Error Analysis—General Chemistry Laboratory
November 13, 2015

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 the value and 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 digits. Significant digits are an approximate approach to treat precision, and they provide rules for addition and multiplication. Significant digits have a lot of rules and less math. A full treatment of uncertainty in measurements has fewer rules with more math. In lab we’ll work towards an understanding of uncertainty in laboratory measurements.

This guide will be sufficient for your work in General Chemistry. More information can be found in John R. Taylor’s An Introduction to Error Analysis¹, a copy of which is on reserve at the Olin Science Library.

Reporting Error
Whenever possible report numerical results with error. In General Chemistry we’ll estimate experimental error either using the standard deviation of the mean from a set of replica experiments or using the errors in slope and intercept in linear fits to experimental data. Both these methods are described below. Error is reported with one significant digit, 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, in lab, I measured the mass of 100-mL of water at 25°C to be 99.6 ± 0.1 g or 9.96 ± 0.01 × 10¹ g in scientific notation. The 0.1 g is the error in our measurement.

Reporting Relative Error
The relative error is the scale of the error with respect to the value of the measurement. Mathematically, it is the absolute error divided by the mean. It’s reported with units of percentage, and it has one significant digit. In the above example, I’d report the mass of 100-mL of water at 25°C is 99.6 g ± 0.1%.

Significance
In General Chemistry, the word “significant” doesn’t mean “important”. We’ll use our error bars to gauge the significance of measurements. If two measured values are contained within each other's error bars, the measured value is the same—we can’t experimentally tell the difference between the two numbers. If two measured values and their error bars are well separated, 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 detected and corrected. This type of error leads to inaccurate measurements of the true value. The best way to check for systematic error is to use different methods to perform the same measurement.

Random error is always present and cannot be corrected. It has to do with the precision of measurements in laboratory; it is the statistical uncertainty of the last digits of precision. An example of random error is that which arises from reading 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. Understanding random error comes from repeated measurements to give a set of replicas. The statistics of the set of replicas give us a way to understand the value and error 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 simple average of your replicas, $x_i$.

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

The standard deviation, $\sigma_x$, is related to the spread in values of replicas of your experiment. We will use the population standard deviation to approximate it:

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}.$$

The standard deviation of the mean, $\sigma_{\bar{x}}$, is also called the estimated standard error. If you were to repeat your full experiment with all your replicas, the spread in your average results should be related to the standard deviation of the mean:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}.$$

Using Excel
More excel help is available on the course website document Excel Help. Here the help is limited to calculating $\bar{x}$, $\sigma_x$, and $\sigma_{\bar{x}}$.

Let’s say I measured 100-mL of water at 25°C using a volumetric flask and weighed it. I repeated
this 5 times, and got the following masses in grams: 99.794, 99.805, 99.198, 99.829 and 99.541. The spreadsheet I used to calculate the mean, standard deviation and standard deviation of the mean is

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 would be 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 it. There are easy rules to follow for multiplication and addition:

#### Multiplication/Division

Multiplication and division scale a quantity, so they scale the error. If I wanted to use the mass of water in the example above to calculate the density of water at 25°C, I can divide my mass by the volume:

\[
99.6 \pm 0.1 \text{g} \times \frac{1}{100\text{-mL}} = 0.996 \pm 0.001 \text{g/mL}. \tag{1}
\]

I simply divided both the value and the absolute error by the volume I measured.

#### Addition/Subtraction

Multiplication and division shift a quantity, so don’t affect error. Say I measured the temperature of the room to be 25.3 ± 0.3°C, and I wanted to convert it to Kelvin:

\[
25.3 \pm 0.3^\circ\text{C} + 273.15 = 298.4 \pm 0.3 \text{K}. \tag{2}
\]

I simply added 273.15 to the measured temperature. This is because the Kelvin and Celcius have the same scale, just a different origin.
Error in Linear Fits with Excel

One last place to get errors from measurements is from the linear fits done by Excel. Using volumetric glassware, I measured different volumes of water at 25°C and weighed them. I make the following graph following the instructions in the Excel Help document:

![Graph showing mass of water at 25°C versus volume with a linear fit equation and R² value.]

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

While it’s good to give the R²-value to give a sense for the precision of the fit, it’s not easy to relate the R²-value to the precision on the slope and intercept. We can use Excel’s =LINEST() function to calculate the errors in these fit parameters. The following spreadsheet demonstrates the output of the =LINEST().

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Volume (mL)</td>
<td>Mass (g)</td>
<td>value</td>
<td>slope</td>
<td>intercept</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>5.0123</td>
<td>0.9941541</td>
<td>0.05646434</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>9.9581</td>
<td>0.0015921</td>
<td>0.08195844</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>25.0673</td>
<td>0.99999231</td>
<td>0.12363158</td>
<td>s_y</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>49.6289</td>
<td>389911.365</td>
<td>4.0548543</td>
<td>residual SS</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>99.5050</td>
<td>5959.70447</td>
<td>0.0458543</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The top two rows of the output from =LINEST() contain the value and errors for the slope and intercept. We can use this data to correct the above figure caption:

Figure 1. The mass of water at 25°C increases linearly with volume giving a slope of 0.994 ± 0.002 g/mL and a y-intercept of 0.06 ± 0.08 g.
Using =LINEST()
To use Excel's =LINEST(),

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, hit .
4. Highlight the cells containing the x-values, hit .
5. Hit again.
6. Type “TRUE)”. Your cell should look like “=LINEST(B3:B7,A3:A7,TRUE)”.
7. Hit +.
8. The output follows the spreadsheet above, and you can label it as such. The only entries you'll need to understand are the values, the errors, and the $R^2$ value. You can look up the meanings of the other quantities, but they’re beyond the scope here.

Reference