



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

Assignment 1 (Group)
SECI2143-07
Probability & Statistical Data Analysis
SEMESTER II, SESSION 2021/2022

Lecturer: Dr. Nor Azizah Ali

Group Name: Little Group

Name	Student ID
TIA SIAW XUEN	A21EC0233
CHUA XIN LIN	A21EC0200
MA ZE JUN	A21EC4009
KEE LE WEI	A21EC0189

1.1 Descriptive: Graphical description (Bar chart Pie chart)

Inferential Statistics: Estimate the population by sample

1.2 Through a sample of 30 students, it is estimated that students with 75 points or above from same university answered the evaluation related questions

1.3

i. understanding the nature of problem

Random collect information about 30 students with 75 points or above

ii: deciding what to measure and how to measure it

for example, the time spent on internet, the purpose of use, and so on.

iii: data collection

through questionnaires, emails, etc.

iv: Data summarization and preliminary analysis

summarize by graph, data and so on.

v. formal data analysis

vi: Interpretation of results

1.4

i: Nominal: what type of internet do you use 1. Google 2. Baidu 3. Sogou

ii: Ordinal: How satisfied are you with online shopping?

1. very Satisfied 2. Satisfied 3. Neutral 4. Unsatisfied.

iii: Interval: How much time do you spend^{on} online shopping one day?

iv: Ratio: How many devices you used for internet?



1.5 Prepare a google form with suitable questions and corresponding possible answers that is able to collect data and generate output as outlined in Figure (1) to (5). Provide a link of the google form in your answer sheet.

Link: <https://forms.gle/D7DZhAcNBabKApYE9> (access using UTM email)



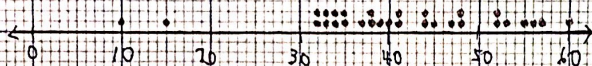
Nama :

Topik :

Tingkatan :

Tarikh :

2.1) a.



2.1) b.

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

Key: 1 | 0 = 10



2.2) 10 15 32 32 33 33 34 34 35 35 37 38 38 39 40 41
41 44 44 45 47 48 48 52 52 53 55 56 57 60

- second fastest student = 15 minutes

$$Y[k] = 15$$

$$\text{- percentile of value 15} = \frac{\text{number of value less than 15}}{\text{total number of values}} (100)$$

$$= \frac{1}{30} (100)$$

$$= 3.33$$

$$= P_4$$



3.1)

Mid-value	Class Interval	Frequency	Cumulative frequency
5	0 - 10	7	7
15	10 - 20	10	17
25	20 - 30	23	40
35	30 - 40	51	91
45	40 - 50	6	97
55	50 - 60	3	100





