what "most likely" means.)

The assignment of probable errors to physical data is not easy. Some sources of error can be estimated fairly accurately; other sources may be difficult or impossible to estimate. The history of physics has many notorious examples of experimenters who have grossly underestimated the errors in their measurements. This is partly the result of human nature–hoping for a precise result–and partly the result of lack of knowledge. Sometimes there are sources of errors that the experiments didn't know about; sometimes they knew about a source of error but didn't know how to properly estimate the effect on the experiment.

Even if you don't go into scientific research as a career, what you learn about probable errors in the course will be useful to you. We are constantly deluged with data and statistics. Sometimes it is interpreted correctly by the "experts"; sometimes the "experts" present an incomplete picture that tends to support their own opinions. Hopefully in this course you will learn enough to help you understand how to evaluate data.

In the following sections, we give a brief summary of definitions and practical information on errors and treatment of data. It is a bit simplified, but it will serve as a useful introduction to the subject. If you are interested in more details, or in the derivations of the equations and conclusions presented, a list of references is given at the end of the handout.

Systematic vs Random Errors.

Random errors (often called statistical uncertainties) are those produced by unknown and unpredictable variations in the experimental situation. *Systematic errors* are errors associated with a particular instrument or experimental technique. The difference is perhaps best illustrated with an example from target shooting.

In Figure 1 below, four targets are shown after many shots have hit. If there were no errors of any kind, all the bullets would hit exactly in the center of the target. We see that random errors cause the bullets scatter about the center of the target with more hits closer the center, and fewer hits farther from the center. Systematic errors, cause the bullets to hit away from the center of the target, but always in the same direction. The random errors may be due to jitter while aiming, variations in the cartridges, etc. The systematic errors may be due to a consistent wind, mis-aligned sights, or a consistent bias while aiming.

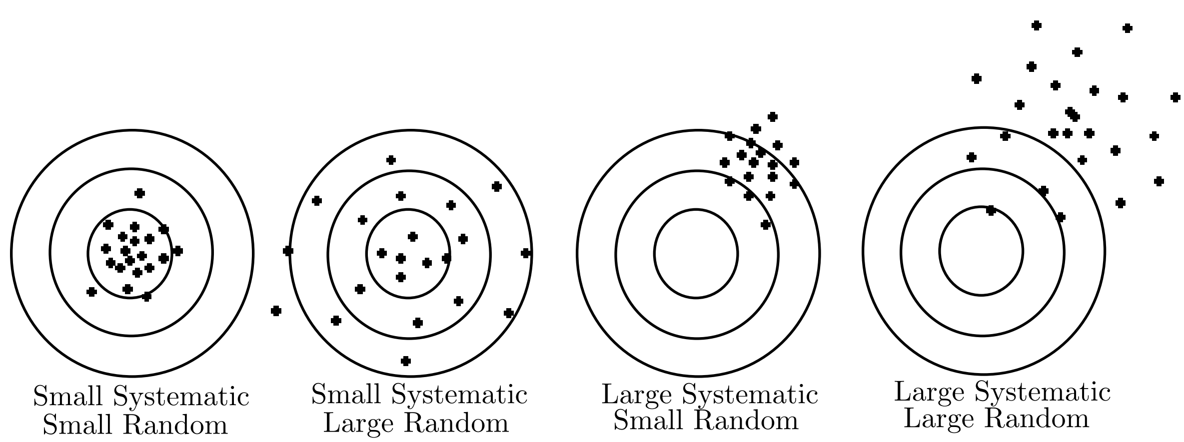

Figure 1: Difference between systematic and random errors.

Note that the effects of random errors can be minimized by taking many measurements and averaging the results. Systematic errors cannot be minimized in this way. On the other hand, if the value of a systematic error can be calculated or measured independently, its effect on the result may be minimized by subtracting it from the result. Random errors cannot be minimized in this way. In most experiments, a combination of random and systematic errors are present at the same time.

