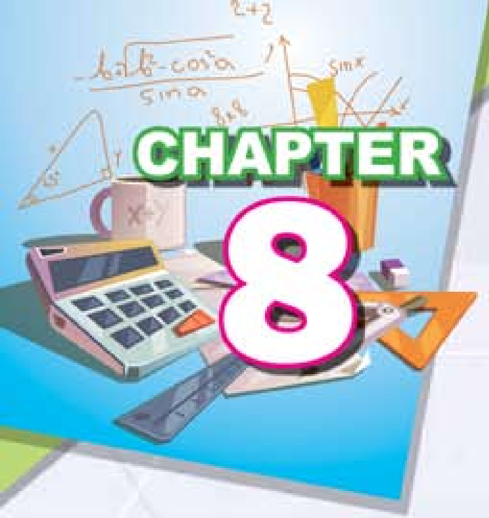


CHAPTER 8

Measures of Dispersion for Ungrouped Data

You will learn

- Dispersion
- Measures of Dispersion

Kuala Lumpur's 29th SEA games were officially held from 19 to 30 August 2017. There were at least 4 646 athletes participated in 404 events. Malaysia became the winner with 145 gold medals.

Do you know how many medals Malaysian athletes had gained according to the events being contested?

Why Study This Chapter?

Statistics help us conduct research effectively, improve on critical and analytical thinking skills, as well as enable us to make the right decisions based on data. Therefore, mastering statistical knowledge enhances our skills in making improvements and data-driven predictions. These skills are extremely important in our career and daily life.

ES

	GOLD	SILVER	BRONZE	TOTAL
AYSIA	145	92	86	323
LAND	72	86	88	246
NAM	58	50	60	168
GAPORE	57	58	73	188
ONESIA	38	[illegible]	[illegible]	[illegible]
IPPINES	24			
NMAR	7			
BODIA	3			
S	[illegible]			
NEI				

WORD BANK

ungrouped data	*data tak terkumpul*
interquartile range	*julat antara kuartil*
range	*julat*
outlier	*pencilan*
stem-and-leaf plot	*plot batang-dan-daun*
box plot	*plot kotak*
dot plot	*plot titik*
standard deviation	*sisihan piawai*
measure of dispersion	*sukatan serakan*
variance	*varians*

Walking Through Time

Sir Ronald Fisher Aylmer
(1890 – 1962)

Sir Ronald Fisher Aylmer had made important contributions to the field of statistics, including pioneering the development of analysis of variance. He is also known as "a genius who created the foundation for modern statistical science".

http://bt.sasbadi.com/m4211

8.1 Dispersion

What is the meaning of dispersion?

Measures of dispersion are important measurements in statistics. The measures of dispersion give us an idea of how the values of a set of data are scattered.

Dispersion is small if the data set has a small range and vice versa.

Learning Standard

Explain the meaning of dispersion.

Measures of dispersion of a set of data are quantitative measures such as range, interquartile range, variance and standard deviation.

Mind Stimulation 1

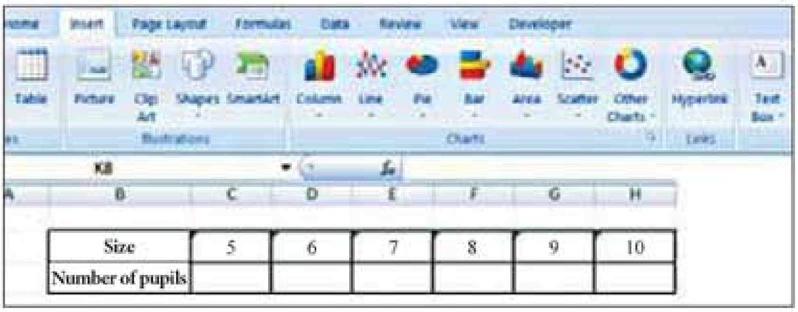
Aim: To explain the meaning of dispersion.

Steps:

1. Divide the class into groups.
2. Using the interview method, collect information about the shoe sizes of your friends in the school. Collect at least 20 data.
3. Using an electronic spreadsheet, construct a table for the information obtained.

Size of shoes	5	6	7	8	9	10
Number of pupils						

4. Complete the following frequency table in the electronic spreadsheet.



Size	5	6	7	8	9	10
Number of pupils						

MY MEMORY

To represent data ethically and avoid confusion
- The scale used in the representations must be uniformed and must start from 0.
- The data displayed must be accurate.

5. Using the same electronic spreadsheet or manual, construct the following charts.
   (a) dot plot (scatter plot) (edit your displayed results if necessary)
   (b) stem-and-leaf plot
6. Print and display your group's results at the Mathematics Corner.

7. During Gallery Walk activity, obtain information from other groups and complete the table below.

Group	1	2	3	4	5	6	7
Largest shoe size							
Smallest shoe size							
Difference in shoe size							

Discussion:

Based on your group's frequency table in Step 7,

(a) state whether differences in shoe size for each group are the same.

(b) discuss the factors that cause these differences in shoe size.

From the activity in Mind Stimulation 1, it is found that:

> The distributions of the data are different. To understand the dispersion of data, the difference between the largest value and the smallest value is taken into consideration. If the difference between the values is large, it indicates that the data is widely dispersed and vice versa.

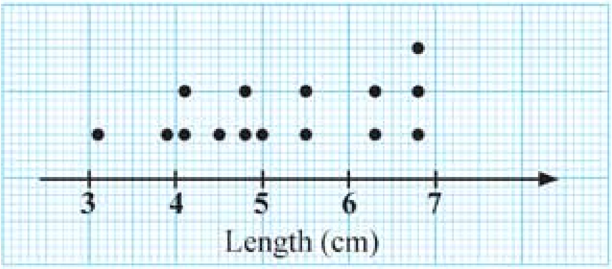
Example 1

(a) The table below shows the masses, in kg, of 20 pupils.

52	60	62	47	55	48	70	67	71	66
48	50	64	51	66	79	62	65	78	72

State the difference in mass, in kg, of the pupils.

(b) The diagram below shows a dot plot of the length, in cm, of a sample of several types of insects.



INFO ZONE

A value is a data or results obtained from an observation.

State the difference in size, in cm, of the longest insect and the shortest insect.

Solution:

(a) Largest mass = 79 kg
Smallest mass = 47 kg
Difference in mass = 79 kg – 47 kg
= 32 kg

(b) Length of the longest insect = 6.8 cm
Length of the shortest insect = 3.1 cm
Difference in length = 6.8 cm – 3.1 cm
= 3.7 cm

Self Practice 8.1a

1. A study about the boiling points, in °C, of 10 types of chemicals is conducted. The results of the study are recorded as below.

112	60	75	81	150	92	108	90	45	132

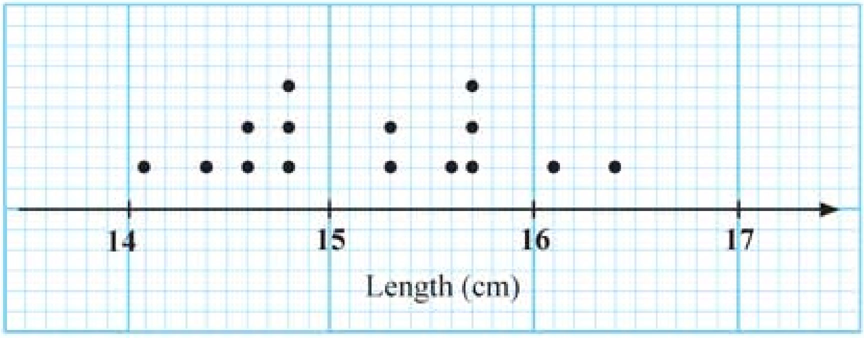
(a) Determine the lowest and the highest temperatures.
(b) Determine the difference between the highest and the lowest temperatures.

2. The durations spent on social media, in minutes, of 12 teenagers on a particular day are recorded as follows.

100	120	80	60	90	30	40	100	60	90	120	60

It is found that the duration spent on social media can be written in the form of $p \leqslant \text{time} \leqslant q$. State the values of p and q.

3. The diagram shows a dot plot of the lengths, in cm, of samples of several leaves.



What is the difference in length between the longest and the shortest leaf?

How do you compare and interpret dispersion of two or more sets of data based on stem-and-leaf plots and dot plots?

Learning Standard

Compare and interpret dispersion of two or more sets of data based on the stem-and-leaf plots and dot plots, and hence make conclusion.

A stem-and-leaf plot is a way to show the distributions of a set of data. Through the stem-and-leaf plot, we can see whether the data is more likely to appear or least likely to appear.

What are the steps to plot a stem-and-leaf plot?

The diagram below shows the marks obtained by a group of 36 pupils in an Accounting test.

27	34	37	39	42	43	46	48	52
29	35	37	40	42	44	47	49	52
31	35	38	40	42	44	47	49	53
32	36	38	41	42	45	47	52	54

If this data is not organised in the table, then we cannot see the dispersion immediately. We set the tens digit as the stem and the units digit as the leaf to plot the stem-and-leaf plot for the given data.

