University of Illinois at Urbana-Champaign Department of Physics
Physics 401 Laboratory

ERROR ANALYSIS, Part I

REFERENCES:

1. E. Wilson – An Introduction to Scientific Research

2. J. Topping – Errors of Observation and Their Treatment

3. A. Worthing and J. Geffner – Treatment of Experimental Data

4. Y. Beers – Introduction to the Theory of Errors

5. H. D. Young – Statistical Treatment of Experimental Data

6. P. R. Bevington – Data Reduction and Error Analysis for the Physical Sciences

7. J. Taylor - An introduction to error analysis

SYSTEMATIC AND RANDOM ERRORS:

Before going into some of the mathematics of statistical error analysis it is important to make the distinction between random and systematic errors. When a given measurement is repeated a number of times the values, in general, do not exactly agree. This variation is the result of random errors which often arise from a number of factors. Some of these are the following:

(1) Errors of judgment: (Estimates of a fraction of the smallest division of a scale on an instrument may vary in a series of measurements.)

(2) Fluctuating Conditions: (Important factors in a given experiment such as temperature, pressure, or line voltage may fluctuate during the measurements, affecting the results.)

(3) Small Disturbances: (Small mechanical vibrations, and the pickup of spurious electrical signals will contribute random errors to some types of measurement.)

Copyright © 1999 The Board of Trustees of the University of Illinois. All rights reserved.

(4) Lack of Definition in the Quantity Measured: (For example, measurements with a micrometer of the thickness of a steel plate having non-uniform surfaces will in general not be reproducible.)

(5) Randomness in the Quantity Measured: (Repeated measurements of the number of disintegrations per second in a radioactive source will give different values because radioactive disintegrations occur randomly in time.)

Thus when random errors are present, deviations of measured values from the true value will vary in a series of repeated measurements. However, when systematic errors are present in the measurement process the deviations of the measured values from the true value will be constant in a series of repeated measurements, provided that the instruments used and the conditions under which they are used remain constant. An example of systematic error would be faulty graduations of a measuring scale. Although repeated length measurements might agree closely among themselves, they will all be in error due to the inaccuracy of the graduations. Systematic errors can be removed by applying corrections (for example, by calibrating an inaccurate scale). They are often present when least suspected. Measurement procedures must therefore be carefully examined for sources of systematic error. [One useful way to examine systematic errors is to measure the same quantity again, using a completely different method or procedure. (This second technique may, of course, contain its own systematic errors.)]

PRECISION AND ACCURACY:

A measurement is regarded as accurate if the measured values cluster closely about the true value. Thus if an experiment has small systematic errors it is regarded as having high accuracy. A measurement is said to be precise if the spread of the measured values is small. A precise measurement requires small random errors. It should be noted that a measurement can be very precise but not accurate, if the systematic errors are large. This should be kept in mind when applying the following error analysis techniques which deal only with random errors.

The analysis of systematic error is far more complex than that of random error for it depends ultimately on the standards and methods of calibration used for measuring equipment and on factors related to the experiment such as line voltage variations, etc. When estimates of error in accuracy are available, they should be included in your error analysis. If they are not, you should think carefully about how such an estimate might be obtained. If time permits, try to obtain it; otherwise, a brief description of how you would check experimental accuracy will be considered a valid part of your discussion of error.

THE NORMAL (GAUSSIAN) DISTRIBUTION:

The normal approximation is a frequently applied mathematical model for experimental values, which are grouped about a mean. It is assumed that if an infinite number of measurements were made and plotted, their plot would be the normal error curve as shown in Figure 1. In this plot, x is the parameter of measurement and $\frac{dN}{N}$ is the fraction of the points which lie between x and x + dx. For the limiting case of many measurements, μ, the mean of the distribution, is assumed to be the true value for the measurement; and σ is a measure of the spread of the distribution about that value due to random error in individual observations.

The treatment which follows is limited to the case of random errors which follow the normal law of errors. We assume that we observe samples drawn from a population which follows the normal law of errors. The term population is used to signify the infinite, continuous frequency distribution that describes all possible observations of the quantity being measured. The equation for a population which follows the normal law of error is:

$$\frac{dN}{N} = \frac{1}{\sqrt{2\pi}\,.\,\sigma}\; e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (1)$$

$\frac{dN}{N}$ is the fraction of the population which lies between the values x and x+dx. μ is the parameter which specifies the <u>population</u> mean. σ is the standard deviation of the <u>population</u> and is a measure of the distribution's spread. The normalization constant is $A = \frac{1}{\sqrt{2\pi}\sigma}$

THE NORMAL ERROR CURVE:

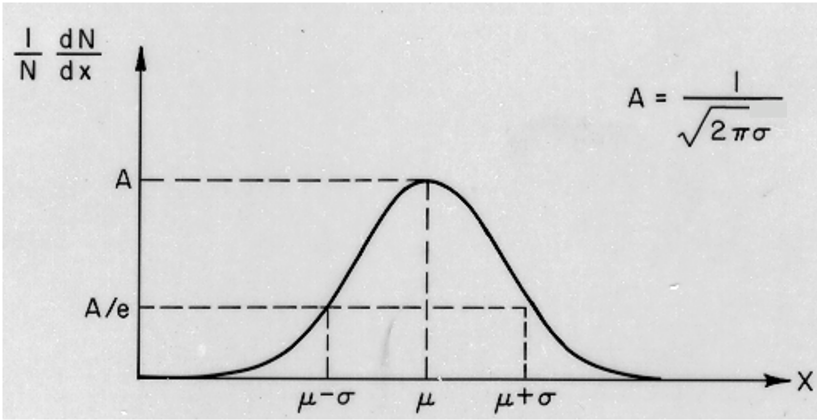

Figure 1. The normal error curve and its description.

The quantity σ in the above distribution is a measure of the population spread. The area under the curve from $x = \mu - \sigma$ to $x = \mu + \sigma$ is approximately 68% of the total area under the curve. In the language of probability this means that the probability for a given measurement of x to fall in this range is 0.68. Sometimes this is spoken of as a "confidence interval." In other words, we may expect values in this range with 68% confidence. But this is only a "correctly stated" confidence when either the σ is known to the experimenter beforehand or in the limit of a large N data set. We'll come back to this point shortly, but one should remark the estimation of errors is just that, an estimation. Although we can define a programatic route to find an acceptable value for error, much is this is based on the philosophical aspect of statistics.

σ is called the standard deviation of the distribution, and the 68% confidence interval is the mean plus or minus one standard deviation. If we include the area under the curve between points two standard deviations from either side of the mean, we find the 95% of the total area is included; so we may call this the 95% confidence interval. Whenever the population parameter σ is *known*, we may calculate any desired confidence interval by including that area under the curve and on either side of the mean which will give us the desired percentage of the total area.

How do we estimate μ and σ from the data (assuming σ is not already known)? The data that one obtains in the experiment is assumed to be a finite, random sample drawn from this infinite population. Assuming that we have a sample of size N, we can calculate two sample parameters:

$$\bar{X} \equiv \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (2)$$

$$S \equiv \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}{N-1}} \qquad (3)$$

where $\bar{X}$ is the sample mean,
S is the sample standard deviation

We must be careful to properly distinguish between the population parameters μ and σ and the sample parameters $\bar{X}$ and *S*. In general, only for very large samples $(N \rightarrow \infty)$ will the sample and the population parameters coincide. The task of statistical error analysis is to ***estimate*** μ and σ from the experimentally determined parameters $\bar{X}$ and *S*.

The primary difference between large sample techniques and small sample techniques is that the distinction between population parameters and sample parameters becomes more important as the size of the sample decreases.

The result of our analysis will be a confidence interval for μ, the population mean. A confidence interval for μ is a range of values of x within which, we can say that the population mean μ lies with a specified confidence (50%, 80%, or whatever we wish). A 50% confidence interval that is from 48.8 to 50.2 cm for the length of a rod means the following. If we repeated the measurement of the length of the rod n times as $n \rightarrow \infty$ we would expect the experimental mean to approach the true value; there is a 50% probability that the experimental mean (as $n \rightarrow \infty$) will fall between 48.8 and 50.2 cm. The experimenter must decide on the degree of confidence, i.e. the probability or percentage, he wishes to use. This decision depends on the precision required for the application of the result.

The size of the confidence interval, or the range of values of x, depends on the standard deviation of the sample, the size of the sample, and the desired degree of confidence. The confidence interval is given by:

$$\bar{X} - \frac{kS}{\sqrt{N}} < \mu < \bar{X} + \frac{kS}{\sqrt{N}} \qquad (4)$$

The value of *k* depends on the sample size and the desired degree of confidence. It can be calculated from Student's T distribution (see reference 1 or some of the others). The following Table 1 gives values of *k*.