The difference between random errors and systematic errors is often expressed using two words that are often confused. The words are *precision* and *accuracy.* Although they both relate to experimental uncertainty, their meanings are quite distinct. *Precision* refers to how large the uncertainty in an experimental quantity is, compared to the size of the quantity itself. *Accuracy* refers to how close an experimental result is to the value of the experimental quantity measured (the accuracy of an experimental result may not be known at the time the experiment is performed). For example, two students set out separately to measure the acceleration due to gravity near the earth's surface, which has the value of 9.80 m/s^2, and they find:

Student A: $g = 9.81 \pm 3.00 m/s^2$
Student B: $g = 8.44 \pm 0.01 m/s^2$.

Student A's result is very accurate but not precise, whereas student B's result is precise but not accurate. If each student has done many trials and averaged the data to obtain their result, then the results tell us something very important. The results of student A's trials differ widely but average to very near the true value. This may indicate that the measurement error is mainly statistical. The results of student B's trials are very consistent, but different from the true value. This may indicate that the measurement error is mainly systematic.

The Normal Distribution, Mean, and Standard Deviation.

Suppose we consider a measurement whose result can take on a continuous range of values. To be concrete let us imagine a very simple experiment. We want to measure the time it takes a ball to fall 1.00 meter. To get an accurate value we use a good stopwatch and repeat the measurement 200 times. Figure 2 shows the results of our hypothetical experiment in the form of a histogram. The vertical height of each bin gives the number of measurements that fell within the range of the bin. For example, there were 29 measurements with fall times between 0.485 s and 0.495 s.

The distribution in Figure 2 is somewhat idealized, but it is typical of what real data from a well designed and executed experiment might look like. The most notable features are:

1. The values are clustered about a well-defined mean value which is close to the most probable value. (the value of t where the distribution has its maximum). The arithmetic mean of the t values in Figure 2 is 0.497 sec.

2. Values which are far from the mean are very unlikely.

3. The distribution is reasonably symmetric about the mean; there is no obvious skewing

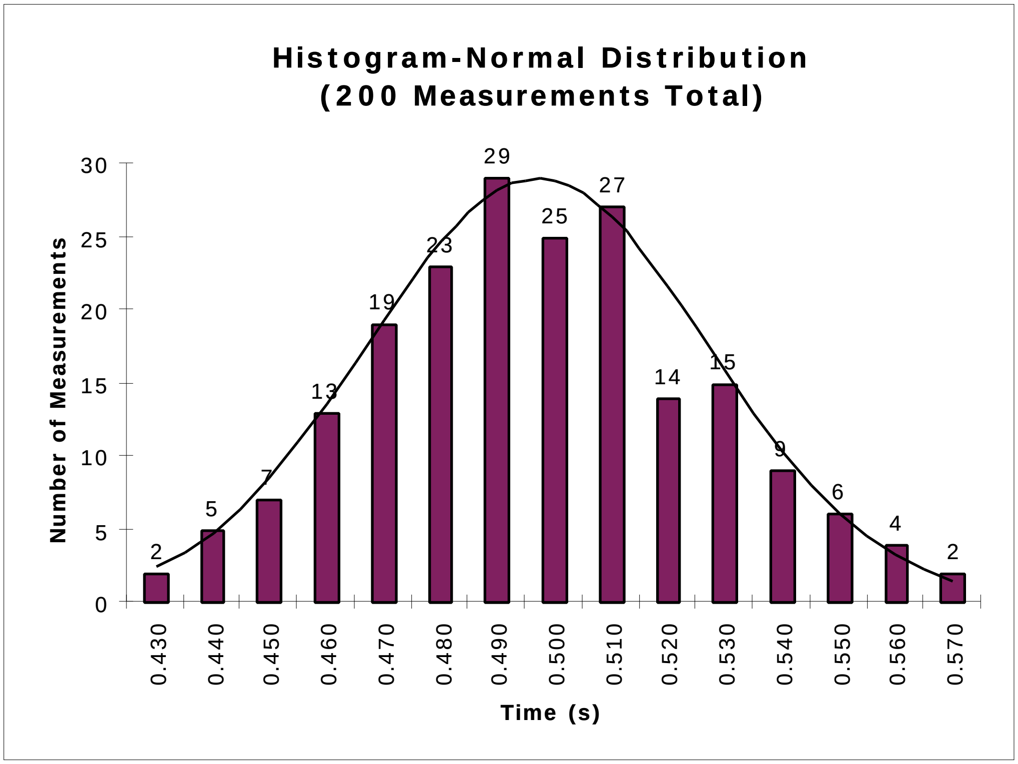

Figure 2: Histogram for a hypothetical experiment.

toward the high or low side.

