Data Analysis: A UT Primer

by Dr. G. Bradley Armen

Department of Physics and Astronomy
401 Nielsen Physics Building
The University of Tennessee
Knoxville, Tennessee 37996-1200

Introduction

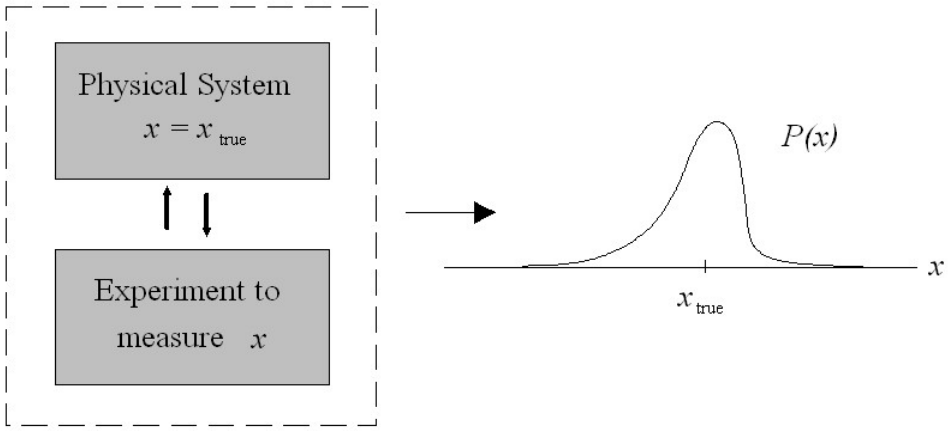
There's no way out of it: experimental physics involves not only measuring something, but also determining how significant that measurement is. Unlike introductory physics laboratory experiments (where the correct answer is known), in doing new research it's inherent that we work on the edge — measuring new quantities using new techniques. In these circumstances our results are usually 'rough'. It's critical in reporting our data to the rest of the scientific community to include a discussion of the significance of our measured values. Alternately, when performing precision measurements, where we experiment to improve the accuracy of some value, it is equally vital to have a deep understanding of error analysis.

The following document is an attempt to briefly outline the most basic ideas of data analysis, which anyone working with experimental data should be aware of. The subject is actually quite subtle, but here we cheerfully ignore anything difficult.

Measurement

Let's begin by considering the measurement of a single quantity x. This might be the length of a bolt or the mass of an elementary particle. Thinking philosophically about the procedure, we see that we have a machine of sorts: we have an isolated physical

system in which the parameter takes on the true value x_{true}, and we have another system (the experiment) which interacts with the first system to measure x.



When we run the machine we get a specific output x_i –a measurement! It's a sad feature of the measurement process that the outcome is uncertain. What we can say about the experiment is that there is a probability distribution $P(x)$ for the outcome. If we run the machine (perform the experiment once) then the probability of getting a result between x and $x + dx$ is given by $P(x)dx$. This distribution is often referred to as the *parent* distribution of the measurement. $P(x)$ contains *all* possible information about our experiment, but is totally unknown. If our design is good then the experiment has good *accuracy*: the peak value of $P(x)$ is near to x_{true} (no *systematic* errors in our procedure). The width of $P(x)$ provides information about the *precision* of our measurement.

Ideally, (if $P(x)$ were somehow known) we could report the entire distribution and revel in a job well done. However, for practical purposes this can seldom be done; instead we report key aspects of the distribution. Two of the most important (conventional) parameters describing a distribution are the *mean* (average)

$$\mu = \int xP(x)dx,$$

and *variance*

$$\sigma^2 = \int (x-\mu)^2 P(x)dx.$$

The mean tells us something about our measurement of x_{true} (assuming good accuracy), and the *standard deviation* σ gives an idea about the uncertainty *inherent* to our experiment (our precision). We say then that our result is

$$x_{true} \cong \mu \pm \sigma .$$

If $P(x)$ were a Gaussian (or 'normal') distribution[1] (usually assumed) then we expect 68% of our measurements to lie within $\pm\sigma$ of μ, and 96% of them within $\pm 2\sigma$. Hence, we expect about one measurement in three to lie more than $\pm\sigma$, but less than $\pm 2\sigma$, from μ.

If the parent distribution is badly asymmetric more parameters can be used to characterize $P(x)$, such as the mode (most probable value of *x*) and median (the value of *x* for which there is equal probability above and below). For a symmetric function, the mean, mode, and median are all equal.

Sampling

If the parent distribution is the goal, but unknown, how can we construct it? In theory this is simple: we make an infinite number of measurements, and a histogram of the results gives $P(x)$. In practice this isn't possible and often we are able to repeat the measurement only a very few times. Suppose we do so, the question becomes what can *N* measurements $\{x_1, x_2, ..., x_N\}$ tell us about $P(x)$?

One obvious thing we can do with our data is to construct the *sample* mean

$$\bar{x} = \frac{1}{N}\sum_i x_i$$

[1] The Gaussian distribution is $P(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$

and the *sample* variance[2]

$$s^2 = \frac{1}{N-1}\sum_i (x_i - \bar{x})^2 .$$

Since the values $\{x_i\}$ are randomly sampled, the observed values of $\bar{x}$ and s are also random. From the parent distribution we can construct the probability distributions $P_N(\bar{x})$ and $P_N(s^2)$, *i.e.* the probabilities that, given we take N samples, the sample mean will be $\bar{x}$ and the sample variance will be s^2. With very few assumptions it's found that these distributions are Gaussian: The function $P_N(\bar{x})$ is centered at μ with a variance of σ^2/N. Similarly, $P_N(s^2)$ is centered at σ^2 with a variance of $\sigma^2/2(N-1)$. Approximating $\sigma \approx s$, our best estimate about the parent distribution is

$$\mu \approx \bar{x} \pm \frac{s}{\sqrt{N}} \qquad \text{and} \qquad \sigma \approx s \pm \frac{s}{\sqrt{2(N-1)}} .$$

Notice that (as expected) when $N \to \infty$ we recover the parameters μ and σ describing $P(x)$ exactly.

Our precision grows only slowly with increasing N: Suppose we make an initial measurement with N samples, but are unhappy with the precision of our result. We can repeat the process, so that we have in total $2N$ samples, but this only decreases our error in our estimate of μ by a factor of $1/\sqrt{2} \approx 0.7$. If we wanted to increase our precision by an order of magnitude, we'd need a total of $100N$ samples! This is an unhappy, but inescapable, fact of experimental work. We'll encounter this same problem when discussing counting experiments.

One final note: When we have very few samples, our error in σ can be large. Hence, in using s as an approximation to σ for our uncertainties, we may be significantly underestimating our error in μ.

[2] Dividing by N-1 instead of N is one of those subtle points we are ignoring.

Correlation

Instead of returning a measurement of a single parameter (x), consider what happens when our experimental machine returns several parameters — say u and v. These might be a simultaneous measurement of length and width, pressure and temperature, and so on. The experiment's parent distribution is then of the form $P(u,v)$. What kind of parameters can we use to describe this joint distribution? If u and v were truly independent parameters we could factor the parent distributions into

$$P(u,v) = P_u(u)P_v(v).$$

In this case we can happily define a mean and variance for each distribution: (μ_u, σ_u) and (μ_v, σ_v).

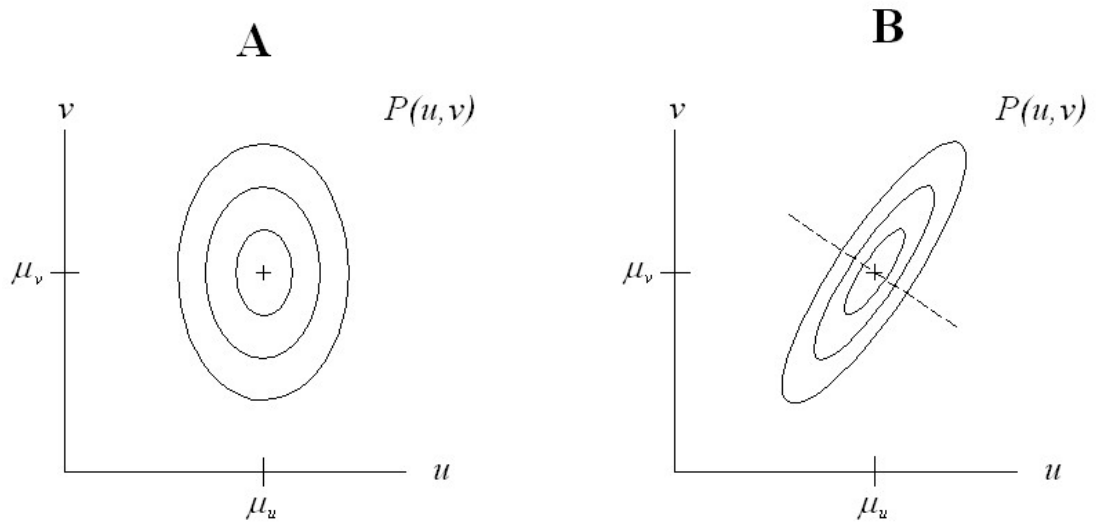
However, it's often the case that u and v aren't independent. Consider the case of measuring an electric potential as a function of distance between two electrodes. Our measurement pair is then (x,V), but we know that V varies with x: $V(x)$. Hence, in trying to measure V at the point x, we know there must be some correlation between the sample points x_i and V_i. What do we do?