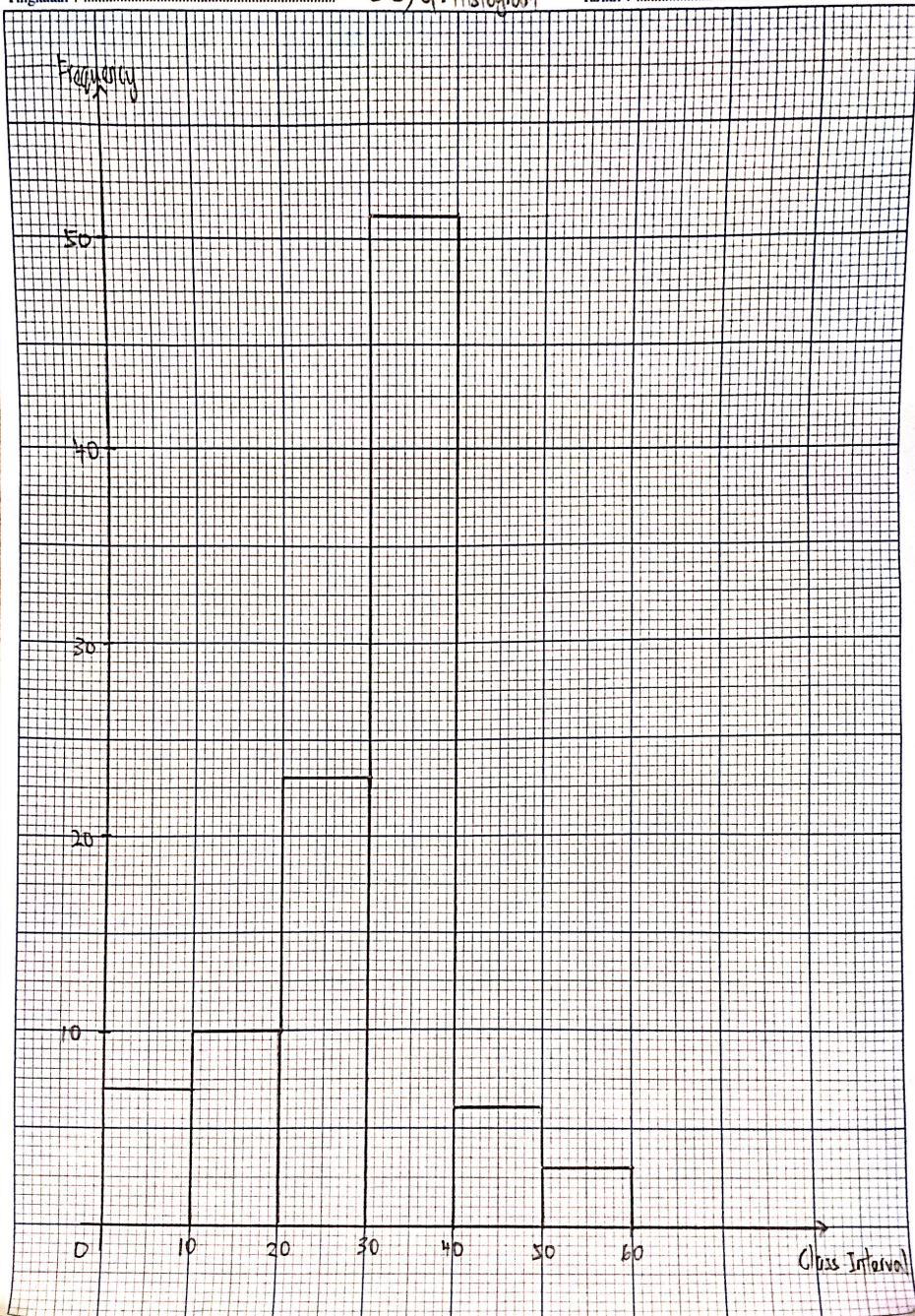
Nama :

Topik :

Tingkatan :

3.2) a. Histogram

Tarikh :





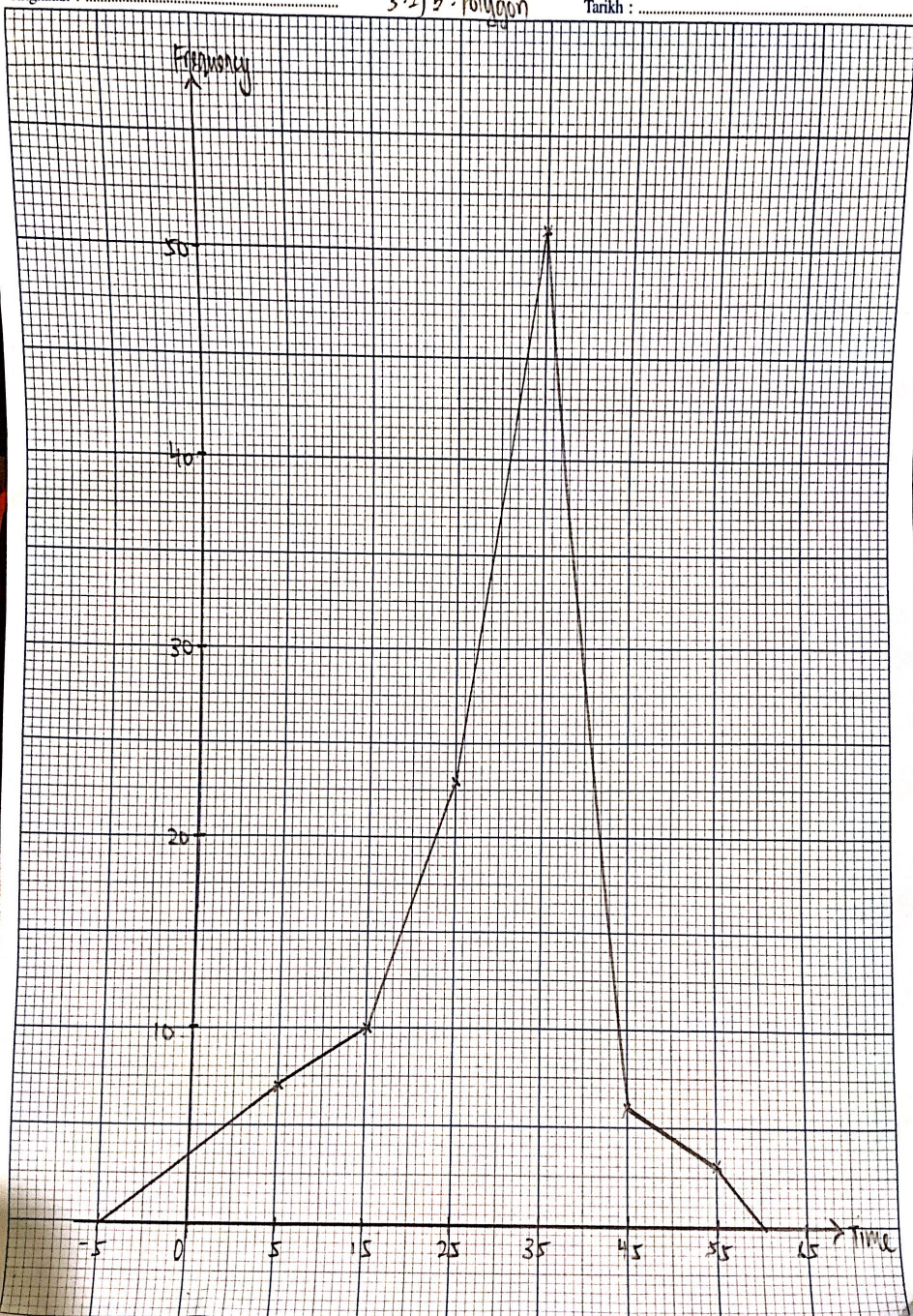
Nama :