If we took many, many measurements and made the bins very fine, our histogram might begin to look like the smooth bell-shaped curve. This curve is the limiting case in an ideal situation. It is referred to as the normal distribution or Gaussian distribution. Measurement errors that follow this distribution are said to be normally distributed. The mathematical theory of of the gaussian curve is important, but beyond the scope of this handout. What is important to us is that if the fluctuations in a given quantity are random, then a distribution resembling a normal distribution is usually found and we can use the results of the mathematical theory to estimate uncertainties in our data.

The bell shaped curve and, to a good approximation. the histogram in Figure 2 can be characterized by two quantities, the mean value (or average) and the width. The mean value (usually denoted with a bar as $\bar{t}$) is just the arithmetic average of the data,

$$\bar{t} = \frac{1}{N}\left[t_1 + t_2 + t_3 + \ldots + t_N\right],$$

where N is the total number of measurements and t_1, $t_2 \ldots t_N$ are the measured times. This equation can be simplified by using the summation notation:

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_i,$$

where $\sum$ stands for sum and i is the index which runs from 1 to N. The most common way to characterize the width of the distribution is called the standard deviation which we will denote by sigma (σ). It is calculated by averaging the squares of the differences between the measured values and the average value. In equation form:

$$\sigma = \left(\frac{1}{(N-1)} \sum_{i=1}^{N} (\bar{t} - t_i)^2 \right)^{\frac{1}{2}}$$

If the measurements follow a normal distribution then about 68% of the measurements will lie within $\pm 1\sigma$ of the mean. For the data in Figure 2, we can estimate the σ by counting off 68 measurements in either direction from the mean. This includes a band of about ± 0.030 sec, so the standard deviation per measurement is about 0.030. Note that the equation for σ is only valid if the data follows a normal distribution so sometimes it is desirable to plot a histogram for a given set of data before using the equation.

The equation for σ shown above gives the probable uncertainty for any one of the time measurements, but often we want to estimate the uncertainty in the mean value, $\bar{t}$. Since we have made many measurements, the uncertainty in the mean is much lower than the uncertainty in any of the individual measurements. We denote the uncertainty in the mean by σ_μ (sigma-mu), and it is given by

$$\sigma_\mu = \sigma / \sqrt{N},$$

where σ is just the standard deviation given above.

In the example above, the mean time was $\bar{t} = 0.497$ s and the standard deviation was $\sigma = 0.03$ s. Thus $\sigma_\mu = 0.03/\sqrt{200} \approx 0.002$ s, so we would write

$$\bar{t} = 0.497 \pm 0.002 \text{ s.}$$

Often called the 68% confidence interval ($\bar{t} \pm 1\sigma$), this means that if the same set of 200 measurements were performed many times, in about 68% of them the average time, $\bar{t}$ would fall between 0.499 s and 0.495 s. (Sometimes also specified is the 95% confidence interval which corresponds to $\bar{t} \pm 2\sigma_\mu$.)

The above discussion assumes that each of the measurements have equal intrinsic accuracy and are therefore equally weighted in the determination of the mean. If some of the measurements are more or less certain than others, more advanced techniques for including a weighting factor when determining $\bar{t}$, σ, and σ_μ exist. These techniques are discussed in the references.

Experiments Whose Outcome is an Integer: The Square Root rule.