We can still use the concept of mean and variance by *marginalizing* the unwanted parameter:

$$\mu_u = \int u \int P(u,v)\, dv\, du$$

$$\sigma_u^2 = \int (u - \mu_u)^2 \int P(u,v)\, dv\, du\,,$$

and similarly for (μ_v, σ_v). Thus μ_u is the average value of the measured u, *given* that we ignore all accompanying information about v. The following figure shows a sketch of two possible parent distributions centered at the point (μ_v, σ_v):



Each distribution in the figure has similar values of the marginalized errors σ_u and σ_v: Discarding any knowledge about *v*, either distribution A or B gives us the same spread in our precision of *u* (and vice versa for *v*). However, distribution B shows a more precise state of knowledge than A, it's much narrower along the dashed diagonal shown. Thus, if we move along certain directions (*i.e.* constrain the 'motion' of *u* and *v* together), our error is somehow less. In this example our measurements of *u* and *v* are positively correlated, a larger (or smaller) result for a measurement of u_i is correlated to some extent with a simultaneously larger (or smaller) result for v_i.

To describe the degree of correlation between *u* and *v*, we use the *covariance*

$$\sigma_{UV} = \int\int (u - \mu_U)(v - \mu_V) P(u,v)\, du\, dv\,.$$

One can immediately see that $\sigma_{UV} = 0$ if *u* and *v* are independent (substitute $P(u,v) = P_U(u)P_V(v)$ into the integral). If $\sigma_{UV} > 0$ *u* and *v* are positively correlated, *i.e.* there is a tendency in the measurements for *u* to increase when *v* does. If $\sigma_{UV} < 0$ *u* and *v* are negatively correlated, *i.e.* if *u* increases there is a tendency for *v* to decrease, and vice-versa.

To determine how significant the correlation is, independent of scale, we examine the so-called cross-correlation coefficient $C = \sigma_{UV} / \sigma_U \sigma_V$. For instance, a totally-dependent linear relation $u = \alpha v$ gives $C = \pm 1$ if $\alpha > 0$ or $\alpha < 0$.

To get some idea of the parent distribution we can again perform our experiment *N* times. This gives *N* pairs of data (u_i, v_i). Ignoring either *u* or *v*, we can calculate the marginalized sample means and variances $(\bar{u}, s_U^2)$ and $(\bar{v}, s_V^2)$ as in the last section. Additionally, we should calculate the sample covariance

$$s_{uv} = \frac{1}{N-1} \sum_i (u_i - \bar{u})(v_i - \bar{v}).$$

If this is comparable to $s_U s_V$, we know there is a strong correlation between the data pairs. If there is a strong correlation, we must think very carefully about how to handle our findings — an important subject, but beyond the scope of this work

Error propagation

It's almost always the situation that the quantity we are interested in is derived from the actual measurements. For instance: we measure the voltage *V* across, and the current *I* through, some circuit element and want to report the resistance *R* = *V*/*I*. Given our uncertainties in *V* and *I*, what can we say about our uncertainty in this measurement of *R*? That is, how do our measurement errors propagate when used in a calculation?