Topik :

Tingkatan :

3.2) b. Polygon

Tarikh :





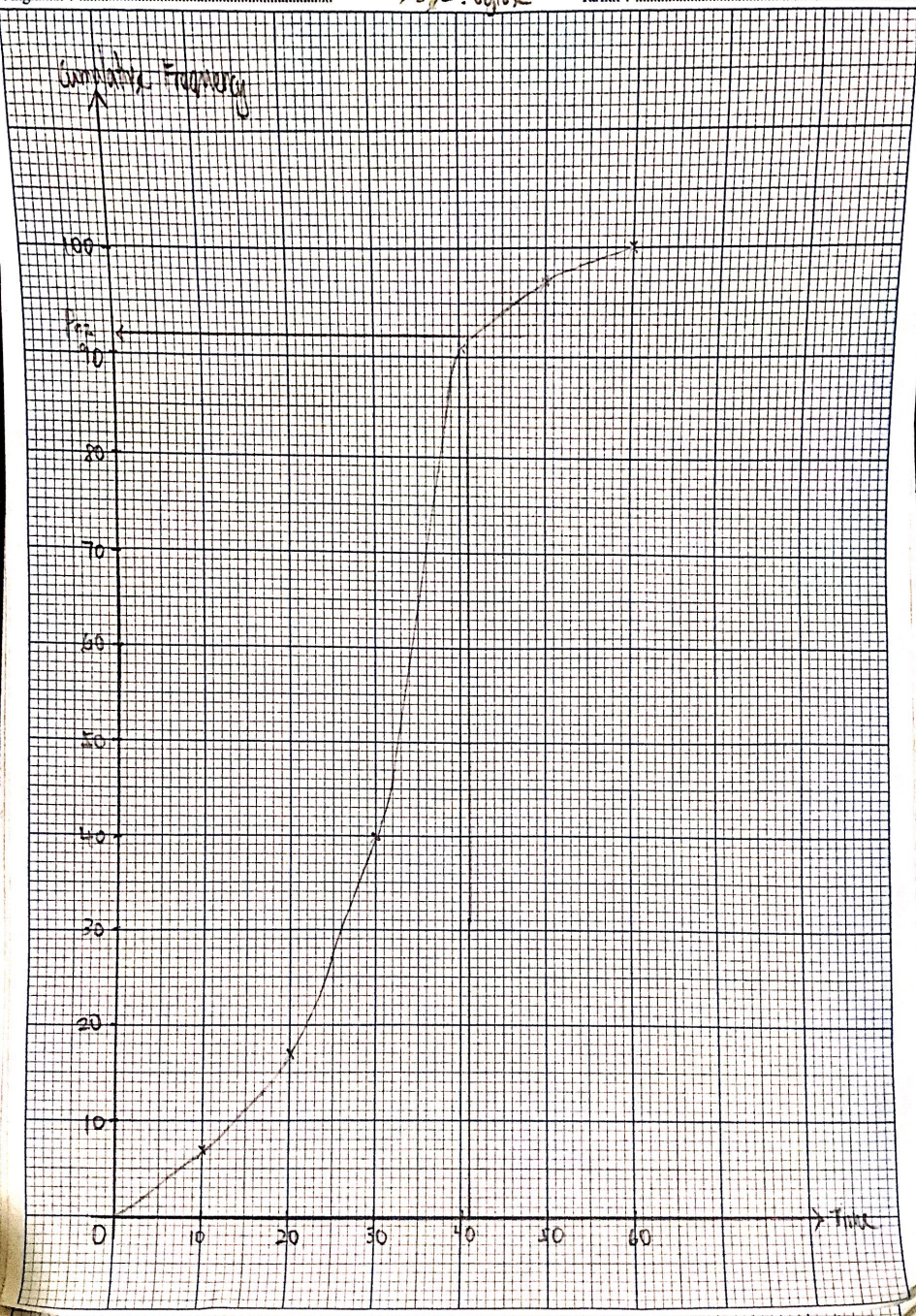
Nama :

