Standard Deviation

Sunday, 3 March 2019 10:11 am

By the end of the lesson I would hope that you have the knowledge and understanding for the following points:

- Know what the standard deviation is
- Know why we use the standard deviation
- Know how to calculate the standard deviation using both
  - Pencil and paper methods and Scientific Calculator
  - CAS Calculator

RECAP:

We are nearing the end of the journey with Statistics.
We have looked at a range of methods to displaying and interpreting data.
We have looked at the **range** and the **interquartile** range.
There is one more measure of spread which we need to know about and it's used a LOT in Australia.

RECAP: The Range

The range is **one measure of spread**.
It looks at the difference between the highest number and the lowest number.

Example:
Find the range between 6 and 47

Range $= 47 - 6 = \underline{\underline{41}}$

Find the range between -3 and 38

Range $= 38 - (-3) = 38 + 3 = \underline{\underline{41}}$

We can look at the two ranges and see which data is more spread out.

The **problem with the range** is when there are outliers.
The outliers really change the range and hence make it a bad measure of spread.

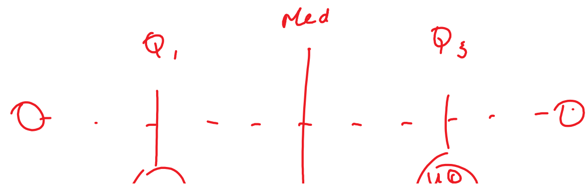
$-6 \quad \boxed{18 - 50} \quad 32 \quad 56$

RECAP: The Interquartile Range

To lessen the effect of outliers on our data, we can find a different **measure of spread**.
This measure looks at the **range** of the middle **50% of our data**.

To be able to do this, we need to split the data in half.
Split the halves into halves and then take the two values away from each other.

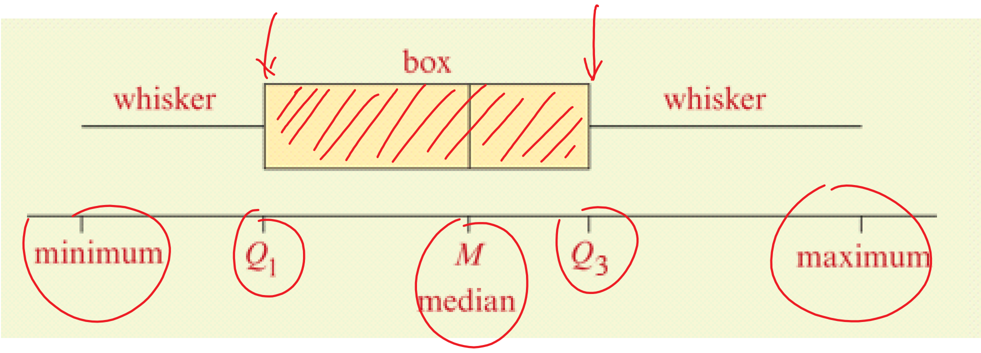
$$IQR = Q_3 - Q_1$$



4

LQ UQ

This can be shown using a box plot too.



Both the **range** and the **interquartile range** are measures of spread which use the **median**.
Sometime, we are interested in looking at the range of values about the mean.

RECAP: The MEAN

The mean is a measure of centre.
We are saying, if I take a lot of numbers the mean is a mathematical way of finding the "middle number"
If we look at the mean using an example shown below:

Example:
Find the mean of the numbers 1, 2, 3, 5, 6, 7
Using the formula we know ...

