

ERROR ANALYSIS REVIEW

Prof. Jeff Yarger, Dept. Chemistry & Biochemistry, Arizona State University, Tempe, AZ 85287-1604

Introduction

In science, the word “error” does not carry the usual connotations of “mistake” or “blunder”. Error in a scientific measurement means the inevitable uncertainty involved in all measurements. As such, errors are not mistakes; you cannot avoid them even when being very careful. The best you can hope to do is minimize errors and to have a reliable estimate of error in any measurement or observation.

When discussing errors in your experimental data or observations, it is customary to distinguish between precision versus accuracy or *systematic* versus *random* errors, respectively. Systematic errors are errors that will make results differ from the “true” values with reproducible discrepancies. Therefore, the accuracy of an experiment is generally dependent on how well we can control or compensate for systematic errors. Common examples include (i) failure to establish a proper instrument calibration, (ii) improper graduation or alignment of an instrument scale, (iii) drift in instrumental property, (iv) incomplete material transfer or leakage, and (v) faulty approximation or theory. The precision of an experiment is dependent on how well we can overcome random errors. These are the fluctuations in observations that yield results that differ from experiment to experiment and often require repeated collection to yield precise results. Accuracy versus precision is further discussed in introductory chemistry and analytical chemistry [1-3]. In experimental physical chemistry, the focus is typically on random errors and will assume systematic errors are minimal.

An important skill in a scientific laboratory is the ability to estimate error or uncertainties in a measurement. Consider for example that you want to measure the volume of a liquid using a graduated cylinder, and find the volume to be 292.5 mL. This number is subject to uncertainty: It is unthinkable that you could know the volume to be exactly 292.5000 mL rather than 292.5001 mL. The difference between the two is 0.1 μL , which would make a tiny drop hardly visible by the human eye. It is clear that you cannot measure with such precision using measuring a graduated cylinder.

Let's suppose the graduate cylinder is graduated every five milliliters. You will have to estimate where the top of the liquid lies relative to the markings of the cylinder, and this necessity causes some uncertainty in the measurement. No physical quantity (a length, time, temperature, etc) can be measured with complete certainty.

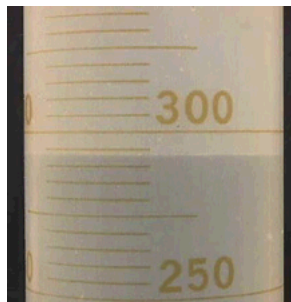


Figure 1 – Picture of a graduated cylinder for measuring the volume of a liquid. The graduation marks on the glass cylinder are calibrated for the volume in milliliters. The meniscus of the liquid is between the ticks for 290 mL and 295 mL. Therefore, we can estimate the volume to be half way between these graduation marks with an estimated error associated with our ability to estimate the volume, $292.5 \text{ mL} \pm 2.5 \text{ mL}$.

Estimating Error

How can we evaluate the magnitude of uncertainty in a measurement? Such evaluation can be fairly complicated, and we will just cover some basic principles. However, this is a very important skill in physical chemistry and is often used because multiple measurements is not a viable option in lots of measurements and scientific procedures. Let's consider first the graduated cylinder shown in figure 1. You could reasonably decide that the volume is $290 \text{ mL} \leq V \leq 295 \text{ mL}$. We will write this statement as:

$$V = 292.5 \pm 2.5 \text{ mL}$$

Note that the estimated error depends on the separation in the markings on the cylinder. This is a 'safe' estimation. Often, you can estimate a quantity with more accuracy than half the separation between markings. A traditional thermometer would be measured in a similar fashion. However, it is more common in modern experimental chemistry to use thermocouples with digital readouts to determine the temperature.

Consider now the thermocouple with a digital display shown in figure 2. The display reads 199.8°F . How can we estimate the uncertainty? The safest way to do this is by referring to the documentation provided with the instrument by the provider. For example, the manual for this particular thermometer says "Accuracy is $\pm 0.06^\circ\text{F}$ from -40 to $+99.99^\circ\text{F}$ ($\pm 0.03^\circ\text{C}$ from -40 to $+99.99^\circ\text{C}$); $\pm 0.1^\circ\text{F}$ from 100 to 257°F ($\pm 0.1^\circ\text{C}$ from 100 to 125°C); $\pm 0.2^\circ\text{F}$ from 257 to 302°F ($\pm 0.2^\circ\text{C}$ from 125 to 150°C)." Thus, we would report this value as

$$T = 199.8 \pm 0.1^\circ\text{F}.$$

If the manual is not available, the uncertainty is commonly estimated by observing the fluctuations in the temperature or making several separate measurement of the temperature using the same thermocouple and digital display. Estimating error from digital readouts and reporting numbers to a reasonable degree of significant figures is a very important skill and keeps us from providing misleading results. For example, if we report the temperature as 99.999999°F , this implies that we have the precision and accuracy to determine temperature to within one one-millionth of one degree. While given the specification of the above thermocouple and most common thermocouples, it is more appropriate to write the temperature as 99.9°F , no matter how many significant figures are given on the display (similar to a scientific calculator). Lastly, what we are asking for here is that all experimental numbers be listed with its estimate or determined uncertainty.

