Notes on Data Analysis and Experimental Uncertainty

Prepared by David B. Pengra, University of Washington, and L. Thomas Dillman, Ohio Wesleyan University

This set of notes has been adapted from ones given to Ohio Wesleyan University physics students in the introductory laboratories. Although they are pitched at a relatively elementary level, they contain many hints that may be useful even to advanced students.

The topics discussed below may be supplemented with pertinent sections in the references listed at the end of the article; the text by Bevington and Robinson (see reference [1]) gives a good account of basic statistical theory and practical computer methods most often used in analyzing data from physics experiments.

1 Types of Uncertainty

There are two basic kinds of uncertainties, *systematic* and *random* uncertainties. Systematic uncertainties are those due to faults in the measuring instrument or in the techniques used in the experiment. Here are some examples of systematic uncertainty:

- If you measure the length of a table with a steel tape which has a kink in it, you will obtain a value which will appear to be too large by an amount equal to the loss in length resulting from the kink. On the other hand, a calibration error in the steel tape itself—an incorrect spacing of the markings—will produce a bias in one direction.
- If you measure the period if a pendulum with a clock that runs too fast, the apparent period will be systematically too long.
- The stiffness of many springs depends on their temperature. If you measure the stiffness of a spring many times, by compressing and decompressing it, the internal friction inside the spring may cause it to warm. You may see this by a systematic trend in your data set; for example, each data point in a data set will be smaller than the previous one.

Random uncertainties are associated with unpredictable variations in the experimental conditions under which the experiment is being performed, or are due to a deficiency in defining the quantity being measured. Here are some examples of random uncertainty:

- Changes in room temperature, electrical noise from nearby machinery, or imperfect connections to the voltmeter probes may cause random fluctuations in the magnitude of a quantity measured by a voltmeter.
- The length of a table may depend on which two points along the edge of the table the measurement is made. The "length" is imprecisely defined in such a case.
- Repeated measurements of the period of a pendulum which are made with a stopwatch vary because it is hard for a person to start and stop the watch at exactly the same point in the pendulum's swing. Note, however, that if the experimenter always starts the watch late, but stops it early, this will lead to a *systematic* error.

Of these two types of uncertainties, random uncertainties are much easier to deal with and to quantify. There is no general procedure for estimating the magnitude of systematic uncertainties as there is for random uncertainties. Only an experimenter whose skill has come through long experience can consistently detect systematic uncertainties and prevent or correct them.

If an experiment has low systematic uncertainty it is said to be *accurate*. If an experiment has low random uncertainty it is said to be *precise*. Obviously an experiment can be precise but inaccurate or accurate but imprecise. When thinking about uncertainty, it is important to remember these associations, so they are worth repeating:

- *Random* uncertainty decreases the *precision* of an experiment.
- *Systematic* uncertainty decreases the *accuracy* of an experiment.

These distinctions are illustrated in Fig. 1. You should avoid falling into the trap of thinking that because the uncertainty of a measurement is always the same, then it is systematic. Systematic uncertainty does *not* mean that the uncertainty is repeatable. What it means is that the uncertainty involves physics that has not been accounted for in the analysis—two very different ideas.

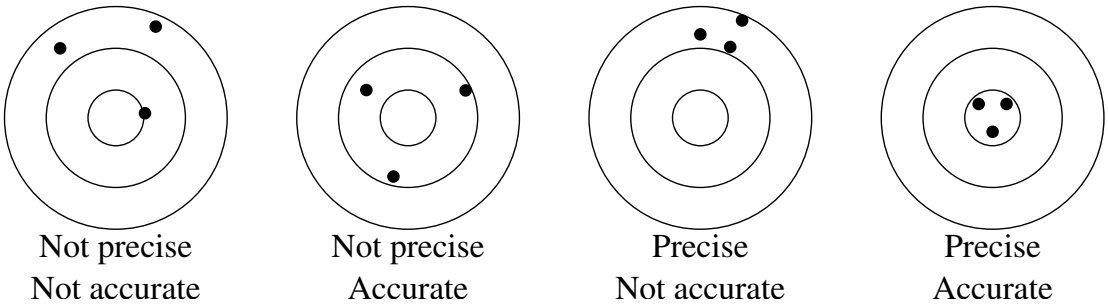

Figure 1: A "bulls-eye" plot showing the distinction between precision and accuracy in a measurement. The black dots represent data points taken in a measurement of a quantity whose true value is at the center of the circles.