Table I *Values of k for equation (4)*

Sample size	k 50% interval	k 80% interval	k 90% interval	k 95% interval	k 99% interval
2	1.00	3.08	6.31	12.71	63.66
3	0.82	1.89	2.92	4.30	9.93
4	0.77	1.64	2.35	3.18	5.84
5	0.74	1.53	2.13	2.78	4.60
6	0.73	1.48	2.02	2.57	4.03
7	0.72	1.44	1.94	2.45	3.71
8	0.71	1.42	1.90	2.37	3.50
9	0.71	1.40	1.86	2.31	3.36
∞	0.68	1.28	1.65	1.96	2.58

the result is generally written:

$$\text{QUANTITY} = \bar{X} \pm \frac{kS}{\sqrt{N}} \qquad \text{(xx\% confidence)}$$

Without giving the statement of the degree of confidence, the result is meaningless. For the lab's, let's use a 95% (colloquially known as a 2−σ) confidence interval unless otherwise stated. Often in physics literature, where it is not stated, one assumes a 1-σ confidence interval (even when it is unclear whether the proper k-value has been applied). Unfortunately, there is much sloppiness about this. On the other hand, error assignment is an art, not necessarily rigorous math.

PROPAGATION OF ERRORS

The techniques for calculating the confidence interval for an indirectly measured quantity depends on having the same degree of confidence for the directly measured quantities. This requires having the same sample size for all quantities if we use standard techniques for the propagation of errors.

Let $y = F(x_1, x_2 \ldots x_n)$

for small, independent variations of the x_i we have:

$$dy = \frac{\partial F}{\partial x_1} dx_1 + \frac{\partial F}{\partial x_2} dx_2 + \ldots + \frac{\partial F}{\partial x_n} dx_n$$

consequently:

$$(dy)^2 = \sum_{i,j} \frac{\partial F}{\partial x_i} \frac{\partial F}{\partial x_j} dx_i dx_j$$

If the errors are strictly random, independent, and symmetric with respect to sign, $dx_i\, dx_j = 0$ for $i \neq j$ (on the average) so:

$$\overline{(dy)^2} = \sum_{i=1}^{n} \left(\frac{\partial F}{\partial x_i} \right)^2 dx_i^2$$

then:

$$\sigma_y^2 = \sum_{i=1}^{n} \left(\frac{\partial F}{\partial x_i} \right)^2 \sigma_{x_i}^2$$

Thus for the sample parameter S_y we have:

$$S_y^2 = \sum_{i=1}^{n} \left(\frac{\partial F}{\partial x_i} \right)^2 S_{x_i}^2$$

If and only if, we have the same confidence level for the given independent variables (x_i), we can write:

$$y = \bar{y} \pm \sqrt{\sum_{i=1}^{n} \left(\frac{\partial F}{\partial x_i} \right)^2 \left[\frac{(k_i S_i)^2}{N_i} \right]} \qquad (5)$$

the confidence level for y will be the same as that of the independent variables

$$\bar{y} = F\,(\bar{x}_1,\ \bar{x}_2\ \ldots.\ \bar{x}_n)$$

and the partial derivatives are evaluated using mean values.

What happens when we want to combine confidence intervals of a quantity measured on several occasions? It's a long story (see ref. 6), but the best answer is that one should merge the raw data sets and find a new confidence for the concatenated data.

DISCUSSION AND EXAMPLES:

Note that the comparison of an experimentally determined mean with a published value in no way constitutes an error analysis. Such a comparison, however, including consideration of the confidence interval for the experimental and published values, is very useful because it may indicate the presence of systematic error. (In either the student's values, the published value, or both.) This comparison may also indicate an error in the error analysis, a blunder, or that the equipment is not measuring what you think it is measuring.

In many cases the application of error estimation techniques is not nearly as clear cut as the above discussion indicates. For example, consider the case of a voltage measurement with a voltmeter. A number of readings would be taken and some variation noted. The manufacturer will state that the accuracy of the meter is 4% of the full-scale deflection. How much of this 4% of full-scale error is of a random nature, and how much of it is systematic (due to scale error, for example) is not known. In addition, the manufacturer probably gives no information on the confidence level associated with the 4% figure, and the time elapsed since the last meter calibration is obviously important. More information about these problems could be obtained by calibrating the meter, but time limitations generally prohibit this.

Let us consider, for an example, the problem of measuring the length and area of a table using a meter stick. Let x_1 denote the meter stick graduation which lines up with one end of the table and x_2 denote the meter stick graduation which lines up with the other end. We estimate between the finest graduations (at 0.1 cm intervals) on the stick, i.e., we estimate to .01 cm. The data is summarized in Table II.