In general, the result of any measurement of a quantity x should be stated as

$$x = x_{\text{best}} \pm \delta x$$



Figure 2 – Picture of a Digi-Sense digital display thermocouple from Cole-Palmer. The display shows the temperature in degrees Fahrenheit ($^\circ\text{F}$).

where x_{best} is the best estimate for the quantity concerned and δx is called the uncertainty, or error, in the measurement of x . We'll discuss in a moment how to get the best estimate.

Reporting Uncertainty

Since the quantity δx is an estimate of an uncertainty, it should obviously not be stated with too much precision. In the example of the graduate cylinder above, it would be absurd to state the result as $V = 292.5 \pm 2.52356$ mL. Uncertainties are most often reported with one or two **significant figures**¹. This difference depends somewhat on the field and personal taste of the researcher. Once the uncertainty in the measurement has been estimated, one must also consider the number of significant figures in the measured value. A statement like $V = 292.5256712 \pm 2.5$ mL is obviously ridiculous. The uncertainty of 2.5 mL means that the digit 2 before the decimal point might really be as small as 0 or as large as 4. Clearly the digits 2 5, 6, 7, ..., after the decimal point have no significance at all and should be rounded off. The last significant figure in any stated answer should be in the same decimal position as the uncertainty. In the last case, this means $V = 292.5 \pm 2.5$ mL.

If the measured number is so small or large that it calls for scientific notation, then it is simpler and clearer to put the answer and uncertainty in the same form (see example below).

In summary:

1. The value and the uncertainty should be reported to the same number of decimal places: (75.63 ± 0.06) kg, not (75.6347 ± 0.06) kg or (75.63 ± 0.0629) kg.
2. The value and the uncertainty should have the same power of 10 if you are using scientific notation. One good format is $(4.73 \pm 0.08) \times 10^{-5}$ J. You make readers do extra work with $(4.732 \times 10^{-5} \pm 8.5 \times 10^{-7})$ J.
3. There is no generally accepted convention for the number of significant figures to use in reporting the uncertainty. Some authors report one, some use two. Give your result to the same number of decimal places as you give its uncertainty. If your calculator gave 752.2083 torr ± 0.0143 torr, you should report (752.208 ± 0.014) torr.
4. The units should be clear. In these examples we have used parentheses to make it clear that the unit applies to both the number and its error, but it's also acceptable to place the unit explicitly on both value and uncertainty: 752.208 torr ± 0.014 torr.

The Mean and Standard Deviation

Suppose we need to measure some quantity x , and have identified and minimized all sources of systematic error. Since all remaining sources of uncertainty are random, we should be able to detect them by repeating the measurement several times. We might, for example, measure a temperature five times, and find the following results:

| | | | | |
|---------|---------|---------|---------|---------|
| 53.4 °C | 53.2 °C | 54.0 °C | 53.9 °C | 53.7 °C |
|---------|---------|---------|---------|---------|

¹ A tutorial on significant figures is available as Appendix 1.

It seems reasonable that the best estimate of the temperature is the **average** or **mean** of the five values we found. In general:

$$x_{best} = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{N}$$

where N is the number of measurements and the bar on top of the quantity x means “average” or “mean”.

In the example above, $T_{best} = 53.64\text{ }^{\circ}\text{C}$. We’ll discuss in a moment how to determine the appropriate number of significant figures we should use to report this value. The **standard deviation** is an estimate of the average uncertainty of the measurements x_1, x_2, \dots, x_n . Given that the mean, \bar{x} , is our best estimate of the quantity x , it is natural to consider the difference $x_i - \bar{x} = d_i$. If these differences are very small, then the measurements are all close together and very precise. If the differences are large, then the measurements are largely scattered and the result will not be very precise. For the example above:

| i | 1 | 2 | 3 | 4 | 5 |
|-----------------------|----------|----------|---------|---------|---------|
| T_i | 53.4 °C | 53.2 °C | 54.0 °C | 53.9 °C | 53.7 °C |
| $d_i = T_i - \bar{T}$ | -0.24 °C | -0.44 °C | 0.36 °C | 0.26 °C | 0.06 °C |

to estimate the uncertainty of the measurements we might naturally try averaging the deviations d_i . However, the average of the deviations is zero (try taking the average of the third row in the table above). Some values of d_i are positive and some negative, so the average of the deviations is not a useful way to characterize the reliability of the measurements. The best way to avoid this is to square all the deviations, which will create a set of positive numbers. The square root of the average of the squares of the deviations is called the **standard deviation**, and is denoted σ_x .

