# Describing the centre and spread of symmetric distributions

Thursday, 21 February 2019 5:19 pm

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

- Know what the mean is
- Know how to calculate the mean
- Know what the formula to calculate the mean looks like and how to use it
- Know the relationship between the mean and the median
- Know when to use the median rather than the mean
- Know what standard deviation means
- Know what the standard deviation formula looks like
- Know how to calculate the standard deviation

## RECAP

In this series of lessons we are looking at data and how to interpret it.
We have, in previous years, looked at the ways to calculate:

- Mean
- Median
- Mode
- Range

This year we have looked at the concept of the Interquartile Range (IQR) and how to calculate it.

The question we ask now is … what is the difference between the mean and the median. Why do we use two measures of centre? Which is the best to use?

## The mean

Yup! The Grinch is fairly mean.
Sadly, Further Mathematics doesn't care about how mean someone is.
It's more interested in the "average" value for a set of data.

In Year 8 an onwards, we told you that the mean was:

$$Mean = \frac{sum\ of\ data\ values}{total\ number\ of\ data\ values}$$

We diligently added all the numbers up and divided by the number of numbers.
But how many of us asked what the mean was? Why are we doing this? What is the point of Mathematics?

OK. We all asked that last question!

**A basic example:**

Find the mean of 2, 3, 4, 5, 6, 10

$$Mean = \frac{2+3+4+5+6+10}{6} = \frac{30}{6} = 5$$

$$Median = 4.5$$

## Barry makes life more complicated.

Barry likes to make things more complicated by throwing in a wonderful formula with lots of Greek letters! But … we're better than this! We can see through his dastardly plan!

To find the mean:

sum

data items

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \sum x$$

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{\sum x}{n}$$

mean

$x$ $\bar{x}$

$$Mean = \frac{sum\ of\ data\ values}{total\ number\ of\ data\ values}$$

**Example:**
*Extracted from Cambridge Further Mathematics Textbook*

The following is a set of reaction times (in milliseconds): 38 36 35 43 46 64 48 25

Write down the values of the following, correct to one decimal place.

**a** $n$

**b** $\sum x$

**c** $\bar{x}$

a) $n = 8$

b) $\sum x = 335$

c) $\bar{x} = \frac{\sum x}{n} = \frac{335}{8} = 41.9$

## Why do we have the mean and the median

If we look at the data set we had before, we can see that the mean and the median are different values.
So, which is the better value for the centre of the distribution?

Find the mean and median of 2, 3, 4, 5, 6, 10 100

Mean = 5

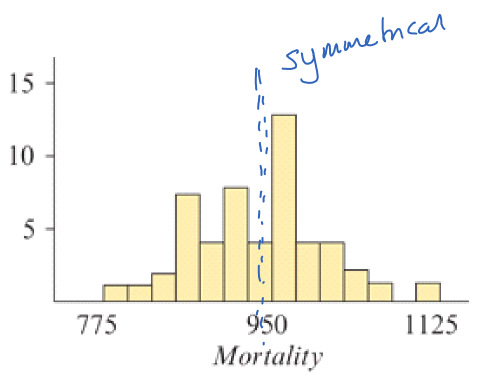
Median = 4.5

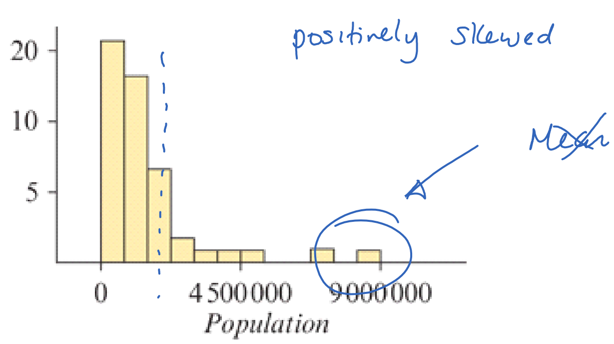
**The median**: This is the central value of the distribution. It's the middle point of the data.

**The mean**: This is the balance point of the distribution.

$\frac{30}{6} = 5$ $\frac{130}{7} = 9.$

Looking at the following two graphs we can see how the shape of the data can change how close the mean and the median are to each other.




The above data is fairly symmetric.
Hence, the middle value and the mean will be roughly the same.

This data is obviously positively skewed.
The centre of the data will be more to the left.
The mean will be affected by the larger values

The above data is fairly symmetric.
Hence, the middle value and the mean will be roughly the same.

This data is obviously positively skewed.
The centre of the data will be more to the left.
The mean will be affected by the larger values dragging it to the right.

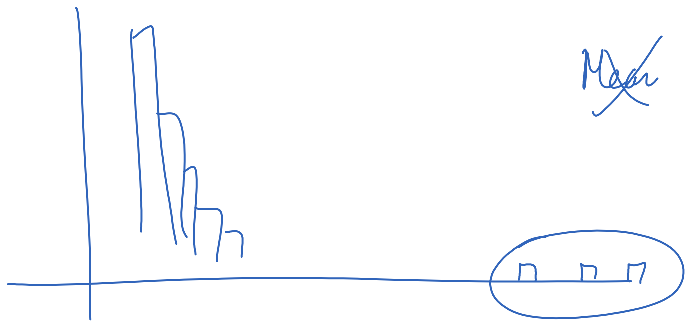
### How do I know which is the best measure of centre?

*i.e. When do I use the mean and when do I use the median?*

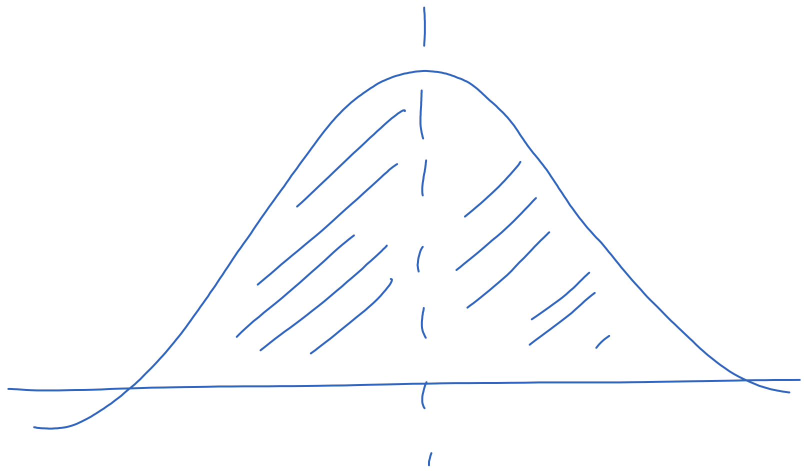
Using the two examples above we can see that the median value is less prone to large values in a distribution. It's known as a **resistant statistic**.

Hence, we use the median when the data is skewed or is likely to contain outliers.

Think about house prices ...




If a distribution is symmetric (with no outliers) then we can use either the mean or the median as a measure of centre. In most cases we prefer to use the mean as it's more familiar to people. We can use the Mean in lots more ways.



### Back to measures of spread

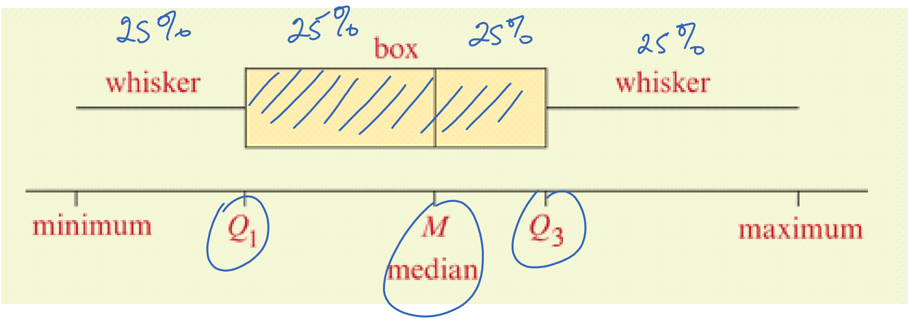
We seem to spend a lot of time looking at measures of centre.
Well, we're about to ramp up our understanding of the measures of spread.

We have two main measures of spread, so far:

- Range
- Interquartile Range

**The Interquartile Range: A recap**

The IQR is a measure of the middle 50% of the data.
It's also known as a measure of spread of data around the median value ($M$).

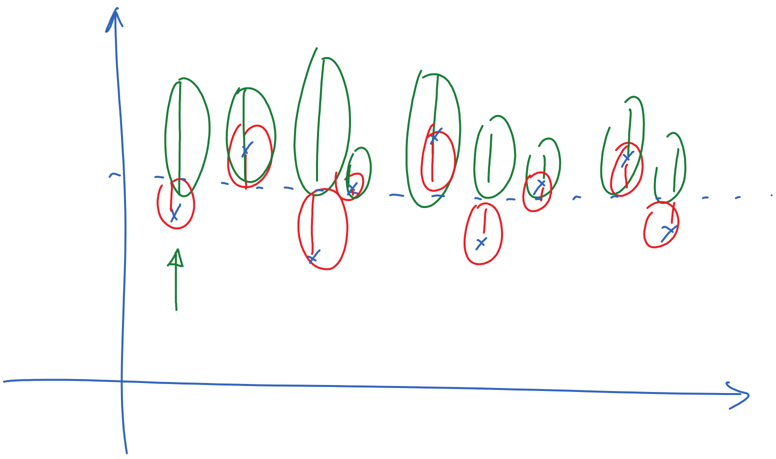



What about if we wanted to look at the spread of the data around the average value or mean ($\bar{x}$)?
*Welcome to the standard deviation.*

## The standard deviation: An average of the squared deviations of each data item from the mean

The hard explanation (which can be ignored but might be interesting to know).
What if we could find some sort of measure of how far each data item is away from the whole group's mean value.

We can do this graphically.



When we square the results, we end up with the units of the data making no sense.
So, we have to change the units back by square rooting the result.

$$\sqrt{\frac{\quad}{\text{num}}}$$

The formula (what you've been waiting for!)

Sum

data item

$$\sum (x - \bar{x})^2$$

mean

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

# Bessel's correction

From Wikipedia, the free encyclopedia

> ? This article includes a list of references, but **its sources remain unclear** because it has **insufficient inline citations**. Please help to improve this article by introducing more precise citations. *(November 2010) (Learn how and when to remove this template message)*

In statistics, **Bessel's correction** is the use of $n - 1$ instead of $n$ in the formula for the sample variance and sample standard deviation, where $n$ is the number of observations in a sample. This method corrects the bias in the estimation of the population variance. It also partially corrects the bias in the estimation of the population standard deviation. However, the correction often increases the mean squared error in these estimations. This technique is named after Friedrich Bessel.

Complicated yes!
How would you do this by hand?
You generally wont.
You will use the CAS.

## Using the CAS to find the Standard Deviation for a set of numbers

*Example extracted from the Cambridge Further Mathematics Textbook Series*

The following are all heights (in cm) of a group of women.

176. 160. 163. 157. 168. 172. 173. 169

Determine the mean and standard deviation of the women's heights correct to two decimal places.

$\bar{x} = 167.25$

$s = 6.67$