Physics 1111 - Term C 2015
Interpreting Measurements

The overall objective of the course's lab series is to show what you do when you are doing real experimental physics. I am doing this with experiments in which you (should) understand more or less everything that is going on. The experiments deal with the problems that you are solving in the homework and hearing about in lecture, so you do have the theoretical background to understand what is happening. However, experimental physics is not the same as theory, not at all. You are going to see something of what experimental work is like.

The purpose of these notes is to discuss what you do with your data after you have acquired them. (Remember: *data* is the plural. The singular is *datum.*)

Suppose you make repeated measurements of the same physical quantity, e.g., the period T of a pendulum of fixed length with a given mass m and maximum amplitude A of the swing. If you did three measurements, the measured values could be symbolized T_1, T_2, and T_3. In general, if you do N measurements, the values of the N measurements could be symbolized $T_1, T_2, T_3, \ldots, T_N$. The structure $, \ldots,$ represents all the values between T_3 and T_N. I did not write down those intervening values of T *explicitly*, meaning I did not write them all down one at a time so you actually see them individually. They are still there.

OK, you have made several measurements of T. If they are all exactly the same, you know what T is. However, if you lean hard on the measurements (this may not work with computerized data acquisition), you will usually find that some of the measurements are slightly larger, and some of them are slightly smaller. If your measurement method is less accurate, it may be painfully obvious that your measurements do not agree with each other. What do you do if the T values are not all the same? There is a standard way to advance, and a warning at the end.

Step One: For each set of measurements, take the average. The average in question is the simple average, so if you made three measurements the average would be

$$\overline{T} = (T_1 + T_2 + T_3)/3. \tag{1}$$

The general form of that equation is

$$\overline{T} = \frac{1}{N}\sum_{i=1}^{N} T_i \tag{2}$$

These are simple averages. These equations implicitly assume that all measurements are equally accurate. In class, we will discuss more complicated weighted averages, but for these experiments we can stop with the two above equations.

The average $\overline{T}$ is not necessarily the correct answer, but it is usually closer to being correct than are the individual measurements, the T_i.

Step Two: For each measurement, take the difference $T_i - \overline{T}$. These differences are the spread in the measurements around the averages.

We now want to use the $T_i - \overline{T}$ to give a description for how accurate each set of measurements are. A simple answer is to take a simple average of a set of $T_i - \overline{T}$. Some of you see what is coming with this idea. Go ahead, take an average. If you did it right, your answer is exactly *zero.* That's fine. If you are trying to figure out how to advance, experimentation is a good answer. This experiment did not work.

A good answer is to notice that the $T_i - \overline{T}$ could average to zero because some of the numbers were positive and some of the $T_i - \overline{T}$ were negative. If you did something to the $T_i - \overline{T}$ to make them all positive, there could be no cancellation. One path would be to take the absolute magnitudes $\mid T_i - \overline{T} \mid$. It turns out – this is why we have courses on statistics – that a better answer is to square the $T_i - \overline{T}$, and take the average of the squares, leading to

$$\overline{\Delta^2} = \frac{1}{N}\sum_{i=1}^{N}(T_i - \overline{T})^2 \tag{3}$$

where Δ is shorthand for $T_i - \overline{T}$.

What are the units of the spread in T? A reasonable choice is that the spread in T should have the same units as does T itself. $\overline{\Delta^2}$ has the wrong units, but you can fix this by taking the square root. The Final outcome is the **Root Mean Square Error** (RMSE) ΔT, namely

$$\Delta T = \left[\frac{1}{N}\sum_{i=1}^{N}(T_i - \overline{T})^2\right]^{1/2} \tag{4}$$

A few minor math warnings: In that equation, you divide by N *before* you take the square root. $[a^2 + b^2]^{1/2}$ does not equal $a + b$.

Step Three: For each set of measurements, calculate ΔT. This will take a few moments for each set of measurements, which is why you have pocket calculators, Mathematica uploaded on your computer, or the like. You now see how accurate your measurement methods were. Some methods were probably more accurate than others. For this discussion, the "best" method is the method that has the smallest ΔT.

The above math steps when applied to all your methods for measuring the period of a pendulum determine for you which of your methods is the *best*, in the sense that you have found which method gives the most reproducible results.

Of course, I could have told you which method to use, but if I did that, you would miss out on an important part of experimental physics. You want to get the best possible measurements, so you have to refine your experimental technique. Refining experimental technique is tedious. Handing out the best method deprives you of an understanding of how to do experiments. You can do theoretical analyses of ways to reduce the scatter in your measurements. You can do experimental tests to see what actually works. As an aside, gazing into the void does not work. Large numbers of people, asked the best way to do this measurement with good electronic sensors replacing human reflexes, will say 'it's obvious' and then propose the wrong answer.