Let's consider the problem of some derived quantity *z* which is a function of two measured quantities *u* and *v*. We have some functional relationship $z = f(u, v)$. Now, the true mean of *z* is given by

$$\mu_z = \iint f(u,v) P(u,v)\, du\, dv.$$

While it's not necessarily the case, it is always assumed that

$$\mu_z \approx f(\mu_u, \mu_v).$$

By doing so, we can then make a Taylor series expansion of *z* about its 'mean':

$$z - \mu_z = (u - \mu_u)\left[\frac{\partial z}{\partial u}\right] + (v - \mu_v)\left[\frac{\partial z}{\partial v}\right] + \ldots,$$

where the derivatives are evaluated at the point (μ_u, μ_v). The variance of *z* is

$$\sigma_z^2 = \iint (z - \mu_z)^2 P(u,v)\, du\, dv.$$

Substituting the Taylor expansion into the integral, expanding the square, and applying a few definitions gives (to lowest order)

$$\sigma_z^2 = \sigma_u^2\left[\frac{\partial z}{\partial u}\right]^2 + \sigma_v^2\left[\frac{\partial z}{\partial v}\right]^2 + 2\sigma_{uv}\left[\frac{\partial z}{\partial u}\right]\left[\frac{\partial z}{\partial v}\right].$$

This equation is used, in lieu of any knowledge about $P(u,v)$, for deriving the propagated uncertainty in *z*. If *u* and *v* are independent $\sigma_{uv} = 0$, we have the more familiar relationship

$$\sigma_z^2 = \sigma_u^2\left[\frac{\partial z}{\partial u}\right]^2 + \sigma_v^2\left[\frac{\partial z}{\partial v}\right]^2 \qquad (u \text{ and } v \text{ independent!}).$$

These relations can be generalized to include more variables, $u \to u_i$ as

$$\sigma_z^2 = \sum_i \sigma_{ui}^2\left[\frac{\partial z}{\partial u_i}\right]^2 \qquad (\text{all } u_i \text{ independent}).$$

The following table lists results for some common situations, assuming independent variables (and constants *a* and *b* which are exact):

Relation $z(u,v)$	Independent error
Scaling: $z = au$	$\sigma_z = a\sigma_u$
Inversion: $z = a/u$	$\frac{\sigma_z}{\mu_z} = \frac{\sigma_u}{\mu_u}$
Addition and subtraction: $z = au \pm bv$	$\sigma_z^2 = a^2\sigma_u^2 + b^2\sigma_v^2$
Multiplication or division: $z = auv$ or $z = au/v$	$\frac{\sigma_z^2}{\mu_z^2} = \frac{\sigma_u^2}{\mu_u^2} + \frac{\sigma_v^2}{\mu_v^2}$
Power: $z = au^{\pm b}$	$\frac{\sigma_z}{\mu_z} = b\frac{\sigma_u}{\mu_u}$
Exponential: $z = ae^{bu}$	$\frac{\sigma_z}{\mu_z} = b\sigma_u$
Natural logarithm: $z = a\ln(bu)$	$\sigma_z = a\frac{\sigma_u}{\mu_u}$

We see that for addition and subtraction the scaled absolute errors are added in quadrature (*i.e*. sum of squares), but for multiplication and division the *relative* errors σ/μ add in quadrature.

A simple example:

Suppose we measure a current of $I_1 = 115 \pm 31$ mA flowing into a circuit junction by one wire, and $I_2 = 241 \pm 63$ mA by another. The total current $I = I_1 + I_2$ is 356 ± 70 mA. Note, that the quadrature error is much smaller than if we had simply added the errors (± 94 mA). This is a consequence of the statistical nature of the two readings: it is very unlikely to have both currents take on their extreme values simultaneously if they are independent.

A less simple example:

Suppose we measure a voltage of $V = 180 \pm 18$ V across a resister and a current $I = .782 \pm .067$ A through it. The resistance is $R = V / I$. Assuming that our measurements of *V* and *I* are independent, the *relative* errors now add in quadature. We have $R = 230 \pm 30\,\Omega$. Here the 8.5% error in current combines with the 10% error in voltage to give a relative error of 13% in *R*.

