

Error Analysis—General Chemistry Laboratory

Error and uncertainty may seem synonymous with trivial mistakes in the lab, but they are well defined aspects of any numerical measurement in any laboratory experiment. A number reported without consideration of precision tells an incomplete story, and a goal in the General Chemistry laboratory is to start thinking about the value and the *precision* of any numerical result—whether it is a laboratory measurement, a survey result or a sports statistic. In the lecture portion of the class, we will spend some time mastering significant figures. Significant figures are an approximate approach to treat precision, and they provide rules for addition and multiplication. In lab we'll work towards an understanding of uncertainty in laboratory measurements. This guide is sufficient for your work in General Chemistry. More information can be found in John R. Taylor's *An Introduction to Error Analysis*¹, a copy of which is on reserve at the Olin Science Library.

Reporting Error

Whenever possible report numerical results with an estimate of the error of the value. In General Chemistry we'll estimate experimental error either using the standard deviation of the mean from a set of replicate experiments or using the errors in the slope and intercept of linear fits to experimental data. Both these methods are described below. By convention error is reported with two digits, and the precision of the error determines the precision of the measurement.

Reporting Absolute Error

The absolute error has the same units as the measurement. For example, the mass of 100 mL of water at 25°C is measured to be 99.62 ± 0.23 g. The 0.23 g is the absolute error in the measurement.

Reporting Relative Error

The relative error is the size of the error as compared to the value of the measurement. Mathematically, relative error is the absolute error divided by the mean. Relative error is reported as a percentage, with two digits. In the above example, the relative error is:

$$\text{relative error} = \frac{\text{absolute error}}{\text{mean}} 100\% = \left(\frac{0.23}{99.62} \right) 100\% = 0.23\%$$

Using relative error, the mass of 100 mL of water at 25°C is reported as $99.62 \text{ g} \pm 0.23\%$.

Significance

In General Chemistry, the word “significant” doesn't mean “important”. Rather significance means “justified considering the inherent errors” in the experiment. We'll use the error as expressed by the standard deviation to gauge the significance of results. The standard deviation is often represented as error bars that span the range $\bar{x} \pm s(x)$, where $s(x)$ is the standard deviation of the mean. If two measured values are contained within the range of each other's standard deviations, the measured values are the same — we can't experimentally tell the difference between the two numbers. If two measured values are not contained within the range of each other's standard deviations, then the two values are unique, and their difference is “significant”. There is a grey area between those two extremes, and some results can be inconclusive. A full

understanding of the significance of inconclusive results requires other statistical tests such as a t-test. More information can be found on page 150 of Taylor.

Random and Systematic Error

In the lab, we'll put a lot of effort into controlling the two types of experimental error: systematic error and random error. **Systematic error** arises from a flaw in experimental design or equipment and can be difficult to detect and correct. Systematic error always has the same effect if the experiment is repeated. Systematic error determines the accuracy of the result. Causes of systematic error are typically instrument calibration or environmental effects, such as a drift in lab temperature. The best way to check for systematic error is to use different methods to perform the same measurement. **Random error** is always present. Random errors are different each time the experiment is repeated. Random errors are sometimes positive and sometimes negative. Random errors are minimized by averaging the results of many trials. Random error determines the precision of measurements in laboratory. An example of random error is the reading of a burette, which is somewhat subjective and therefore varies at each reading. Note: "error" and "uncertainty" are sometimes used interchangeably to mean "random error". The phrase "error in a measurement" is synonymous with "uncertainty in a measurement". In an experiment, we aim to *eliminate* systematic error and *minimize* random error to obtain a high degree of both accuracy and precision. A goal of the General Chemistry laboratory is to practice thinking about the largest contributors to both types of error in our experiments.

Random Error in Lab Experiments

Systematic error is corrected for in the lab procedure. Repeated measurements are used to estimate random error. The statistics of the set of replicates give us a way to understand the errors in our measurement. The statistical tools we'll use are the Mean, the Standard Deviation, and the Standard Deviation of the Mean.

Mean, Standard Deviation, and Standard Deviation of the Mean

The **mean**, \bar{x} , is the average of n-replicate measurements, x_i :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

The sample **standard deviation**, $s(x)$, is a measure of the spread in values of replicates of the experiment:

$$s(x) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad \text{with the sum over all data values } i = 1 \text{ to } n$$

The **standard deviation of the mean**, $s(\bar{x})$, is also called the estimated standard error. If you were to repeat your full experiment with all your replicas, the spread in your averages is based on the standard deviation of the mean:

$$s(\bar{x}) = \frac{s(x)}{\sqrt{n}}$$

Using Excel

More excel help is available on the course website document *Excel Help*. Here the help is limited to calculating \bar{x} , $s(x)$, and $s(\bar{x})$. The mass of 100-mL of water at 25°C is determined using a volumetric flask. The measurement is repeated 5 times, giving the following masses in grams: 99.794, 99.805, 99.198, 99.829 and 99.541. A spreadsheet to calculate the mean, standard deviation and standard deviation of the mean is:

	A	B
1		
2	Replica	mass (g)
3	1	99.794
4	2	99.805
5	3	99.198
6	4	99.829
7	5	99.541
8	mean	99.6 g
9	SD	0.3 g
10	SDOM	0.1 g

	Replica	mass (g)
1		99.794
2		99.805
3		99.198
4		99.829
5		99.541
mean	=AVERAGE(B3:B7)	g
SD	=STDEV(B3:B7)	g
SDOM	=B9/SQRT(COUNT(B3:B7))	g

with the actual spreadsheet to the left and the equations to calculate these quantities on the right. The number of decimal places shown for the mean, SD, and SDOM were adjusted using the buttons in the toolbar.