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

Key: 2 | 7 means 27 marks

INFO ZONE

Leaf digits are arranged from the smallest value to the largest value.

From the stem-and-leaf plot, we find that the marks which appear the most are in the range of 40 to 49.

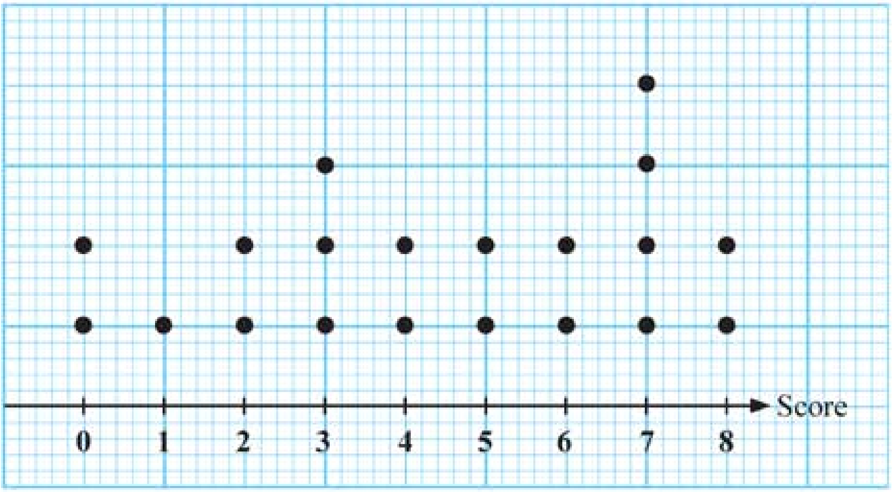
What do you understand about a dot plot?

A dot plot is a statistical chart that contains points plotted using a uniform scale. Each point represents a value.

The following data shows the scores obtained by a group of 20 pupils in a Biology quiz.

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

The diagram below shows a dot plot for the above data. Each point represents a value.



INFO ZONE

The dot plot is appropriate for small data set.

In the above dot plot, the highest value is 8, the lowest value is 0 and most pupils get a score of 7.

How do you compare and interpret the dispersion of two sets of data based on a stem-and-leaf plot?

Mind Stimulation 2

Aim: To compare and interpret dispersion of two sets of data based on a stem-and-leaf plot.

Steps:

1. Divide the class into groups.
2. The following data shows the marks scored by Class 4 Budi pupils in April and May monthly History tests.

Marks of April monthly test

32	41	44	51	58
35	42	46	53	58
35	43	48	54	58
39	43	48	54	60
41	44	49	56	61

Marks of May monthly test

34	46	55	63	69
38	46	55	65	71
40	49	55	66	73
40	52	59	68	75
43	53	59	68	77

3. Complete the following stem-and-leaf plot.

Marks of April monthly test		Marks of May monthly test
9 5 5 2	3	4 8
	4	
	5	
	6	
	7	

Discussion:

Based on the above stem-and-leaf plot, which test shows better pupils' achievement? Justify your answer.

From the activity in Mind Stimulation 2, it is found that:

The pupils' achievement in May monthly test is better because there are more pupils scoring higher marks in the test compared to April monthly test.

In general,

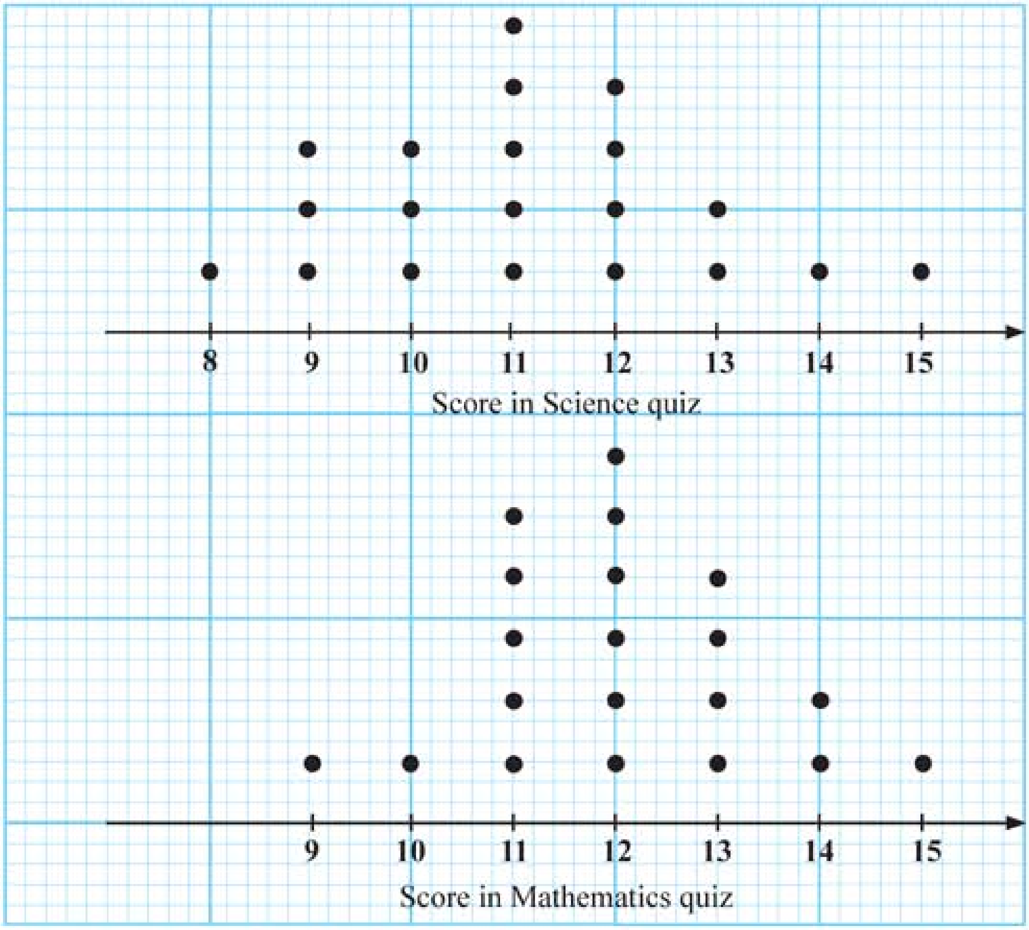
When two sets of data are plotted using stem-and-leaf plots, we can compare the patterns of the two plots.

How do you compare and interpret two sets of data based on dot plots?

From the dot plots, we can compare and interpret the shapes, the dispersion of points, the frequencies as well as the values concentrated on the left or on the right of the two sets of data.

Example 2

A Mathematics quiz and a Science quiz are held at SMK Bestari in conjunction of the Mathematics and Science Week. 20 pupils take part in the quizzes. The diagram below shows two dot plots of the scores obtained by the pupils.



(a) Which set of data shows a wider dispersion? Justify your answer.
(b) Which quiz has a higher difference in value?
(c) Between the scores of Science and Mathematics quizzes, which pupils' achievement is better?

Solution:

(a) The dispersion in Science quiz is larger because it has larger difference in value.
(b) Science quiz has a higher difference in value, that is 7.
(c) The pupils score better in Mathematics quiz because most of the values are concentrated on the right of the dot plot.

Self Practice 8.1b

1. The data below shows the masses, in kg, of two groups of pupils.

Group A

58	47	68	63	61
60	54	70	63	45
69	54	52	41	82
70	53	70	60	52
81	67	56	50	76
86	66	62	73	75
44	46	62	72	49
84	76	82	68	64

Group B

72	54	76	49	64
58	76	40	64	65
58	69	48	52	42
40	70	66	57	83
41	70	69	56	61
78	52	75	63	84
46	82	55	59	57
83	42	56	60	59

Draw a stem-and-leaf plot to show the distribution of masses of the two groups of pupils. Comment on the distribution of masses of the two groups of pupils.

2. The data below shows the marks obtained by a group of 25 pupils in two monthly Geography tests.

April monthly test

35	56	42	56	48
65	51	58	42	60
46	61	46	48	62
54	50	41	50	55
50	69	57	51	50

May monthly test

63	42	47	52	52
48	45	51	41	50
55	48	61	42	49
39	63	54	56	54
46	50	51	57	53

Draw two dot plots using the same scale to show the difference of the distribution of marks of the two monthly tests. Comment on the achievement of the group of pupils in the two monthly tests.

3. The data below shows the shoe size of 20 pupils in two different classes.

Class Rose

8.0	8.0	12.5	7.5	8.0
6.5	6.0	8.0	11.0	8.5
7.0	7.5	10.5	8.0	7.0
8.0	10.5	7.5	6.5	6.5

Class Lotus

8.5	9.0	9.5	10.5	9.0
9.5	11.0	9.0	9.5	9.5
10.5	9.5	8.5	9.0	9.0
9.0	9.0	10.5	9.5	11.0

(a) Construct two dot plots using the same scale.

