Group 3

1. Heng Xing Yu A21EC0183
2. Gan Jia Hui A21EC0177
3. Tan Kai Yuan A21EC0230
4. Sondos Tamer Mustafa A21EC0277

Question 1

1.1 In this study, I will explain descriptive and Inferential stat related to the study. For descriptive, it used to describe the features of the data such as histogram and pie chart. For inferential statistics, it is used to draw inferences of the population which is students in Malaysian and a sample which is Computer science students.

1.2 The population in this study is the students in Malaysia. The population of the study is the whole collection of the individuals in the study. The sample in this study is Computer science students in Malaysia which is a subset of the population in this study.

1.3 1) Understanding the nature of the problem

- In this step, we need to understand and the goal about study "Internet usage among Computer science students in Malaysia". We have a clear direction about this study before gathering data to avoid being unable to answer the question using the data collect.

2) Deciding what to measure and how to measure it.

- In this study, we decide to use time spent on the internet, purpose of using internet, the problem when using internet, use of internet services and satisfaction with internet facilites as our measure and question. We must define the question and studied for determining the values carefully.

3) Data collection

- We use Google form to generate the primary data. If a decision made to use secondary data, it is important to understand how the data collected.

4) Data summarization and preliminary analysis.

- In this study, we summarize the data graphically and numerically like histogram and pie chart. This analysis can provide guidance for further analysis.

5) Formal data analysis

- After step 4, we select and apply the appropriate inferential statistic methods. such as point estimation, interval estimation, hypothesis testing and prediction of data.

6) Interpretation of results.

- We can know the knowledge of Internet usage among Computer science students in Malaysia. Next, we can determine what problem or question about internet usage from the analysis. Lastly, our study can guide for futere research.

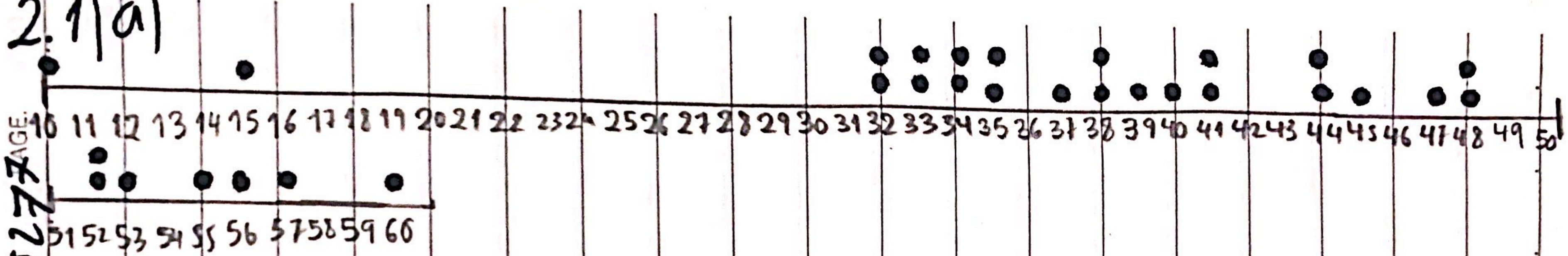
1.4 .Nominal → Gender

- Ordinal → satisfaction of the internet. connection.
- Interval → Temperature of device when using internet
- Ratio → Internet speed.

1.5. https://forms.gle/Ne3AdTMMsa6yt38v8

PSDA Assignment 1 Sondos Tamer Maarteh
Q2 A21EC0277

2.1 (a)



b)

Stem	leaf
1	0,5
3	2,2,3,3,5,5,7,8,8,9
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) Second fastest = 15

$$\text{Percentile} = \frac{\text{vals less than 15}}{\text{total no. of vals}} \times 100$$

