Standard Deviation

The following mass and volume data was collected by using the water displacement method for six different pieces of copper. The density was calculated for each trial and is shown below.

Mass	Volume	Density
(g)	(cm^3)	(g/cm^3)
8.60	1.00	8.60
15.20	1.70	8.94
18.60	2.00	9.30
22.10	2.60	8.50
31.90	3.80	8.39
42.20	4.70	8.98

When analyzing a set of data, it is common to show the distribution of that data. Three factors showing distribution are shape, center point, and spread of the data. The shape of the data in graphical format refers to the data being symmetrical or asymmetrical, the center point is usually given by the average (mean) and the spread is given by the minimum and maximum data values.

A common way to express the center point is using the arithmetical average (mean) and is given by:

$$x_{ave} = (x_1 + x_2 + \ldots + x_n)/n \text{ also written as } x_{ave} = 1/n\Sigma x_i$$

The mean is calculated by summing the data values and dividing by the number of data values.

Standard deviation is used to measure the spread by showing how far the data values are from the mean and is given by:

$$s = (1/(n-1)\Sigma(x_i - x_{ave})^2)^{½}$$

Standard deviation measures the spread about the mean in the original units of measure. As the data becomes further from the mean, s, gets larger. Another thing to keep in mind is that the standard deviation is strongly influenced by outliers.

To calculate the standard deviation, s, of the densities:

- Calculate the mean density, D_{ave}.

  $D_{ave} = (D_1 + D_2 + D_3 + D_4 + D_5 + D_6)/6$

  $D_{ave} = (8.60 + 8.94 + 9.30 + 8.50 + 8.39 + 8.98)\ g/cm^3/6 = 8.78\ g/cm^3$

- Subtract the mean from each density value as shown in the table below.

- Square each of the differences.

- Divide the sum of the squares by the number of density values, minus one (n-1).

- Take the square root of the result in the previous step.

$D - D_{ave}$	$(D – D_{ave})^2$
(g/cm^3)	$(g/cm^3)^2$
-0.18	0.032
0.16	0.026
0.52	0.270
-0.28	0.078
-0.39	0.150
0.20	0.040

The formula for standard deviation is given by:

$$s = (1/(n-1)\Sigma(x_i – x_{ave})^2)^{½}$$

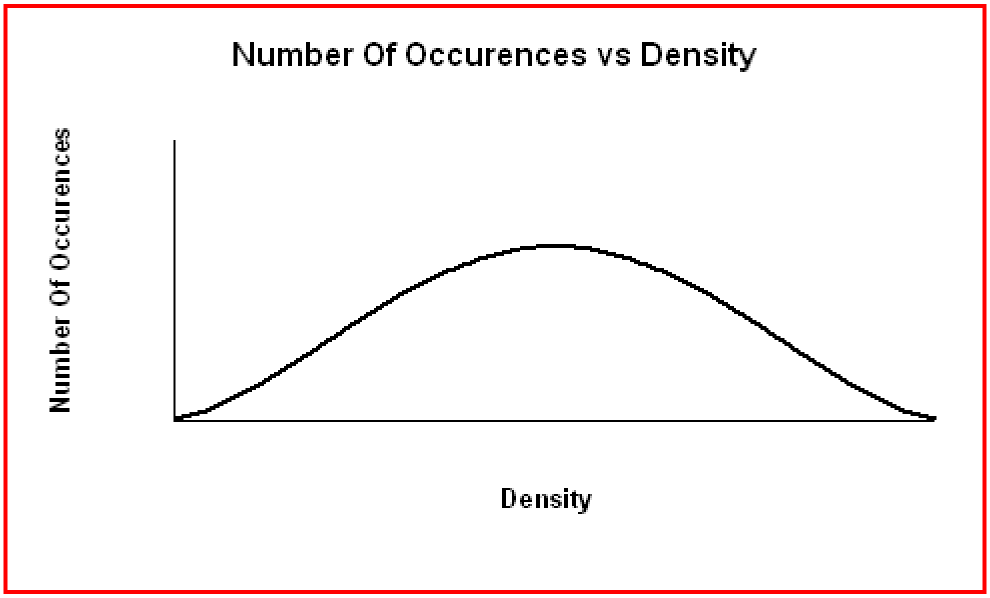
and substituting the table values yield:

$s = (0.0322+0.0262+0.272+0.0782+0.152+0.0402+/5)^{½} = 0.345\ g/cm^3$

The smaller the standard deviation, s, the closer the density values are to the average density while a large standard deviation means the density values lie further from the mean.