(b) Compare the two dot plots obtained in (a). Which class has a greater difference in shoe size? Justify your answer.

8.2 Measures of Dispersion

How do you determine the range, interquartile range, variance and standard deviation as a measure to describe dispersion of an ungrouped data?

Learning Standard

Determine the range, interquartile range, variance and standard deviation as a measure to describe dispersion of an ungrouped data.

In statistics, the range of an ungrouped data is the difference between the largest and the smallest values. This measure shows how data is distributed. For example, in a Bahasa Melayu test, a class that has a larger range of marks means that pupils' achievement varies greatly.

Range = Largest value – Smallest value

Example 3

Given a set of data 34, 23, 14, 26, 40, 25, 20, determine the range of this set of data.

Solution:

34, 23, $\boxed{14}$, 26, $\boxed{40}$, 25, 20

Smallest value

Largest value

Range = 40 – 14
= 26

How do you determine the range from the frequency table?

Example 4

The table below shows the scores obtained by a group of pupils in a Chemistry test.

Score	2	3	4	5	6	7	8
Number of pupils	3	5	6	8	12	7	3

Determine the range of the above data.

Solution:

Smallest value

Largest value

Score	$\boxed{2}$	3	4	5	6	7	$\boxed{8}$
Number of pupils	3	5	6	8	12	7	3

Range = 8 – 2
= 6

How do you determine the interquartile range of a set of ungrouped data?

When the values of a set of data are arranged in an ascending order, the first quartile, Q_1 is the value of data that is at the first $\frac{1}{4}$ position while the third quartile, Q_3 is the value of data that is at the $\frac{3}{4}$ position from the data arranged in order.

Use the following steps to determine Q_1 and Q_3 from a set of ungrouped data.

1. Arrange the data in ascending order.
2. Determine the median of the data.
3. Divide the set of data into two parts, the part of data before the median and the part of data after the median.
4. Q_1 is the value of data in the middle position of the part of data before the median.
5. Q_3 is the value of data in the middle position of the part of data after the median.
6. If the number of data is odd, the median, Q_1 and Q_3 can be identified directly from the the data.
7. It the number of data is even, the median, Q_1 and Q_3 can be obtained by finding the average of the two values in the middle positions of the parts of data before or after the median.
8. Interquartile range = $Q_3 - Q_1$

Example 5

Determine the interquartile range of each of the following sets of data:

(a) 4, 4, 6, 8, 10, 13, 15, 16, 20, 22, 23
(b) 11, 14, 14, 17, 19, 20, 20, 21, 22
(c) 2.3, 2.5, 2.5, 2.6, 2.8, 2.9, 3.0, 3.2

INTERACTIVE ZONE

Data can also be arranged in a descending order. In this case, how do you obtain Q_1 and Q_3? Discuss.

Solution:

(a)
$$\underbrace{4 \quad 4}_{\text{2 values}} \quad \underset{\underset{Q_1}{\downarrow}}{6} \quad \underbrace{8 \quad 10}_{\text{2 values}} \quad \underset{\underset{\text{median}}{\downarrow}}{13} \quad \underbrace{15 \quad 16}_{\text{2 values}} \quad \underset{\underset{Q_3}{\downarrow}}{20} \quad \underbrace{22 \quad 23}_{\text{2 values}}$$

$$\begin{aligned}\text{Interquartile range} &= 20 - 6 \\ &= 14\end{aligned}$$

Determine the position of the median first.

MY MEMORY

Median is the value in the middle when a set of data is arranged in an ascending or descending order.

(b)
$$\underbrace{11 \quad 14}_{\text{2 values}} \underset{\underset{Q_1}{\downarrow}}{\,} \underbrace{14 \quad 17}_{\text{2 values}} \quad \underset{\underset{\text{median}}{\downarrow}}{19} \quad \underbrace{20 \quad 20}_{\text{2 values}} \underset{\underset{Q_3}{\downarrow}}{\,} \underbrace{21 \quad 22}_{\text{2 values}}$$

$$\begin{aligned}\text{Interquartile range} &= \frac{20 + 21}{2} - \frac{14 + 14}{2} \\ &= 6.5\end{aligned}$$

INFO ZONE

Second quartile, Q_2 is also the median.

(c)
$$\underbrace{2.3 \quad 2.5}_{\text{2 values}} \underset{\underset{Q_1}{\downarrow}}{\,} \underbrace{2.5 \quad 2.6}_{\text{2 values}} \underset{\underset{\text{median}}{\downarrow}}{\,} \underbrace{2.8 \quad 2.9}_{\text{2 values}} \underset{\underset{Q_3}{\downarrow}}{\,} \underbrace{3.0 \quad 3.2}_{\text{2 values}}$$

$$\begin{aligned}\text{Interquartile range} &= \frac{2.9 + 3.0}{2} - \frac{2.5 + 2.5}{2} \\ &= 0.45\end{aligned}$$

INFO ZONE

First quartile, Q_1 is also known as lower quartile while third quartile, Q_3 is known as upper quartile.

Example 6

Determine the interquartile range of the following frequency table.

Score	2	3	4	5	6	7	8
Number of pupils	3	5	6	8	12	7	3

Solution:

11th value of score

33rd value of score

Score	2	3	4	5	6	7	8
Number of pupils	3	5	6	8	12	7	3
Cumulative frequency	3	8	14	22	34	41	44

$$Q_1 = \text{the } \left(\frac{1}{4} \times 44\right) \text{th value}$$
$$= \text{the 11th value}$$
$$= 4$$

$$Q_3 = \text{the } \left(\frac{3}{4} \times 44\right) \text{th value}$$
$$= \text{the 33rd value}$$
$$= 6$$

Interquartile range = 6 – 4
= 2

INFO ZONE

Cumulative frequency of a given data is determined by adding its frequency and all the frequencies before it.

What do you understand about variance and standard deviation?

Variance and standard deviation are the measures of dispersion commonly used in statistics. The variance is the average of the square of the difference between each data and the mean.

The standard deviation is the square root of the variance which also measures the dispersion of a data set relative to its mean; measured in the same units of the original data.

How do you determine the variance and standard deviation for a set of ungrouped data?

Variance of a set of ungrouped data can be obtained using the formula of variance.

$$\sigma^2 = \frac{\Sigma(x - \bar{x})^2}{N} \quad \text{or} \quad \sigma^2 = \frac{\Sigma x^2}{N} - \bar{x}^2$$

Standard deviation, $\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N}}$ or $\sigma = \sqrt{\frac{\Sigma x^2}{N} - \bar{x}^2}$

MY MEMORY

$\bar{x}$ = mean of data

$$\bar{x} = \frac{\Sigma x}{N}$$

Example 7

Determine the variance of the set of data 2, 4, 5, 5, 6.

Solution:

$$\text{Mean, } \bar{x} = \frac{2 + 4 + 5 + 5 + 6}{5}$$
$$= 4.4$$

$$\text{Variance, } \sigma^2 = \frac{(2-4.4)^2 + (4-4.4)^2 + (5-4.4)^2 + (5-4.4)^2 + (6-4.4)^2}{5}$$
$$= 1.84$$

Alternative Method

$$\sigma^2 = \frac{2^2 + 4^2 + 5^2 + 5^2 + 6^2}{5} - 4.4^2$$
$$= 1.84$$

Example 8

Determine the standard deviation of the set of data 5, 7, 8, 8, 10, 13, 15, 16, 16, 20.

Solution:

$$\text{Mean, } \bar{x} = \frac{5 + 7 + 8 + 8 + 10 + 13 + 15 + 16 + 16 + 20}{10}$$
$$= 11.8$$

$$\text{Variance, } \sigma^2 = \frac{5^2 + 7^2 + 8^2 + 8^2 + 10^2 + 13^2 + 15^2 + 16^2 + 16^2 + 20^2}{10} - 11.8^2$$
$$= 21.56$$

$$\text{Standard deviation, } \sigma = \sqrt{21.56}$$
$$= 4.643$$

INFO ZONE

σ^2 = variance
σ = standard deviation

Example 9

The table below shows the number of books read by a group of pupils in a particular month.

Number of books	0	1	2	3	4
Number of pupils	3	5	8	2	2

Calculate the variance and standard deviation for the number of books read.