Often the result of an experiment or measurement is an integer; for example, the number of mice out of an initial sample of 100 that die within one year or the number of radioactive nuclei out of a sample that decay in one second. The standard deviation of the number of such "events" (i.e., deaths, decays, etc.) can be estimated by the square root rule. If N is the number of events, then the standard deviation is given by

$$\sigma_\mu = \sqrt{N}$$

For this to be an accurate estimate, the following conditions must be satisfied. (The better they are satisfied the better the estimate of σ_μ.)

1. The number of events N must be large. (Some would argue that $N > 10$ is enough.)
2. The probability that one specific event (i.e. the death of a specific mouse, or the decay of a specific nuclei) occurring during the measurement must be small. If, for example, we did an experiment to see how many mice out of 100 will die in 75 years, the answer would be that all the mice would surely die or 100 ± 0. The uncertainty would not follow the square root rule because the chances of any specific mouse dying are not small. On the other hand, if we start with 10^8 radioactive nuclei which decay at a rate of 10^3 per second, in a 10 second experiment $N = 10,000 \pm 100$. The square root rule should work very well because $N >> 1$ and the probability of any given nucleus decaying is 10^{-4} during the experiment.

Error Propagation.

Usually we cannot make a direct measurement of the quantity we are interested in. We must measure another quantity or quantities and calculate the desired quantity from them. In our example of the falling ball, we might be interested in the acceleration of gravity g. It can be calculated from the fall time t and the distance of fall y from the equation

$$g = 2y/t^2.$$

We now ask how the uncertainty in g can be found if the uncertainty in t and y are both known.

The rules for error propagation can be readily derived using calculus. We merely state the results here. In this section we shall denote the quantity we wish to measure as Q and its uncertainty as δQ. Also of interest is the ratio $\delta Q/Q$, the fractional uncertainty. To distinguish δQ from $\delta Q/Q$, δQ is often called the absolute uncertainty. A, B, and C will denote the quantities that are measured directly.

Rule #1 If $Q = cA$ where c is a constant (or a quantity with negligible fractional error), then

$$\frac{\delta Q}{Q} = \frac{\delta A}{A} \quad \text{or} \quad \delta Q = c\delta A \tag{1}$$

Rule #2 If $Q = cA^m$ where m is some power (positive, negative, integer, or fraction), then

$$\frac{\delta Q}{Q} = m\frac{\delta A}{A} \quad \text{or} \quad \delta Q = cmA^{m-1}\delta A \tag{2}$$

If Q depends on two quantities A and B, the following rules are useful:

Rule #3 If $Q = A + B$ or $Q = A - B$ then

$$\delta Q = \sqrt{(\delta A)^2 + (\delta B)^2} \tag{3}$$

Rule #4 If $Q = cA^mB^n$ where m and n are powers (positive, negative, integer, or fraction) and c is constant, then

$$\frac{\delta Q}{Q} = \sqrt{\left(\frac{m\delta A}{A}\right)^2 + \left(\frac{n\delta B}{B}\right)^2} \tag{4}$$

With hand calculators available, it becomes practical to calculate the uncertainty in Q by the "brute force" method. This method is helpful when the functional form for Q is not covered by the simple examples above. For this method, simply calculate Q for the mean value of A then recalculate Q for $A + \delta A$. The difference between the two results will give δQ. If Q is a function of more than one variable, vary each separately; then combine the separate terms in quadrature. For example, if $Q = f(A, B, C)$, then the uncertainty in Q is

$$\delta Q = \sqrt{(\delta Q_A)^2 + (\delta Q_B)^2 + (\delta Q_C)^2} \tag{5}$$

where δQ_A represents the change in Q when A is varied by δA. Often one or two of the terms will dominate; the others will give negligible contribution to δQ.

