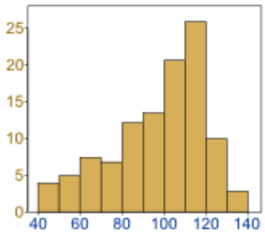
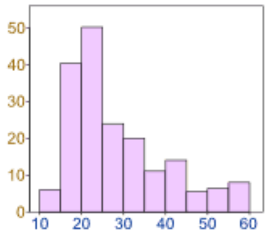
Normal Distribution

Data can be "distributed" (spread out) in different ways.

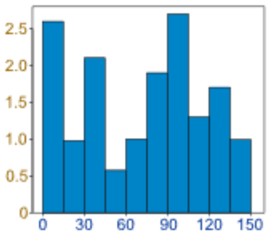
It can be spread out more on the left



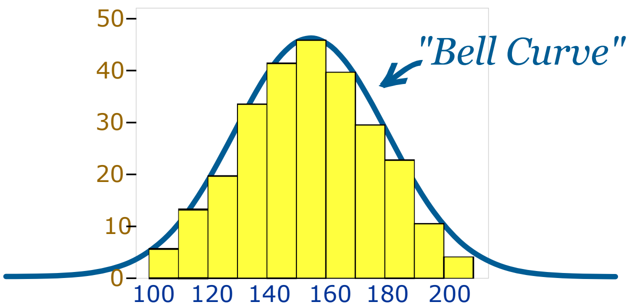
Or more on the right



Or it can be all jumbled up



But there are many cases where the data tends to be around a central value with no bias left or right, and it gets close to a "Normal Distribution" like this:



A Normal Distribution

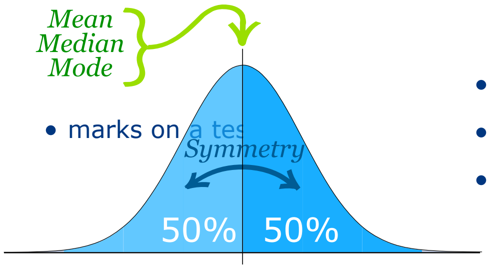
The "Bell Curve" is a Normal Distribution.
And the yellow histogram shows some data that follows it closely, but not perfectly (which is usual).

It is often called a "Bell Curve" because it looks like a bell.

Many things closely follow a Normal Distribution:

- heights of people
- size of things produced by machines
- errors in measurements
- blood pressure

We say the data is "normally distributed":



The **Normal Distribution** has:

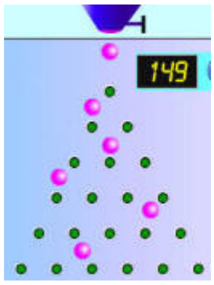
- mean = median = mode
- symmetry about the center
- 50% of values less than the mean and 50% greater than the mean

Quincunx

You can see a normal distribution being created by random chance!

It is called the Quincunx and it is an amazing machine.

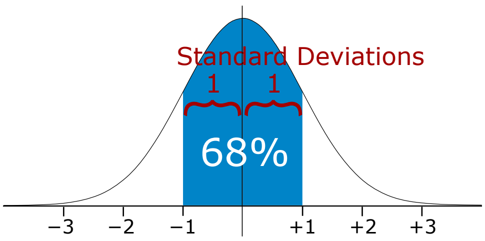
Have a play with it!



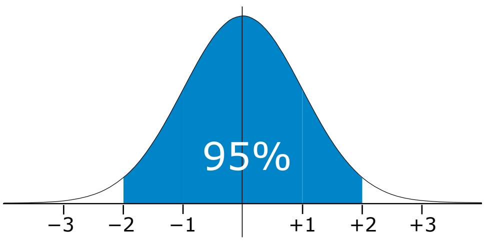
Standard Deviations

The Standard Deviation is a measure of how spread out numbers are (read that page for details on how to calculate it).

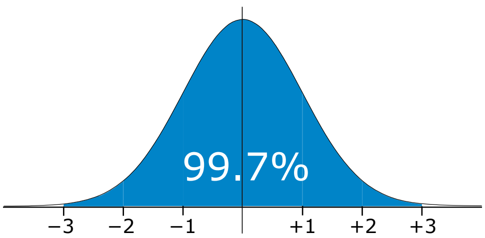
When we calculate the standard deviation we find that **generally**:



68% of values are within
1 standard deviation of the mean



95% of values are within
2 standard deviations of the mean



99.7% of values are within
3 standard deviations of the mean

Assuming this data is **normally distributed** can you calculate the mean and standard deviation?

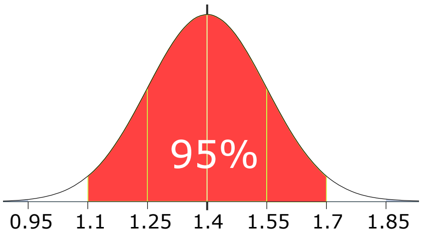
The mean is halfway between 1.1m and 1.7m:

Mean = (1.1m + 1.7m) / 2 = **1.4m**

95% is 2 standard deviations either side of the mean (a total of 4 standard deviations) so:

1 standard deviation = (1.7m-1.1m) / 4
= 0.6m / 4
= **0.15m**

And this is the result:



It is good to know the standard deviation, because we can say that any value is:

- **likely** to be within 1 standard deviation (68 out of 100 should be)
- **very likely** to be within 2 standard deviations (95 out of 100 should be)
- **almost certainly** within 3 standard deviations (997 out of 1000 should be)

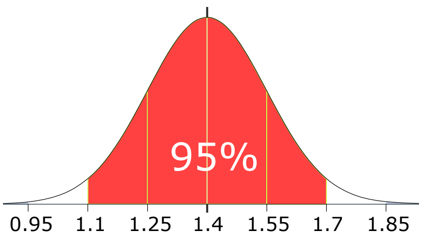
Standard Scores

The number of **standard deviations from the mean** is also called the "Standard Score", "sigma" or "z-score". Get used to those words!

Example: In that same school one of your friends is 1.85m tall

You can see on the bell curve that 1.85m is **3 standard deviations** from the mean of 1.4, so:

Your friend's height has a "z-score" of 3.0



It is also possible to **calculate** how many standard deviations 1.85 is from the mean

How far is 1.85 from the mean?

It is 1.85 - 1.4 = **0.45m from the mean**

How many standard deviations is that? The standard deviation is 0.15m, so: