Data Analysis:

1. Introduction.

Scientists need very precise and accurate measurements for many reasons. We may need to test (distinguish) rival theories which make almost the same prediction (e.g for the track of an asteroid headed for earth, or the effect of relativistic time-dilation on a pilot's clock in a high speed aircraft) or to figure out what is the smallest number of ballots which need to be re-counted in order to separate tied candidates in an election. Perhaps the most common use of the very powerful methods you are about to learn, is to work out how long we need to collect data (e.g. particle counts in a scattering experiment, or the exposure time for an image recorded on mars) in order to discern a pattern against a noisy background.

The best book which explains this material simply and well is the paperback by P. Bevington and D. Robinson, "Error reduction and error analysis in the physical sciences" . McGraw-Hill. It includes a disk with programs, and clear explanations of how they work. This book is strongly recommended for all students if you will be working in science. Get a personal copy.

We can get a more accurate measurement of the length of a table by taking several measurements and taking the average μ. Each measurement will be slightly different due to random fluctuations (e.g in the position of our eye), but the more measurements we take the more precise the average becomes. The spread of values σ (e.g. the largest minus the smallest) is proportional to **the error in one measurement**. But **the error in the average of many is equal to the spread divided by the square root of the number of measurements N.**

There are two kinds of errors. If the ruler we are using is the wrong length, the mean will converge to the wrong result. The difference between this wrong result and the (unknowable !) true length is called a **systematic** error. Systematic errors are usually fixed errors of which we are unaware, and they affect the **accuracy** of the result. The second kind of error is called a random error, which effects the **precision** of the result. This error is due to random fluctuations, such as the slightly different results we get each time we read the ruler, or the variation in the number of counts per second from a Geiger counter. This error can be reduced by averaging over more measurements. Another example concerns targeting of bombs - if all the bombs land very close together in the wrong place, we have high precision and low accuracy.

So remember this: More measurements takes the mean closer to the true value (if there are no systematic errors) and reduces the random error inversely as sqrt(N).

If you have trouble with this material you could ignore initially all the discussion of histograms and distributions. Then just think of a standard deviation σ (SD), which measures the width of a distribution, as the difference between the largest and the smallest measurement, divided by 2, which is approximately correct.

2. Histograms and distributions.

We can make a histogram consisting of all our measurements of the length of the table. The height of each column in the histogram is the number of times we measured a length which fell within a certain small range (the width of the column). Take this column width to be one millimeter, for example (for very rough measurements). The horizontal axis is then marked in units of length in one millimeter increments. If the table is about 100 cm long, the first vertical axis value (column height) could be the number of length measurements which fell between 99 and 99.1 cm, the next between 99.1 and 99.2, and so on. For subtle and profound reasons (explained in Bevington), it happens that, if the measurements are subject to random errors (as they are in this case), the resulting histogram will consist of columns whose heights fall on a curve, known as a Poisson distribution, which usually also happens to be a Bell-shaped Gaussian peak, or Normal distribution. As N increases and the width of the columns decreases, the data will fall more and more closely onto this Gaussian curve. (The gaussian function is y = exp(- a x^2)). You can now see that the width of this curve will give an indication of the spread of values, and hence is proportional to the error we are looking for. The smooth function which fits our histogram for large N is called a probability distribution function. Its vertical axis has units “number of measurements per unit length”. It is important to understand that as we make more measurements, the histogram does not become narrower, instead it becomes smoother. But the error in the mean, σ/sqrt(N) , decreases.

We often hear that “the probability of rain tomorrow is 40%”. This means that on 40% of previous occasions when conditions had been similar, it rained. Similarly with dice-throwing - we expect the probability (or likelihood) of rolling a three to be 1/6, because, in a large number of throws (“trials”) the dice will come up about an equal number of times on each of its six faces. **So probability just**

means how frequently something happens if all the trials occur under similar conditions. Probability is often described as the "fraction of successful outcomes from similar trials", success meaning getting the result you want (raining, or a three).

Let's say we make a hundred measurements of the length of our table. If the height of the first column in the histogram is 10 (meaning that we measured a length of between 99 and 99.1 cm ten times), we can say that "the probability of making a measurement of between 99 and 99.1 cm is 10/100 =0.01 ". This is just the fraction of "successful" outcomes - the trials in this case are measurements, and success means getting an answer between 99.0 and 99.1. Probability is always a number between 0 and 1. P = 1 means guaranteed success. To simplify things, we usually "normalize" a distribution, making the total area under the curve unity. Then areas can be interpreted directly as probabilities - for example half the area under the curve will measure 0.5, which is the probability of making a measurement greater than the average (the maximum value).

3. Modelling.

A model is like a theory, but, unlike a theory, it may have no physical basis - it is just an inspired guess for a formula which fits our data. For example, the Wright brothers "modelled" (i.e. guessed) a formula to fit the dependence of lift force on wind speed, wing area, and angle of attack when refining wing design in their wind-tunnel. Few of us will ever develop a new "law" of physics, but most of us will need to develop and/or apply "models" for more specific behavior such as the resistance of a MOSFET transistor vs. its gate voltage, or the yield strength of a superhard alloy vs. annnealing treatment. The most important recent use of models has been to replace environmentally damaging testing of nuclear weapons with computer simulation, based on models. Another is the use of "derivatives" on the stock exchange to predict trends.