Before proceeding further it may be useful to point out that *blunders* are not a source of uncertainty. They can always be *eliminated completely* by careful work. In your laboratory reports never list misreading the instrument or getting the wrong units as a source of uncertainty!

2 The Mean, Standard Deviation, and Standard Deviation of the Mean

Random uncertainty is often associated with the concept of *standard deviation*. This is best illustrated by an example. Suppose ten students each measure the diameter of a steel ball with a micrometer caliper. For a variety of reasons we do not expect all the measurements to be identical. The sources of error include:

- some students tighten the micrometer caliper more than others.
- the steel ball may not be perfectly round.

- some students may not exercise care to be sure they are measuring a "great diameter"—the ball is not centered between the jaws.
- the temperature of the steel ball may change with time as the ball is handled and hence its diameter may change slightly through thermal contraction or expansion.
- there may be varying amounts of corrosion on the steel ball.

Exercise 1 *Which of the above sources of error contribute to systematic uncertainty? Which contribute to random uncertainty? Explain how you came up with your answers.*

The obvious question to ask is, "What is the *best* value for the diameter of the steel ball?" If the sources of error are random, that is, they give values for the diameter which vary randomly above and below the "true" value, but do not skew all of the values in one particular direction, then an obvious procedure to get the best value for the diameter is to take the average or arithmetic mean. The mean of a set of numbers is defined as the sum of all the numbers divided by the number of them. In mathematical language, if we have N observations and x_i represents any one of the observations (i.e. i can have any integer value from 1 to N), then the arithmetic mean, which we designate by the symbol, $\overline{x}$, is given by

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{1}$$

Having obtained a mean or "best" value, $\overline{x}$, it is important to have a way of stating quantitatively how much the individual measurements are scattered about the mean. For a precise experiment we expect all measurements to be quite close to the mean value. The extent of scatter about the mean value gives us a measure of the precision of the experiment, and thus, a way to quantify the *random* uncertainty.

A widely accepted quantitative measure of scatter is the sample standard deviation, s. For the special case where all data points have equal weight, the sample standard deviation is defined by the equation,

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}}. \tag{2}$$

Although this equation may not be intuitive, inspection of it reveals that s becomes larger if there is more scatter of the data about the mean. This is because $(x_i - \overline{x})^2$ for any particular i will *on the average* increase with greater scatter of the data about the mean so that $\sum(x_i - \overline{x})^2$ increases. Note that s has the same units as x_i or $\overline{x}$ since the square root of the sum of squares of differences between x_i and $\overline{x}$ is taken.

The standard deviation s defined by Eq. (2) provides the random uncertainty estimate for any *one* of the measurements used to compute s. Intuitively we expect the *mean value* of the measurements to have *less* random uncertainty than any *one* of the individual measurements. It can be shown that the standard deviation of the *mean* value of a set of measurements σ_m, ("sigma-em") when all measurements have equal statistical weight, is given by

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N(N-1)}} = \frac{s}{\sqrt{N}}. \tag{3}$$

Note that σ_m is necessarily smaller than s. When we speak of the uncertainty σ of a set of measurements made under identical conditions, we mean that number σ_m and not s. It is most important that the student distinguish properly between standard deviation associated with individual data points, s, and standard deviation of the mean of a set of data points, σ_m.

Exercise 2 *Five students measure the mass of an object by making two separate measurements each. These measurements, in grams: 9.80, 9.87, 9.89, 9.95, 9.91, 9.98, 9.92, 10.05, 9.97, 9.84.*

1. *Calculate the* **mean**, *the* **standard deviation**, *and* **standard deviation of the mean**, *using your calculator and the above formulas. Show how you made the calculations.*

2. *Do the same calculations as in part 1 but using the statistical package on your calculator. Refer to your calculator's manual for instructions. If you have lost your manual, you may be able to find the instructions at the manufacturer's website; for example, Texas Instruments has copies of their manuals at* `http://www.ti.com/`. *Usually, in lab you will use your calculator to find means and standard deviations rather than doing the calculations by hand, so it is important to know how this is done. Write the results down, along with a* brief *description of how you performed them.*

3 Stating Results with Uncertainty

There are two common ways to state the uncertainty of a result: in terms of a σ, like the standard deviation of the mean σ_m, or in terms of a percent or fractional uncertainty, for which we reserve the symbol ϵ ("epsilon"). The relationship between ϵ and σ is as follows. Let the quantity of interest be x, then, by definition,

$$\epsilon_x \equiv \frac{\sigma_x}{x}\,. \tag{4}$$

When stating a result and its uncertainty in a report, one typically uses the form $x \pm \sigma_x$, with the units placed last. For example, if the mass of an object is found to be 9.2 g and the uncertainty in the mass is 0.3 g, one would write

$$m = 9.2 \pm 0.3 \text{ g}\,.$$

Sometimes one will present uncertainty in terms of ϵ, but in this case, ϵ is usually multiplied by 100, so that one would say, "The mass of the object is 9.2 grams with an uncertainty of 3 percent." Unless otherwise instructed, you should state all of your measurements following the first form, using σ. There is one important distinction between σ and ϵ when stating results: *σ_x always has the same units as x, while ϵ is always unitless.* Failure to be conscious of this difference typically costs students many points.

4 Comparing Quantities with Uncertainty

Frequently one wants to know whether two numbers obtained by two different methods but hypothetically referring to the same physical quantity *agree*. The term "agreement" means something very specific in an experiment. If uncertainties for one or both numbers (expressed by an associated σ) have been calculated, one can say that the two numbers agree with each other if they *overlap within their uncertainties*. For example, if a theory predicts that the density of an object should be

10.0 ± 0.1 g/cm^3, and a measurement gives a value of 9.8 ± 0.3 g/cm^3, then we can say the two values agree within the experimental uncertainty. But if the measurement gave instead 9.81 ± 0.02 g/cm^3, then we would be forced to admit that the two values did *not* agree.

In the case of disagreement, the experimenter faces a problem: what effects have not been accounted for? There could be a source of additional random error that has not been appreciated, or more vexing, there may be a source of *systematic* error that is fouling the *accuracy* of the measurement. Generally, sources of random error are easier to track down and rectify; but in so doing, one may uncover other sources of systematic error that were previously invisible!

You will often be asked to determine what the dominant source of error is in a particular experiment. In general, this is a subtle problem as there is no general method for determining systematic error. However, one important clue can be used when comparing measurements with each other, or with theory: *if the measured quantity including the uncertainty calculated from random sources of error does not overlap with another expected value (either from another experiment or theory) then you can assume that the systematic error in the experiment dominates the experimental error.* This is especially true when comparing against theoretically calculated values, as the theory almost always assumes some simplifications in order to make the calculation reasonable (for example, neglecting the weight of a string or assuming that friction is zero). To reiterate: systematic error comes into an experiment when the experimenter neglects some important physics in the analysis.

In quick measurements, we may not always calculate uncertainties for the quantities we measure. In these cases, the best we can state is that two values disagree by some amount. This disagreement is usually presented as a percent of the value of the quantity. For example, if we did not have uncertainties calculated for the above two density values, we could say that they disagree by

$$\left| \frac{9.8 - 10.0}{10.0} \right| \times 100 = 2\% \tag{5}$$

The general rules for comparing results in lab reports are these:

- If uncertainties exist, state the quantities with their uncertainty, and see if they overlap. If they do, they agree. If not, they don't, and you should try to explain why, that is, discuss the physics of the experiment and try to come up with some sources of systematic error.
- If uncertainties do not exist, calculate a percent disagreement. If the percent disagreement is less than a few percent, the results are probably in agreement. If the disagreement is more than ten percent, they are probably not in agreement, and you should try to explain why.