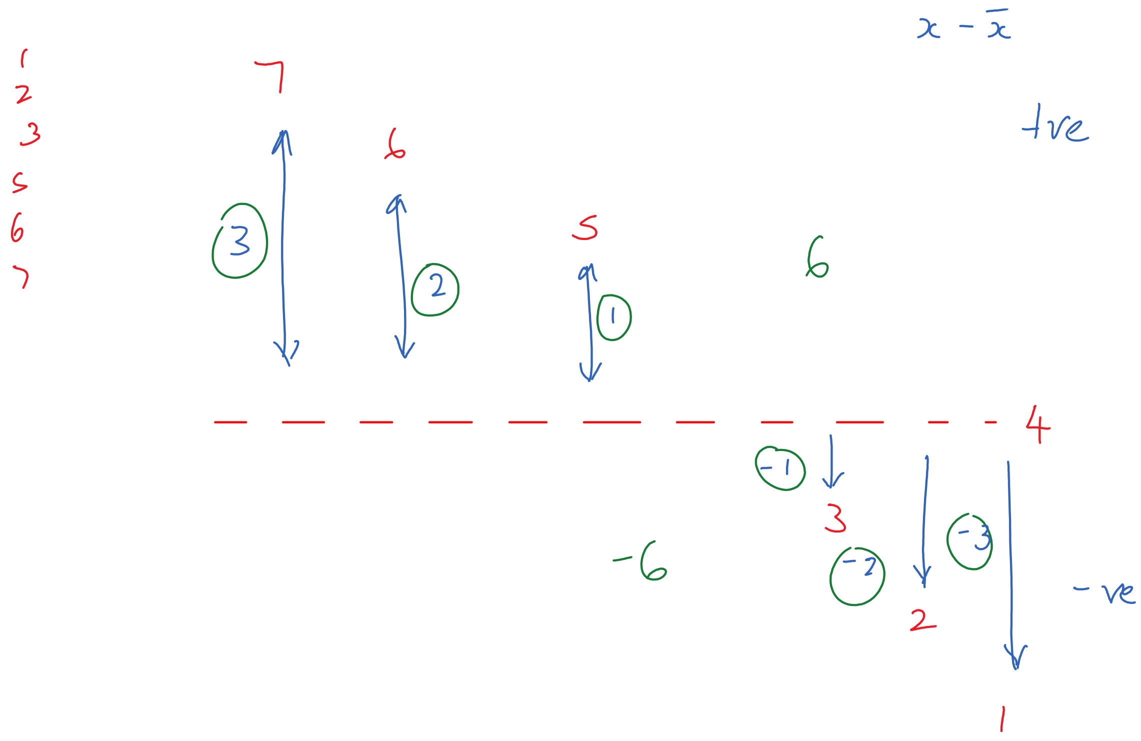
4

$$\text{Mean } (\bar{x}) = \frac{1+2+3+5+6+7}{6} = \frac{24}{6} = 4$$

The mean is the "mathematical middle"
Let's show this using a diagram:

$x - \bar{x}$

$\frac{1}{2}$ 7



Let's look at the differences between each of the numbers from the mean.
We can see the differences are both positive and negative.
What if we add all the values together?
Why do we get zero?

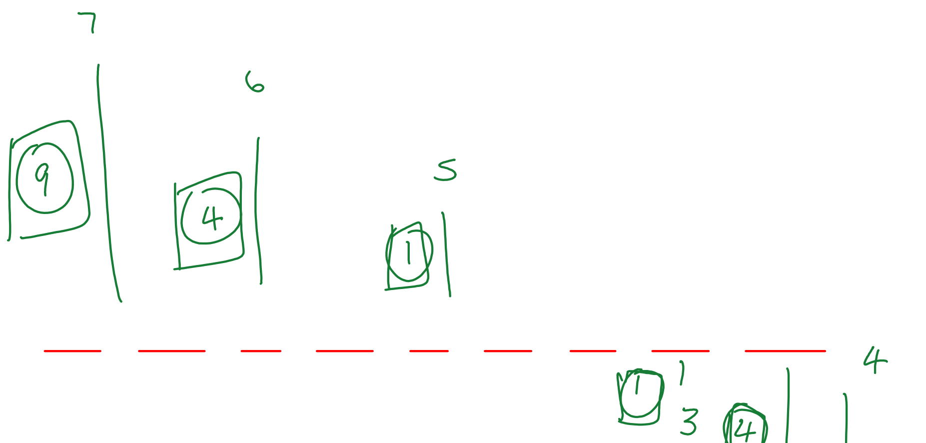
What if I wanted to find the average difference away from the mean?
Negative numbers can be turned positive by **squaring them**.
We can now turn our distances positive.

$(-1)^2 = 1$

$(-2)^2 = 4$

Square

$(-3)^2 = 9$



Average distance

$$= \frac{9+4+1+1+4+9}{6}$$

$= 4.67$ $s = \sqrt{4.67}$

If I want to find the average distance away from the mean I can now do this.

The problem is, we squared all the distances to make them positive.
This means we have changed the problem.
We need to undo the squaring and do this using a square root sign.

You have just found the **standard deviation**.

BARRY IS AT IT AGAIN

Just when we thought it was safe to get into the water ....
Just when we thought we understood this ...
Barry had to go and write this as a formula!

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

It looks gross, but it isn't.

Example questions:

There are two ways of doing this.
There is the **hard way** and the **easy way**.
Let's look at the hard way first.

Question has been used, with permission, from the *Cambridge Essentials Textbook Series*.

Calculate the mean and sample standard deviation for this small data set, correct to one decimal place.

2, 4, 5, 8, 9

$$\bar{x} = \frac{2+4+5+8+9}{5} = 5.6$$

The key to doing this is being able to unpack the standard deviation formula and drawing a table:

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

x	$x-\bar{x}$	$(x-\bar{x})^2$
2	−3.6	12.96
4	−1.6	2.56
5	−0.6	0.36
8	2.4	5.76
9	3.4	11.56
Total		33.2

$$s = \sqrt{\frac{33.2}{4}}$$

$$s = 2.88$$

Now … the easy way.
Using a CAS calculator

Second example (using a CAS)

Partly extracted from the *Cambridge Essentials Textbook Series*

This back-to-back stem-and-leaf plot shows the distribution of distances that 17 people in Darwin and Sydney travel to work. The means and standard deviations are given.

Darwin Leaf	Stem	Sydney Leaf	
			Sydney $\bar{x} = 27.9$
8 7 4 2	0	1 5	$s = 15.1$
9 9 5 5 3	1	2 3 7	
8 7 4 3 0	2	0 5 5 6	Darwin
5 2 2	3	2 5 9 9	$\bar{x} = 19.0$
	4	4 4 6	$s = 10.1$
	5	2	

3 | 5 means 35 km

Using the data above, use your CAS to check that the mean ($\bar{x}$) and standard deviation (s) is as shown