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

| i | 1 | 2 | 3 | 4 | 5 | |
|-----------------------------|----------|----------|---------|---------|---------|------------------------------|
| T_i | 53.4 °C | 53.2 °C | 54.0 °C | 53.9 °C | 53.7 °C | |
| $d_i = T_i - \bar{T}$ | -0.24 °C | -0.44 °C | 0.36 °C | 0.26 °C | 0.06 °C | $\sum_{i=1}^5 d_i = 0$ |
| $d_i^2 = (T_i - \bar{T})^2$ | 0.0576 | 0.1936 | 0.1296 | 0.0676 | 0.0036 | $\sum_{i=1}^5 d_i^2 = 0.452$ |

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} = \sqrt{\frac{0.452}{5}} = 0.30066...$$

If you choose to report errors with one significant figure, you would write: $\sigma_x = 0.3$, so each measured temperature would be reported as:

| | | | | |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| $(53.4 \pm 0.3) ^\circ\text{C}$ | $(53.2 \pm 0.3) ^\circ\text{C}$ | $(54.0 \pm 0.3) ^\circ\text{C}$ | $(53.9 \pm 0.3) ^\circ\text{C}$ | $(53.7 \pm 0.3) ^\circ\text{C}$ |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|

Note that the standard deviation describes the uncertainty of each of the five individual measurements. However, the uncertainty of the mean is lower as we'll see in the next section. Before we continue, you should be aware of an alternative definition of the standard deviation:

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

The two definitions give a similar result when N is large. The distinction is important in statistics, but it is beyond the scope of this class [references].

For many cases where multiple measurements are made, you will want to find the **standard deviation of the mean** value. If x_1, x_2, \dots, x_n are the results of N measurements of the same quantity x , the best estimate for the quantity x is the mean. The standard deviation characterizes the average uncertainty of the separate measurements x_1, x_2, \dots, x_n . However, \bar{x} is more reliable than any individual measurement, so it is expected that its uncertainty is smaller than the corresponding uncertainty for the individual measurements. The uncertainty in \bar{x} turns out to be:

$$\sigma_{\bar{x}} = \sigma_x / \sqrt{N}$$

Going back to our previous example:

| i | 1 | 2 | 3 | 4 | 5 |
|-------|---------|---------|---------|---------|---------|
| T_i | 53.4 °C | 53.2 °C | 54.0 °C | 53.9 °C | 53.7 °C |

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = 53.64$$

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} = \sqrt{\frac{0.452}{5}} = 0.30066... \Rightarrow \sigma_{\bar{x}} = \frac{0.30066...}{\sqrt{5}} = 0.13446...$$

Thus, you would report the mean value as $T = (53.6 \pm 0.1)^{\circ}\text{C}$ or $T = (53.64 \pm 0.13)^{\circ}\text{C}$ depending on whether you choose to report the error with one or two significant figures.

In summary, the standard deviation (σ_x) represents the uncertainty of each individual measurement, while the standard deviation of the mean (σ_x / \sqrt{N}) is a measure of the uncertainty of the mean value.

Rejection of Data

Every now and then, you will come across an experimental result that simply does not seem to make sense. For example, let's say you perform a sixth measurement of temperature and obtain

| i | 1 | 2 | 3 | 4 | 5 | 6 |
|-------|---------|---------|---------|---------|---------|---------|
| T_i | 53.4 °C | 53.2 °C | 54.0 °C | 53.9 °C | 53.7 °C | 54.9 °C |

The last measurement seems a bit off, and you may be tempted to throw it out of the set on aesthetic grounds alone. However, you must never throw out a result from a data set unless you have a statistical or chemical reason to do so. First, you should use your common sense: did something unusual happen during that measurement? Sometimes one realizes that a particular determination wasn't performed properly, and it would make sense to 'flag it' in your notebook and eventually discard it if it is off. Statisticians have devised many rejection tests for the detection of non-random errors. We will describe only one, the **Q test**, which works well in cases where $3 < N < 10$. The quantity Q is defined by

$$Q = |(\text{suspect value}) - (\text{value closest to it})| / [(\text{highest value}) - (\text{lowest value})]$$

In our example above,

$$Q = (54.9 - 54.0) / (54.9 - 53.2) = 0.53$$

We now compare this value of Q with the critical value Q_c corresponding to the number of observations:

| Q_c (90% confidence) | 0.94 | 0.76 | 0.64 | 0.56 | 0.51 | 0.47 | 0.44 | 0.41 |
|------------------------|------|------|------|------|------|------|------|------|
| N | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

If Q is equal to or larger than Q_c , the suspect measurement can be rejected. In our case, $N = 6$, so $Q < Q_c = 0.56$. Thus, the suspect measurement cannot be rejected.