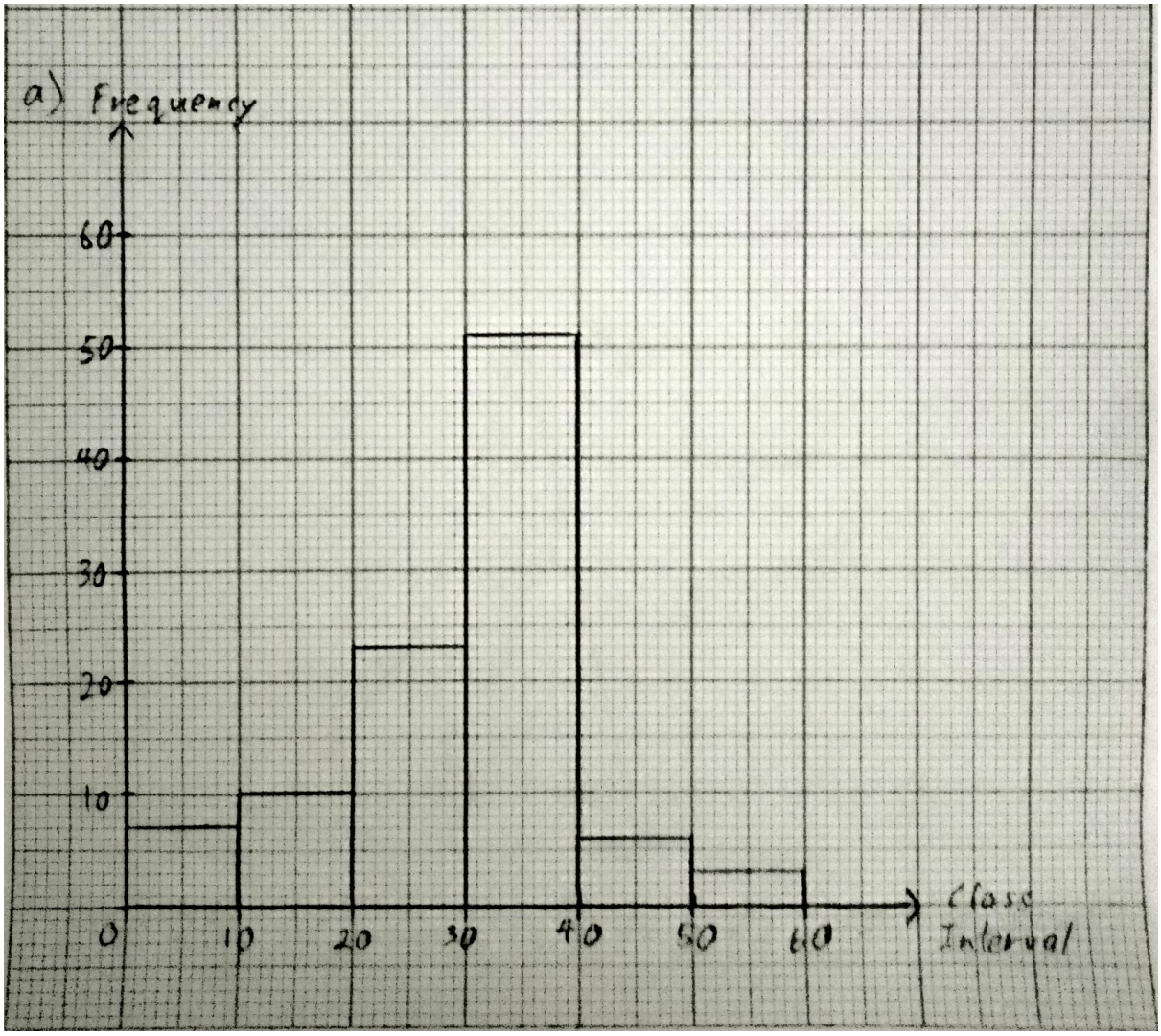
$$= \frac{1}{30} \times 100$$

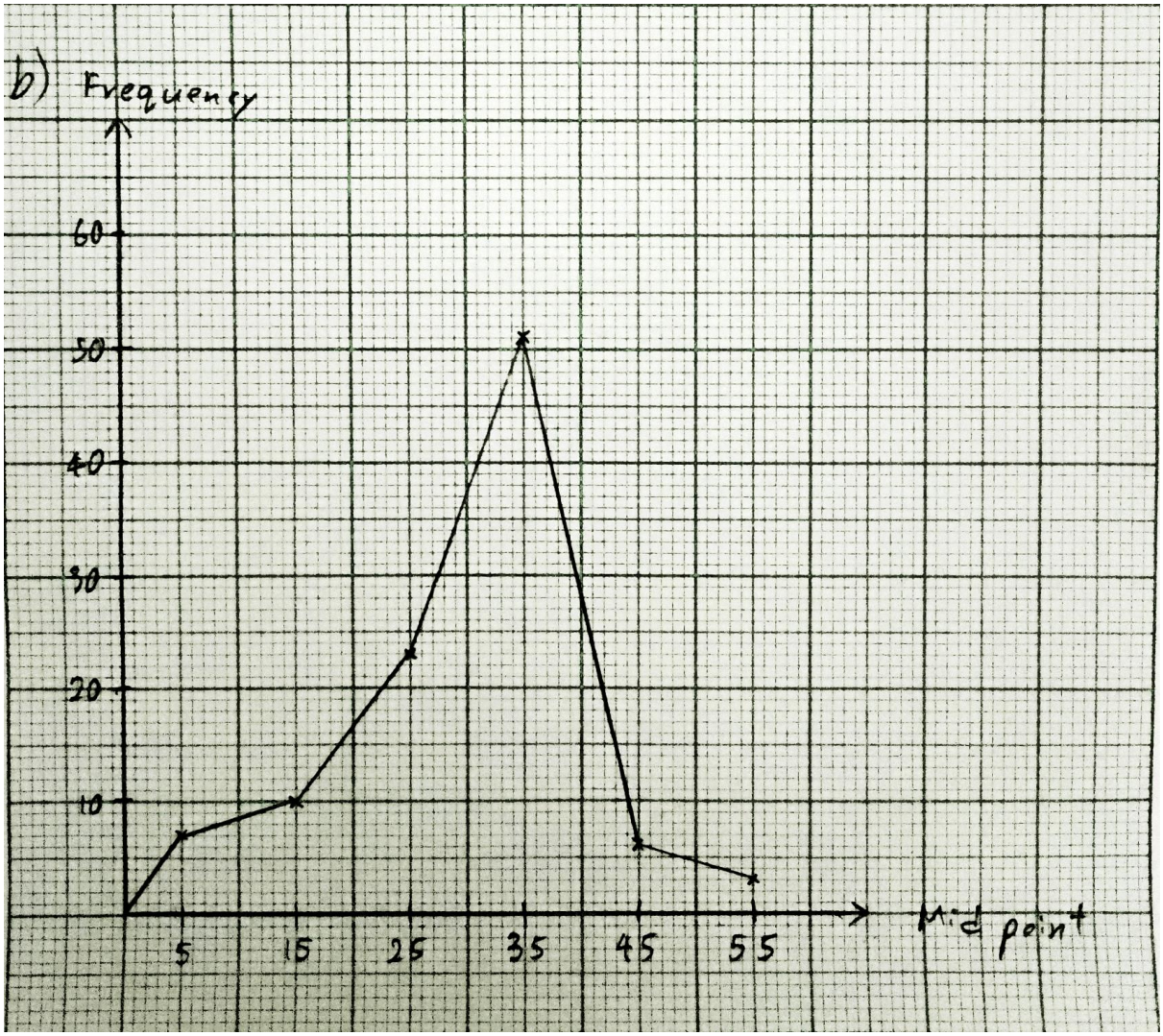
$$= 3.33 \text{ Percentile}$$

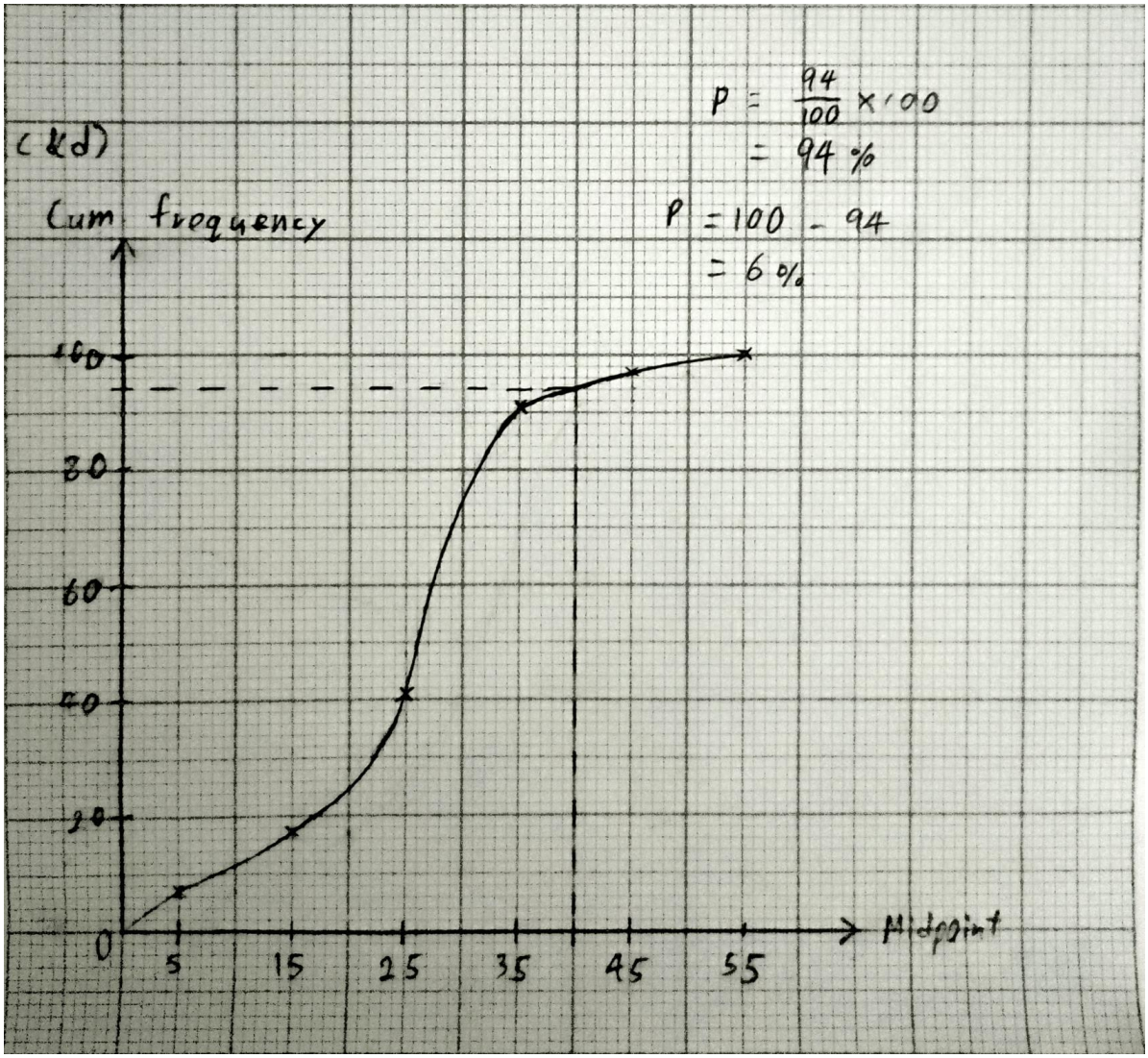
3.1)

Class Interval	Midpoint	Frequency, f	Cum. f	fx
0 ~ 10	5	7	7	35
10 ~ 20	15	10	17	150
20 ~ 30	25	23	40	575
30 ~ 40	35	51	91	1785
40 ~ 50	45	6	97	270
50 ~ 60	55	3	100	165
Total		100		2980

3.2)




b) Frequency
60
50
40
30
20
10
5
15
25
35
45
55
Mid point



3.3)

$$\bar{x} = \frac{2980}{100}$$
$$= 29.8$$

$$\text{Median} = 30 + \left(\frac{50 - 40}{51}\right) 10$$
$$= 31.96$$

$$\text{Mode} = 30 + \left(\frac{28}{28 + 45}\right) 10$$
$$= 33.84$$

④ First sample of data (in Q2)

first Q ↑ median ↓ third Q ↓

10, 15, 32, 32, 33, 38, 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 sample of data (in Q4)

median ↓ third Q ↓

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

↑ first quartile

4.1) ⓐ range (first sample) = 60 - 10

= 50

range (second sample) = 90 - 48

= 42

rangle of first sample is bigger than second sample.