So a model is a formula, which predicts something new in terms of some measured quantities (eg the wing lift in terms of area, speed and angle). But there will be random errors in the measurements of area, speed and angle. An important result we will derive below is how to calculate the error that these errors in "experimental parameters" produce in the lift force. This is called the study of the propagation of errors.

Students often consider "error" to be the difference between measured vs. "true value", assuming that the they will always know, or be given, the true value by their boss. This thinking belies the entire philosophy of this course, and of science in general! In real life, the "true value" is generally not known, but we may have a model formula we wish to test. If it fits, it can then be used to make predictions for a much wider range of conditions. So we need to know if our model falls within the error bars around our data.

The error bar is the only way to assess the validity of your assumed "model" for your particular situation.

It is only by having well defined error bars that we can answer (with high probability) such questions as: Will hurricane Harry hit the coast? How many trucks can that freeway overpass safely carry? Will that turbine blade crack? Were you speeding? How long will that disk drive last? How long will the sun last? On a simpler level, we need good error bars to assess and improve an experiment. Imagine the delight of your manager if you could state confidently and correctly during your first week at Motorola, that they could replace a set of $10,000 timers with $10 Casio watches! This might be possible by careful design of the project. In the following sections we will see how to do exactly that.

4 Uncertainty in single variables (read this several times carefully).

We begin by considering a simple experiment involving measurement of a single variable. For example, we'll measure the acceleration of gravity by timing the fall of a ball. The "answer" is known, namely g=981 cm/sec^2. This is stated to 3 significant figures (see Appendix on significant figures). It is possible to measure "g" much more accurately. Indeed, geologists search for subterranian oil or water by measuring subtle variations in local g, at the level of 5 significant figures. We will use a known model whereby the time to free-fall a distance h is given by

$$h = 0.5gt^2 \qquad \text{eq. 1.4.1}$$

This can also be written as

$$g(h,t) = 2h/t^2 \qquad \text{eq. 1.4.2}$$

which shows explicitly that g can be determined from measurements of h and t. For now we consider h to have no errors, so the task is to measure t and its errors, then find g and its errors.

Three scientists (Larry, Moe, Curly) take 12 trials each with results as shown in table 1. They use a digital timing meter that shows time as xxx.xxx seconds, and we assume it is accurate to the least significant digit, ie $\pm$ 0.001 sec. They start and stop the timer manually. Note the format of this table carefully. The original data in the lab notebook should be organized like this. The table includes a descriptive title, units with ±, and values recorded with the appropriate number of digits. Larry realized half way through that his random errors were much larger than the 3 decimals accuracy of the meter, so he shortened his data to only 2 decimals. Moe and Curly adopted this procedure also.

These data could now be used to produce a bell-shaped histogram, showing the mean value, and indicating the spread of values by its width. The width of the columns in the histogram should be small enough so that there are several trials within each column. In this case we really need more trials to make a good histogram, but an interval of 0.05 might do.

Trial #	Larry	Moe	Curly
1	1.639	1.42	1.51
2	1.451	1.38	1.54
3	1.413	1.46	1.50
4	1.549	1.63	1.27
5	1.468	1.40	1.34
6	1.62	1.58	1.53
7	1.56	1.44	1.54
8	1.55	1.55	1.39
9	1.41	1.48	1.36
10	1.53	1.59	1.26
11	1.45	1.61	1.36
12	1.59	1.47	1.52
sum	18.241	18.005	17.123
count	12	12	12
average	1.52	1.50	1.43
SD	0.08	0.09	0.11
SDM	0.024	0.026	0.032
frac (%)	1.6	1.7	2.3

Table 1. Raw data for ball drop experiment. All times in (sec)

At the bottom, we have tabulated the mean value and "errors" for each person. This can/should be done with a calculator or computer. All students are strongly urged to learn to use a spreadsheet, such as EXCEL. Mean value and "errors" are defined as follows. Suppose we measure a quantity x a number of times, N. The mean value μ is defined as

$$\mu = \frac{x_1 + x_2 + ... + x_N}{N} = \left(\sum_1 x_1 \right) / N \qquad \text{eq.1.4.3}$$

This is readily done in the spreadsheet using the SUM() and COUNT() functions. Note that COUNT() ignores non-numeric entrees, as it should. You can also use the AVERAGE() function directly, giving AVERAGE() =SUM()/COUNT().

Standard deviation σ (SD) is a very important quantity, which measures the spread of the values in the data. It is similar to the difference between the largest and smallest measurement we mentioned above. It is proportional to the width of a histogram based on table 1, and to the gaussian, bell-shaped normal distribution function for our random errors which fits it. (In fact the "full width at half-maximum height" of the gausssian is 2.354 σ). Many hand calculators have a SD function. It is defined as

$$\sigma = \sqrt{\left[\frac{1}{(N-1)}\sum_{1}(x_1-\mu)^2\right]} \quad \text{eq.1.4.4}$$

This function too is available in EXCEL as STDEV(). It is essentially the square root of the average of the individual deviations squared (root-mean-square or RMS). The factor (N-1) rather than N, as used for average, can be thought of as the number of differences in the set of values.