Exercise 3 *The manufacturer of the mass that was measured by the students in Exercise 2 claims that the mass is 10 g, within 0.4%. Is this a valid claim? Discuss whether your result* agrees *with the manufacturer's claim, following the guidelines above.*

5 Significant Digits

Most students learn the idea of significant digits in high school, at about the same time that they learn "scientific notation". But the results of uncertainty analysis complicates matters: What if the uncertainty is very large? What are the significant digits for σ itself? What is the uncertainty of a result that is measured repeatedly with a digital instrument (like a voltmeter), and the same number is recorded every time? These questions cause much confusion.

Here are some guidelines:

- The uncertainty σ should have, at most, 2 digits, and more commonly 1 digit. Remember, all uncertainty calculations are *estimates*; there is no such thing as an "exact uncertainty". Use this rule: if the first digit of σ is 1, use 2 digits for sigma, e.g., $\sigma_x = 0.14$ g, or $\sigma_x = 0.3$ g, but *not* $\sigma_x = 0.34$ g.

  The result itself should be stated to the same precision as σ_x. For example, you should write 9.5 ± 0.3 g, or 9.52 ± 0.14 g, but not 9.52 ± 0.3 g.

- If σ is especially large, you will *lose* significant digits. For example, suppose that multiple measurements are made with an instrument that is precise to 3 digits, and mean value of 9.52 g is found, but for other reasons the data points varied so that the standard deviation of the mean was 2 g. The result would have to be reported as 9 ± 2 g.

  If the measurement is so bad that σ is larger than the value itself, you may have *no significant digits*, but only know the order of magnitude. This case is most common when the quantity in question is expected to be close to zero—such measurements may only give an upper or lower bound on the quantity.

- If σ is calculated to be much smaller than the smallest digit of your measurement, then assume that σ is equal to "1" of the smallest digit. For example, if a measurement of a mass gives exactly 9.52 g ten times, the result should be stated as $m = 9.52 \pm 0.01$ g. Thus you may need to round your uncertainty up to the least significant digit in your measurement.

- Do not confuse round-off errors with uncertainty. With calculators and computers, there is no reason to prematurely truncate a result, just because it is found to be uncertain. If properly used, the formulas for propagating uncertainty will take care of the uncertainty in the final result. So keep you extra digits (at most one or two extra) as you go, but make sure to adjust the final result when you present your measurements for comparison.

6 *Propagation of Uncertainty*

The method of computing the uncertainty in a result which depends on several variables, each with its own uncertainty, is called *propagation of uncertainty*, or casually *error propagation*. Suppose we have measured the length and width of a table and have computed the standard deviation of the mean value for both the length and the width. Our aim is to determine the area of the table and an associated standard deviation of the area. It can be shown that the best estimate of the area is simply the mean length times the mean width.

What uncertainty should we associate with this same area? The answer is not obvious and in fact we can distinguish two distinct extreme cases:

- The uncertainties in the length and width are *completely independent*.
- The uncertainties in the length and width are *completely dependent*.

If the uncertainties are completely independent, the possibility of compensation occurs. That is, if the uncertainty in the length causes the area to be too large, then the uncertainty in width *may* be such as to cause the area to be too small. *On the average*, the total uncertainty in the area will

Table 1: Common formulas for propagating uncertainty. These equations can be combined in the cases of more complicated formulas, or the student may work directly from equation (6).

Functional Form	Formula	Uncertainty formula
Product or Quotient	$f = xy$ or $f = x/y$	$\epsilon_f = \sqrt{\epsilon_x^2 + \epsilon_y^2}$
Sum or Difference	$f = x + y$ or $f = x - y$	$\sigma_f = \sqrt{\sigma_x^2 + \sigma_y^2}$
Product of factors raised to powers	$f = x^m y^n$	$\epsilon_f = \sqrt{m^2\epsilon_x^2 + n^2\epsilon_y^2}$
Constant multipliers	$f = Kx$ (K =constant)	$\sigma_f = K\sigma_x$
Logarithmic functions	$f = \log_e(x)$	$\sigma_f = \epsilon_x$
	$f = \log_{10}(x)$	$\sigma_f = \log_{10}(e)\epsilon_x = 0.4343\epsilon_x$
Exponential functions	$f = e^x$	$\epsilon_f = \sigma_x$
	$f = 10^x$	$\epsilon_f = \log_e(10)\sigma_x = 2.303\sigma_x$

