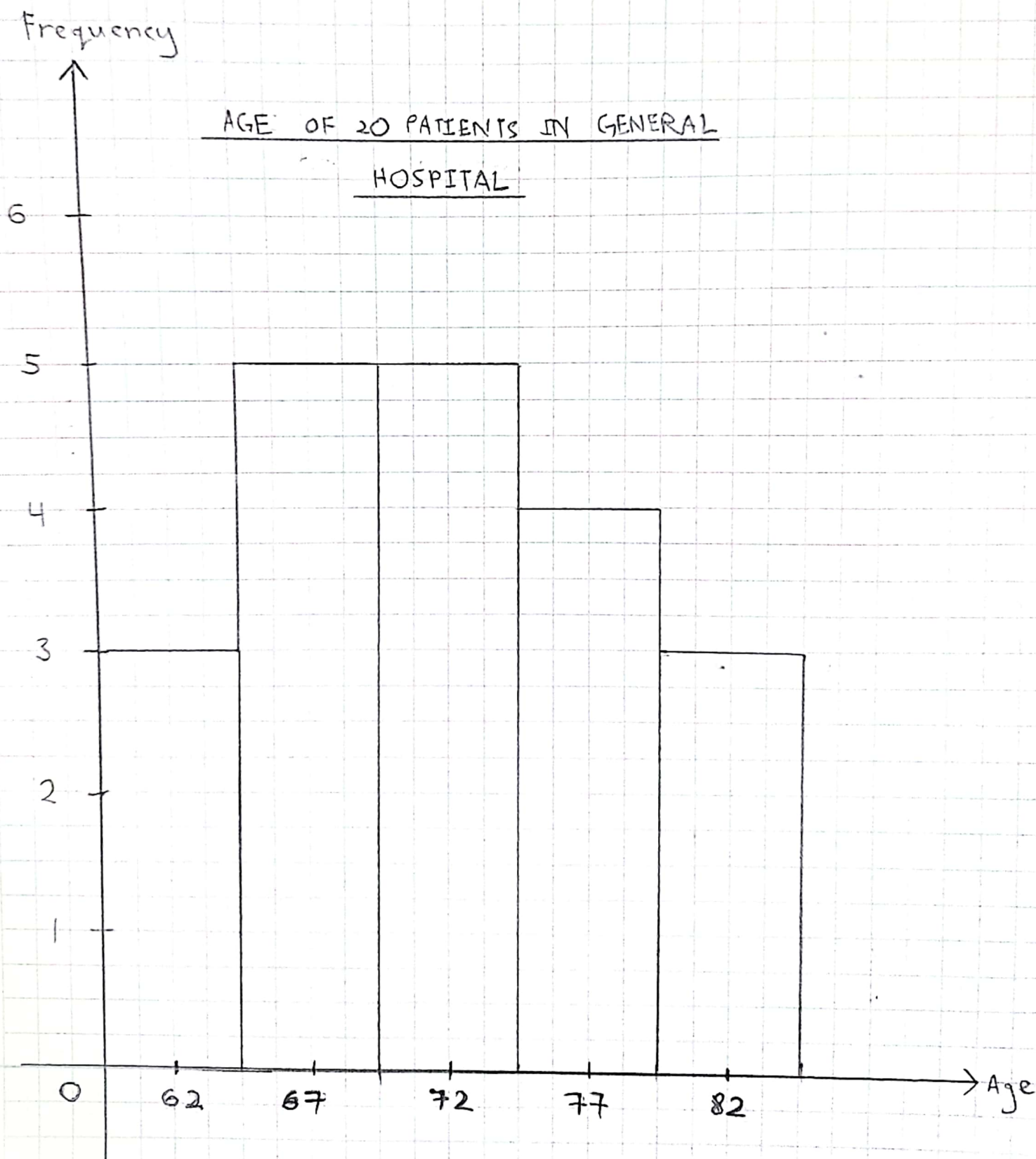
2.)

class interval	category	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

|||| |||| ||
 |||| ||
 |

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	20



Question 6

Time	f	Midpoint, m	f.M.	Cumulative f
16-30	3	23	69	3
31-45	13	38	494	16
46-60	30	53	1590	46
61-75 _{mc}	25	68	1700	71
76-90	14	83	1162	85
91-105	8	98	784	93
106-120	4	113	452	97
121-135	2	128	256	99
136-150	1	143	143	100
	$\Sigma n = 100$		$\Sigma fm = 6650$	

* ← modal class

* ← median 47-71

$$\begin{aligned}\text{Mean, } \bar{X} &= \frac{\Sigma fm}{\Sigma n} \\ &= \frac{6650}{100} = 66.5\end{aligned}$$

$$\begin{aligned}\text{Median} &= \frac{1}{2} (100) \\ &= 50\text{th element} \\ &= L + \frac{\frac{N}{2} - cf}{f_{med}} (w)\end{aligned}$$

$$\begin{aligned}\text{Mode} &= 16m_0 + \left(\frac{D_1}{D_1 + D_2} \right) i \\ &= 45.5 + \left(\frac{17}{17+5} \right) (15) \\ &= 57.09 \\ &\approx 57.1\end{aligned}$$

$$\begin{aligned}&= 61.5 \\ &= 60.5 + \left(\frac{50 - 46}{25} \right) (15) \\ &= 62.9\end{aligned}$$

- 2) b) Graph is positively skewed and mesokurtic
- c) Graph shows an average of delayed time is 66.5 minutes and the most frequent delayed time is 57.1 minutes.