Let us work to an 80% confidence level. From Table I we see k = 1.4, so (actually k = 1.42, but we've rounded off)

$$L=\left(57.26 \pm \frac{2.6 \text{ x } 10^{-2} \text{ x } 1.4}{\sqrt{8}}\right)cm \qquad \text{[80\% confidence]}$$

or

$$L=(57.26 \pm .01)cm \qquad \text{[80\% confidence]}$$

The above treatment has considered only random errors. We now must consider systematic and instrumental errors. The instrumental error on a meter stick is usually about its least count. (Least count ≡ *smallest* scale graduations, which is 0.1 cm in this case.) However, we don't really know what confidence level this represents. Let us assume it is about 50% confidence.

Table II

Measurement No.	x_1	x_2	$L = x_2 - x_1$ cm	$L - \overline{L}$ cm	$(L-\overline{L})^2$ $(cm)^2$
1	9.01	66.24	57.23	–.03	9×10^{-4}
2	6.50	63.78	57.28	+.02	4×10^{-4}
3	3.25	60.54	57.29	+.03	9×10^{-4}
4	7.68	64.90	57.22	–.04	16×10^{4}
5	8.76	66.04	57.28	+.02	4×10^{-4}
6	9.22	66.46	57.24	–.02	4×10^{-4}
7	8.11	65.36	57.25	–.01	1×10^{-4}
8	10.15	67.40	57.25	–.01	1×10^{-4}
		Sum	458.04	–.04	48×10^{-4}

Assuming sources of systematic error are small, i.e. meter stick expansion due to temperature is negligible (we could calculate this to prove it), the estimate for the table length is

$$L = (57.26 \pm .10) \text{ cm}, \qquad [\sim 50\% \text{ confidence}]$$

where we simply used the least count as the error estimate. Which is the best estimate of the error, the one based on instrumental error or the statistical one? Answer: in this case of negligible systematic error, the statistical error is the best to use if one's data set is large. The above is a simple example for showing how to use error analysis techniques, pointing out the problem that is probably the most vexing in trying to estimate errors: One finds that one can rely on techniques of statistics, rather than good judgment and experience, <u>*only*</u> when random errors are the dominant source of errors.

Let us assume that we measured the Width W of the table, and using similar techniques, found that:

$$W = (40.79 \pm .01)cm \qquad \text{[~50\% confidence]}$$

We want to find the area of the table, using the formula $A = L \bullet W$.

The first question is: does this theoretical equation fit the system? If the table has rounded corners or the length or width is not uniform, it is obvious that it doesn't. Assuming that the formula is correct, we have, using equation (5), as an estimate:

$$A = L \bullet W \pm \sqrt{L^2 (\Delta W)^2 + W^2 (\Delta L)^2}$$

$$A = (57.26)(40.79)cm^2 \pm \sqrt{(57.26)^2 (.01)^2 + (40.79)^2 (.01)^2}$$

$$A = 2335 \ cm^2 \pm 1 \ cm^2 \qquad \text{[~50\% confidence]}$$

The example that we will consider below is probably more meaningful.

Let us now consider measuring the distance between two spectral lines on a glass photographic plate (glass plates are used in order that systematic errors due to distortion of the emulsion during development can be minimized). We are interested in the distance between the centers of the lines, and we measure this distance with a microscope that is fitted with a stage that adjusts with a micrometer. Let x_1 be the micrometer reading when the first line's center is under the cross hairs, and let x_2 be the reading when the second line's center is under the cross hair. The data is

Measurement No.	x_1	x_2	$D = x_2 - x_1$ (cm)	$D - \overline{D}$ (cm)	$(D - \overline{D})^2$ $(cm)^2$
1	.275	.643	.368	+.004	16×10^{-6}
2	.218	.593	.375	+.011	121×10^{-6}
3	.210	.562	.352	–.012	144×10^{-6}
4	.230	.591	.361	–.003	9×10^{-6}
		Sum	1.456	0	290×10^{-6}

The mean, using formula (2) is:

$$\overline{D} = \sum_{i=1}^{4} \frac{D_i}{4} = \frac{1.456}{4} = 0.364\,cm$$

Using formula (3) to find the standard deviation:

$$S = \sqrt{\frac{\sum\left(D-\overline{D}\right)^2}{4-1}} = \sqrt{\frac{290 \text{ x } 10^{-6}}{3}} \cong 10 \text{ x } 10^{-3}$$

Let us work to a 95% confidence level. From Table I we have $k = 3.18$. Therefore:

$$D = \left(0.364 \pm \frac{0.010 \text{ x } 3.18}{\sqrt{4}}\right)cm = \left(0.364 \pm 0.016\right)cm \qquad \text{[95\% confidence]}$$

Assuming that systematic errors are negligible, and that the instrumental error is ~0.001 cm, we see that our result is

$$D = \left(0.364 \pm 0.016\right)cm \qquad \text{[95\% confidence]}$$