A fun (and wise) thing to do is to program the mathematical definitions for the mean and standard deviation into Excel to verify that the “=AVERAGE()” and “=STDEV()” functions do what they’re supposed to do.

Multiplying and Adding with Error

Once we’ve made a measurement, we’ll usually do some dimensional analysis with the results. There are easy rules to follow for multiplication and addition:

Multiplication/Division

Multiplication or division *scale* a quantity, so they *scale* the error. Using the mass of water in the example above to calculate the density of water at 25°C, the density is the mass divided by the volume. Assuming the volume is accurately known, compared to the mass:

$$99.6 \pm 0.1 \text{ g} \times \frac{1}{100.0 \text{ mL}} = 0.996 \pm 0.001 \text{ g mL}^{-1}$$

Both the value and the absolute error are divided by the volume.

Addition/Subtraction of a Constant

Addition or subtraction of a constant *shift* a quantity, so don’t change the size of the error. For example, to convert the temperature of the room, $25.3 \pm 0.3^\circ\text{C}$, to kelvins:

$$(25.3 \pm 0.3^\circ\text{C}) + 273.15 \text{ K} = 298.4 \pm 0.3 \text{ K}$$

The value 273.15 is added to the measured temperature. This simple conversion results because the Kelvin and Celsius scales have the same degree size, just a different origin.

Error in Linear Fits with Excel

The best way to estimate errors from measurements is using linear curve fits done by Excel. For example, using volumetric flasks of different sizes, the masses of water at 25°C are determined are given in the spreadsheet below. The following graph is constructed using the instructions in the *Excel Help* document:

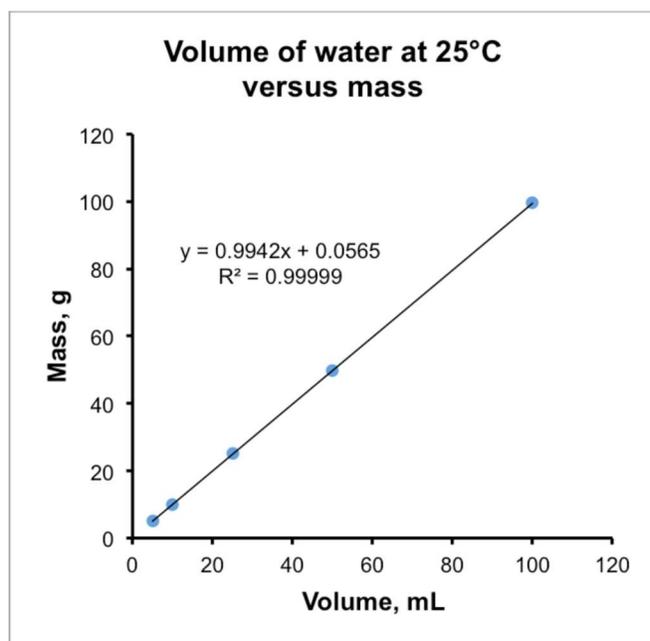


Figure 1. The mass of water at 25°C increases linearly with volume giving a slope of 0.9942 g mL^{-1} and a y-intercept of 0.0565 g.

While it is good to report the R^2 -value to give a sense of the precision of the fit, a better measure of the goodness-of-fit is the standard deviations of the slope and intercept. We can use Excel's =LINEST() function to calculate the standard deviations of these fit parameters. The following spreadsheet demonstrates the output of the =LINEST().

	A	B	C	D	E	F	G
1							
2	Volume(mL)	Mass(g)			slope	intercept	
3	5	5.0123		value	0.9941541	0.05646434	
4	10	9.9581		error	0.0015921	0.08195844	
5	25	25.0673		r^2	0.99999231	0.12363158	s_y
6	50	49.6289		F	389911.365	3	DOF
7	100	99.5050		regression S	5959.70447	0.0458543	residual SS

The top two rows of the output from =LINEST() contain the value and standard deviations of the slope and intercept giving a slope of $0.9942 \pm 0.0016 \text{ g/mL}$ and a y-intercept of $0.056 \pm 0.082 \text{ g}$.

Using =LINEST()

To use Excel's =LINEST() function:

1. Highlight an empty block of cells 2 columns by 5 rows in size.
2. Immediately type "=LINEST(".
3. Highlight the cells containing the y-values, and then enter a comma.
4. Highlight the cells containing the x-values, and then enter a comma.
5. Type in TRUE, and then a comma.
6. Type in "TRUE)". Your cell should look like "=LINEST(B3:B7,A3:A7,TRUE,TRUE)".
7. Press the `command` ⌘ + `return` keys simultaneously.
8. The output follows the spreadsheet above, and you can label the =linest() parameters accordingly. The only entries you'll need to understand are the fit values, the standard deviations, and the R^2 value. You can look up the meanings of the other quantities, but they're beyond the scope here.

Reference

1. Taylor, J. R. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, 2nd Ed.*; University Science Books: Sausalito, CA, 1997.