An explicit calculation of STDEV is shown for Moe's data in table 2. Note that the individual deviation dev = $x_i - \mu$, averages to zero by definition, while $(dev)^2$ does not.

It is not expected that you produce such a table for your data sets - you may simply use the STDEV() function directly.

Trial #	Value	Dev	Dev^2
1	1.42	-0.079	0.0063
2	1.38	-0.122	0.0149
3	1.46	-0.039	0.0015
4	1.63	0.127	0.0161
5	1.40	-0.098	0.0096
6	1.58	0.080	0.0063
7	1.44	-0.061	0.0038
8	1.55	0.048	0.0023
9	1.48	-0.025	0.0006
10	1.59	0.085	0.0072
11	1.61	0.114	0.0130
12	1.47	-0.028	0.0008
sum	18.00	0.000	0.082
count	12		

average	1.500		
sigma	0.087		

Table 2. Explicit tabulation of standard deviation for Moe.
All times in (sec).

The meaning of σ is well defined in probability theory, which deals only with the smooth continuous functions, not real data which they approximate. It can be shown that:

> ***For a Gaussian distribution, 68% (about 2/3) of events lie within ± 1 σ of the mean, and 95% lie within ± 2σ of the mean.***

In our example, we see that for Moe, 9 out of 12 (75%) of the values lie within $\mu \pm 1\sigma$.

The standard deviation (SD) gives the spread in the values of the data, but it does not give the error in the mean. (This is a subtle point beyond the scope of our course - for details, see Bevington). You can probably see that we can locate the center of the Bell-shaped curve much more accurately (if N is large) than by using its width or SD, so the error in the mean will be much less than σ (which gives the error in just one measurement, not from the average of a lot). It is shown in textbooks that the error in the mean is given by the standard deviation of the mean σ_m defined as

$$\sigma_m = \sigma/\sqrt{N} \qquad \text{eq. 1.4.5}$$

Note that the precision of the mean increases with N, which is reasonable, whereas σ remains constant. (The spread of values - eg difference between largest and smallest , or the SD, does not change as you take more measurements - the SD σ is fixed by the square root of the mean as discussed in section 1.7). This has the curious effect that your final result (mean value) improves the more often you miss! Indeed, truly random errors can be essentially eliminated by taking a very large number of measurements.

In our example, we have for Larry, $\sigma_m = 0.080/\sqrt{12} = 0.024$ sec. If this were the end result of the experiment, it would be properly stated as:

"The drop time was found to be t = 1.520 ± 0.024 sec (1.6%)."

Note the formatting: the sentence stands out clearly, is grammatically correct and includes value, error, units and the appropriate number of significant figures, and a % ratio of error/value. (See appendix on significant figures). The error is determined by the value of σ_m, which is only known after the analysis. This result means that if Larry repeats the 12 trials and calculates a new mean, it will lie within $\pm 1\sigma_m$ of his previous result (with 68% probability). This is essentially what happens when Moe repeats the run. Indeed, his mean value lies very close to Larry's.

Results for the 3 partners are summarized in table 3. At this point, since σ_m is known, you should be careful to use the appropriate number of significant figures. You can be less rigorous about formatting during intermediate calculations, especially if done by computer. Note that computers carry out calculations with full accuracy regardless of the display format. Technical note: If you paste EXCEL tables into a report, do this as formatted text, so you can do further edits in WORD, which allows full control of formatting, such as symbols, subscripts and number of digits.

	value	(%)	Comment
t1	1.52 ± 0.02 sec	1.3%	Larry, 12 trials
t2	1.50 ± 0.03 sec	2.0%	Moe, 12 trials
t3	1.43 ± 0.03 sec	2.1%	Curly, 12 trials

Table 3. Summary of mean drop times.

Based on these data, we can say that Larry and Moe's results agree, but are different from Curly's. More specifically, Larry and Moe obtain the same mean value (within error bars at the 68% confidence level), while Curly's mean value is distinctly different. This is probably the situation which occurred in the 2000 Presidential election in Florida - with Dole and Bush in the situation of Larry and Moe. The difference between the mean count for each of them (if the ballots had been counted many times) was less than the standard deviation, so that the histograms or distributions overlapped.

These results suggest systematic errors in the experiment, although it is possible that the apparent discrepancy is simply a statistical fluctuation. This is not impossible, since all 3 do overlap within $\pm 2\sigma_m$. This could be checked by making enough additional trials until σ_m is small enough to clearly distinguish each person's result. Assuming there is indeed a systematic error (ie they disagree within errors), it is tempting to blame Curly. Yet, this is not obvious, unless we have a model or "true value" to judge results, or can independently "guess" at likely systematic effects in the procedure. This brings up the next step of analysis, which is to determine "g" from the drop times. This requires understanding the propagation of errors.