But wait! We just used the fact that *V* and *I* are related by Ohms law. Perhaps the variations in *V* and *I* are correlated, at least to some extent! What happens to our error in *R*? Going back to our data samples we calculate the covariance s_{VI} and find there is a degree of correlation $s_{VI} = 0.5 s_V s_I$ (note the positive correlation: an increase in *I* produces an increase in *V*). Using the general formula above we find (after some manipulation)

$$\frac{\sigma_R^2}{\mu_R^2} = \frac{\sigma_V^2}{\mu_V^2} + \frac{\sigma_I^2}{\mu_I^2} - 2\frac{\sigma_{VI}}{\mu_V \mu_I}.$$

We see that the positive correlation decreases our error in *R*. This is because not all of the variation in our samples was statistical, *i.e.* a low current reading I_i is correlated to a low voltage reading V_i so that the actual spread in the values of $R_i = V_i / I_i$ is smaller than we would expect were *V* and *I* independent. Plugging in the numbers gives a revised estimate of 9% relative error, or $R = 230 \pm 22\,\Omega$.

Counting statistics

In many types of measurements our data arises by counting something[3]. It might be anything from the number of neutrons detected in a given amount of time *t* to the number of kids on a school bus. In either case, we assume there is some quantity

[3] In most digital measurements we are actually counting.

governing the result of our measurement of n units: this could be the neutron ejection rate R ($n \approx N = Rt$) or, in the case of the bus, simply the class size ($n \approx N = N_{class}$). In either case, there is uncertainty in how many counts we actually will measure. In the neutron example, the outcome is *inherently* uncertain due to quantum mechanics. For the bus driver, the kids might be running around while he counts. Hence, there is some discrete parent distribution $P(n)$ which describes the experimental process.

It turns out that, for a number of fairly fundamental reasons, *with no further information*[4] than that there is some such expected quantity *N*, the best parent distribution to use is Poissonian[5]. The Poisson distribution has the property that its average value is *N*, ($\mu_n = N$) as is its variance ($\sigma_n^2 = N$). Thus, if we take a measurement and get the value *n*, we can say from our previous considerations that

$$N \approx n \pm \sqrt{n}$$

We see that our estimate of *N* is *relatively* more precise the larger *n* is *i.e.* $\sigma_n / n = 1/\sqrt{n}$. We encountered this same scaling when considering the error in the mean of a measurement sampled *n* times. Obviously, the larger the counts the higher precision of our measurement. The sad result is that the relative error decreases slowly with increasing *n*:

Counts	Relative error (%)
10	31.6
50	14.1
100	10.0
500	4.5
1000	3.2

[4] The bus driver actually has a lot more than this to go on, but let's keep the analogy going for purposes of discussion.

[5] The Poisson distribution is $P(n)=\frac{N^n e^{-N}}{n!}$.

5000	1.4
10,000	1.0
100,000	0.3
1,000,000	0.1

To increase precision we need to acquire more counts. The neutron scientist can do this by increasing the counting time *t*, assuming that the rate is sufficiently time independent. To increase the precision by an order of magnitude, the sample time must be increased by a factor of 100. Alternately, the experiment can be repeated a number of times (with the original counting time *t*). The bus driver is stuck with the latter option, doing recounts.

Suppose we redo the experiment *M* times, producing *M* sample counts $\{n_i\}$. The average value of these samples is

$$\bar{n} = \frac{1}{M}\sum_i n_i = \frac{n_{total}}{M},$$

where n_{total} is simply the total of all counts collected. If our experiment is accurate, we expect $N \approx \bar{n}$. Using our method for error propagation, we can estimate the error in $\bar{n}$ by

$$\sigma_{\bar{n}} = \sqrt{\sum_i \sigma_i^2 \left[\frac{\partial \bar{n}}{\partial n_i}\right]^2} = \sqrt{\sum_i \sigma_i^2 \left[\frac{1}{M}\right]^2}.$$

Since $\sigma_i^2 = n_i$, this simply gives $\sigma_{\bar{n}} = \sqrt{n_{total}} / M$. Thus, our best estimate for *N* is

$$N \approx \frac{1}{M}\left(n_{total} \pm \sqrt{n_{total}}\right).$$

This is reassuring: it should make no difference to our results whether we average the samples, or simply use the total collected with its associated error. What matters for the precision of our measurement is the total number of counted items. The $\sqrt{n_{total}}$ limit in our precision is fundamental to counting experiments, and usually the frustrating design parameter when we plan an experiment.

