Standard Deviation

How to Calculate Standard Deviation

Standard deviation (σ) is a statistical measure of how precise your data is. It is calculated using the following equation, where $\bar{x}$ is the data average, x_i is the individual data point, and N is the number of data points:

$$\sigma = \sqrt{\frac{\sum\limits_{i=1}^{N}(x_i - \bar{x})^2}{(N-1)}}$$

To calculate the standard deviation, you would begin with calculating the quantity ($x_i - \bar{x}$), which is the deviation of each data point from the average. You would square square each one, then add them up and divide by one less than the number of data points. Finally, you would find the square root of this value.

EXAMPLE:
The volume of a liquid was measured five times, the results being 4.5 mL, 4.4 mL, 4.5 mL, 4.2 mL and 4.3 mL. The deviations ($x_i - \bar{x}$) are shown below:

Measurements	Deviations ($x_i - \bar{x}$)
4.5 mL	(4.5 – 4.4) mL = 0.1 mL
4.4 mL	(4.4 – 4.4) mL = 0.0 mL
4.5 mL	(4.5 – 4.4) mL = 0.1 mL
4.2 mL	(4.2 – 4.4) mL = –0.2 mL
4.3 mL	(4.3 – 4.4) mL = –0.1 mL

Average ($\bar{x}$) = 4.4 mL

$$\sigma = \sqrt{\frac{\left[(0.1)^2 + (0.0)^2 + (0.1)^2 + (-0.02)^2 + (-0.1)^2\right]\text{mL}^2}{(5-1)}} = \sqrt{\frac{0.07\,\text{mL}^2}{4}} = 0.13\,\text{mL} = 0.1\,\text{mL}$$

We generally would rather go on and calculate the *relative* standard deviation, so that we can see whether 0.1 mL is a small or large quantity compared to the average value (4.4 mL). This done by finding the percent standard deviation:

$$\text{relative standard deviation} = \frac{\text{standard deviation}}{\text{average value}} \times 100$$

$$= \frac{0.1\,\text{mL}}{4.4\,\text{mL}} \times 100 = 2\,\%$$

ANSWER: The standard deviation (σ) is 0.1 mL, and the relative standard deviation is 2 %

What Exactly Does Standard Deviation Tell Us?

Inconsistent errors that cannot be attributed to any known cause but are due to unavoidable errors in measurements made by human beings are called indeterminate errors. For a large collection of such data, the measurements would form a bell curve, known as the normal distribution curve, as shown below. The x-axis represents the magnitude of magnitude of error and the y-aixs represents the frequency of error. The correct value should be the one that is most frequently obtained.
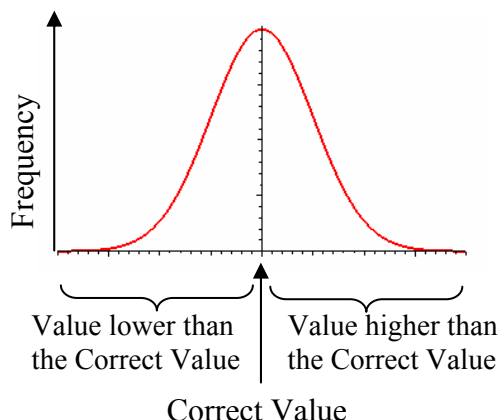
For such data, which occurs only for large number of samples (N>20), the standard deviation has the following meaning:

68 % of the data would fall within the range of **one** standard deviation of the average
95% will fall within the range of **two** standard deviations
99% will be within **2.5** standard deviations.

In the example given above, $\sigma = 0.1$ mL and the average is 4.4 mL.
This means that 68 % of the data is within 4.4 mL $\pm$ 0.1 mL (or between 4.3 ml and 4.5 mL).
95 % of the data is within 4.4 mL $\pm$ 2(0.1mL) = 4.4 mL $\pm$ 0.2 mL (or between 4.2 and 4.6 mL).

The smaller the standard deviation, the narrower is the range, which translate to a higher reproducibility. A small standard deviation means the experimental values are clustered together tightly (i.e. a higher precision).