Solution:

$$\text{Mean, } \bar{x} = \frac{3(0) + 5(1) + 8(2) + 2(3) + 2(4)}{3 + 5 + 8 + 2 + 2}$$
$$= 1.75$$

$$\text{Variance, } \sigma^2 = \frac{3(0-1.75)^2 + 5(1-1.75)^2 + 8(2-1.75)^2 + 2(3-1.75)^2 + 2(4-1.75)^2}{20}$$
$$= 1.2875$$

Alternative Method 1

$$\sigma^2 = \frac{3(0)^2 + 5(1)^2 + 8(2)^2 + 2(3)^2 + 2(4)^2}{20} - 1.75^2$$
$$= 1.2875$$

INFO ZONE

If the ungrouped data given in a frequency table, then

$\bar{x} = \dfrac{\Sigma fx}{\Sigma f}$ and

$\sigma^2 = \dfrac{\Sigma f(x-\bar{x})^2}{\Sigma f}$ or

$\sigma^2 = \dfrac{\Sigma fx^2}{\Sigma f} - \bar{x}^2$

Alternative Method 2

x	f	fx	x^2	fx^2
0	3	0	0	0
1	5	5	1	5
2	8	16	4	32
3	2	6	9	18
4	2	8	16	32
	$\sum f = 20$	$\sum fx = 35$		$\sum fx^2 = 87$

$$\text{Mean, } \bar{x} = \frac{35}{20}$$
$$= 1.75$$

Variance, σ^2

$$= \frac{87}{20} - 1.75^2$$
$$= 1.2875$$

Standard deviation, $\sigma = \sqrt{1.2875}$

$$= 1.1347$$

Checking Answer

1. Press the Mode key 2 times
   Display SD REG BASE
   1 2 3
2. Press 1 to select SD
3. Enter the data
   Press 0, followed by M+ 3 times
   Press 1, followed by M+ 5 times
   Press 2, followed by M+ 8 times
   Press 3, followed by M+ 2 times
   Press 4, followed by M+ 2 times
   Display n =
   20
4. Press shift 2
   Display $\bar{x}$ $x\sigma n$ $x\sigma n\text{-}1$
   1 2 3
5. Press 2, then =
   Display 1.134680572
6. Press button x^2 , then =
   Display 1.2875

Self Practice 8.2a

1. Determine the range and interquartile range of each of the following sets of data.
   (a) 3, 7, 5, 9, 4, 4, 8, 7, 6, 2, 5
   (b) 13, 15, 19, 22, 17, 14, 15, 16, 18, 19, 11, 10, 20
   (c) 2.3, 2.2, 3.1, 2.8, 2.7, 2.4, 2.5, 2.3

2. Determine the range and interquartile range of each of the following sets of data.

(a)

Pocket money (RM)	4	5	6	7	8	9
Number of pupils	5	10	7	6	3	1

(b)

Score	10	11	12	13	14	15
Number of pupils	3	10	13	4	8	6

3. For each of the following sets of data, determine the variance and standard deviation.
   (a) 5, 7, 6, 9, 12, 10, 10, 13
   (b) 32, 40, 35, 39, 44, 48, 42

4. The table below shows the number of goals scored by 20 football players in a match.

Number of goals	0	1	2	3	4
Number of players	3	5	6	5	1

Determine the variance and standard deviation of the distribution.

What are the advantages and disadvantages of various measures of dispersion?

Learning Standard

Explain the advantages and disadvantages of various measures of dispersion to describe ungrouped data.

Measures of dispersion measure the distribution of a set of data. The range is the measure of dispersion easiest to calculate. However, the range cannot provide a good overview of how data is distributed.

In cases where there is an outlier or extreme value, the interquartile range would be the more appropriate measure of dispersion to show the distribution of the data.

Standard deviation is usually used to compare two sets of data. In general, a low standard deviation indicates that the data is dispersed close to the mean while a high standard deviation indicates that the data is dispersed far from the mean.

Example 10

Calculate the range and interquartile range of the set of data 10, 11, 11, 14, 18, 20, 21, 21, 25 and 40. Determine the most appropriate measure of dispersion to be used, to measure the distribution of the data set.

Solution:

$$\begin{aligned}\text{Range} &= 40 - 10\\ &= 30\end{aligned}$$

$$\begin{aligned}\text{Interquartile range} &= 21 - 11\\ &= 10\end{aligned}$$

Interquartile range is the most appropriate measure of dispersion because the existence of the outlier, 40.

Example 11

The table below shows the achievements of two pupils in 5 Physics tests.

	Test 1	Test 2	Test 3	Test 4	Test 5
Aiman	40	70	90	85	64
John	80	65	73	58	73

One of the pupils will be selected to represent the school in a Physics quiz competition. Determine who is eligible to be selected and justify your answer.

Solution:

Aiman

$$\text{Mean marks} = \frac{40 + 70 + 90 + 85 + 64}{5} = 69.8$$

$$\text{Standard deviation, } \sigma = \sqrt{\frac{40^2 + 70^2 + 90^2 + 85^2 + 64^2}{5} - 69.8^2} = 17.67$$

John

$$\text{Mean marks} = \frac{80 + 65 + 73 + 58 + 73}{5} = 69.8$$

$$\text{Standard deviation, } \sigma = \sqrt{\frac{80^2 + 65^2 + 73^2 + 58^2 + 73^2}{5} - 69.8^2}$$
$$= 7.574$$

John will be selected to participate in the Physics quiz competition. The standard deviation of John's scores is lower because John's achievement is more consistent.

Checking Answer

1. Press the Mode key 2 times.
   Display SD REG BASE
   1 2 3
2. Press 1 to select SD
3. Enter the data
   Press 80, followed by M+
   Display $n = 1$
   Press 65, followed by M+
   Display $n = 2$
   Press 73, followed by M+
   Display $n = 3$
   Press 58, followed by M+
   Display $n = 4$
   Press 73, followed by M+
   Display $n = 5$
4. Press Shift 2
   Display $\bar{x}$ $x\sigma n$ $x\sigma n\text{-}1$
   1 2 3
5. Press 2
6. Press =
   Display 7.573638492

Self Practice 8.2b

1. Calculate the range and interquartile range of the set of data 2, 16, 16, 19, 20, 20, 24, 26, 27, 27, 29. Determine which measure of dispersion is more appropriate to be used to measure the distribution of the set of data. Justify your answer.

2. Mr Rakesh wants to choose one of his two pupils to represent his school in a National Business Quiz competition. Mr Rakesh gave 5 tests to the two pupils. The following are the results of the 5 tests.

	Test 1	**Test 2**	**Test 3**	**Test 4**	**Test 5**
Pupil A	32	61	75	82	90
Pupil B	50	67	70	73	80

Determine the appropriate measure of dispersion to be used by Mr Rakesh to make the selection. State who Mr Rakesh should select based on the measures of dispersion calculated.

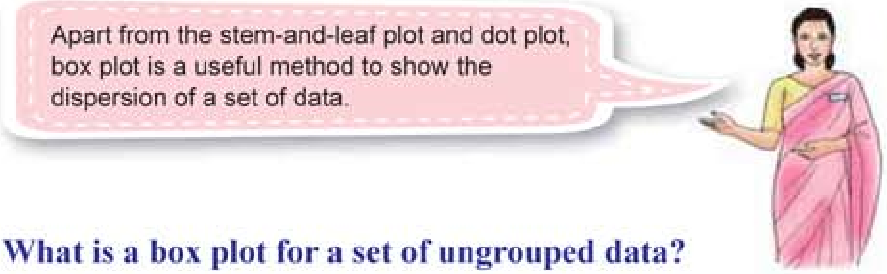
3. The table below shows the monthly salaries of seven employees of Texan Company.

Employee	1	2	3	4	5	6	7
Salary (RM)	900	920	950	1 000	1 100	1 230	3 000

(a) Calculate the range, interquartile range and standard deviation of the monthly salaries of the employees in Texan Company.

(b) State the most appropriate measure of dispersion to be used in order to show the distribution of the monthly salaries of the employees in Texan Company.

How do you construct and interpret the box plot for a set of ungrouped data?



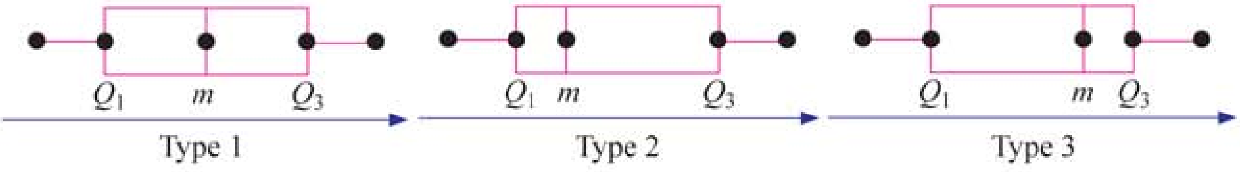
Learning Standard

Construct and interpret the box plot for a set of ungrouped data.

What is a box plot for a set of ungrouped data?

Box plot is a way of showing the distribution of a set of data based on five values, namely the minimum value, first quartile, median, third quartile and the maximum value of the set of data. The box plot can show whether the set of data is symmetric about the median. The box plot is often used to analyse a large number of data.

Some of the types of box plots are as follows.



Type 1 Type 2 Type 3

How do you construct a box plot?

The box plot is constructed on a number line.
Consider the data below.

The data below shows the marks obtained by 15 pupils in a Chinese Language test.

