## Standard Deviation

http://www.theseashore.org.uk/theseashore/Stats%20for%20twits/Additional%20Material/SDeviation.htm

Standard Deviation is a measure of the spread of the data about the arithmetic mean value. It represents the average amount that each value differs from the mean.

You collect the following data:

| **Category** | | | | | **Beard lengths in metres** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Ecologists: | 0.1 | 0.9 | 0.4 | 0.5 | 0.5 | 0.5 | 0.5 | 0.6 | 0.5 | 0.5 |
| Druids: | 0.4 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.6 |
| Eco-Warriors: | 0.1 | 0.2 | 0.5 | 0.8 | 0.5 | 0.7 | 0.4 | 0.5 | 0.6 | 0.7 |

The mean beard length of all groups is the same (0.5m). The data for each group are however different in the way that individual data points are scattered about the mean.

How can we describe the differences in the 3 data sets? What is needed is a measure that indicates the average amount that each piece of data is different from the mean. The standard deviation is just such a number. Here is how it's calculated:

### Standard deviation of Ecologist's data

| The raw data | Each item of data minus the mean | Each item of data minus the mean squared |
|---|---|---|
| $x$ | $(x - \bar{x})$ | $(x - \bar{x})^2$ |
| 0.1 | -0.4 | 0.16 |
| 0.9 | 0.4 | 0.16 |
| 0.4 | -0.1 | 0.01 |
| 0.5 | 0.0 | 0.0 |
| 0.5 | 0.0 | 0.0 |
| 0.5 | 0.0 | 0.0 |
| 0.5 | 0.0 | 0.0 |
| 0.6 | 0.1 | 0.01 |
| 0.5 | 0.0 | 0.0 |
| 0.5 | 0.0 | 0.0 |

We sum the right-hand column to get a value of 0.34 that represents the total deviation of all our pieces of data from the mean. We divide by one less than the number of samples. In our case that = 10 – 1 = 9. Statisticians call this value (n – 1) the degrees of freedom. You will always use n – 1 to calculate standard deviation.

0.34/9 = 0.037

We now have a number (0.037) representing the scatter or spread of our Ecologist's beard lengths about the mean value. Statisticians call this number the **variance** of the data.

The calculation we have just done for variance can be represented by this formula:

$$\frac{\Sigma (x - \overline{x})^2}{n - 1}$$

To complete our calculation of standard deviation we must take the square root to convert the number back to its original units.

$\sqrt{}$ 0.037 = 0.192 = **standard deviation** of Ecologist's beard lengths in meters

We find that:

- S.D. (Ecologists) = 0.192
- S.D. (Druids) = 0.047
- S.D. (Eco-Warriors) = 0.221

Notice that the Druids (whose beards were all clustered around the mean) have a very small value. The Ecologists (whose beards were mostly clustered around the mean) had a bigger value and the Eco-Warriors whose beard lengths were all over the place (i.e. widely scattered about the mean) have the biggest value.

**Standard Error and 95% confidence intervals**

http://www.theseashore.org.uk/theseashore/Stats%20for%20twits/Additional%20Material/SError.htm

How confident are we that our sample mean is a reasonable approximation of the true mean?

Let's say you decide to take another set of eleven measurements and calculate the mean of them. You would probably get an answer slightly different to the answer you got from the original measurements.

If you did this many times, you would end up with a large number of means, all a bit different from each other.  If you took all of these separate means and calculated an overall mean for the whole lot, you would end up with a value that was the same as the population mean. We could work out the standard deviation of this set of sample means. It would however take ages and we haven't got ages. Happily, we don't need to collect multiple samples because we can calculate it using just one sample like this:

$$\frac{\text{Standard deviation of sample}}{\sqrt{\text{Number of observations}}} = \frac{s}{\sqrt{n}}$$

This standard deviation of lots of means is now called the **STANDARD ERROR (S.E.)** of the mean.

Now, because statisticians say so, (you can look this up in any statistics book if you really want to), it can be shown that we can be 95% confident that the population mean will fall within 1.96 X SE of a sample mean.

Work out the standard error of the mean for each category of beardie using the standard deviation and the number of observations:

Eco-Warriors: $SE = \frac{s}{\sqrt{n}} = \frac{0.221}{\sqrt{10}} = 0.07$

Druids: = 0.01

Ecologists: = 0.06

(Calculation shown only for the Eco-Warriors)

Next, work out 1.96 X SE for each category:

Eco-Warriors: 0.07 X 1.96 = 0.14

Druids: 0.01 X 1.96 = 0.02

Ecologists: 0.06 X 1.96 = 0.12

This tells us that:

We are 95% confident that the population mean of Eco-Warriors is 0.5 +/- 0.14

We are 95% confident that the population mean of Druids is 0.5 +/- 0.02

We are 95% confident that the population mean of Ecologists is 0.5 +/- 0.12

We can now make a bar graph with error bars, which extend 1.96 SEs on either side of the mean (+/- 0.14 in the case of the Eco-Warriors):