As a numerical example, suppose $g = 2y/t^2$ and we know from a series of measurements that $\bar{y} = (1.010 \pm 0.014)$ m and $\bar{t} = 0.454 \pm 0.008$ s. Clearly, the best estimate for g is $g = 2(1.010)/(0.454)^2 = 9.80$ m/s^2, but what is the uncertainty in g? Using the brute force technique we find that g changes by approximately 0.36 m/s^2 if we change t by 0.008 s, and by 0.14 m/s^2 if we change y by 0.14 m. Then from Eq. 5, the overall uncertainty in g is the square root of the sum of the squares of the contributions of changing t and changing y,

$$\delta g = \sqrt{(.14)^2 + (.36)^2} = 0.38 \text{ m/s}^2$$

If instead, we calculate δg from Eq. 4, we have c =2, $A = 0.014$, $m = 1$, $B = 0.454$, $\delta B = 0.008$, $n = -2$ and

$$\frac{\delta g}{g} = \sqrt{\left(\frac{0.014}{1.01}\right)^2 + \left(\frac{(-2)(0.008)}{0.454}\right)^2} = 0.038$$

Therefore, $\delta g = 0.038g = 0.37$ m/s^2. Thus either method gives an overall uncertainty in the measurement of g of about 0.38 m/s^2 and we would quote our result for g as

$$g = 9.80 \pm 0.38 \text{ m/s}^2.$$

The fractional error in g is $\delta g/g = 0.38/9.80 \approx 0.04$ or 4%.

Estimating Uncertainties.

In most of the experiments you will be asked to estimate the uncertainties in the quantities you measure, such as a distance or time interval. This requires some practice so let us explore an example. In reading a ruler or a meter scale you can usually interpolate between the divisions on the scales. A reasonable minimum error might be about one third of the smallest division on the ruler. For example, if you are measuring the distance, x between two points on a meter stick whose smallest division is 0.001 m (or 1 mm) and each point is known to about one-third of a mm, we must subtract the two numbers to get the length (even if you put one of the points at 0, both points are still uncertain by the same amount). Using rule #3 above for subtraction we have

$$\delta x = \sqrt{(0.33)^2 + (0.33)^2} \approx 0.47 \approx 0.5 \text{ mm}$$

and the net error in the measurement turns out to be about one-half of the smallest division.

Notice that the absolute error in the measurement for a given meter stick is constant regardless of how large x is. The fractional error can change dramatically. If x is 105.0 mm then the fractional error is $\delta x/x = 0.5/105.0 \approx 0.5\%$, but if x were only 5.0 mm then the fractional error would be 10%. This is why we use special tools with smaller divisions to measure very small distances.

Rounding Off.

The number of significant figures you give in your answer should be consistent with the precision of your answer. It doesn't make sense to give too many significant figures (e.g., 9.5876543±0.67789798), or too few significant figures (e.g., 10±0.014). Instead write: 9.58±0.68 and 10.000±0.014.

Graphs.

Graphs are a very useful means of exhibiting data so that it can be readily visualized. Graphs can be used for smoothing out data and often provide a convenient means of analyzing data to provide a "best fit" value of some physical quantity. We shall assume you have had some experience with graphing data and restrict our comments to the finer points.

Choosing the best scale– This is a bit of an art. Often you will need to make more than one try. The size of your graph should reflect the accuracy of the data. If the errors are 1% then you will want a large graph, whereas if they are 20% there is no point in drawing your graph very large. On the other hand, if you have 0.1% it would be silly to do the analysis on a graph which is only accurate to 1%. In that case, a numerical analysis routine such as a linear regression (available on nearly all scientific calculators) may be more appropriate.

Error Bars– Usually the quantities you are plotting will have an uncertainty associated with them. The uncertainty is indicated by drawing a line through the point to indicate the uncertainty. For example, if the velocity is measured as $v = 1.80 \pm 0.15$ m/s, you would plot the point at 1.80 and the vertical bar would extend from 1.65 to 1.95. Figure 3 on the next page gives an example of linear experimental data with error bars. If there is uncertainty in both the x and y coordinate, you would generally show this with both a horizontal and vertical error bar. Usually the error bars indicate one standard deviation.