35 40 42 45 52 52 53 57 62 62 66 68 73 73 75

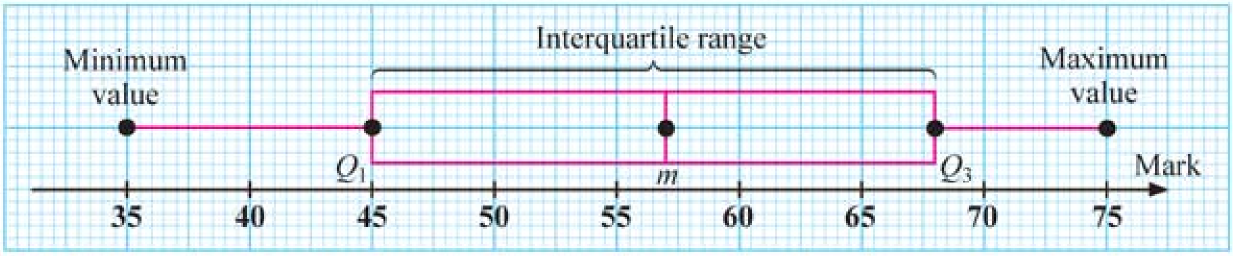
For the set of data, determine the values shown in the table below first.

Minimum value	35
First quartile	45
Median	57
Third quartile	68
Maximum value	75

INFO ZONE

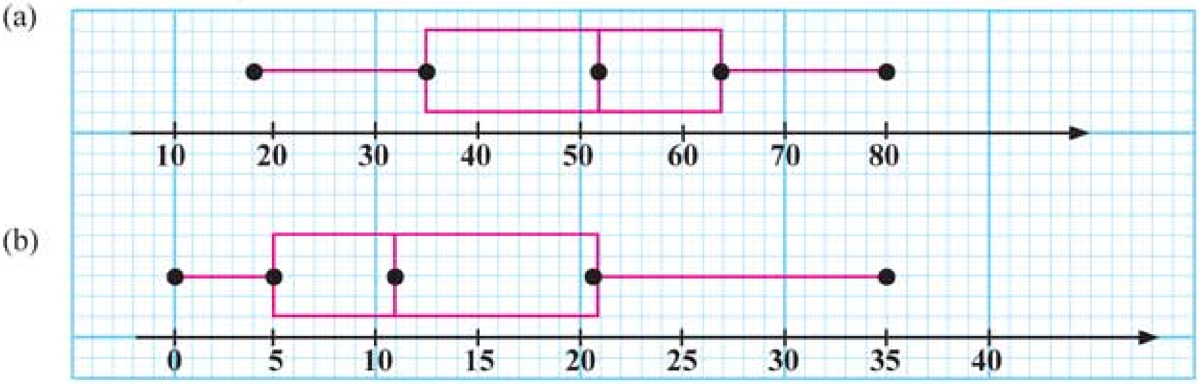
The box plot can also be constructed vertically.

Thus, the box plot for the above set of data is as follows.



Example 12

Two box plots are given below.



For each box plot, state

(i) the median. (ii) the first quartile. (iii) the third quartile.
(iv) the interquartile range. (v) the minimum value. (vi) the maximum value.
(vii) the range.

Solution:

(a) (i) median = 52
(ii) first quartile = 35
(iii) third quartile = 64
(iv) interquartile range = 29
(v) minimum value = 18
(vi) maximum value = 80
(vii) range = 62

(b) (i) median = 11
(ii) first quartile = 5
(iii) third quartile = 21
(iv) interquartile range = 16
(v) minimum value = 0
(vi) maximum value = 35
(vii) range = 35

Self Practice 8.2c

1. For each of the following sets of data, construct a box plot.

(a)

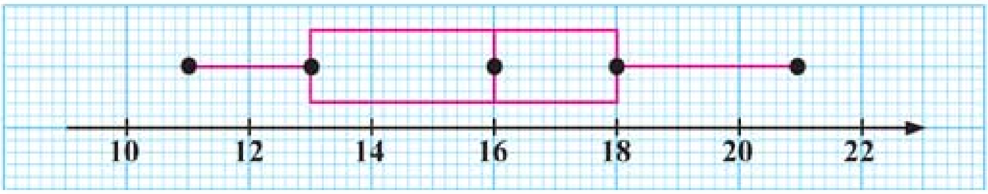
40	43	40	35	44	25
38	46	64	48	40	31
35	35	43	34	30	40

(b)

70	52	43	62	56	66	52
63	51	48	43	56	52	40
65	50	70	55	43	57	70

2. For the box plot below, state

(a) the minimum value. (b) the maximum value. (c) the first quartile.
(d) the third quartile. (e) the interquartile range. (f) the median.



What is the effect of data changes on dispersion?

The data changes include the following cases.

(a) Each data is added or subtracted by a constant uniformly.
(b) Each data is multiplied or divided by a constant uniformly.
(c) The existence of outliers or extreme values in the set of data.
(d) Certain values are removed from the set of data or added into the set of data.

Learning Standard

Determine the effect of data changes on dispersion based on:
- the value of measure of dispersion
- graphical representation

What are the effects on the measures of dispersion when each data is changed by adding or subtracting a constant?

Consider this case.
Initially, I have a set of data 5, 8, 3, 4, 1.
If each value of the set of data is added by 5, then the data above will be 10, 13, 8, 9, 6. What will happen to the measures of dispersion such as range, interquartile range, variance and standard deviation for the new set of data?

Mind Stimulation 3

Aim: To identify the effect on the range, interquartile range and standard deviation when each value of a set of data is added or subtracted by a constant.

Steps:

1. Divide the class into groups.
2. Each group is given cards A1, A2 and A3. (Refer to the QR Code)
3. Each group is required to complete the information stated on the card.

Discussion:

What is the effect on the measures of dispersion when each value of the set of data is added or subtracted by a constant?

Scan the QR Code to download card A1, card A2 and card A3.
http://bt.sasbadi.com/m4228

From the activity in Mind Stimulation 3, it is found that:

There are no changes to the measures of dispersion when each value of a set of data is added or subtracted by a constant.

What are the effects on the measures of dispersion when each value of a set of data is multiplied or divided by a constant?

Consider this case.
Initially, I have a set of data 5, 8, 3, 4, 1.
If each value of the set of data is multiplied by 2, then the data above will be 10, 16, 6, 8, 2. What will happen to the measures of dispersion such as range, interquartile range, variance and standard deviation for the new set of data?

Mind Stimulation 4

Aim: To identify the effect on the range, interquartile range and standard deviation when each value of a set of data is multiplied or divided by a number.

Steps:

1. Divide the class into groups.
2. Each group is given cards B1, B2 and B3. (Refer to the QR Code).
3. Each group is required to complete the information stated on the card.

Discussion:

What are the effects on the measures of dispersion when each value of the set of data is multiplied by 3? What are the effects on the measures of dispersion when each value of the set of data is divided by 2?

Scan the QR Code to download card B1, card B2 and card B3.
http://bt.sasbadi.com/m4229

From the activity in Mind Stimulation 4, it is found that:

(a) When each value of a set of data is multiplied by a constant k, then
  (i) new range $= k \times$ original range
  (ii) new interquartile range $= k \times$ original interquartile range
  (iii) new standard deviation $= k \times$ original standard deviation
  (iv) new variance $= k^2 \times$ original variance
(b) When each value of a set of data is divided by a constant k, then
  (i) new range $= \dfrac{\text{original range}}{k}$
  (ii) new interquartile range $= \dfrac{\text{original interquartile range}}{k}$
  (iii) new standard deviation $= \dfrac{\text{original standard deviation}}{k}$
  (iv) new variance $= \dfrac{\text{original variance}}{k^2}$

What are the effects on the measures of dispersion when an outlier is added or removed from the set of data?

(a) **Range**
The range will change dramatically when an outlier is added or removed from the set of data.

(b) **Interquartile range**
The interquartile range is less affected when an outlier is added or removed from the set of data.

(c) **Variance and standard deviation**
The variance and the standard deviation will increase significantly when an outlier is added into the set of data.

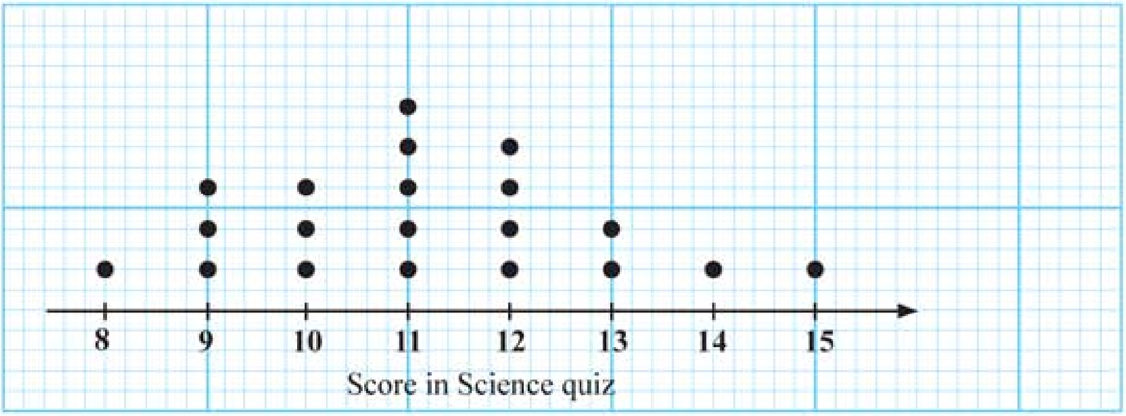
If the difference between the new value of the data and the mean is small, then the new standard deviation will be smaller and vice versa.

If the difference between the value that is removed and the mean is small, then the new standard deviation will be larger and vice versa.

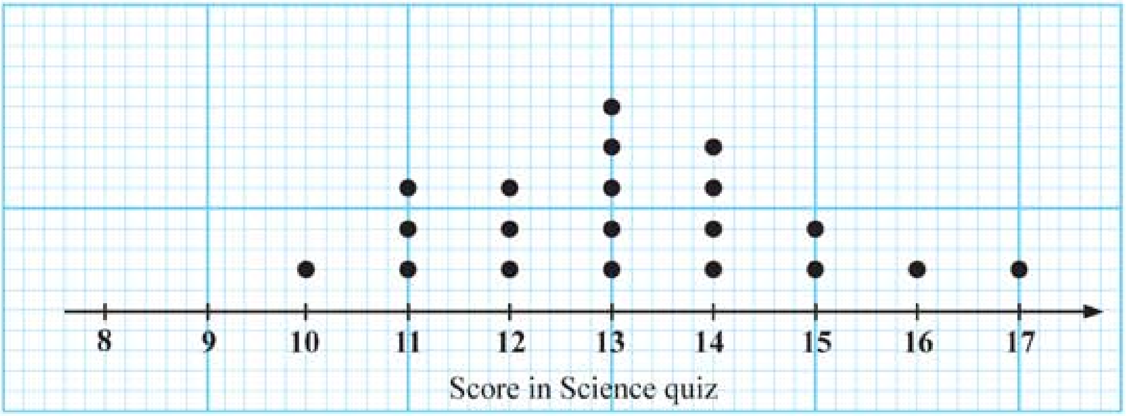
How do you determine the effect of data changes on dispersion based on graphical representation?

The effect of data changes can be graphically displayed by using dot plot, stem-and-leaf plot or box plot.

(a) The data is changed uniformly by adding or subtracting a constant.



When each value of a set of data is added by 2



In the above diagram, when each value of a set of data is added by 2, it is found that:

(i) The whole dot plot is shifted to the right by 2 units where the dispersion remains unchanged.

(ii) Original score range = 15 – 8
= 7

New score range = 17 – 10
= 7

The range of both sets of the data is the same.

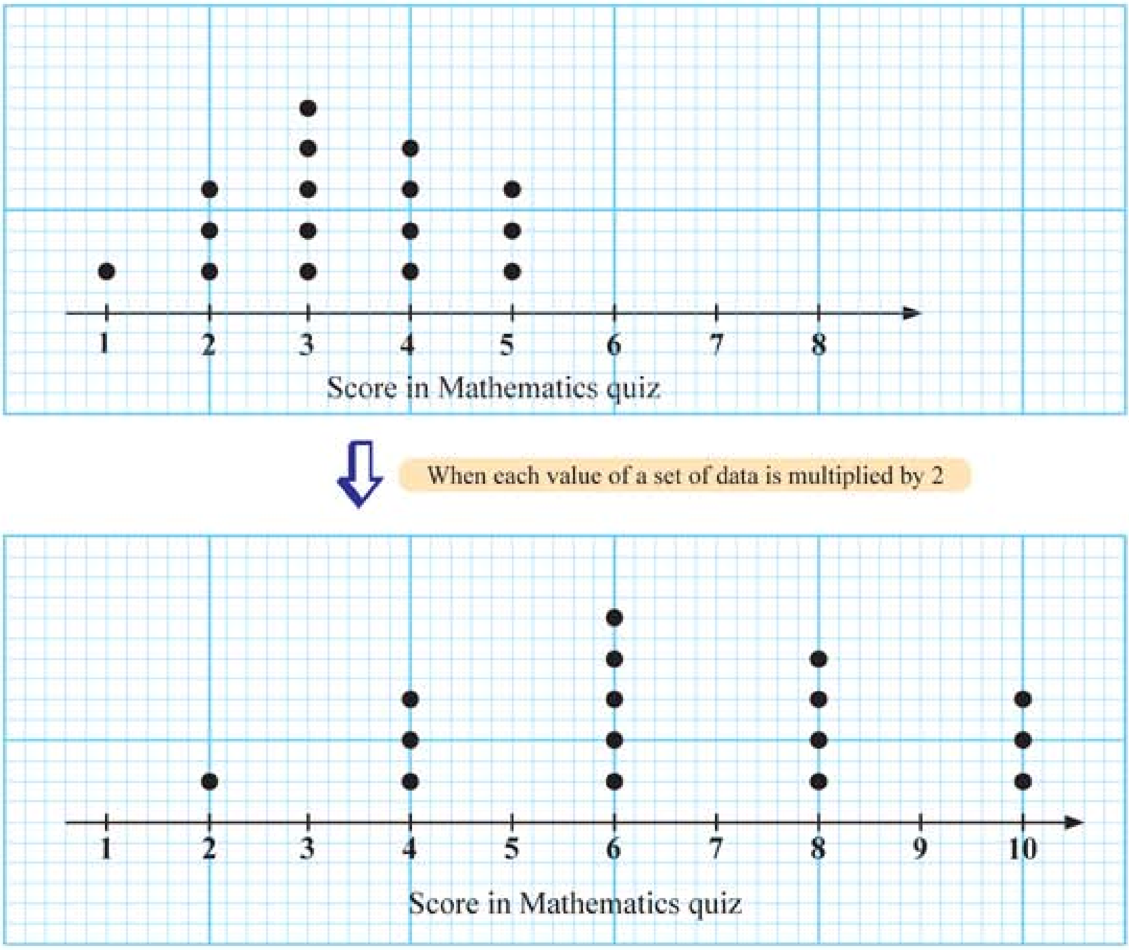
In general,

When each value of a set of data is added by a constant, the measures of dispersion remain unchanged.

INTERACTIVE ZONE

Discuss what will happen to the measures of dispersion if each value of a set of data is subtracted by a constant.

(b) The data is changed uniformly by multiplying or dividing by a constant.



In the above diagram, when each value of a set of data is multiplied by 2, it is found that:

(i) The dispersion of the whole dot plot becomes twice as wide.

(ii) Original score range = 5 – 1
= 4

New score range = 10 – 2
= 8

The range of the new set of data is two times the range of the original set of data.

When each value of a set of data is multiplied by a constant, k,

New range = k × original range

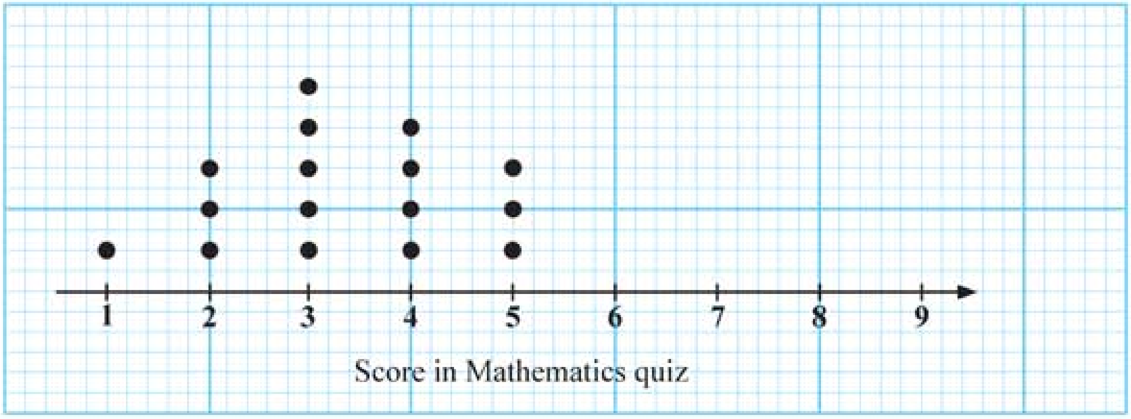
INTERACTIVE ZONE

Discuss what will happen to the measures of dispersion if each value of a set of data is divided by a constant.

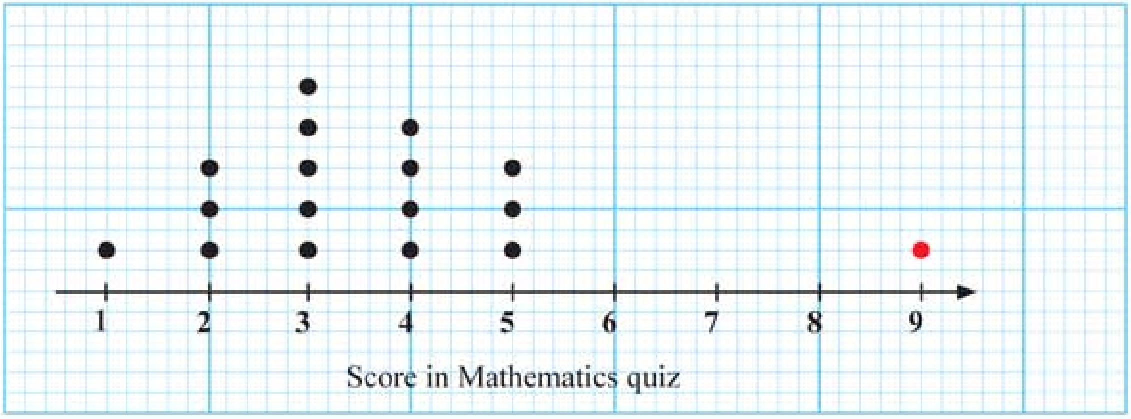
In general,

When each value of a set of data is multiplied by a constant, k, the measures of dispersion will change.

(c) The value added or removed is an outlier.



When an outlier 9 is added



In the set above diagram, when an outlier is added to a set of data, it is found that:

(i) The dispersion of the whole dot plot becomes wider.

(ii) Original score range = 5 – 1
= 4

New score range = 9 – 1
= 8

The range of the new set of data becomes larger compared to the range of the original set of data when an outlier is added.

In general,

The measures of dispersion of the new set of data become larger compared to the measures of dispersion of the original set of data when an outlier is added.

INTERACTIVE ZONE

Discuss what will happen to the range when an outlier is removed from the set of data.

INTERACTIVE ZONE

Discuss what will happen to the measures of dispersion when an outlier is removed from the set of data.

Example 13

Given a set of data 4, 6, 7, 7, 9, 11, 12, 12, calculate the standard deviation of the set of data. Hence, calculate the new standard deviation when

(a) 4 is removed. (b) 7 is added. (c) 18 is added.

Solution:

$$\text{Mean} = \frac{4+6+7+7+9+11+12+12}{8}$$
$$= 8.5$$

$$\text{Standard deviation, } \sigma = \sqrt{\frac{4^2+6^2+7^2+7^2+9^2+11^2+12^2+12^2}{8} - 8.5^2}$$
$$= 2.784$$

(a) $$\text{New mean} = \frac{6+7+7+9+11+12+12}{7}$$
$$= 9.143$$

$$\text{New standard deviation, } \sigma = \sqrt{\frac{6^2+7^2+7^2+9^2+11^2+12^2+12^2}{7} - 9.143^2}$$
$$= 2.356$$

When the value that has a greater difference from mean is removed, the new standard deviation will be smaller.

(b) $$\text{New mean} = \frac{4+6+7+7+7+9+11+12+12}{9}$$
$$= 8.333$$

$$\text{New standard deviation, } \sigma = \sqrt{\frac{4^2+6^2+7^2+7^2+7^2+9^2+11^2+12^2+12^2}{9} - 8.333^2}$$
$$= 2.668$$

When the value that is close to the mean is added, the new standard deviation will be smaller.

(c) $$\text{New mean} = \frac{4+6+7+7+9+11+12+12+18}{9}$$
$$= 9.556$$

$$\text{New standard deviation, } \sigma = \sqrt{\frac{4^2+6^2+7^2+7^2+9^2+11^2+12^2+12^2+18^2}{9} - 9.556^2}$$
$$= 3.974$$

When the value that has a greater difference from mean is added, the new standard deviation will be larger.

Self Practice 8.2d

1. The interquartile range and the standard deviation of a set of data are 6 and 3.2 respectively. Calculate the new interquartile range and the new standard deviation if each value of the set of data is added by 4.

2. The interquartile range and the standard deviation of a set of data are 3 and 1.8 respectively. Calculate the new interquartile range and the new variance if each value of the set of data is multiplied by 3.

3. Given a set of data 4, 6, 7, 8, 10, 12, 16, calculate the standard deviation of the set of data. Explain the change on the standard deviation if 70 is added to the set of data.

4. (a) Determine the standard deviation of the set of data 3, 4, 7, 9, 10.
   (b) Use the answer in (a) to determine the new standard deviation of each of the following sets of data.
      (i) 6, 8, 14, 18, 20
      (ii) 1.5, 2, 3.5, 4.5, 5

5. The range and the variance of a set of data are 25 and 2.7 respectively. If each value of the data is added by 2 then multiplied by 4, determine the new range and the new variance.

6. The interquartile range and the standard deviation of a set of data are 2.7 and 3.6 respectively. If each value of the data is divided by 3 then subtracted by 2, determine the new interquartile range and the new standard deviation.

How do you compare and interpret two or more sets of ungrouped data based on the appropriate measures of dispersion?

The measures of central tendency may not give us enough information to do comparisons. We need to determine the measures of dispersion of a set of data that can give an impression about the features of the data. Thus, the conclusion made is valid and acceptable.

Learning Standard

Compare and interpret two or more sets of ungrouped data, based on the appropriate measures of dispersion, and hence make conclusion.

Example 14

The table below shows the time taken by 20 employees in two factories to complete a job.

Time (hours)	5	6	7	8	9
Factory A	3	2	9	2	4
Factory B	1	5	7	5	2

Calculate the mean and the standard deviation for the time taken to complete the job. Hence, determine which factory's employees are more efficient to complete the job. Justify your answer.

Solution:

Factory A

$$\text{Mean, } \bar{x} = \frac{(5 \times 3) + (6 \times 2) + (7 \times 9) + (8 \times 2) + (9 \times 4)}{3 + 2 + 9 + 2 + 4}$$
$$= 7.1$$

$$\text{Standard deviation, } \sigma = \sqrt{\frac{(5^2 \times 3) + (6^2 \times 2) + (7^2 \times 9) + (8^2 \times 2) + (9^2 \times 4)}{3 + 2 + 9 + 2 + 4} - 7.1^2}$$
$$= 1.261$$

Factory B

$$\text{Mean, } x = \frac{(5 \times 1) + (6 \times 5) + (7 \times 7) + (8 \times 5) + (9 \times 2)}{1 + 5 + 7 + 5 + 2}$$
$$= 7.1$$

$$\text{Standard deviation, } \sigma = \sqrt{\frac{(5^2 \times 1) + (6^2 \times 5) + (7^2 \times 7) + (8^2 \times 5) + (9^2 \times 2)}{1 + 5 + 7 + 5 + 2} - 7.1^2}$$
$$= 1.044$$

The employees in Factory B are more efficient because its standard deviation is smaller.

Self Practice 8.2e

1. The table below shows the record of time in seconds, for a running event of 100 metres in 5 trials of two school athletes.

Athlete A	12.78	12.97	12.56	12.34	13
Athlete B	12.01	13.03	12.98	12.84	12.79

Using the appropriate measures of dispersion, determine which athlete is more consistent in his achievement.

2. The table below shows a study conducted on the effects of two types of fertilisers to the amount of tomatoes yield, in kg, for 10 tomato trees respectively.

Fertiliser A	Fertiliser B
12, 18, 25, 30, 36, 36, 40, 42, 50, 54	25, 28, 30, 32, 32, 38, 40, 40, 42, 45

Using the measures of dispersion, determine which fertiliser is more effective in improving the yield of tomatoes.

How do you solve problems involving measures of dispersion?

Learning Standard

Solve problems involving measures of dispersion.

Example 15

The table below shows the information of the masses of two groups of pupils.

Group	Number of pupils	Mean	Variance
A	18	52	2.5
B	12	56	1.8

All the pupils from group A and group B are combined. Determine the standard deviation of the mass of the combined groups of pupils.

Solution:

Understanding the problem

Determine the standard deviation of the mass of the combined groups of pupils.

Planning a strategy

Identify Σx and Σx^2 of the two groups of pupils and then calculate Σx and Σx^2 of the combined groups of pupils.

Implementing the strategy

Group A

$$\bar{x} = \frac{\Sigma x}{N}$$
$$52 = \frac{\Sigma x}{18}$$
$$\Sigma x = 936$$

$$\sigma^2 = \frac{\Sigma x^2}{N} - \bar{x}^2$$
$$2.5 = \frac{\Sigma x^2}{18} - 52^2$$
$$\Sigma x^2 = 48\,717$$

Group B

$$\bar{x} = \frac{\Sigma x}{N}$$
$$56 = \frac{\Sigma x}{12}$$
$$\Sigma x = 672$$

$$\sigma^2 = \frac{\Sigma x^2}{N} - \bar{x}^2$$
$$1.8 = \frac{\Sigma x^2}{12} - 56^2$$
$$\Sigma x^2 = 37\,653.6$$

Total of $\Sigma x = 936 + 672 = 1\,608$

Total of $\Sigma x^2 = 48\,717 + 37\,653.6 = 86\,370.6$

Mean, $\bar{x} = \frac{1\,608}{30} = 53.6$

$$\text{Standard deviation, } \sigma = \sqrt{\frac{86\,370.6}{30} - 53.6^2}$$
$$= \sqrt{6.06}$$
$$= 2.462$$

Conclusion

The new standard deviation is 2.462.