4.2) ⓑ mean (first sample) = $\frac{1228}{30}$ = 40.93

mean (second sample) = $\frac{1711}{30}$

= 57.03

mean (first sample) is smaller than mean (second sample)

ⓒ position of median = $\frac{30+1}{2}$

= 15.5

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

= 40.5

∴ median (first sample is smaller than second sample)

median (second sample) = $\frac{54+55}{2}$

= 54.5

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

mode (second sample) = 49, 50, 54, 57

* first sample is UNIVERSITI CANGGIH
* Second sample is UNIVERSITI MUNISHAH MALAYSIA

ⓔ third Quartile (first sample) = 48

third Quartile (second sample) = 59

ⓕ First Quartile (first sample) = 34

First Quartile (second sample) = 50

Interquatile range (first sample) = 48 - 34

= 14

Interquartile range (second sample) = 59 - 50

= 9

g)

Sample 1

x	$(x-\bar{x})^2$
10	956.66
15	672.36
32	79.74
32	79.74
33	62.88
33	62.88
34	48.02
34	48.02
35	35.16
35	35.16
37	15.44
38	8.58
38	8.58
39	3.72
40	0.86
41	0.00
41	0.00
44	9.42
44	9.42
45	16.56
47	36.84
48	49.98
48	49.98
52	122.54
52	122.54
53	145.68
55	197.96
56	227.10
57	258.24
60	363.66
	3727.72

Sample 2

x	$(x-\bar{x})^2$
48	81.54
48	81.54
49	64.48
49	64.48
49	64.48
50	49.42
50	49.42
50	49.42
51	36.36
52	25.30
52	25.30
53	16.24
54	9.18
54	9.18
54	9.18
55	4.12
55	4.12
56	1.06
57	0.00
57	0.00
57	0.00
59	3.88
59	3.88
60	8.82
63	25.64
65	63.52
70	168.22
72	224.10
73	255.04
90	1087.02
	2494.94

variance (first sample) $= \frac{3727.72}{30-1}$

$= 128.54$

variance (second sample) $= \frac{2494.94}{30-1}$

$= 86.03$

h) standard deviation (first sample)

$= \sqrt{128.54}$

$= 11.34$

standard deviation (second sample)

$= \sqrt{86.03}$

$= 9.28$

Sample 1

X	$(x-\bar{x})^3$
10	-29589.65
15	-17434.42
32	-712.12
32	-712.12
33	-498.68
33	-498.68
34	-332.81
34	-332.81
35	-208.53
35	-208.53
37	-40.70
38	-25.15
38	-25.15
39	-7.19
40	-0.80
41	0.00
41	0.00
44	28.93
44	28.93
45	67.42
47	223.65
48	353.39
48	353.39
52	1356.57
52	1356.57
53	1758.42
55	2786.37
56	3422.47
57	4150.00
60	6935.09
	-27827.14

Sample 2

X	$(x-\bar{x})^3$
48	-736.31
48	-736.31
49	-517.78
49	-517.78
49	-517.78
50	-347.43
50	-347.43
50	-347.43
51	-219.26
52	-127.26
52	-127.26
53	-65.45
54	-27.82
54	-27.82
54	-27.82
55 → 55	-8.37 ← -8.37
56	-1.09
57	-0.00
57	-0.00
57	-0.00
59	7.65
59	7.65
60	26.20
63	212.78
65	506.26
70	2181.83
72	3354.79
73	4013.00
90	35839.08
	41500.47

i) skewness (first sample)

$$= \frac{-27827.14}{(30-1)(11.34)^3}$$

$$= -0.66$$

(negative skew)

left skew

skewness (second sample)

$$= \frac{41500.47}{(30-1)(9.28)^3}$$

$$= 1.79$$

(positive skew)

right skew

j)

sample 1

X	$(X-\bar{X})^4$
10	915207.73
15	452074.56
32	6359.25
32	6359.25
33	3954.51
33	3954.51
34	2306.39
34	2306.39
35	1236.57
35	1236.57
37	238.54
38	73.70
38	73.70
39	13.87
40	0.75
41	0.00
41	0.00
44	88.83
44	88.83
45	274.40
47	1357.55
48	2498.49
48	2498.49
52	15017.25
52	15017.25
53	21224.09
55	39190.10
56	51576.64
57	66690.43
60	132252.16
	1743170.8