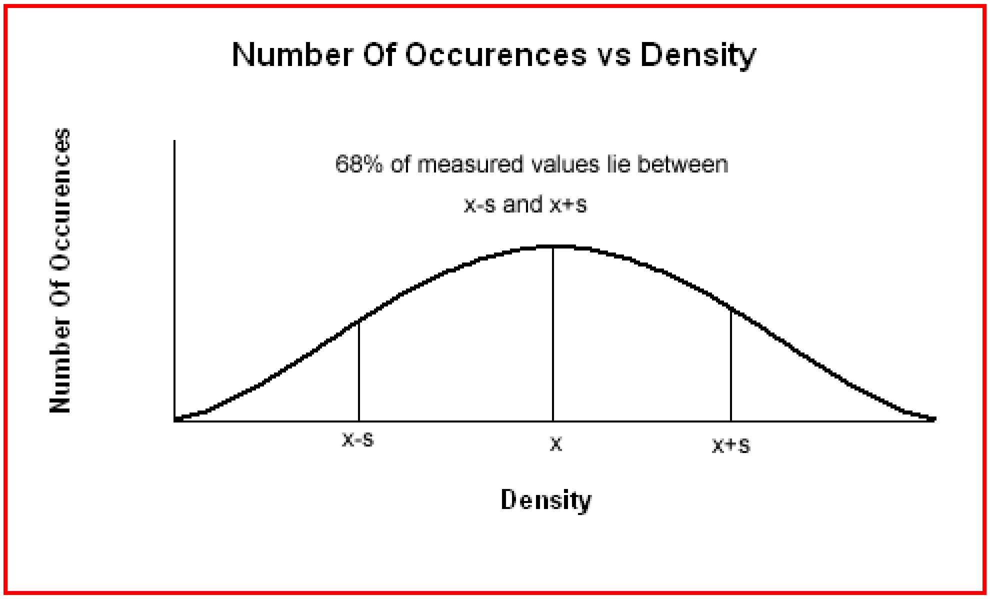
For purposes of illustration, we will assume **normal distributions**. Normal distributions are marked by curves that are symmetrical, single-peaked, and bell-shaped.

The graph of Number Of Occurences vs Density shown below produce an error curve for low precision data. When errors are large resulting in low precision data, the data are more spread out and the curve is broader and less sharp.



For high precision data, most values will be clustered around the average (although some will lie further away) and the curve is narrower and sharper.

Standard deviation is a measure of the width of the error curve and indicates the precision of the data. The smaller the standard deviation, s, the closer the density values are to the average density while a large standard deviation means the density values lie further from the mean. The graph on the next page shows that when errors are large (less precise data), the data are more spread out and the error curve is broader and less sharp.



The standard deviation for the density data is 0.345 g/cm^3. Standard deviation always has the same units as the original data. A standard deviation of 0.345 g/cm^3 means that 68% of subsequent density measurements will lie within ± 0.346 g/cm^3 of the average value (± 1 standard deviation). The values will lie in the range:

$8.43\ g/cm^3 \leq 8.78\ g/cm^3 \leq 9.13\ g/cm^3$

It also means that 95% of further density values will lie within ± 2(0.346 g/cm^3) or 0.652 g/cm^3 of the average value (± 2 standard deviations). The values will lie in the range:

$8.09\ g/cm^3 \leq 8.78\ g/cm^3 \leq 9.47\ g/cm^3$