Simple Statistics Used In Chemistry

Your study of chemistry will involve many instances where statistical calculations are necessary. The calculators you are required to furnish for class can easily perform all of the calculations below. You may ask your instructor to demonstrate these features on your particular model of calculator. However, since there are many new models, each of which operates differently, he/she may not be able to help you with your particular model. For help with various calculators, go to one of the following links. In particular, the Texas Instruments link contains a wealth of information, including online manuals.

Texas Instruments: http://education.ti.com
Casio: http://www.casio-usa.com/calculators
Sharp: http://www.sharp-usa.com/products/FunctionLanding/0,1050,4,00.html
Hewlett-Packard: http://www.hp.com/calculators

The statistical functions shown below are routinely used in the analysis of data, where x_i represents each of the individual data points and n represents the total number of data points.

- Mean ($\bar{x}$): Calculated by summing together all values and divide by the total number of data points.

$$\bar{x} = \frac{\sum x_i}{n}$$

- Error: Calculated by subtracting the theoretical value from the experimental mean. Result may be positive or negative.

$$\text{Error} = \bar{x} - x_{theoretical}$$

- Relative error (percent error): To calculate, divide error by the theoretical value, and then multiply by 100. Result may be positive or negative.

$$\text{Relative Error} = \frac{\bar{x} - x_{theoretical}}{x_{theoretical}} \times 100\%$$

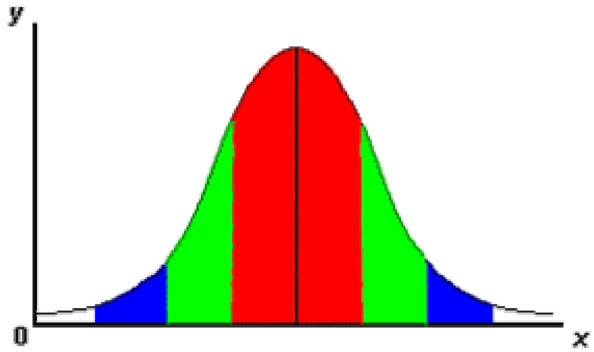
- Deviation (d): Calculated by subtracting the mean from an experimental data point (each data point has a deviation from the mean). Result may be positive or negative.

$$d = x_i - \bar{x}$$

- Standard deviation (σ): To calculate, square each of the deviations and sum them together. Divide the result by the number of samples minus 1. Finally, take this result and calculate the square root (explained in greater detail below). The standard deviation is represented as plus or minus a value, as in $\pm$ 0.05 g.

$$\sigma = \sqrt{\frac{\sum d^2}{n-1}}$$

The standard deviation is a useful tool, as it represents the precision, or scatter, of data around the mean. The graph is a representation of the standard deviation as it relates to the mean. In a large pool of data, 68% of all data should lie within plus or minus one standard deviation from the mean ($\pm\sigma$), represented by the area shaded in red. 95% of all data should lie within $\pm 2\sigma$ of the mean, represented by the green and red shaded areas together. Finally, 99% of all data lies within $\pm 3\sigma$ from the mean, represented by all of the shaded areas. Any reference you may have heard to the so-called "bell-shaped curve" refers to the distribution of data around the mean.



Consider the following example to show how the standard deviation would be calculated. In this example, the mass of water delivered by a 5.00-mL pipet was measured in four separate trials. The masses obtained were 4.9871 g, 4.9638 g, 5.0008 g, and 4.9711 g.

H_2O	mass (g)	d (g)	d^2 (g^2)
sample 1	4.9871	0.0064	0.0000410
sample 2	4.9638	-0.0169	0.0002856
sample 3	5.0008	0.0201	0.0004040
sample 4	4.9711	-0.0096	0.0000922
mean	4.9807		
std dev	0.0166		
error	-0.0193		
percent error	-0.386%		

By definition, for a measurement to have the correct number of significant figures, there should only be one uncertain digit at the end of the number. The standard deviation in the example above is ± 0.0166 g. Since this shows three uncertain digits, the standard deviation should be rounded to ± 0.02 g. Then the mean should be rounded to 4.98 g. You should then report the average mass delivered as 4.98 ± 0.02 g. The example below shows how the above table should be corrected to reflect the correct number of significant figures. This correction is very easily done using spreadsheets. You will learn more about spreadsheets in the following two labs.

H_2O	mass (g)	*d* (g)	d^2 (g^2)
sample 1	4.9871	0.0064	0.0000410
sample 2	4.9638	-0.0169	0.0002856
sample 3	5.0008	0.0201	0.0004040
sample 4	4.9711	-0.0096	0.0000922
mean	4.98		
std dev	0.02		
error	-0.02		
percent error	-0.4%		

Since it is assumed you will have access to scientific calculators and spreadsheets, it is not necessary to calculate the deviation or deviation squared. However, if you do not yet know how to use your scientific calculator or spreadsheets, you will need to perform the calculations longhand to arrive at your results.

PROBLEM

Try the following example to see how well you understand. A student measured the volume of a metal cylinder in four separate trials, obtaining these results: 7.525 cm^3, 7.194 cm^3, 7.311 cm^3, and 7.838 cm^3. Calculate the mean and standard deviation. Report the results with the correct number of significant figures. To check your answer, go to http://webs.anokaramsey.edu/chemistry, click on “Chem 1061”, then click on “Statistics Answer”.