



Measures of spread

**Year 11 General Maths
Units 1 and 2**

www.maffsguru.com

Learning Objectives

By the end of the lesson I would hope that you have an understanding and be able to apply to questions the following concepts:

- Be able to understand what the range, interquartile range and standard deviation are when related to measures of spread.
- Know when to use each as a measure of spread for a particular distribution
- Be able to use the CAS to calculate summary statistics



Recap

In the previous lesson we looked at how we can find measures of centre and which one to use for a particular data set. This lesson we are going to look at measures of spread.

In a past lesson we said that it is important to compare how spread out data is.

In Years 7 to 10, we used the range (which is the highest value minus the lowest value). But does this always make sense to use it?



The problem with the range: Outliers

Wherever I go, I feel like a loner. I'm an outsider. I seem to be a bit of an **outlier**.

An outlier is something which doesn't fit near a set of data. It's out on its own.

In many cases, this will be either a really **large number** or really **small number**.

If we find the range of the numbers 1, 2, 3, 4, 5 it will be 4.

There are no **obvious outliers**.

When I find the range of 1, 2, 3, 4, 57 it will be 56.

This one number has really blown out the range and makes it not really reliable.



$$\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & \\ \text{Range } 5 - 1 = 4 & & & & & \\ 1 & 2 & 3 & 4 & \boxed{57} & \\ \text{Range } = 57 - 1 = \underline{\underline{56}} & & & & & \end{array}$$



Example: Finding the range

Consider the marks, for two different tasks, awarded to a group of students:

Task A							
2	6	9	10	11	12	13	22
23	24	26	26	27	33	34	
35	38	38	39	42	46	47	47
52	52	56	56	59	91	94	
Task B							
11	16	19	21	23	28	31	31
33	38	41	49	52	53	54	
56	59	63	65	68	71	72	73
75	78	78	78	86	88	91	

Find the range of each of these distributions.

$$A: \text{Range} = 94 - 2 = \underline{\underline{92}}$$

$$B: \text{Range} = 91 - 11 = \underline{\underline{80}}$$

There are so many things to be careful of here! Make sure the data is ordered! Ask yourself are there any obvious outliers?



How to deal with outliers: Interquartile range

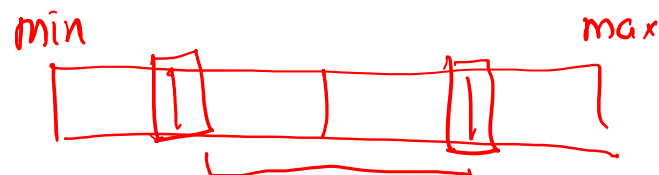
If we look at the stem plot on the right, we can now see that there are clearly some outliers.

We can't just ignore them ... so how can we find a "better" measure of spread??

We use something called the **Interquartile range**.

This is where we split the data into **quarters** and then subtract the upper quarter from the lower quarter.

This gives us the range of the middle 50% of the data and hence will not be subject to problems presented by outliers.



$\boxed{2} \mid 6 \ 9 \quad 10 \ 11 \mid 12 \ 13 \ 22 \ 28 \ \dots \mid \dots \ 56 \mid 56 \ 59$
 $91 \ \boxed{94}$

↓

Task A	key: 1 2 represents 12 marks
0	2 6 9
1	0 1 2 3
2	2 3 4 6 6 7
3	3 4 5 8 8 9
4	2 6 7 7
5	2 2 6 6 9
6	} gap
7	
8	
9	1 4



Finding the interquartile range

It's probably best to show by example.

But ... to help ... think of cutting a cake to serve 4 people.

You would probably halve it and then halve each of the halves. This is what you do when finding the interquartile range.

Let's do it for both data sets.

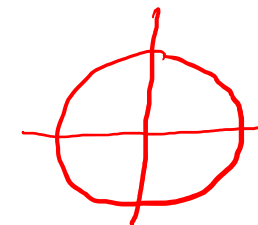
Task A key: 1 | 2 represents 12 marks

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

$$LQ = \underline{\underline{22}}$$

$$UQ = \underline{\underline{47}}$$

$$\begin{aligned} IQR &= UQ - LQ \\ &= 47 - 22 \\ &= \underline{\underline{25}} \end{aligned}$$



$$\frac{30 + 1}{2} = 15.5$$

$$\frac{15 + 1}{2} = 8$$



Finding the interquartile range

It's probably best to show by example.

But ... to help ... think of cutting a cake to serve 4 people.