Topik :

Tingkatan :

3.3) c. Ogive

Tarikh :



5.2) d. P_{q_2}

$$3.3) \sum f_{ix} x_i = (5 \times 7) + (15 \times 10) + (25 \times 23) + (35 \times 51) + (45 \times 6) + (55 \times 3) \\ = 2980$$

$$\text{mean} = \bar{x} = \frac{\sum f_{ix} x_i}{N} = \frac{2980}{106} = 29.80$$

$$\text{median} = L + \frac{\frac{N}{2} - c_{f_1}}{f_{\text{med}}}(w) = 30 + \frac{104/2 - 40}{51}(10) = 30.20$$

$$\text{mode} = L + h \times \frac{(f_1 - f_0)}{(2f_1 - f_0 - f_2)} \\ = 30 + 10 \times \frac{(51 - 23)}{(2 \times 51 - 23 - 6)} \\ = 33.84$$



Question 4 (Data in Question 2) (Universiti Canggih Malaysia)

4.1) a) range = $60 - 10$
 $= 50$

b) mean = $\frac{(35+33+38+10+52+48+34+44+41+56+53+41+39+32+37+60+34+45+33+57+38+48+40+32+55+52+35+47+44+15)}{30}$
 $= 40.93$

c) 10 15 32 32 33 33 34 34 35 35
37 38 38 39 40 41 41 44 44 45
47 48 48 52 52 53 55 56 57 60

median = $\frac{(40+41)}{2}$
 $= 40.5$

d) mode = 32, 33, 34, 35, 38, 41, 44, 48, 52

1) first Quartile (Q_1)

$$i = \frac{25}{100} \times 30 = 7.5$$

$$k = 8$$

$$Y[8] = 34$$

$$Q_1 = 34$$

e) third Quartile (Q_3)

$$i = \frac{75}{100} \times 30 = 22.5$$

$$k = 23$$

$$Y[23] = 48$$

$$Q_3 = 48$$

f) Interquartile range = $Q_3 - Q_1$
 $= 48 - 34$
 $= 14$

g) variance

$$= \frac{[(10-40.93)^2 + (15-40.93)^2 + (32-40.93)^2 + (32-40.93)^2 + (33-40.93)^2 + (33-40.93)^2 + (35-40.93)^2 + (35-40.93)^2 + (37-40.93)^2 + (38-40.93)^2 + (38-40.93)^2 + (39-40.93)^2 + (40-40.93)^2 + (41-40.93)^2 + (41-40.93)^2 + (44-40.93)^2 + (44-40.93)^2 + (45-40.93)^2 + (47-40.93)^2 + (48-40.93)^2 + (48-40.93)^2 + (52-40.93)^2 + (52-40.93)^2 + (53-40.93)^2 + (55-40.93)^2 + (56-40.93)^2 + (57-40.93)^2 + (60-40.93)^2]}{(30-1)}$$

