

Physics 145 Data Analysis Reference Materials

Experimentation: An Overview

Before describing any detailed experiments, we first give an overview of the activities involved in any experimental laboratory work in the physical sciences. These ideas will be referred to in most of the weekly laboratory outlines and should become well understood during this course. The essential elements or steps of any experimental study are:

1. Formulate the procedures and design/build the apparatus used to make the measurements of interest.
2. Perform measurements and record the data, generally in numerical form.
3. Organize, summarize, graph, tabulate, dissect, correlate, and otherwise analyze the data to extract useful physical information that the data may contain. This should include estimates of the accuracy or reliability with which the measurements have been made, the identification of sources of error, and corrections for known extraneous effects where possible.
4. Interpret and present the data.

Experiment design: Before initiating an experimental study it is of prime importance to clarify and solidify in your mind (and usually on paper) precisely what you want to measure and the data you expect to get. Only then can you give meaningful direction to the process. Is the information sought qualitative or quantitative in nature? How many physical variables are involved? Which variables are fixed and which will vary. How can potential sources of error be isolated or eliminated? How will the data be presented? For each laboratory experiment that you conduct as part of our course, you will be provided with a selection of equipment. Instead of receiving step by step instructions on how to use this equipment, you generally be given broad objectives and sufficient time to figure out how to best design your own experiment. Additional information will occasionally be provided when necessary.

Data collection: In either case, a good experiment will begin with a series of simple tests to make sure that the equipment is working properly. An apparatus will often need to be calibrated against standards in order to correlate a raw output parameter like transducer voltage with a physical quantity of interest like mass or force. Reproducibility is also very important if you want your results to be considered valid by others. Care must be taken in following a good plan, all experiment procedures and parameters should be documented, all raw data should be accurately recorded and described along with subjective observations, motivations and insights.

Data analysis: While some experiments may involve only a single measurement (single-point data set), much of physical science and engineering involve discovering relationships amongst physical variables. Newton's second law of mechanics, for example, states that the acceleration of an object is the product of the net force on that object and its inertial mass: $F = ma$. Discovering the details of a functional relationship amongst physical variables is often more challenging than a single-point measurement, but also more interesting. An experiment aimed at characterizing a relationship requires a series of single-point measurements in which one or more experimental input parameters are varied from one data point to the next. Three types of data analysis that you will commonly encounter are graphical/tabular, qualitative and quantitative, which are generally performed in that order.

1. Begin by displaying your data in either a **graphical** or **tabular** form, which is probably the simplest representation of a relationship amongst variables.
2. Identify **qualitative** mathematical expressions that "match" or "describe" the experimentally observed relationship. These theoretical expressions may be based on fundamental laws or postulates, or they

may be purely empirical (i.e. they appear to match the data well but lack fundamental motivation). In fact, many of the fundamental laws themselves are simply empirical equations that have been "canonized" (i.e. widely accepted by the scientific community). While finding a mathematical form is not always necessary, it does represent a higher level of knowledge than that embodied in a simple graph or table, especially if grounded in physical law or intuition rather than purely empirical observation.

3. Perform a **quantitative** fit of the theoretical model to the experimental data to extract meaningful physical parameters. This includes estimating reasonable errors in the experimentally measured data, propagating them to the fitted physical parameters, and carefully interpreting the data in light of the errors. While quantitative information is vastly more difficult to obtain than qualitative information, it is again a much higher and more valuable form of knowledge.

While your lab exercises will require you to perform all three types of analyses, the remainder of these notes primarily emphasize the more difficult issues associated with quantitative data analysis.

Data interpretation and presentation: While we occasionally enjoy saying that "good data speaks for itself", this statement is generally not true in practice. Labels and units should be clearly visible on every graph axis or table column. Captions should clearly explain what is seen in a figure or table, including any special symbols, lines, colors, annotations, panels, insets, or other content. Interpretation and conclusions, on the other hand, are usually reserved for document text, where one should clearly explain what experiment corresponds to each dataset presented, what important qualitative trends were observed, what model (if any) was used in the interpretation, and what quantitative physical parameters were extracted including error estimates. Beginning students often struggle with these issues.