Propagation of Error

Most physical quantities cannot be measured in a single direct measurement. Instead, they are calculated from quantities measured in different steps. For example, to measure the density of a substance one would measure its volume, its mass, and then calculate the density (ρ) from the definition:

$$\rho = \frac{\text{mass}}{\text{Volume}} = \frac{m}{V}$$

To calculate the uncertainty in the density, one must first determine the uncertainty in the measured quantities, m and V , and **propagate** through the calculation.

Suppose that x, \dots, z are measured with uncertainties $\delta x, \dots, \delta z$, and the measured values are used to compute a function $q(x, \dots, z)$. If the uncertainties in x, \dots, z are independent and random, then the uncertainty in q is

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x} \delta x\right)^2 + \dots + \left(\frac{\partial q}{\partial z} \delta z\right)^2}$$

In this equation, the term $\frac{\partial q}{\partial x}$ represents the partial derivative of q with respect to x . If you need help calculating partial derivatives please see Appendix II or a general calculus book. You will need to be able to calculate partial derivatives in order to understand the remainder of this section.

For example, let's assume your goal is to calculate the density of a substance. You determine its mass $m \pm \delta m$ and its volume $V \pm \delta V$. You would calculate the density as $\rho = m/V$. But what about the uncertainty in the density? Is $\delta \rho = \delta m / \delta V$?

Using the equation for error propagation we have:

$$\delta \rho = \sqrt{\left(\frac{\partial \rho}{\partial m} \delta m\right)^2 + \left(\frac{\partial \rho}{\partial V} \delta V\right)^2}$$

The partial derivatives are:

$$\frac{\partial \rho}{\partial m} = 1/V \text{ and } \frac{\partial \rho}{\partial V} = -m/V^2$$

so

$$\delta \rho = \sqrt{\left(\frac{\delta m}{V}\right)^2 + \left(-\frac{m}{V^2} \delta V\right)^2}$$

For instance, if $m = (32.52 \pm 0.15) \text{ g}$ and $V = (35.17 \pm 0.52) \text{ mL}$

$$\rho = \frac{32.52}{35.17} \text{ g/mL} = 0.92465... \text{ g/mL}$$

$$\delta\rho = \sqrt{\left(\frac{0.15 \text{ g}}{32.52 \text{ mL}}\right)^2 + \left(-\frac{32.52 \text{ g}}{35.17^2 \text{ mL}^2} 0.52 \text{ mL}\right)^2} = 0.014428... \text{ g/mL}$$

If we choose to report our error with two significant figures, we would report the calculated density as:

$$\rho = (0.925 \pm 0.014) \text{ g/mL}$$

Note that we used the same number of decimal places for the value and error.

Least-Squares Fitting (working)

A- Data that should fit a straight line

Often, an experiment involves measuring several values of two different physical variables in order to investigate the mathematical relationship between the two. For instance, one might measure the volume of a mole of gas at 1 atm as a function of temperature (in °C) and see if the volumes and temperatures are connected by the expected relation:

$$V = 0.082 \text{ L/}^\circ\text{C} \cdot (T + 273.15 \text{ }^\circ\text{C})$$

This relationship is *linear*, that is, if we plot V as a function of T we expect a straight line.

Let's consider a general case where two physical variables y and x are connected by a linear relation of the form

$$y = A + Bx$$

where A and B are constants. A plot of y against x should give a straight line with slope B and y -intercept A .

Going back to the example above, imagine we make a series of measurements of volume and temperature. Since we know that the two variables are linearly related, we can ask ourselves which is the straight line $V = A + BT$ that best fits the measurements. The analytical method of finding the best straight line to fit a series of experimental points is called *linear regression*, or *least-squares fit*.

In this method, the sum of the squares of the residuals is minimized. Residuals are defined as the difference between the predicted and observed values. This concept is better understood graphically comparing the three graphs shown below. All three plots have the same data points, but different straight lines that were drawn according to the equations shown in each plot. Which of the three straight lines would you consider a better fit to the experimental data?