Drawing the best fit curve–In practice, graphical analysis is often used to see if your data are consistent with a theory and perhaps to determine the value of some constant in the theoretical expression which best fits your data. For example, an object dropped from rest is expected to have a speed at time t of

$$v = gt$$

where g is the acceleration due to gravity. This equation describes a straight line, and g could be determined by graphing different experimental values for v and t then determining the slope of the resulting line (described in the next section). Suppose, however, that your data consisted of various values of position y and time t of a ball dropped from rest. Then we expect

$$y = \frac{1}{2}gt^2$$

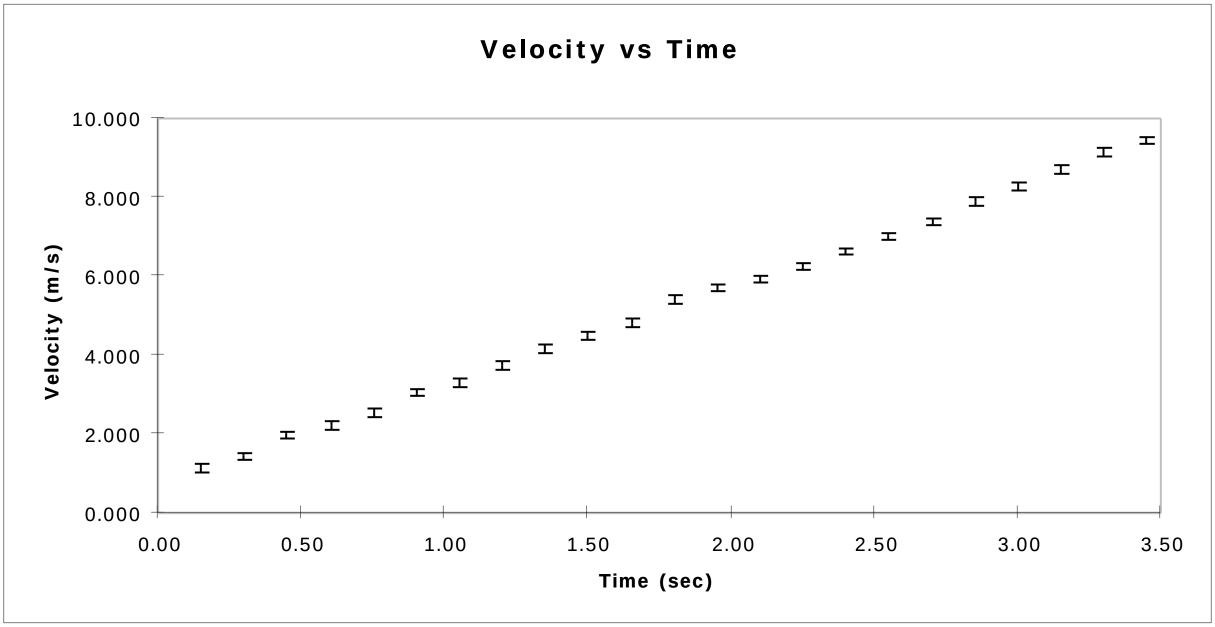

Figure 3: Plot of Linear Data.

This is the equation for a parabola which would be much more difficult to fit graphically. One way to to determine g in this situation, is to linearize the equation, that is, find the proper parameters to graph so that the resulting function is a straight line. In the above example, plotting y vs t^2 results in a straight line with slope equal to $g/2$.

A linearizing technique for two common functions that at first glance may appear to be non-linearizable appears in the next section. Some functions cannot be linearized and must be solved with advanced techniques. You will find for all of the labs for 221 and 222 linear fits are possible.

To find the best fit straight line to your data, lay a straight edge along the data points and move it until it passes through as many points as possible (the most common mistake here is to simply connect the first and last point). If the error bars are realistically estimated and the data really do follow a linear relationship then on average you should be able to draw a straight line through about 68% of the error bars. If you can draw your best fit straight line through all of the error bars, then either you have been quite lucky, or more likely, you have overestimated the error bars. If your line misses substantially more than two-thirds of your error bars, either you have underestimated your error bars, or the data is not truly linear.