Fitting Mathematical Equations to Experimental Data

Fitting errors: One often needs to fit a mathematical function to an experimental dataset. It will have one or more undetermined parameters that can be adjusted to obtain a best "fit" to the data. The function of choice may convey a relationship that is based on a physical model or it may be purely empirical (an intelligent "guess" of the correct "form" of the relationship). Before attempting to fit a function $y(x)$ to a set of N points on a graph: (x_1, y_1) , (x_2, y_2) , ..., (x_N, y_N) , always plot the raw data first to get a visual "feel" for the trend, and then make an intelligent guess regarding a suitable functional form. In many cases, theoretical arguments will suggest a function that is likely to work. Once you have selected a function, the next step is to determine the values of its undetermined parameters that will best "fit" your data. If the function $y(x)$ is a perfect fit, then $y_i = y(x_i)$ for every data point. If the fit is not perfect, there will be discrepancies between the observed y_i and the calculated $y(x_i)$. In this case, a good measure of the overall discrepancy can be computed as

$$E(a, b, \dots) = \sum_{i=1}^N (y_i - y(x_i))^2.$$

Notice here that the **summed squared error** E is a function of each of the unknown fitting parameters. In practice, the best fit is obtained by employing a computer algorithm to vary each of the fitting parameters until the summed squared error has been minimized, a process commonly known as **least squares minimization**. The most important functional form to become familiar with is the straight line ($y = a + bx$). A linear fit to any dataset produces an exact analytical solution for the fitting parameters a and b when subjected to least-squares minimization. Most calculators and spreadsheet programs have a **linear regression** function that will calculate this solution. More complicated functions that appear

frequently include the power function and the exponential. Both of these forms can be readily manipulated into linear form so as to take advantage of straight-line simplicity. Other more complex functions can also be used represent a variety of simple and complex physical models. These typically require a more general approach called **non-linear least squares (NLSQ)** fitting. There are many general algorithms for NLSQ fitting, which tend to be classified according to their method of exploring the available **fitting-parameter space**.

Power function: $y = b+ax^m$, where a , b , and m are fitting parameters. Assuming that b , the y -intercept, can be approximated with reasonable accuracy from a smooth hand-drawn curve through the data points and extended to the y -axis, one can apply the natural logarithm to both sides of the expression to obtain $\ln(y-b) = \ln(a) + m \ln(x)$. Defining $u = \ln(y-b)$ and $v = \ln(x)$ produces linear function $u(v) = \ln(a)+mv$, which has a u -intercept of $\ln(a)$ and a slope of m . By calculating $u_i = \ln(y_i-b)$ and $v_i = \ln(x_i)$ at each data point, we then fit $u(v)$ to the (u_i, v_i) data rather than directly fitting $y(x)$ to the more complicated (y_i, x_i) data.

Exponential function: $y = a e^{mx}$, where a and m are fitting parameters. As in the previous case, the expression can be put into linear form by taking the natural logarithm of both sides: $\ln(y) = \ln(b) + mx$. By defining $u = \ln(y)$, we arrive at a new linear expression of the form $u(x) = \ln(b) + mx$, with a u -intercept of $\ln(b)$ and a slope of m . After calculating $u_i = \ln(y_i)$ for each data point, $u(v)$ is then fitted to the (u_i, x_i) data.

These examples illustrate that creativity and deductive power help in determining what manipulations will best simplify the fitting process. When linearization is possible, seeing the modified data fall along a straight line provides a nice visual verification of a good fit. With functions that cannot be manipulated into linear form, one must be content to see the data fit nicely along a more complex curve.

Experimental Error: Accuracy and Precision

It is very common to hear the word "error" used loosely in reference to experiments, when in fact, there are different types of error that should be carefully distinguished. Whenever using the word error in your lab reports, you should be careful to explicitly indicate the type and the source of the error that you are referring to.