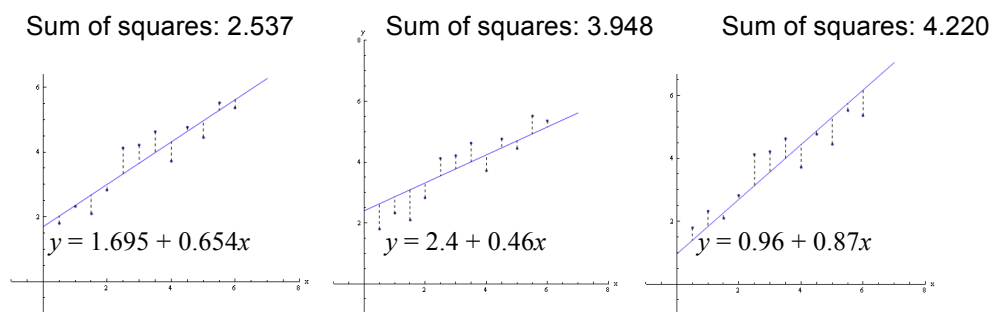


Figure 3. Three different straight lines superimposed to the same set of data points. The dotted lines represent the residuals.

In the least-squares method, we first define the residuals as the difference between the experimental value and the y -value predicted by the straight line. Residuals are shown as dotted lines in the plots in Figure 3. Note that residuals can be positive or negative, depending on whether the experimental point is on top or below the straight line. If we sum the residuals, we might get a small number just because the positive numbers cancel out with the negative values, so a small sum of residuals is not necessarily a good indicator of a good fit. To work around this problem, we'll calculate the sum of the squares of the residuals. We'll square all residuals first (to assure that all values are positive), and then calculate the sum of these values. A small value of the sum of the squares is an indication of a good fit, while a large value indicates a poor fit. In a least-squares method, the best straight line is calculated by minimizing the sum of the squares of the residuals. The plot on the left shows the result of the least-squares analysis for this data set, while the other two plots show straight lines that give a larger sum of squares (and thus represent poorer fits). One could imagine changing the slope and intercept by trial and error until the lowest sum of squares is obtained, but in fact there is an analytical solution to this problem.² That is, there are formulas that will give the best intercept and slope for a given set of experimental data.

The two graphs in Fig. 4 show two different data sets that give the same linear fit. That is, the y -intercept and slope obtained by performing linear regression is the same in both cases. However, there is a big difference between the two data sets: if we change the slope or intercept by a small amount, the sum of the squares changes more dramatically in the example on the right. The two graphs on the top row represent the best fit, which is the same for both sets of data. If we change the slope and intercept (bottom plots), the fit gets much worse for the data set on the right than what it does for the data set on the left. In other words, both the intercept and slope determined by the least-squares procedure have more uncertainty for the data set on the left, and are determined more accurately for the data set on the right. This is not surprising given that the points on the right are less scattered than the points on the left.

Thus, it is important to calculate the uncertainties in both the slope and intercept, and report them together with their values. The least-squares fit of the data set on the left gives a value of the slope $B = (0.082 \pm 0.029) \text{ L}^\circ\text{C}$ and a value of the intercept $A = (22.4$

² See for example "The Chemistry Maths Book" by Eric Steiner, or "Mathematical Methods for Scientists and Engineers" by Donald McQuarrie

± 1.6) L. The corresponding values for the data set on the right are: $B = (0.0819 \pm 0.0081) \text{ L}/^\circ\text{C}$ and $A = (22.41 \pm 0.43) \text{ L}$. Note that since the error is smaller for the data set on the right, we report the slope and intercept with more significant figures.