$$= \frac{3631.82}{29}$$

$$= 125.24$$

$$h) \text{ standard deviation} = \sqrt{125.24}$$

$$= 11.19$$

i) Skewness

$$= \frac{[(10-40.93)^3 + (15-40.93)^3 + (32-40.93)^3 + (33-40.93)^3 + (33-40.93)^3 + (35-40.93)^3 + (35-40.93)^3 + (37-40.93)^3 + (38-40.93)^3 + (38-40.93)^3 + (39-40.93)^3 + (40-40.93)^3 + (41-40.93)^3 + (41-40.93)^3 + (44-40.93)^3 + (44-40.93)^3 + (45-40.93)^3 + (47-40.93)^3 + (48-40.93)^3 + (48-40.93)^3 + (52-40.93)^3 + (52-40.93)^3 + (53-40.93)^3 + (55-40.93)^3 + (56-40.93)^3 + (57-40.93)^3 + (60-40.93)^3]}{(30-1)(11.19)^3}$$

$$= \frac{-26449.39}{40633.88}$$

$$= -0.65$$

j) Kurtosis

$$= \frac{[(10-40.93)^4 + (15-40.93)^4 + (32-40.93)^4 + (33-40.93)^4 + (33-40.93)^4 + (35-40.93)^4 + (35-40.93)^4 + (37-40.93)^4 + (38-40.93)^4 + (38-40.93)^4 + (39-40.93)^4 + (40-40.93)^4 + (41-40.93)^4 + (41-40.93)^4 + (44-40.93)^4 + (44-40.93)^4 + (45-40.93)^4 + (47-40.93)^4 + (48-40.93)^4 + (48-40.93)^4 + (52-40.93)^4 + (52-40.93)^4 + (53-40.93)^4 + (55-40.93)^4 + (56-40.93)^4 + (57-40.93)^4 + (60-40.93)^4]}{(30-1)(11.19)^4}$$

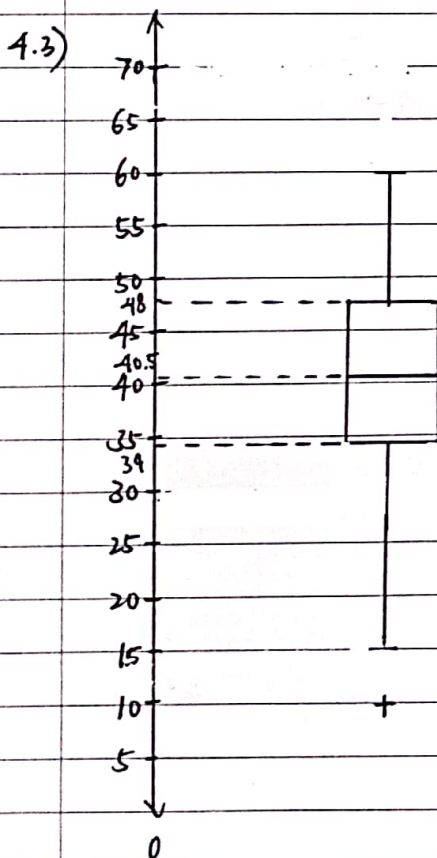
$$= \frac{1732198.77}{454693.08}$$

$$= 3.81$$

$$\begin{aligned} 4.2) \text{ Lower limit} &= Q_1 - 1.5 \times IQR \\ &= 34 - 1.5 \times 14 \\ &= 13 \end{aligned}$$

$$\begin{aligned} \text{Upper limit} &= Q_3 + 1.5 \times IQR \\ &= 48 + 1.5 \times 14 \\ &= 69 \end{aligned}$$