Accuracy Limitations: The accuracy of a measurement refers to the difference between the experimentally measured value and the ideal theoretical value arising from **systematic errors**, or in other words, an uncorrected discrepancy between the experiment performed and the theory used to describe that experiment. We must be careful in all experimental work to distinguish carefully between the actual experiment and the idealized thinking which fills our minds when we design and perform experiments. For example, we talk of massless ropes, point masses, point charges, perfect electrical and thermal insulators, frictionless surfaces, etc. Such idealizations serve a very important role in developing the theoretical structure of science, but must be extended in the final product. Real experiments often require one to take into account systematic non-idealities such as extended mass, or heat loss, friction. Does a given systematic error indicate a flaw in the theory or in the experiment? This is really a matter of perspective. If one intentionally performs a complex experiment with many experimental variables, but fails to use a sufficiently complex theory to account for those variables, then we might say that the theory is flawed. However, if one intends to measure the influence of a single variable, but fails to eliminate or minimize the influence of other variables, we could say that the experiment is flawed. The solution is to carefully isolate systematic errors, and then to include them in the theoretical model or else eliminate

them from the experimental data. All too often, this requires the redesign and execution of an entire experiment.

Precision Limitations: Even with a theory that accurately represents your experiment, there will always be a limit to the precision of your measurements. Precision refers to the smallest increment in the experimentally measured value that can be reproducibly controlled. In principle at least, if all systematic errors have been eliminated, the measurement accuracy becomes equal to the measurement precision.

1. One factor that limits precision is **statistical error** -- if you repeat the same experiment many times using a sufficiently sensitive technique, you will find that the outcome varies from one measurement to the next due to randomness, producing a distribution of measured values. If, for example, you use a hand-held millisecond stop-watch to time the fall of a mass from a 1-meter height, and repeat the measurement several times in a row, you will almost certainly get a different result each time due to random errors in your motion perception and reaction speed. If you make a large number of such measurements, the average of the resulting distribution approximates the mean fall time, and its standard deviation (i.e. statistical uncertainty) approximates the effective experimental precision of your overall time-interval measurement technique.
2. Another factor that commonly limits experimental precision is **instrument resolution**, which is caused by experimental input parameters that can only be adjusted or otherwise determined to within a finite range of values. A ruler within only millimeter resolution, for example, can't be expected to provide 1 micron of precision, and a digital bathroom scale with a 0.1 kg resolution can't be expected to provide 1 gram of precision.

Error estimates: Suppose that we measure speed of an object as the ratio of a measured distance and a measured time interval. In that case, both time and distance are raw data variables, while speed is a derived quantity. To determine the error in the measured speed, we must first estimate the errors in the raw time and distance measurements, and then propagate those errors to the calculated speed. For this reason, you should always estimate and record the errors associated with each raw data variable measured. Systematic errors that cannot be avoided must be well characterized so that they can either be processed out of the data or else added to the theoretical model. Instrument resolution limitations are usually easy to quantify based on simple observations. Statistical errors can either be predicted based on known probability distributions (e.g. Poisson or Gaussian) or else measured through experiment repetition.

Statistical Error Analysis

Multiple Measurements and Mean Values: Suppose that we make N measurements of the distance between two points along a straight line. These measurements, which could be made by the same individual or by several different people with different measuring instruments or procedures, will yield a distribution of values rather than a single distance due to random errors. If we assume that the random errors can be either positive or negative with equal probability, the average of all N measured values would be the best estimate of the true distance. Of course, any systematic errors would add an offset (also called a bias) to the statistical errors, and you must separately consider such errors. But we assume that our errors are purely statistical, and designate the i^{th} measured distance as x_i and the mean of these measured values by the symbol \bar{x} . In mathematical form:

$$\bar{x} = \frac{1}{N}(x_1 + x_2 + x_3 + \dots + x_N) = \frac{1}{N} \sum x_i. \quad (1)$$

Now suppose that we accurately measure the diameters of N individual hairs from your head, once again arriving at a distribution of results since each hair has a slightly different diameter. If you designate each individual measurement as x_i , and calculate the mean using Eq. 1 above, \bar{x} will now represent your best estimate of the size of any specific hair of your head selected at random. In this case, however, there is no such thing as the “true” diameter.