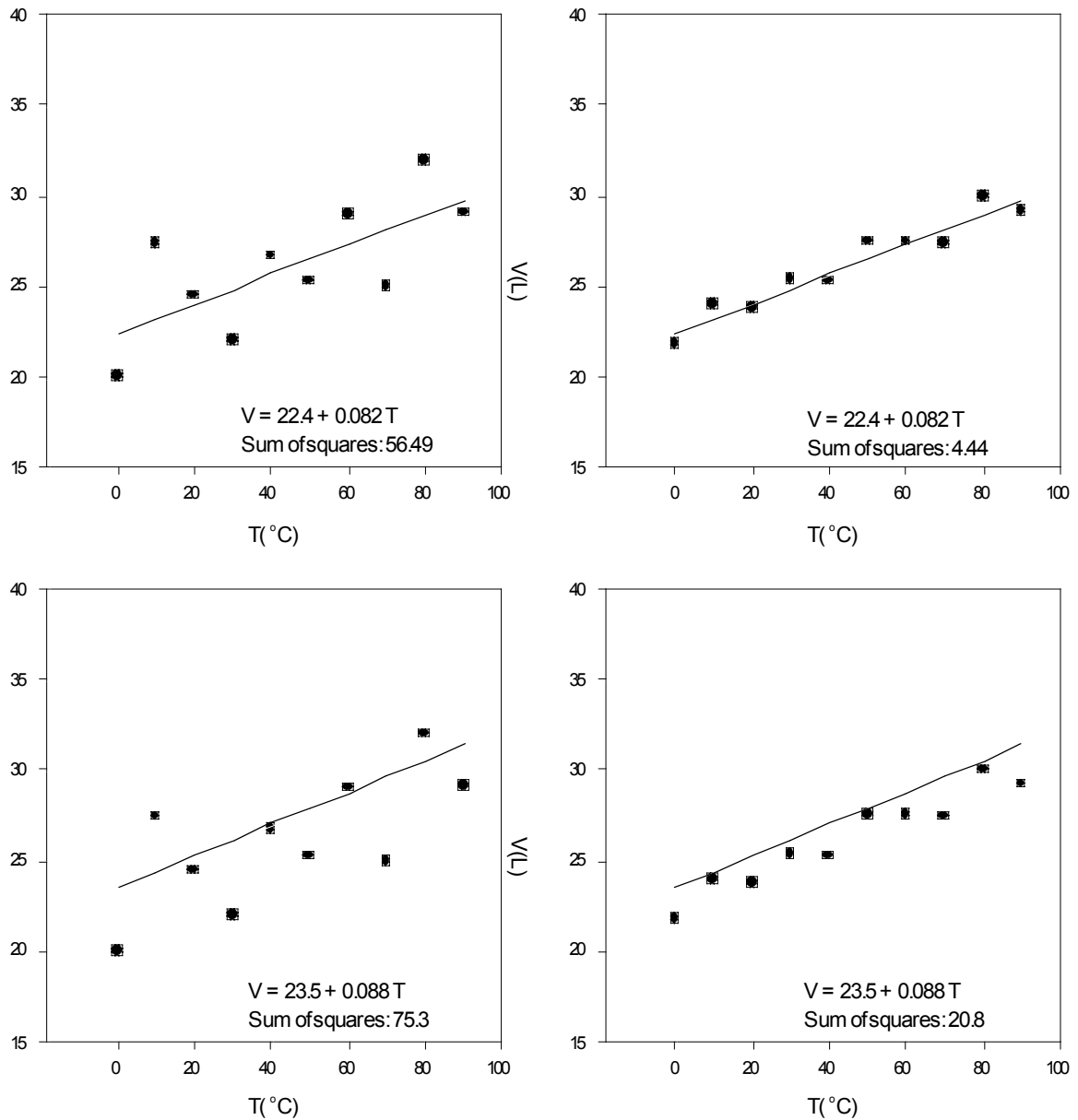


Figure 5. Two different data sets that are fitted by the same straight line. Top: result of the least-squares analysis. Although the slope and intercept are the same, the sum of squares of the residuals is not. Bottom: the same data sets plotted with an arbitrary straight line which is not the best fit. Both fits get worse, but the fit on the left is less sensitive than the fit on the right. Note that all data points are below this arbitrary straight line for the case in the right, while the points on the left are still scattered around the line.

B- Least-squares fits for other curves

We can extend the previous discussion to cases where the relationship between the variables is not linear. For example, consider an experiment where the concentration of a reactant is followed as a function of time. Let's assume that we know that for that particular reaction the concentration is expected to decay exponentially as:

$C = C_0 e^{-kt}$. Here, C_0 is the initial concentration of reactant, t is the time, and the parameter k is known as the rate constant. Since we know the initial concentration of reactant we used in the experiment, we can plot C/C_0 as a function of time and obtain the value of k through a least-squares analysis.

Figure 6 represents an example of experimental data obtained in such an experiment. The line in red represents the equation $C/C_0 = e^{-kt}$ with $k = 0.404 \text{ min}^{-1}$. But how do we know this is the best fit? Would the fit be better or worse for $k = 0.420 \text{ min}^{-1}$? To address this question, we can calculate the sum of the squares of the residuals for both cases and select the value with the lowest value. The residuals are again defined as the distance between the experimental point and the red line, and we take the squares to assure that we deal with positive values. This procedure is shown on graph 6B, which shows that the lowest value is obtained for $k = 0.404 \text{ min}^{-1}$. Thus, the red curve in figure 6A is the best fit for this set of experimental data.