Self Practice 8.2f

1. The table below shows the scores obtained by seven participants in a contest. The scores are arranged in an ascending order.

Participant	A	B	C	D	E	F	G
Score	10	h	12	14	17	k	23

(a) The interquartile range and the mean score obtained are 7 and 15 respectively. Calculate the values of h and k.

(b) Calculate the standard deviation of the scores obtained by the participants.

2. A set of data contains 20 numbers. The mean and the standard deviation of the numbers are 9 and 2 respectively.

(a) Calculate the values of Σx and Σx^2.

(b) Some numbers in the set of data are removed. The sum of the numbers removed is 96 and the mean is 8. Given the sum of the squares of the numbers removed are 800, calculate the variance of the new set of data.

Comprehensive Practice

1. Calculate the range and the interquartile range of each set of data.

(a) 8, 25, 16, 11, 24, 18, 22

(b) 27, 33, 45, 18, 62, 50

(c) 3.4, 2.8, 2.7, 4.3, 3.8, 3.2, 3.0, 2.9

(d) 12, 19, 17, 18, 15, 12, 17, 20, 22, 30, 32, 16, 18

2. Calculate the range and the interquartile range of each set of data.

(a)

Diameter (cm)	6.0	6.2	6.4	6.6	6.8	7.0	7.2
Number of limes	6	9	12	18	20	10	5

(b)

Age (years)	13	14	15	16	17	18
Number of participants	12	18	21	20	21	8

3. Calculate the variance and the standard deviation of the following sets of data.
(a) 7, 9, 11, 8, 3, 7
(b) 50, 72, 63, 58, 55, 50, 70, 62, 66, 64
(c) 3.2, 4.4, 3.9, 4.1, 5.2, 4.8, 5.2
(d) 20, 27, 32, 47, 50, 38, 42, 40, 33, 37, 30

4. A set of data contains seven numbers. The sum of the seven numbers is 84 and the sum of the squares of the numbers is 1 920. Calculate the variance and the standard deviation of the set of data.

5. The range and the standard deviation of a set of numbers x_1, x_2, x_3, ..., x_{10} are 10 and 5.2 respectively. Calculate
(a) the range and the standard deviation of the set of numbers
$2x_1, 2x_2, 2x_3, ..., 2x_{10}$.
(b) the range and the standard deviation of the set of numbers
$\frac{x_1-1}{4}, \frac{x_2-1}{4}, \frac{x_3-1}{4}, ..., \frac{x_{10}-1}{4}$.

6. The masses of a group of eight pupils have a mean of 45 kg and a variance of 2.5 kg^2. Calculate
(a) the sum of the masses of the eight pupils.
(b) the sum of the squares of the masses of the pupils.

7. The mean of a set of numbers $(m-4)$, m, $(m+2)$, $2m$, $(2m+3)$ is 10.
(a) Calculate
  (i) the value of m.
  (ii) the standard deviation.
(b) Each number in the set of data is multiplied by 3 and then added by 2. Calculate the variance of the new set of data.

8. The table below shows the values of n, Σx and Σx^2 of a set of data.

n	Σx	Σx^2
12	66	1 452

(a) Calculate the variance.
(b) A number p is added to the set of data and it is found that the mean is increased by 0.5. Calculate
  (i) the value of p.
  (ii) the standard deviation of the new set of data.

9. Calculate the variance of the set of data $(p-4)$, $(p-2)$, $(p-1)$, p, $(p+4)$, $(p+9)$.

10. The table below shows the masses of players in two *sepak takraw* teams.

Team	Mass (kg)
A	48, 53, 65, 69, 70
B	45, 47, 68, 70, 75

(a) Calculate the mean, range, variance and standard deviation for the masses of the players in both teams.

(b) Is the range appropriate to be used as a measure of dispersion to represent the data above? Justify your answer.

(c) Determine the mass of which team has a greater dispersion from the mean.

11. The sum of a set of 10 numbers is 180 and the sum of the squares of the set of numbers is 3 800.

(a) Calculate the mean and the variance of the numbers.

(b) The number 19 is added to the set of numbers. Calculate the new mean and the new variance.

12. The table below shows the time taken, in hours, by 32 pupils to do revision in a week.

Time (hours)	1	2	3	4	5	6	7	10
Number of pupils	2	5	6	9	6	2	1	1

(a) Calculate the range, interquartile range, variance and standard deviation of the distribution.

(b) State the most appropriate measure of dispersion to show the time spent on revision by the pupils.

PROJECT

1. You will learn how to draw different graphs using a dynamic geometry software.
2. Scan the QR Code to carry out this project.
3. Print and display your drawing at the Mathematics Corner.

Scan the QR Code to carry out this activity.
https://www.geogebra.org/classic/h4frqvzj

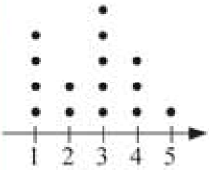
CONCEPT MAP

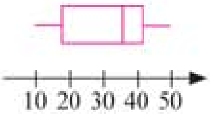
Measures of Dispersion

- Data Representation
  - Stem-and-Leaf Plot
  - Dot Plot
  - Box Plot

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

Key:
2 | 1 means 21





Range
= Largest value – Smallest value

Quartile

- First quartile, Q_1
- Third quartile, Q_3

Interquartile range
$= Q_3 - Q_1$

Variance

$$\sigma^2 = \frac{\Sigma(x - \bar{x})^2}{N}$$

$$\sigma^2 = \frac{\Sigma x^2}{N} - \bar{x}^2$$

$$\sigma^2 = \frac{\Sigma f(x - \bar{x})^2}{\Sigma f}$$

$$\sigma^2 = \frac{\Sigma f x^2}{\Sigma f} - \bar{x}^2$$

Standard deviation

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N}}$$

$$\sigma = \sqrt{\frac{\Sigma x^2}{N} - \bar{x}^2}$$

$$\sigma = \sqrt{\frac{\Sigma f(x - \bar{x})^2}{\Sigma f}}$$

$$\sigma = \sqrt{\frac{\Sigma f x^2}{\Sigma f} - \bar{x}^2}$$

Self Reflection

Circle your answers in the given word search.

1. ☐ is the difference between the largest value and the smallest value.

2. The data that divides a set of data is known as quartile. The ☐ quartile is the value of data at the $\frac{1}{4}$ position while the ☐ quartile is the value of data at the $\frac{3}{4}$ position from the data arranged in order.

3. ☐ is the measure of dispersion that refers to the difference between the third quartile and the first quartile.

4. The range, interquartile range and standard deviation are known as measures of ☐.

5. ☐ plot can show the distribution of a set of data.

6. The data with a smaller standard deviation is said to be more ☐.

7. ☐ is the measure of dispersion that measures how the data disperse around the mean of the set of data.

N	Q	D	Y	M	D	M	K	Z	D	Z	R	L	Q	Y	J	Y	J	E
O	G	R	R	J	Y	L	J	R	M	R	G	Y	K	L	W	L	G	R
I	U	A	R	T	I	L	T	H	I	R	D	A	K	R	Y	N	P	Q
T	J	D	N	B	M	D	M	N	L	D	L	Y	V	L	A	D	C	G
A	X	N	O	I	S	R	E	P	S	I	D	Z	X	R	J	O	V	N
I	L	U	D	D	W	V	T	D	T	L	G	M	E	J	N	Z	Y	T
V	G	L	A	X	K	Q	L	D	Q	D	J	L	M	S	M	J	B	Y
E	R	L	M	R	L	Y	Z	L	G	J	I	N	I	B	J	J	M	T
D	V	L	G	B	T	D	L	N	Q	T	X	S	B	Y	T	Y	D	W
D	D	J	T	N	B	I	B	V	R	J	T	K	L	G	M	T	R	T
R	B	M	K	L	L	M	L	A	R	E	L	N	R	L	P	G	L	L
A	L	Y	N	K	L	Z	U	F	N	A	Z	J	T	M	V	D	Z	B
D	Y	B	M	N	X	Q	B	T	I	N	N	T	Z	Q	L	T	Y	P
N	G	M	K	R	R	O	Q	R	T	R	R	G	D	W	R	T	B	N
A	B	V	R	E	V	N	B	Y	L	L	S	J	E	D	L	L	J	G
T	B	M	T	R	L	B	L	T	B	J	T	T	R	J	B	K	Z	T
S	L	N	Y	R	M	G	T	J	O	D	N	R	M	L	Z	D	P	B
K	I	V	G	R	D	N	D	R	M	I	N	R	M	A	R	W	B	D

Mathematics Exploration

1. Divide the class into groups.

2. Each group is required to obtain information about the months of birth of all their classmates.

3. Based on the information obtained, construct an appropriate data representation.

4. Using the data, determine
   - (a) the range.
   - (b) the first quartile.
   - (c) the third quartile.
   - (d) the interquartile range.
   - (e) the variance.
   - (f) the standard deviation.

5. Conduct Gallery Walk activity to see the work of other groups.