Linear regression

One of the most common tasks in data analysis is to fit a straight line to data. Suppose we have a paired set of *M* sampled means $(\bar{x}_i, \bar{y}_i)$ with sample errors (s_{xi}, s_{yi}). We think there is some underlying relationship between *x* and *y* of the form $y = Ax + B$, and so want to determine the values of *A* and *B* that are *most* consistent with the data. Since these parameters usually have physical significance, we not only want their values but also an idea of their errors (something an Excel trendline doesn't provide).

To make the problem simple, we consider the case where our errors in the $\bar{x}_i$ are negligible[6]. Thus we consider the $\bar{x}_i$ to be precisely known, and for each such value we think $y \approx \bar{y}_i \pm s_i$ (where we have defined $s_i \equiv s_{yi}$). For the i^{th} set we then know that $y = A\bar{x}_i + B$. The error between our measurement $\bar{y}_i$ and the expected y is thus $\varepsilon_i = \bar{y}_i - A\bar{x}_i - B$. If our guess of *A* and *B* is wrong, this error is significant compared with the expected error s_i. A measure of the overall "goodness-of-guess" is the so-called chi-squared statistic

$$\chi^2 = \sum_i \left(\frac{\varepsilon_i}{s_i} \right)^2 .$$

[6] If we can't make this assumption, the problem becomes much more difficult.

For a good fit, we expect $\chi^2 \approx M$. To find the values of A and B for which χ^2 is a minimum[7], we set

$$\frac{\partial \chi^2}{\partial A} = 0 \quad \text{and} \quad \frac{\partial \chi^2}{\partial B} = 0 .$$

Some algebra provides the solution

$$A = \frac{1}{\Delta}\left[\sum \frac{1}{s_i^2} \sum \frac{\bar{x}_i \bar{y}_i}{s_i^2} - \sum \frac{\bar{x}_i}{s_i^2} \sum \frac{\bar{y}_i}{s_i^2} \right]$$

$$B = \frac{1}{\Delta}\left[\sum \frac{\bar{x}_i^2}{s_i^2} \sum \frac{\bar{y}_i}{s_i^2} - \sum \frac{\bar{x}_i}{s_i^2} \sum \frac{\bar{y}_i \bar{x}_i}{s_i^2} \right]$$

$$\text{where } \Delta = \left[\sum \frac{1}{s_i^2} \sum \frac{\bar{x}_i^2}{s_i^2} - \left(\sum \frac{\bar{x}_i}{s_i^2} \right)^2 \right].$$

From our earlier considerations, we can find the estimated errors in these derived quantities propagated via their dependence on the measurements $\{\bar{y}_i\}$. For example

$$\sigma_A^2 = \sum_i \sigma_i^2 \left[\frac{\partial A}{\partial \bar{y}_i} \right]^2 \approx \sum_i s_i^2 \left[\frac{\partial A}{\partial \bar{y}_i} \right]^2 .$$

The results are

$$\sigma_A^2 \approx \frac{1}{\Delta} \sum_i \frac{1}{s_i^2},$$

$$\sigma_B^2 \approx \frac{1}{\Delta} \sum_i \frac{x_i^2}{s_i^2} .$$

[7] Often referred to as the method of least-squares.

How good is the fit between our data and a line? Conventionally, one uses the reduced chi-squared

$$\chi_r^2 = \frac{\chi^2}{M-2}.$$

A good fit should give a reduced chi-squared near one[8].

Putting it all together: An example

Let's analyze some data from a decay experiment. After some preliminary measurements, we decide that the decay half-life $t_{1/2}$ of our specimen is on the order of an hour. We intend to measure $t_{1/2}$ by counting the number of decays counted in some time interval δt at various times t_i: $n_i = n(t_i)$. Since the sample's activity is low, we want to count as long as possible for each sample, but we don't want to introduce too much error by the fact that the count rate actually changes over the sample time.