be algebraically less than the sum of the separate contributions to the uncertainty in the area. On the other hand, for completely dependent uncertainties, we must take into account the fact that the uncertainties are always correlated. This leads to complications involving a quantity called the covariance of two correlated quantities which we do not discuss in this elementary account. The case of completely independent uncertainties is nearly approached in many experimental situations, and we confine our attention to this case.

Without going into the derivations (see [1], pp. 36–41 for further details), the theory of error analysis gives a general formula for the uncertainty when a result is found by a calculation from a collection of measurements. The formula is based on the idea of a first-order Taylor series expansion of functions of many variables. It is valid when the various uncertainties σ_i of the i different variables are small compared to the values of the quantities and on the requirement that the uncertainties are uncorrelated with each other. Specifically, if the desired result is a well-behaved function $f(x, y, z, \ldots)$ of the physical variables $x, y, z, \ldots$ which have uncertainties $\sigma_x, \sigma_y, \sigma_z, \ldots$, then the uncertainty in the value of the result σ_f is given by the formula

$$\sigma_f^2 = \sigma_x^2 \left(\frac{\partial f}{\partial x}\right)^2 + \sigma_y^2 \left(\frac{\partial f}{\partial y}\right)^2 + \sigma_z^2 \left(\frac{\partial f}{\partial z}\right)^2 + \ldots , \tag{6}$$

where the partial derivatives are all evaluated at the best known values of $x, y, z, \ldots$

We give the equations required to propagate uncertainty for a number of simple cases. All of the formulas in Table 1.1 may be derived from equation (6) and the functional form that is listed. See [1, pp. 44–47] for derivations.

We expect students to learn how to propagate uncertainty for simple cases which are covered by the above equations. For example, suppose you have an equation for some physical quantity, say F which is related to another physical quantity, say r by the formula

$$F = -\frac{2\pi K}{r^2} , \tag{7}$$

where K is a physical constant with no uncertainty. This equation is of the form $f = x^m y^n$ where x corresponds to $2\pi K$, and y corresponds to r; in this case, $m = 1$ and $n = -2$, so the uncertainty formula is

$$\epsilon_F = \sqrt{m^2\epsilon_{2\pi K}^2 + n^2\epsilon_r^2} = \sqrt{1^2 \cdot 0^2 + (-2)^2 \cdot \epsilon_r^2} = 2\epsilon_r . \tag{8}$$

Note that because $2\pi K$ has no uncertainty, $\epsilon_{2\pi K} = 0$, and it drops out of the equation for ϵ_F. To find σ_F, you would then simply calculate $\sigma_F = |F\epsilon_F|$. The quantities σ_f and ϵ_f are always positive, as should be evident from the defining formulas (3) and (6).

The formulas in Table 1 are not just math; we can read some physics into their form. For example, in the case of constant multipliers, if one scales each data point by a constant, then the uncertainty in the mean value should also scale up proportionally. More interestingly, if you were to measure the area of a square by two *independent* measurements of the length x and the width y, you would use the product or quotient formula to propagate the uncertainty, as $A = xy$. In this case the errors in the two measurements may be opposite, leading to a better final result. But if you used only the length measurement x, and *assumed* that the width y were the same as the length, then the possibility of two measurement errors working against each other would be lost, and you would have to use the formula $A = x^2$. In this case, the exact same measurement is used twice, so the errors in width versus length can't cancel. To propagate the uncertainty, you would use the formula for the product of factors raised to powers, which gives a larger final uncertainty than in the former case.