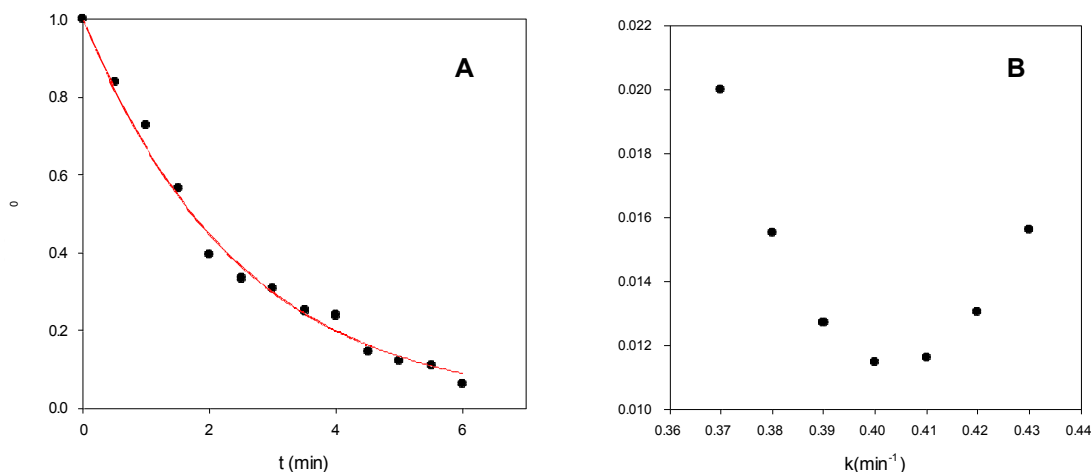


Figure 6. A- Experimental data (•) and best fit (—) according to the equation $C/C_0 = e^{-kt}$. The value of k obtained by least-squares was 0.404 min^{-1} . B- value of the sum of the squares of the residuals as a function of the value of the rate constant, k .

Finding the best value of k was relatively easy because one can easily calculate the sum of the squares of the residuals for many different values of k and identify the value for which this sum is lowest. This would not work for more complicated cases that involve more than one parameter. For instance, imagine that you study a more complex reaction for which C/C_0 decays as the sum of two exponentials:

$$C/C_0 = \alpha e^{-k_1 t} + (1 - \alpha) e^{-k_2 t}$$

In this case, we need to determine the best values of α , k_1 and k_2 , so performing a manual minimization of the sum of the squares of the residuals becomes a very complicated task.

Nonlinear regression is a general technique to fit data to any equation that defines Y as a function of X and one or more parameters. It finds the values of those parameters that generate the curve that comes closest to the data (that is, the parameters that minimize the sum of the squares of the vertical distances between the data points and curve). In contrast to the linear regression case, it is not possible to directly derive an equation to compute the best-fit values from the data. Instead nonlinear regression requires a computationally intensive, iterative approach. Luckily, there are many software programs that do this type of analysis. Choose a program that gives the statistical uncertainties as well, so you can report the best parameters and their corresponding errors. A tutorial on using *Kaleidagraph* for this purpose is available at <http://www.asu.edu/~mlevitus/chm343>.

Appendix: Reference Resources in Physical Chemistry.

You should always cite resources that are authoritative and credible.

- *Physical Chemistry Textbooks*

Any Physical Chemistry textbook can be used as a source of data. See especially the appendices in the back of the book for a wealth of thermodynamic and physical data. Cite books following this format:

Author 1; Author 2; etc. *Book Title*; Publisher: Place of Publication, Year; Inclusive Pagination.

For example:

Dodd, J.S. *The ACS Style Guide*, 2nd. ed.; ACS: Washington, D.C., 1997, 173-229.

- *The CRC Handbook of Chemistry and Physics*

CRC Press is a publishing company that specializes in printing specialty reference books. The most famous and widely used CRC Handbook in physical chemistry is the *Handbook of Chemistry and Physics*. This reference book contains a wealth of information, from melting points and boiling point for a variety of compounds to data on radioactive decay to thermochemical information. The CRC handbook is updated annually. ASU has access to the electronic version of this handbook:

<http://www.asu.edu/lib/resources/db/crcchem.htm>

The screenshot shows the ASU Libraries website interface. At the top is the ASU Libraries logo with the text "ARIZONA STATE UNIVERSITY". Below the logo is a navigation bar with links: ASU Home, ASU A-Z Index, ASU Interactive, My ASU, Colleges & Schools, Directory, Map. Below this is another navigation bar with links: All ASU Libraries, My Accounts, Hours, Library Catalog, Research Databases, Course Reserves, Get Help. The main content area is divided into three columns. The left column has a section "Ask a Librarian" with icons for email, phone, and chat, followed by a list of links: About the Libraries, Find Articles, Find Books & Media, Specialized Collections, Services, For Faculty & Staff, For Students, For Visitors. The middle column is titled "CRC Handbook of Chemistry and Physics" with a "Connect" button. Below this is a "Database Information" section with a "Full-text" button. The "Description" states: "(88th Edition) Contains tables and reference sections on all aspects of chemistry and physics. Includes all the figures, data, and tables found in the print version." The "Date Range" is "88th Edition; 2007-2008". The "Updated" status is "Yearly". The "Includes" section lists "full text and tables". The "Subject Coverage" is "Biochemistry; Chemistry; Physics". The "Types of Materials Covered" is listed. The right column is titled "Conditions of Use" and contains text about electronic resources being owned by independent providers and protected by copyright and other laws. It also states "You must:" followed by a bullet point "Use the resources for".

