**Standard Deviation**

**What is it?**

The standard deviation is a measure of statistical dispersion.

In plain English it's a way of describing how spread out a set of values are around the mean of that set. For example, if you have a set of height measurements, you can easily work out the arithmetic mean (just sum up all the individual height measurements and then divide by the number of those measurements). However, knowing the mean (or average, as it's more commonly called), doesn't tell you about the spread of those heights. Were all the people in your group the same height, or did you have some tall and some short, or was there one really tall person who towered over everybody else? It's possible that you could have exactly the same average height from wildly different groups. Knowing about how spread out those heights are compared to the mean gives you extra information over and above the mean value.

**How do I calculate it?**

The basic idea of the standard deviation is that you're measuring variations around the mean value. Some of those values will be below the mean, some above and sometimes you'll have some that are equal to the mean. In other words some of the differences between the individual measurements will be positive (more than the mean), some will be negative (below the mean) and some will be zero (directly equal to the mean). Now just adding these differences up is dangerous because the positive and negative values will cancel each other out. For example, to take an incredibly simplistic case, if you've got a sample of two values, one of 9 and one of 11, the mean is equal to 10. The differences are -1 and +1, adding these together gives us a total variation of 0. But we know that there's not zero variation around that mean value!

So, to get round this problem each of the variations around the mean is squared. When you square a negative value you get a positive value. So, to work out the standard deviation we square all of the differences from the mean, add them all up and divide by one less than the number of values in our set. This new number is called the variance. Now we take the square root of the variance (which is reversing the squaring we did earlier so that our number is closer to the original differences), and that's the standard deviation.

Here's the recipe again:



1. Work out the average (mean value--*M*) of your set of numbers
2. Work out the difference between each number and the mean
3. Square the differences (X-M)2
4. Add up the square of all the differences:
5. Divide this by one less than the number of numbers in your set - this is called the **variance**
6. Take the square root of the variance and you've got the standard deviation

That's it for the most part. A set of values that are closely clustered near the mean will have a low standard deviation, a set of numbers that are widely apart will have a higher standard deviation and a set of numbers that are all the same will have a standard deviation of zero (because they're all equal to the mean anyway).

**What's this 'normal distribution' stuff?**

If your data is normally distributed, in other words you've got most data near the mean and the further away you get from the mean the fewer measurements you have, then the standard deviation gives you extra information. A lot of real world measurements are normally distributed - height, weight, test scores, wages… - and if you graph the data it has a *bell-shaped* curved. For data that is normally distributed the standard deviation gives us the following information:

* Around 68% of data are within one standard deviation of the mean
* Around 95% of data are within two standard deviations of the mean
* Around 99% of data are within three standard deviations of the mean

So, if the mean of your data set is 100 and the standard deviation is 10, you can expect to see about 68% of values in the range 90-110, and 95% to be in the range 80-120. And anything below 70 or above 130 is going to be very rare…

http://www.techbookreport.com/tutorials/stddev-30-secs.html