Reporting and Propagating Error

In science error is a necessary fact of life rather than a sin. It is uncertainty in measurement. Here we will assume that error and imprecision are identical. The only way to know about inaccuracy is to have some alternative measure, which is not universally the case. Moreover, when there are two methods that disagree, it is not always clear which is the more accurate.

Significant figures

In everyday life, writing the number 1.3 means just that. In science, it means that you are reasonably sure that the true value (precise value, assuming that your measurement tool is working accurately) lies between 1.25 and 1.349999... When you publish any number, be sure that you can support the precision implied; don t simply copy all the digits from your calculator or instrument. Do not add extra zeros after the decimal point: 4.60 implies that the true value lies between 4.595 and 4.6049999 Think about what the number of significant figures implies; reporting one significant figure implies a fractional uncertainty (the ratio between your uncertainty, δx, and your best estimate, x, |δx|/x) of 10-100% (roughly 50%), reporting two significant figures implies a fractional uncertainty of roughly 5%, and three, roughly 0.5%. Avoid arbitrary rounding by quoting rational number, for example 45/87 eggs proved infertile (instead of 0.52 of the individuals (n = 87) proved infertile).

Both measurement and error should have the same dimensions AND units. Always include the leading zero (– 0.4 and not – .4). The decimal point is too easy to lose when it is out front by itself. The following are all acceptable: 55.9 – 0.4; 56 – 4; 50 – 40. If you get disparate precision in calculations of the best estimate and its imprecision, round the more precise one to be compatible with the less precise one, after you have made sure that you can support even that number of significant figures. In scientific notation, place the uncertainty before the power of 10, *e.g.,* $(3.43 – 0.02)\text{x}10^{-7}$ not $3.43 \times 10^{-7} – 2 \times 10^{-9}$. There are mild exceptions to these rules when the situation warrants. For example, if your best estimate of the mean length of adult shrimp in a population is 38.6 mm, and you measured to the nearest whole millimeter, it is arguably less misleading to write 38.6 – 1 mm than to write 39 – 1 mm. Similarly, if you are doing a string of calculations, you should carry extra digits to the extent that rounding before calculation could introduce unnecessary errors before you report the final result with the appropriate number of significant figures.

Propagating errors

Consider tow variables, x and y, and a third variable q such that $q = f(x, y)$ (f means a function of). Let s denote by subscript b the best estimate of a variable and by δ its uncertainty. Here we look at a function of 2 variables. All formulas can be directly generalized to more variables.

Thus $x = x_b – \delta x$. The fractional uncertainty is given as $\frac{|\delta x|}{x_b}$ (implying $x = x_b\left(1 \pm \frac{|\delta x|}{x_b}\right)$).

1. In addition or subtraction problems, ($q = x + y$, $q = x - y$), uncertainties add: δq † $\delta x + \delta y$. If x and y are independent and random, $\delta q = \sqrt{(\delta x)^2 + (\delta y)^2}$.

2. In multiplication or division ($q = xy$, $q = x/y$), ***fractional*** uncertainties add: $\frac{|\delta q|}{q} \le \frac{|\delta y|}{y_b} + \frac{|\delta x|}{x_b}$.

If x and y are independent and random, $\frac{|\delta q|}{q} = \sqrt{\left(\frac{|\delta y|}{y_b}\right)^2 + \left(\frac{|\delta x|}{x_b}\right)^2}$.

3. In multiplication by a constant k ($\delta k = 0$ [That is, there is no uncertainty in the value of the constant.], $q = kx$), $\delta q = k/\delta x/$, uncertainty varies proportionally.

4. For a power function ($q = x^n$, $\delta n = 0$), $\frac{|\delta q|}{q} = n\frac{|\delta x|}{x_b}$, error multiplies with the power.

5. For the general known functional dependence $q = f(x, y)$, we use the chain rule:

$$\delta q = \left|\frac{\partial q}{\partial x}\right|\delta x + \left|\frac{\partial q}{\partial y}\right|\delta y.$$

6. When the relationship between dependent (q) and independent variables (x, y) is very complicated or unknown (for example, the dependence of the weather prediction of a general circulation model on the uncertainty in the temperature at a given location) a probabilistic approach is often adopted (sometimes called a Monte Carlo approach based on the reputation of MC for its casinos). Variables are sampled randomly from their likely values (knowledge of the error statistics is needed) and their effect on the outcome is monitored until the statistics of the outcome are obtained. For example, let $q = f(x, y)$, $x = x_b - \delta x$, $y = y_b - \delta y$, with δx and δy being the standard deviation of many observation of x and y and x_b and y_b being the mean (here we assume both x and y to be normally distributed, or Gaussian). Using a random number generator we sample randomly from a Gaussian distribution around x_b and y_b with δx and δy being the standard deviation, getting a series of numbers $\{x_1, x_2, \; x \;\; _N\}$ and $\{y_1, y_2, \; y \;\; _N\}$ which we substitute into our function (or complex model) and generate a series of outcomes: $\{q_1, q_2, \; q \;\; _N\}$. If we did enough replication, we have generated a representative distribution of qs from which we can compute q_b δq. (*i.e.,* the mean and standard deviation of q also follow a normal distribution).

Reference:

Taylor, J. R.,1997, An introduction to error analysis, 2nd edition, University Science Books, Sausalito, California. Ch:1-3.

Homework:

1. Rewrite the following results in their clearest forms, with suitable numbers of significant figures:

   a. measured height = 5.03 – 0.04329 m

   b. measured time = 1.5432 – 1 s

   c. measured charge = -3.21×10^{-19} – 2.67×10^{-20} C

   d. measured wavelength = 0.000,000,563 – 0.000,000,07 m

   e. measured momentum = $3.267\text{x}10^{3}$ – 42 g cm/s

   f. x = 3.323 – 1.4 m

   g. t = 1,234,567 – 54,321 s

   h. $l = 5.33 \times 10^{-7}$ – $3.21\text{x}10^{-9}$ m

2. In Hwk 5, what is the uncertainty in the drag coefficient you computed for one body at a variety of towing speeds (assume perfect knowledge of the kinematic viscosity, feel free to use data from the answer sheet)?

3. In Hwk 5, what is the uncertainty in the drag coefficient you computed for one body at a variety of towing speeds (assume perfect knowledge of the density, feel free to use data from the answer sheet)?

4. What is the uncertainty in the hull speed for the homework of lab 6?

Reference:

Taylor, J. R.,1997, An introduction to error analysis, 2nd edition, University Science Books, Sausalito, California. Ch:1-3.

' Jumars and Boss, 2003