Finding the slope and its uncertainty—Often you need to find the slope of the best fit straight line in order to find the quantity of interest in the experiment. Usually you also want to estimate the uncertainty in the slope. To do this, simply vary the slope of your best fit straight line within the limits of the error bars for the data, and recalculate the slope of the new line(s). This takes some judgment. As a rule of thumb, you'll want to draw the new line(s) near the top edge of the error bars on the first few points and near the bottom edge of the error bars on the last few points (or visa versa).

Exponential and Power Law Functions.

In some cases the quantity that is to be determined by experiment is in the exponent. Two common examples of this are radioactive decay (exponential) and period of a simple pendulum (power law). In radioactive decay, the number of counts per second, R, is expected to vary with time as

$$R = C \exp(-\lambda t)$$

where C and λ are quantities to be determined by measuring R as a function of time. In Figure 4a, typical data for this type of experiment is shown. Fitting the data directly to determine λ would be tedious, but taking the logarithm of both sides of the above equation yields

$$\ln R = -\lambda t + \ln C.$$

Thus, if we use our calculators to determine the the ln of R for all values of t and replot the data as in Figure 4b, we see that the data follow a straight line with slope equal to $-\lambda$, and y-intercept equal to $\ln C$. Note that the error bars in the new plot are not symmetric. To find the error bars for the new graph you must take the logarithm of both $R + \delta R$ and $R - \delta R$.

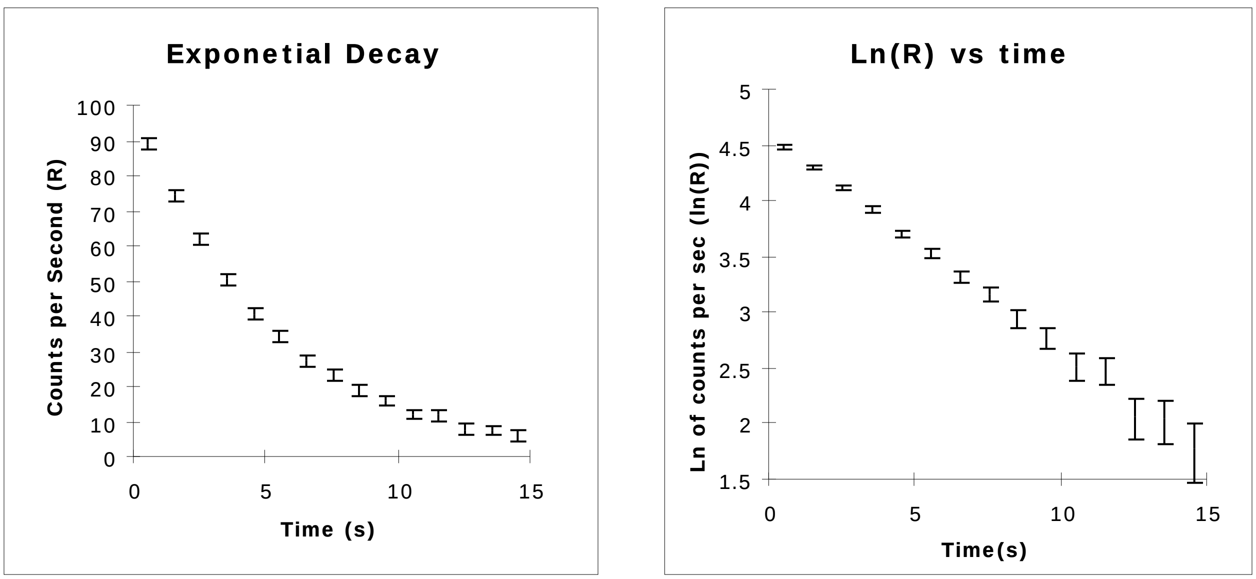

Figure 4: (a) Raw experimental data for an exponential decay. (b) Linear plot of same data using logarithms.

In the case of a power law, the function is of the form

$$y = cx^b$$

where the quantities c and b are to be determined. In this situation, taking the logarithm of both sides yields

$$\ln y = b \ln x + \ln c,$$