A print copy will be available in the laboratory. Cite this source indicating its edition and publisher. For example: *CRC Handbook of Chemistry and Physics*, 87th ed.; CRC: Boca Raton, FL, 2006.

Other Reference Books

Kirk-Othmer Encyclopedia of Chemical Technology:

<http://www.asu.edu/lib/resources/db/kirkothm.htm>

Ullmann's Encyclopedia of Industrial Chemistry

<http://www.asu.edu/lib/resources/db/eindchem.htm>

Specialized Journals

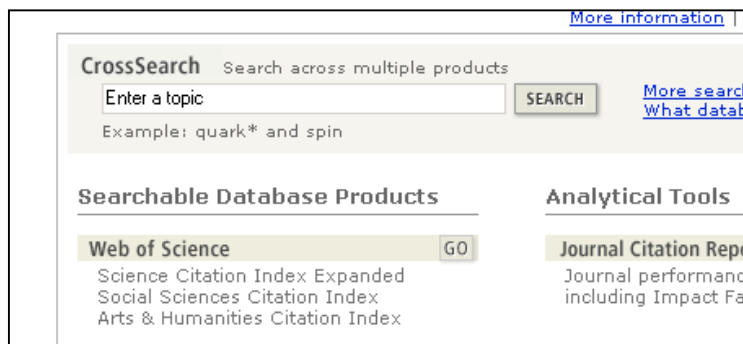
a) *The Journal of Physical and Chemical Reference Data* is a publication that aims to provide critically evaluated physical and chemical property data, fully documented as to the original sources and the criteria used for evaluation.

<http://jpcrd.aip.org>

b) Other Journals. Use the Web of Science or similar search engine to search the scientific literature.

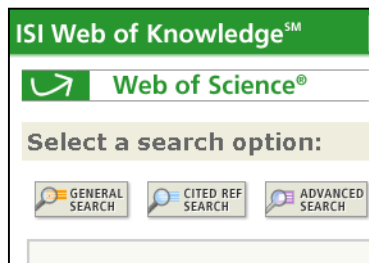
<http://www.asu.edu/lib/resources/db/webscience.htm>

To search, click first on the “GO” button next to “Web of Science”.



The screenshot shows a web interface titled 'CrossSearch' with the subtitle 'Search across multiple products'. It features a search input field with the placeholder 'Enter a topic' and a 'SEARCH' button. Below the input field is an example: 'Example: quark* and spin'. To the right of the input field are links for 'More search' and 'What datab'. Below the search area, there are two columns: 'Searchable Database Products' and 'Analytical Tools'. Under 'Searchable Database Products', 'Web of Science' is highlighted with a 'GO' button next to it. Below 'Web of Science' are three sub-options: 'Science Citation Index Expanded', 'Social Sciences Citation Index', and 'Arts & Humanities Citation Index'. Under 'Analytical Tools', 'Journal Citation Reports' is listed with a description: 'Journal performance including Impact Fa'.

And then on “General Search”



The screenshot shows the 'ISI Web of Knowledge' logo at the top. Below it is the 'Web of Science' logo. Underneath is a section titled 'Select a search option:' with three buttons: 'GENERAL SEARCH', 'CITED REF SEARCH', and 'ADVANCED SEARCH'. Below these buttons is a search input field.

There is no universally accepted way of citing journal articles. The style depends on the particular journal or publisher. Independently on the style, you should always include the authors, journal title, year, volume and pages. Journal titles are usually abbreviated, and the title of the article is sometimes included.

Examples:

"Kinetic and spectroscopic studies on the quercetin 2,3-dioxygenase from *Bacillus subtilis*," M.R. Schaab, B.M. Barney, W.A. Francisco, *Biochemistry* **45** 1009-1016 (2006)

Vieregg J, Cheng W, Bustamante C, Tinoco I Jr. "Measurement of the Effect of Monovalent Cations on RNA Hairpin Stability". *J Am Chem Soc*, 129, 4966-14973 (2007).

Websites

Use only authoritative and credible sources. Wikipedia might be useful as a starting point, but should NEVER be used as a reference for obvious reasons. Use only websites that contain information about the author and his/her affiliation. Avoid citing websites whenever the same information can be found in a book or edited volume.

To cite a website, use the following syntaxes:

Author (if any). Title of Site. URL (accessed date).

Example:

Notre Dame Radiation Laboratory (NDRL) Data Center, Quantum Yields for the Photosensitized Formation of the Lowest Electronically Excited Singlet State of Molecular Oxygen in Solution, <http://allen.rad.nd.edu/compilations/Qy/QY2146.HTM>, (accessed January 13, 2006)