Kurtosis (first sample)

$$= \frac{1743170.8}{(30-1)(11.34)^4}$$

$$= 3.63 \text{ (leptokurtic)}$$

sample 2

X	$(X-\bar{X})^4$
48	6648.92
48	6648.92
49	4157.79
49	4157.79
49	4157.79
50	2442.43
50	2442.43
50	2442.43
51	1322.12
52	640.14
52	640.14
53	263.77
54	84.29
54	84.29
54	84.29
55	16.98
58→ 56	1.13 ←16.98
57	0.00
57	0.00
57	0.00
59	15.06
59	15.06
60	77.81
63	1270.27
66	4034.90
70	28298.27
72	50221.21
73	65045.86
90	1181614.44
	1366872.24

Kurtosis (second sample)

$$= \frac{1366872.24}{(30-1)(9.28)^4}$$

$$= 6.36 \text{ (leptokurtic)}$$

4.2)

First Sample

Lower Limit = 34 − (1.5 × 14)

= 13

Upper Limit = 48 + (1.5 × 14)

= 69

Outlier = 10

Second Sample

Lower Limit = 50 − (1.5 × 9)

= 36.5

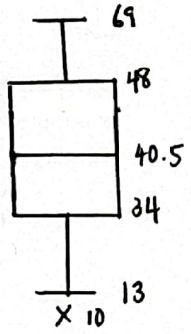
Upper Limit = 59 + (1.5 × 9)

= 72.5

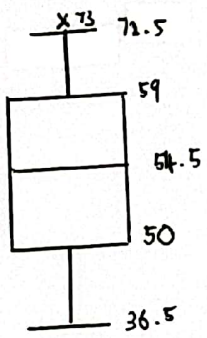
Outlier = 73, 90

4.3)

First sample Modified boxplot



Second Sample Modified boxplot



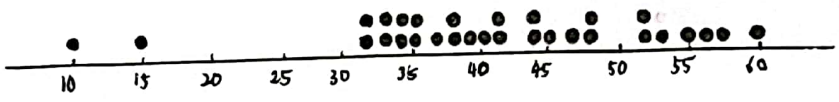
4.4) a) - mean of sample 1 is smaller than sample 2
- median of sample 1 is smaller than sample 2
- mode of sample 1 has 9 and mode of sample 2 has 4.

b) - range of sample 1 is bigger than sample 2
- variance and standard deviation of sample 1 bigger than sample 2

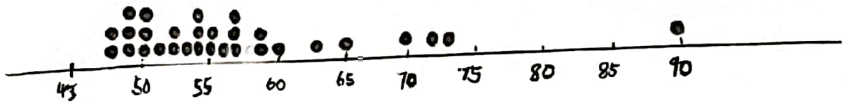
c) - skewness of sample 1 is skewed to left but skewness of sample 2 is skewed to right
- kurtosis of sample 1 and sample 2 is leptokurtic.

4.5) a) Graphical method : dotplots

b) **Sample 1**



Sample 2



4.5) d) - Sample 2 is right skewed and sample 1 is slightly left skewed because the skewness value of sample 1 is near to 0.

- Both sample 1 and sample 2 is leptokurtic.

4.6) - The speed of students from UNIVERSITI CANGGIH is faster than student from UNIVERSITI MUHIBBAH MALAYSIA because the mean of sample 1 is smaller than sample 2. Median also one of the reason. Median of sample 1 also smaller than sample 2.

- Next, the lower limit and upper limit of sample 1 is smaller than the lower limit and upper limit of sample 2. It represent the speed of Universiti Canggih is faster. (using shorter time)