The activity is given by

$$A(t) = A(0)e^{-t/\tau},$$

where the time constant $\tau = t_{1/2}/\ln(2)$. From this we find that the relative change in activity (and thus counts) is

$$\frac{\delta A}{A} \approx \frac{1}{A}\frac{dA}{dt}\delta t = -\frac{\delta t}{\tau}.$$

If we arbitrarily decide that we don't want the activity to change by more than 2%, this gives us a sample time of

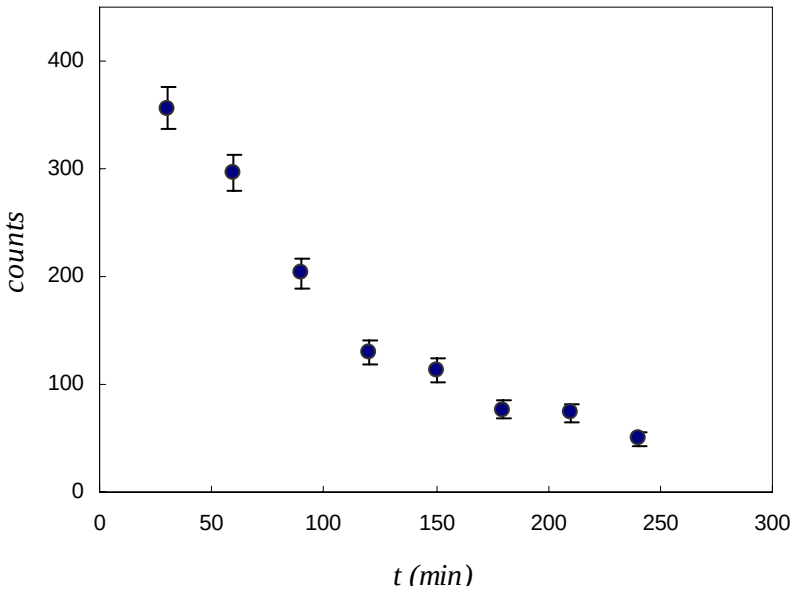
[8] How near? One can argue about this using the *chi-squared distribution*.

$$\delta t \approx 0.02 \times 60\,\text{min} \times \ln(2) = 0.8\,\text{min}$$

We decide to take counts for 1 minute every half hour. We note that we should keep this source of error in mind, and proceed with the experiment. Our results are the eight measurements:

t_i (min)	n_i	$s_i = \sqrt{n_i}$
30	356	18.9
60	296	17.2
90	203	14.2
120	129	11.4
150	113	10.6
180	77	8.8
210	74	8.6
240	49	7.0

A graph of the raw data shows the expected exponential decay:



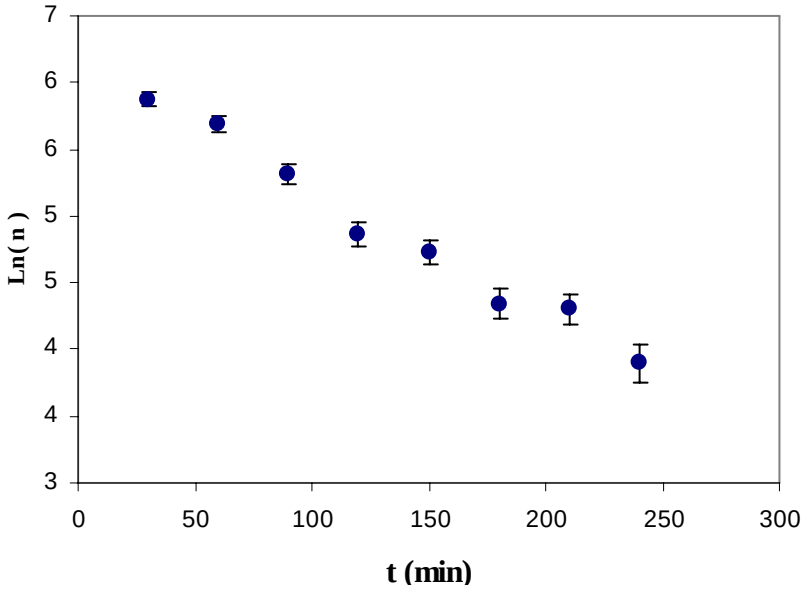
Of course, the scheme is to measure $t_{1/2}$. To do so we recognize that the number of counts expected in δt is proportional to the activity. Taking the logarithm gives

$$\ln[n(t)] = \ln[n(0)] - \frac{1}{\tau} t \ .$$

If we fit a straight line to ln(*n*) the slope will gives us the time constant and so the half-life. In preparing our data for this step, we find from our error-propagation table that the absolute error in ln(*n*) is the relative error of the data ($1/\sqrt{n}$). We have

t_i (min)	$\ln[n_i]$	$s_{\ln(n)} = 1/\sqrt{n_i}$
30	5.8757	0.0530
60	5.6912	0.0581
90	5.3121	0.0702
120	4.8609	0.0880
150	4.7307	0.0939
180	4.3415	0.1141
210	4.3013	0.1164
240	3.8980	0.1424