Histograms and Probability Distributions: When you make a reasonably large number of measurements (call this number N), a histogram-type graph of the results can help you to visualize the amount of variance in the data and to check for abnormalities in the overall distribution. The **histogram** (see Fig. 2) is a bar graph obtained by grouping together all measurements that fall within predetermined intervals of the random variable x . You must judiciously select the size of the **interval Δx** , such that a suitably large number of measurements get grouped into each interval. If you slice the intervals too finely, and only wind up with 0 or 1 counts n in each interval, the shape of the distribution is difficult to discern. As N grows larger, the size of the intervals can be reduced while retaining a modest number of counts in each interval, so that the histogram gradually takes on the true shape of the distribution. As N becomes very large, the interval width approaches zero and the histogram curve becomes smooth rather than discrete, as illustrated in Fig. 2. Many physical processes exhibit intrinsically random variations that obey a symmetrical bell-shaped distribution commonly referred to as the **Gaussian** or **normal distribution**, which has the mathematical form:

$$f(x) = e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} / \sqrt{2\pi\sigma^2}$$

In Fig. 2, the Gaussian distribution is contrasted with an abnormal asymmetric distribution.

To know the probability distribution of a random variable is to know everything that can be known about the statistical properties of that variable. In practice, however, the distribution of a random measurement quantity is generally not known beforehand unless you are dealing with a process that is well known or previously characterized. When necessary, you can map out the distribution by making a large number of measurements and plotting your own histogram. Unless you have good reason to suspect otherwise, you can usually safely assume that a random variable obeys a Gaussian distribution. In such a case, you need only measure the mean and the variance to completely characterize the distribution.

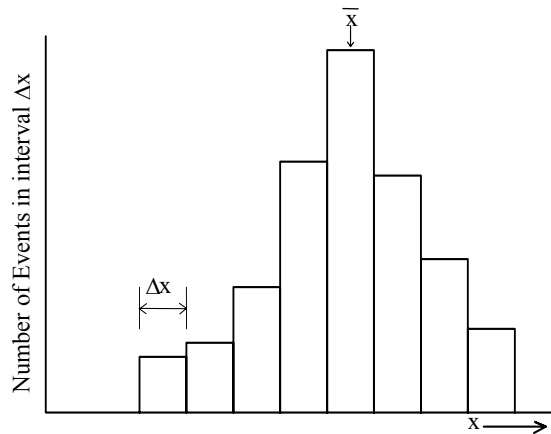


Figure 1

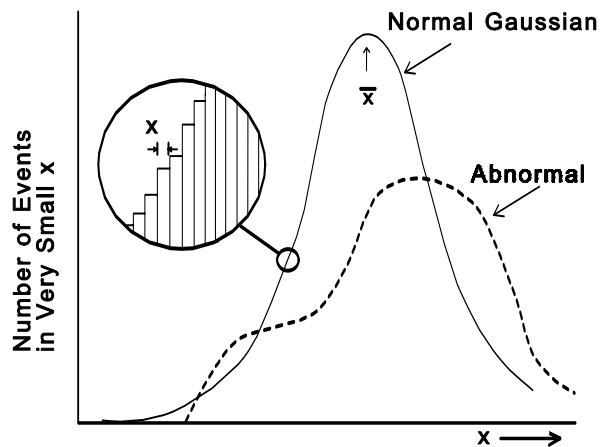


Figure 1

The **estimated standard deviation (esd)** σ of a data set is a quantitative measure of the spread or scatter of the measurements around the average value \bar{x} . The square of the esd (i.e. σ^2) is called the **variance**. For a large number of measurements, σ is associated with the effective width of the associated probability distribution. To further develop the idea of the standard deviation, return to the example of N individual measurements. Given a dataset, one can easily calculate \bar{x} and then the individual deviation from the mean: $d_i = \bar{x} - x_i$. Some numerical measure of the average deviation is desirable, though the true average deviation $\bar{d} = \sum d_i / N \propto \sum (\bar{x} - x_i) \propto \bar{x} - \bar{x}$ is precisely zero. One common approach is to compute the average of the absolute values of the deviations, a measure known as the **probable error**:

$$P.E. = \frac{|\bar{x} - x_1| + |\bar{x} - x_2| + |\bar{x} - x_3| + \dots}{N} = \frac{\sum |d_i|}{N}. \quad (2)$$