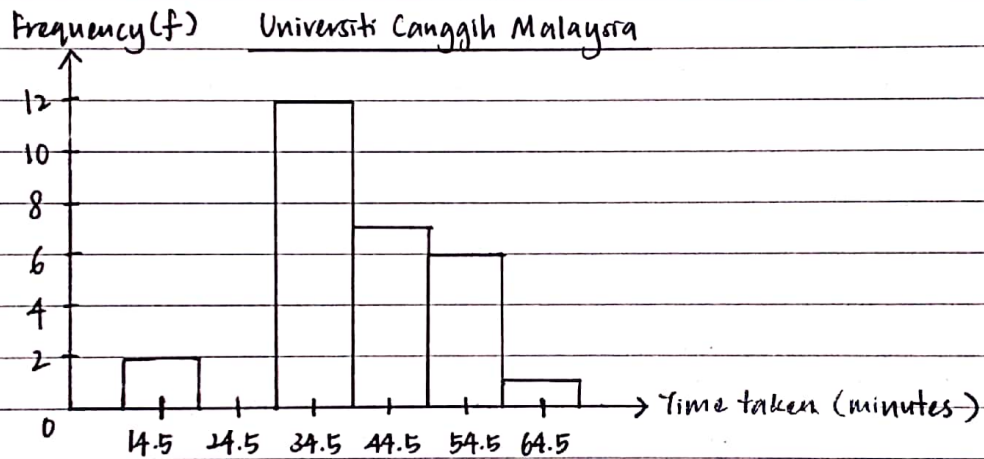
Yes. It has an outlier in this data sets which is 10. This is because the lower limit in this data set is 13 and value 10 is less than 13. Thus, 10 is an outlier in this data sets.



4.5) a) Histograms

b)

Time taken (minutes)	Midpoint	Frequency (f)
10-19	14.5	2
20-29	24.5	0
30-39	34.5	12
40-49	44.5	7
50-59	54.5	6
60-69	64.5	1
		$\Sigma f = 30$



- 4.4) a) The box plot of Universiti Muhibbah Malaysia is located which is near to the lower whisker. Thus, we conclude that it is positive skewed. This can be shown by the value of central tendency. In this data set, mode is smaller than the median and median is smaller than the mean. ($\text{mode} < \text{median} < \text{mean}$). While in the data set of Universiti Canggih Malaysia, the box plot is located near to the upper whisker. Thus, it is negative skewed. In this data set, mean is slightly higher than the median and median is smaller than mode. ($\text{mean} < \text{median} < \text{mode}$).
- c) The box plot of Universiti Muhibbah Malaysia is showed positive skewed. Median is located at the middle of the box and the it is near to the lower whisker. The box plot of Universiti Canggih Malaysia is showed negative skewed. It is located near to the upper whisker.

No:

Date:

University Muhibbah Malaysia

(4.1)

48	48	49	49	49	50	50	50	51	52
52	53	54	54	54	55	55	56	57	57
57	59	59	60	63	65	70	72	73	90

$$(a) \text{ Range} = 90 - 48$$

$$= 42$$

$$(b) \text{ mean} = \frac{2(48) + 3(49) + 3(50) + 51 + 52(2) + 53 + 3(54) + 2(55) + 56 + 3(57) + 2(59) + 60 + 63 + 65 + 70 + 72 + 73 + 90}{30}$$

$$= 57.03$$

$$(c) k = \frac{1}{2}(30)$$

$$= 15$$

$$\text{median} = \frac{Y[15] + Y[16]}{2}$$

$$= \frac{54 + 55}{2}$$

$$= 54.5$$

$$(d) \text{ mode} = 49, 50, 54, 57$$

$$(e) k = \frac{1}{4}(30)$$

$$= 7.5 \approx 8$$

$$\text{First quartile} = 50$$

$$(f) k = \frac{3}{4}(30)$$

$$= 22.5 \approx 23$$

$$\text{Third quartile} = 59$$

$$(g) \text{ interquartile range} = Q_3 - Q_1$$

$$= 59 - 50$$

$$= 9$$

No.

Date

$$\begin{aligned}
 (h) \quad s^2 &= (48-57.03)^2 + (48-57.03)^2 + (49-57.03)^2 + (49-57.03)^2 + (49-57.03)^2 \\
 &\quad + (50-57.03)^2 + (50-57.03)^2 + (50-57.03)^2 + (51-57.03)^2 + (52-57.03)^2 \\
 &\quad + (52-57.03)^2 + (53-57.03)^2 + (54-57.03)^2 + (54-57.03)^2 + (54-57.03)^2 \\
 &\quad + (55-57.03)^2 + (55-57.03)^2 + (56-57.03)^2 + (57-57.03)^2 + (57-57.03)^2 \\
 &\quad + (57-57.03)^2 + (59-57.03)^2 + (59-57.03)^2 + (60-57.03)^2 + (63-57.03)^2 \\
 &\quad + (65-57.03)^2 + (70-57.03)^2 + (72-57.03)^2 + (73-57.03)^2 + (90-57.03)^2 \\
 &\quad \quad \quad 30-1
 \end{aligned}$$