Which graphed looks like



Note that that the error in the large-time (low-count) data is much larger than for the low-time points. These points will not be weighted as heavily in our least-squares fit: Recall that the χ^2 which we minimize weights each point by the reciprocal of that point's variance. Points with small error are weighted more heavily than those with large error. The following table gives an idea of the various data point's relative importance to the fit:

Point t_i (min)	Weight $\frac{1}{s_i^2}$	Relative weight $\frac{1/s_i^2}{\sum_j s_j^2}$
30	356.3	0.2745
60	296.2	0.2283
90	202.8	0.1563
120	129.1	0.0995
150	113.4	0.0874
180	76.8	0.0592
210	73.8	0.0569
240	49.3	0.0380

To proceed with the linear regression we set $y_i = \ln(n_i)$, $x_i = t_i$, and $s_i = s_{\ln(ni)}$ and use the results of the previous section.

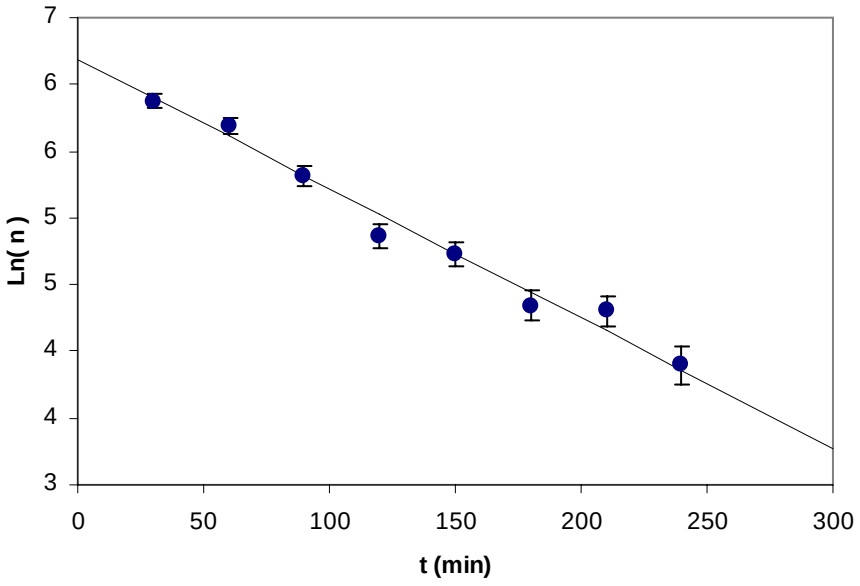
Some intermediate results are:

$\sum 1/s_i^2$	1.298×10^3
$\sum x_i / s_i^2$	1.204×10^5
$\sum x_i^2 / s_i^2$	1.602×10^7
$\sum x_i y_i / s_i^2$	5.895×10^5
$\sum y_i / s_i^2$	6.864×10^3
$\sum y_i^2 / s_i^2$	3.677×10^4
Δ	6.304×10^9

Putting these together we have the results

Slope: $$A = \frac{\ln(2)}{t_{1/2}} = (-9.707 \pm .454) \times 10^{-3}$$

Intercept: $$B = \ln[n(0)] = 6.189 \pm .050$$



Since the reduced chi-squared $\chi_r^2 = 1.38$ is reasonably close to one, we're comfortable with the linear model for our data.

We can finally figure our sample's half life

$$t_{1/2} = 71.4 \pm 3.3 \text{ min}.$$

Note the *relative* error in $t_{1/2}$ is the same as that for *A* (error propagation again). This error is about 5%, which is roughly the error of our most precise measurement ($n_1 = 356$ counts). This makes sense; generally you wouldn't expect the precision of a derived result to be much better than that of your data. For further experiment, we should also bear in mind that the precision to which n_1 is measured is coming close to the 2% error we estimated in *n* due to its variation over the sample time δt.