You would probably halve it and then halve each of the halves. This is what you do when finding the interquartile range.

Let's do it for both data sets.

Task B key: 1 | 2 epeents 12 maks

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

$$LQ = 31$$

$$UQ = 73$$

$$IQR = 73 - 31 \\ = \underline{\underline{42}}$$

$$n = \frac{30 + 1}{2} = 15.5$$

$$n = \frac{15 + 1}{2} = \underline{\underline{8}}$$

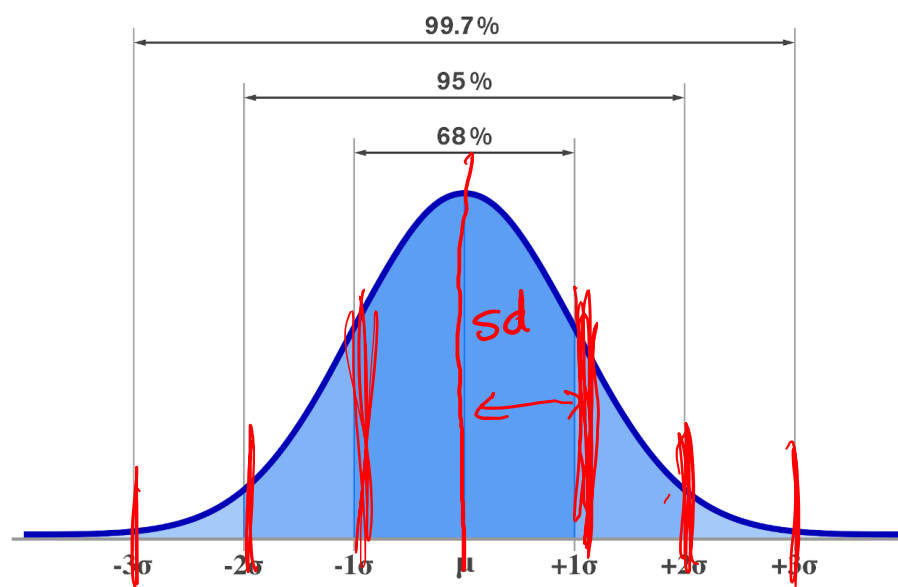


Understanding standard deviation

This really is a whole video in itself and can be a little complex.

Ultimately, we are able to show that data can be split into a number of slices. Each of these slices is called a standard deviation and, in most cases, we are interested in knowing how far away data is from the mean value.

VCE scores are distributed on a bell curve. This can then be split into standard deviations.



Finding the standard deviation

There is a formula we can use to find the standard deviation for a set of data.

Note: This is not normally required in an exam as you will use your CAS.

The formula to find the standard deviation of a set of data is:

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

Write this in your summary books just in case.

$\bar{x}$ = mean

n = num of items



Example: Finding the standard deviation

Use the following formula to find the standard deviation of the numbers 2, 3 and 4

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

The best way is to draw a table by breaking down the formula!



$$(x - \bar{x})$$

$$\bar{x} = \frac{2 + 3 + 4}{3} = \underline{\underline{3}}$$

$$\therefore sd = 1$$

x	$x - \bar{x}$	$(x - \bar{x})^2$
2	-1	1
3	0	0
4	1	1
		<u>2</u>

$$\sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

$$\frac{2}{2} = 1$$

$$\sqrt{1} = \underline{\underline{1}}$$



Summary statistics

When doing a statistical analysis we generally use a set of numbers called the summary statistics.

These can include:

- Mean
- Standard deviation
- Median
- Interquartile range
- Range

Remember: These numbers on their own are pretty useless. But, when we find them for different data sets, we can use them to compare the data. This is where the power of statistics comes in!

$$\begin{aligned} \text{IQR} &= \text{UQ} - \text{LQ} \\ &= 4 - 2 \\ &= \underline{\underline{2}} \end{aligned}$$

$$\begin{aligned} \text{Range} &= 4 - 2 \\ &= \underline{\underline{2}} \end{aligned}$$



Using the CAS to find summary statistics

It is much too complex to use the formula for large data sets and so we can use the CAS.

I will show how to do it using the two most popular versions of the CAS.

Example:

The table shows the monthly rainfall figures for a year in Melbourne.

Month	J	F	M	A	M	J	J	A	S	O	N	D
Rainfall (mm)	48	57	52	57	58	49	49	50	59	67	60	59

$$\bar{x} = 55.42$$

$$sd = 5.807$$