Exercise 4 *In an experiment with an air track, an experimenter wishes to determine the average speed of an air track cart between two photogates. The distance Δx between the photogates is given by $\Delta x = 1.000 \pm 0.003$ m, and the time of travel Δt between these two points is $\Delta t = 2.3 \pm 0.1$ s. Calculate the average speed $s = \Delta x/\Delta t$, the fractional uncertainty ϵ_s, and the absolute uncertainty σ_s given these data. Show your work, and state the results using correct significant digits, and following the format given in the section Stating Results with Uncertainty.*

Exercise 5

1. *To measure the density of a rectangular object, an experimenter measures the object's volume and mass. The volume is given by the formula $V = LWH$, where L is the length, W is the width, and H is the height. The density ρ is given by $\rho = m/V$, where m is the object's mass. If the measurement of the mass is uncertain by 2%, and each of L, W and H is uncertain by 4%, what is the uncertainty, in percent, of the density ρ? Show your work.*

2. *The experimenter conducts the same density measurement with a second sample that is spherical in shape. The mass is again uncertain by 2%. The diameter d of the sphere is measured to a precision of 4%. The volume V of a sphere is given by the formula*

$$V = \frac{4\pi}{3}\left(\frac{d}{2}\right)^3$$

What is the percent uncertainty of the density ρ in this case? Show your work. Why is the uncertainty in this case different than in the case of a rectangular object? What is the underlying reason (not just how are the formulas different)?

7 Least Squares Curve Fits

The method of least squares is often applied to determine the "best" curve through a set of data points that is suspected to exhibit a functional relationship. If all the points have nearly the same weight/uncertainty then we can try to arrange the curve so that as many points fall above the line as below. However, it is not always clear how to "eyeball" for the best curve, thus analytical techniques become necessary. The least-squares technique may be described as follows: Suppose

you have a functional form $f(x; a, b, \ldots)$ which you would like to represent the data set as well as possible, and where $a, b, \ldots$ are adjustable parameters that can be varied in order to produce the best fit curve. The function may be a line ($f(x) = mx + b$, where the parameters are m and b) a higher-order polynomial (with more parameters), or some complicated function like a sine curve (with the amplitude, frequency and phase as parameters).

For each data point (x_i, y_i) the least-squares technique is to compute $y_i - f(x_i; a, b, \ldots)$, and then to calculate a quantity known as χ^2 ("chi-square"), which is given by

$$\chi^2 = \sum_i \frac{[y_i - f(x_i; a, b, \ldots)]^2}{\sigma_i^2} \,, \tag{9}$$

where the σ_i is the uncertainty of each data point. The best fit is found by adjusting the parameters $a, b, \ldots$ and calculating χ^2 until the minimum value is achieved. If there are N data points and n adjustable parameters, one can calculate the "reduced chi-square",

$$\chi^2_\nu = \frac{\chi^2}{\nu} \equiv \frac{\chi^2}{N - n} \,. \tag{10}$$

The quantity ν is known as the "degrees of freedom" in the problem. If one is able to adjust the parameters so that $\chi^2_\nu \approx 1$, then a "good fit" can be asserted: in this case, the difference between the fit curve and the data is, on average, about as big as the uncertainty in the data itself.

The general theory of curve fitting is very subtle, and beyond the scope of this article. But we do have computer programs that can make such fits at our disposal. These programs not only find the best fit parameters, but also produce the uncertainties in the parameters. When propagating uncertainty, you should use the values of σ given by the computer programs in the error propagation formulas.

References

[1] Bevington, Philip R., and D. Keith Robinson *Data Reduction and Error Analysis for the Physical Sciences, 3rd edition*, McGraw-Hill, New York, 2003.

[2] Barford, N. C., *Experimental Measurements: Precision, Error and Truth*, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1967.

[3] Beers, Yardly, *Introduction to the Theory of Error*, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1953.

[4] Hawkins, C. E., and Niewahner, J. H., *Data Analysis, Graphing and Report Writing*, 1st ed., Mohican Publishing Co., Loudonville, Ohio, 1983.

[5] Meyer, Stuart L., *Data Analysis for Scientists and Engineers*, John Wiley and Sons, Inc., New York, 1975.

[6] Young, Hugh D., *Statistical Treatment of Experimental Data*, McGraw-Hill Book Company, Inc., New York, 1962.