$$= \frac{356.52 + 209.92 + 69.08 + 9.30 + 52.22 + 1797.90}{29}$$

$$= 86.03$$

$$\begin{aligned}
 (i) \quad s &= \sqrt{s^2} \\
 &= \sqrt{86.03} \\
 &= 9.28
 \end{aligned}$$

$$\begin{aligned}
 (j) \quad \text{skewness} &= (48-57.03)^3 + (48-57.03)^3 + (49-57.03)^3 + (49-57.03)^3 + (49-57.03)^3 \\
 &\quad + (50-57.03)^3 + (50-57.03)^3 + (50-57.03)^3 + (51-57.03)^3 + (52-57.03)^3 \\
 &\quad + (52-57.03)^3 + (53-57.03)^3 + (54-57.03)^3 + (54-57.03)^3 + (54-57.03)^3 \\
 &\quad + (55-57.03)^3 + (55-57.03)^3 + (56-57.03)^3 + (57-57.03)^3 + (57-57.03)^3 \\
 &\quad + (57-57.03)^3 + (59-57.03)^3 + (59-57.03)^3 + (60-57.03)^3 + (63-57.03)^3 \\
 &\quad + (65-57.03)^3 + (70-57.03)^3 + (72-57.03)^3 + (73-57.03)^3 + (90-57.03)^3 \\
 &\quad \quad \quad (30-1)(9.28)^3
 \end{aligned}$$

$$\begin{aligned}
 &= (-3025.97) + (-1388.81) + (-276.17) + (-17.82) + 254.27 \\
 &\quad + 45954.96 \\
 &\quad \quad \quad (29)(9.28)^3
 \end{aligned}$$

$$\begin{aligned}
 &= 41500.46 \\
 &\quad 23176.18
 \end{aligned}$$

$$= 1.79$$

$$\begin{aligned}
 (k) \text{ kurtosis} &= \frac{(48-57.03)^4 + (48-57.03)^4 + (49-57.03)^4 + (49-57.03)^4 + (49-57.03)^4}{(30-1)(9.28)^4} \\
 &+ \frac{(50-57.03)^4 + (50-57.03)^4 + (50-57.03)^4 + (51-57.03)^4 + (52-57.03)^4}{(30-1)(9.28)^4} \\
 &+ \frac{(52-57.03)^4 + (53-57.03)^4 + (54-57.03)^4 + (54-57.03)^4 + (54-57.03)^4}{(30-1)(9.28)^4} \\
 &+ \frac{(55-57.03)^4 + (55-57.03)^4 + (56-57.03)^4 + (57-57.03)^4 + (57-57.03)^4}{(30-1)(9.28)^4} \\
 &+ \frac{(57-57.03)^4 + (59-57.03)^4 + (59-57.03)^4 + (60-57.03)^4 + (63-57.03)^4}{(30-1)(9.28)^4} \\
 &+ \frac{(65-57.03)^4 + (70-57.03)^4 + (72-57.03)^4 + (73-57.03)^4 + (90-57.03)^4}{(30-1)(9.28)^4}
 \end{aligned}$$

$$= \frac{25771.20 + 9289.53 + 1156.77 + 35.09 + 1378.20 + 1329214.69}{(29)(9.28)^4}$$

$$= \frac{1366845.48}{215074.99}$$

$$= 6.36$$

$$(4.2) \quad Q_1 = 50, Q_3 = 59$$

$$\text{lower limit} = Q_1 - 1.5 \times IQR$$

$$= 50 - 1.5 \times 9$$

$$= 36.5$$

$$\text{upper limit} = Q_3 + 1.5 \times IQR$$

$$= 59 + 1.5 \times 9$$

$$= 72.5$$

$\therefore$ Yes. When lower limit = 36.5, the lowest value from data is 48, $(36.5 < 48)$ no outlier here but when upper limit = 72.5, from the data, $73 > 72.5$ and $90 > 72.5$. Hence, has two outlier.