The most common method of computing the esd, however, is to square the individual deviations, average the squares, and then take the square root of the average. From this process, we obtain the so-called **root-mean-square (rms)** average:

$$\sigma_N = \sqrt{\frac{(\bar{x} - x_1)^2 + (\bar{x} - x_2)^2 + (\bar{x} - x_3)^2 + \dots}{N}} = \sqrt{\frac{\sum d_i^2}{N}}. \quad (3)$$

For a set of measurements taken on items that have an actual variation of the variable within the group, as in the example of hairs given above, the standard deviation σ_N is appropriate. However for multiple measurements of a single item in which the variations are associated with statistical measurement errors, a more thorough analysis than given here require a slight modification -- replace N with $N-1$ due to the fact that we have no indication of the uncertainty when we make only one measurement. For large values of N , the two definitions are essentially equivalent.

$$\sigma_{N-1} = \sqrt{\frac{\sum d_i^2}{N-1}}. \quad (4)$$

Here we list some **useful properties of the Gaussian distribution**. These results, which are proved in formal statistics courses, are very easy to use and provide a convenient means of evaluating experimental data. Note, however, that these rules will break down if the parent distribution is not a truly Gaussian distribution.

1. If a large set of measurements are taken, 68.3% of measured x_i values will lie within the range $\bar{x} \pm \sigma$, and 31.7% will lie outside this range. This result implies that a single measurement has a 68.3% chance of falling within one standard deviation of the mean.
2. It can be shown that the probable error is $P.E. = 0.674\sigma$. This relationship implies that if you calculate σ , you immediately know the probable error. The probable error has the very simple interpretation that 50% of the measured values lie within the range $\bar{x} \pm P.E.$. Thus, for a single measurement, there is a 50% chance that measured value will lie within one $P.E.$ of the mean.
3. Similarly, there is a 95.5% that a given measurement will lie within the range $\bar{x} - 2\sigma$ and a 99.7% chance that it will lie within the range $\bar{x} - 3\sigma$.

Counting statistics and the Poisson Distribution: Repeated measurements of a random counting variable will produce what we call a Poisson distribution, which is described by the function $f(n) = N^n e^{-N} / n!$ and only defined for integer values of n . For example, the number of cars that pass through an intersection during a given 10-minute interval of time is a random variable subject to Poisson

statistics. Assuming that the average number of cars is $N = 18.7$, then the probability that 12 cars will pass during a given 10-minute interval is $f(12) = 18.7^{12} e^{-18.7} / 12! = 0.0289 \approx 3\%$. The Poisson distribution has the interesting property that its variance is equal to its mean. Thus for a mean value of N counts, the standard deviation is $\sigma = \sqrt{N}$. If the average number of radioactive decays detected during a one-minute interval is $N = 1600$, the standard deviation (i.e. statistical uncertainty) will be $\sigma = \sqrt{1600} = 40$. In other words, 1600 ± 40 counts are expected during an arbitrary one-minute interval. This type of error estimate is often referred to as **counting statistics**. For large means, the Poisson distribution is approximated by a Gaussian distribution with the same property.

Quoting Experimental Uncertainties: In reporting experimental results, data are usually listed with a numerical value for the uncertainty. For example, the best value of the mass of an electron currently is $(5.48579902 \pm 0.00000013) \times 10^{-4}$ atomic mass units. Since uncertainty can be defined in many different ways, how do we know what number to assign? For the purposes of most analyses, and in particular for the purposes of this class, the rules for assigning uncertainty follow:

1. For statistical uncertainty, use one standard deviation (unless otherwise specified).
2. For instrumental uncertainty, estimate the uncertainty as well as you can. Although this uncertainty isn't strictly Gaussian, it will usually be treated as if it were Gaussian in subsequent error analysis.
3. For known systematic errors (not really an uncertainty), either the data or the theory should be appropriately corrected in a clearly-stated fashion.