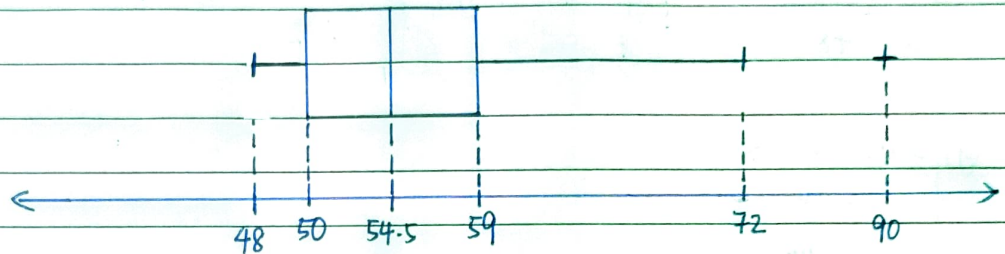
(4.3) maximum observation = 90

minimum observation = 48

$Q_1 = 50$

median = 54.5

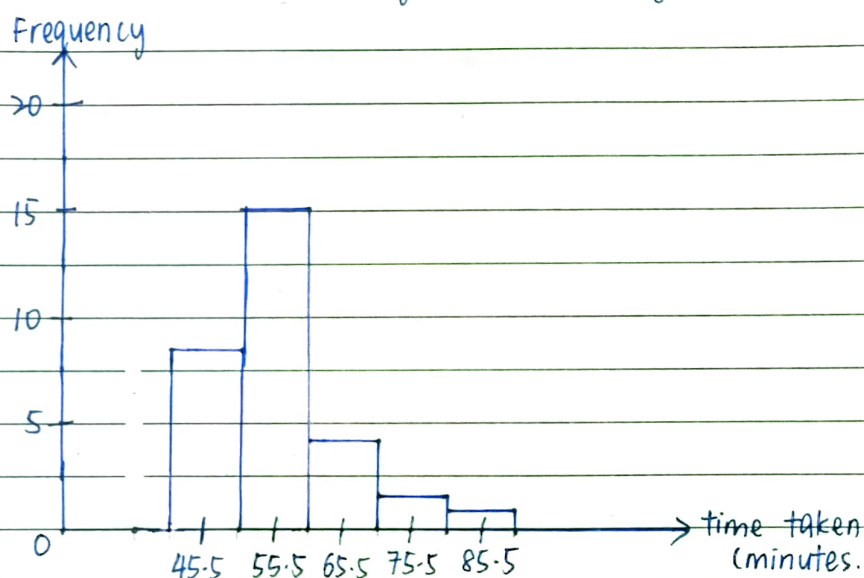
$Q_3 = 59$



4.4 b) The spread of the data can be observed through the range, variance and standard deviation. The range of University Canggih (50) is greater than University Muhibbah (45). The variance of University Canggih (125.24) is greater than University Muhibbah (86.03). The standard deviation of University Canggih is greater than University Muhibbah, which is $11.19 > 9.28$. Hence, the spread of University Canggih is larger than University Muhibbah.

4.5 b)	Cell boundaries	Midpoint	frequency
	41 - 50	45.5	8
	51 - 60	55.5	15
	61 - 70	65.5	4
	71 - 80	75.5	2
	81 - 90	85.5	1

University Muhibbah Malaysia



- 4.5c) - University Muhibbah Malaysia's histogram is skewed to the right. This means that histogram graph of university muhibbah is positively skewed. According to the data calculated for university muhibbah Malaysia, $\text{mean} < \text{median}$. Besides that, the data is highly concentrated on the left.
- University Canggih histogram skewed to the left. This means that the shape of the histogram graph of university canggih is negatively skewed. The histogram shows that data is more concentrated on the right.

4.6) Yes, the answering speed of students from both university has significant difference. According to the boxplot drawing of University Canggih, the boxplot is near to the upper whisker while boxplot drawing of University Muhibbah Malaysia, the boxplot is near to the lower whisker. Hence, the average writing speed of University Muhibbah is faster than University

No:

Date:

Canggih and this shows a significant difference. The size of the boxplot (IQR) for University Muhibbah is shorter than University Canggih. The boxplot for University Muhibbah has two outliers while boxplot for University Canggih does not have any outliers. This shows that data or the spread of university Canggih is more compact than University Muhibbah.