Propagation of Measurement Errors to Derived Quantities: One often needs to propagate the statistical uncertainties of experimentally measured parameters to other quantities derived from those parameters. In the language of statistics, this is the problem of determining the distribution of a function of random variables. If we take an arbitrary function f of measured variables a , b and c , which have measurement uncertainties Δa , Δb and Δc , respectively, the corresponding uncertainty Δf is computed using the **Gaussian error-propagation formula**:

$$(\Delta f)^2 = \left(\frac{\partial f}{\partial a} \right)^2 (\Delta a)^2 + \left(\frac{\partial f}{\partial b} \right)^2 (\Delta b)^2 + \left(\frac{\partial f}{\partial c} \right)^2 (\Delta c)^2 \quad (5)$$

We see that there are two contributions from each variable to the total uncertainty in f , one from the partial derivative and one from the component uncertainty. Clearly, larger measured uncertainties will result in a larger derived uncertainty. Also, if f depends strongly on a given variable, that variable will contribute more strongly to Δf via the partial derivative. Because uncertainties sum nicely after squaring each one, we say that Gaussian errors **add in quadrature** like the components of a vector or the sides of a right triangle. If you have never before seen a partial derivative, the concept is really quite simple. While the function f depends on several variables, a partial derivative is a derivative with respect to only one variable, treating all of the other variables like constants. So to evaluate $\partial f / \partial a$, we treat b and c as constants. Once again, for better or worse, we have assumed that all relevant variables, including the derived quantity, have Gaussian distributions. If and when more care is required, these assumptions can be relaxed at the expense of a more complicated analysis.

Now consider a specific example: the calculated density ρ of a circular metal cylinder:

$$\rho = \frac{\text{mass}}{\text{volume}} = \frac{M}{\pi R^2 L}, \quad (6)$$

Given a Vernier caliper and an electronic scale we find that $M = 185.2$ g, $L = 6.23$ cm and $R = 1.05$ cm, from which we calculate the density to be $\rho = 8.58$ g/cm³. Based on the precision of the measuring instruments, you deduce that associated errors are $\Delta M = 0.1$ g and $\Delta L = \Delta R = 0.01$ cm. Applying Gaussian error propagation we obtain

$$\frac{\partial \rho}{\partial M} = \frac{1}{\pi R^2 L} = \frac{\rho}{M} \quad \frac{\partial \rho}{\partial L} = \frac{-M}{\pi R^2 L^2} = -\frac{\rho}{L} \quad \frac{\partial \rho}{\partial R} = \frac{-2M}{\pi R^3 L} = -\frac{2\rho}{R}$$

from which we see that

$$\left(\frac{\Delta \rho}{\rho}\right)^2 = \left(\frac{\Delta M}{M}\right)^2 + \left(\frac{\Delta L}{L}\right)^2 + 4\left(\frac{\Delta R}{R}\right)^2. \quad (7)$$

Using equation (7) to solve for $\Delta \rho$, we obtain:

$$\Delta \rho = \rho \sqrt{\left(\frac{\Delta M}{M}\right)^2 + \left(\frac{\Delta L}{L}\right)^2 + 4\left(\frac{\Delta R}{R}\right)^2} = \left(8.58 \frac{\text{g}}{\text{cm}^3}\right) \sqrt{\left(\frac{0.1}{185.2}\right)^2 + \left(\frac{0.01}{6.23}\right)^2 + 4\left(\frac{0.01}{1.05}\right)^2} = 0.16 \frac{\text{g}}{\text{cm}^3}$$

Thus we can now say that the density of the cylinder is $\rho = (8.58 \pm 0.16)$ g/cm³. By looking at the above expressions, which quantity would you like to measure with greater accuracy and why?

The Gaussian error propagation formula in Eq. (5) is a general formula that can be applied to any non-pathological function. In the following two special cases (additive and multiplicative errors), it can be reduced to an even simpler form.

1. If $f = \pm a \pm b \pm c$, then

$$(\Delta f)^2 = (\Delta a)^2 + (\Delta b)^2 + (\Delta c)^2. \quad (8)$$

2. If $f = a^{\pm i} b^{\pm j} c^{\pm k}$ (e.g. abc or a/bc , etc.), then

$$\left(\frac{\Delta f}{f}\right)^2 = i^2 \left(\frac{\Delta a}{a}\right)^2 + j^2 \left(\frac{\Delta b}{b}\right)^2 + k^2 \left(\frac{\Delta c}{c}\right)^2. \quad (9)$$

The quantity $\Delta f / f$ is termed the fractional or relative error in f , while the quantity $(\Delta f / f) \times 100$ is termed the percent error in f .