Extracting Fitting Parameters

We now turn to the next part of analyzing your measurements, namely converting them to physical parameters in which you are interested.

Dimensional analysis (homework problem) leads to the supposition that the period of of a pendulum is related to the force of gravity g, the mass m of the pendulum bob, and the length ℓ of the pendulum string via

$$T = Ag^q m^r \ell^s \tag{5}$$

where A, q, r, and s are numerical constants. An objective of this experiment is to determine r and s experimentally. (Determining q would be a challenge. How could you vary the force of gravity?)

How do we determine r or s from a set of measurements?

First, we will look at a math problem. Then we will show that the problems of determining r or s have the same solution as the solution to the problem we first discuss.

The math problem: We have an experiment in which we can reach in and set the value of some variable x. After each time we set x, we make one or more measurements of a variable y. Example: x is the length of the string, and y is the period of the pendulum. You choose different values for x, and for each x you make one or several measurements of y.

At the end, for a series of values of x, which you call $x_1, x_2, \ldots$, you have a matching series of values of y, namely y_1, y_2, $\ldots$. The next correct step is to make a scatter plot of the y_i against the x_i, to see what the dependence of y on x looks like. Under most conditions, you make the plot first, and then try to analyse farther.

In this hypothetical problem, you find from your scatter plot that y depends at least approximately linearly on x, so that

$$y = mx + b \tag{6}$$

looks like it should describe your data. Here m and b are the slope and the intercept. They are the parameters you are trying to find. How do you find the "best" values of m and b?

The answer is a mathematical approach known as *linear least squares*. The method is called "linear" because y is linearly proportional to the parameters m and b. If y depended non-linearly on m or b, you would need to use a non-linear least squares method such as the simplex functional fit. We will not cover method that in this course.

How do we advance? For each value x_i of x, and any particular value for m and b, you can calculate a theoretical value Y_i for y, namely

$$Y_i = mx_i + b. \tag{7}$$

You now try to find the values of m and b that give the "best" agreement between your calculated curve, equation 6, and your measurements.

What is "best"? A reasonable measure of the agreement between the fitting equation 6 and the measurements is the mean-square difference D between the experimental points and the fitting equation, namely

$$D = \sum_{i=1}^{N} (y_i - (mx_i + b))^2. \tag{8}$$

The "best" m and b are the values of m and b that make D as small as possible.

How do we make D as small as possible? We find the minima, which we can do using the derivative tests, namely at most minima

$$\frac{dD}{dm} = 0 \tag{9}$$

and

$$\frac{dD}{db} = 0. \tag{10}$$

The first derivative gives us

$$0 = \frac{dD}{dm} = 2\sum_{i=1}^{N}(y_i - mx_i - b)(-x_i) \tag{11}$$

while the other derivative yields

$$0 = \frac{dD}{db} = -2\sum_{i=1}^{N}(y_i - mx_i - b). \tag{12}$$

These two equations can be rewritten as

$$\sum_{i=1}^{N} x_i y_i - m\sum_{i=1}^{N}(x_i)^2 - b\sum_{i=1}^{N}(x_i) \tag{13}$$

and

$$\sum_{i=1}^{N} y_i - m\sum_{i=1}^{N} x_i - Nb. \tag{14}$$

Each of the sums (including $\sum_{i=1}^{N} 1 = N$) is just a number, so you have two equations in two unknowns that you can solve for m and b.

OK, we have now seen how to extract best values for the parameters m and b. However, the equation we care about, equation 5, is not linear in the parameters that we want to determine. In the case here, you can modify the equation of interest to make it into a linear equation, namely if you take the logarithm of equation 5 you get

$$\log(T) = \log(A) + q\log(g) + r\log(m) + s\log(\ell). \tag{15}$$

This equation shows that $\log(T)$ is given by a linear equation in $\log(m)$, so a plot of $\log(t)$ against $\log(m)$ will give a more-or-less straight line whose slope is the exponent r, and similarly for $\log(t)$ against $\log(\ell)$

Outliers: Sometimes you will do a series of measurements that all more or less agree with each other, except one or two are way off. The points that are way off are called *outliers.* In real science, it is understood that every so often an experiment does not work right, and therefore you should not include its outcome in further analysis. For example, if you have the time required for five pendulum swings, you might get as answers 5.1, 5, 4.9, 5, and 6.0. Further repeats of the experiment all give numbers very close to 5. What do you do with the 6.0? You might suspect, for example, that you accidentally let the pendulum swing 6 rather than 5 times, but the experiment is done and gone, one with the snows of yesteryear, so you cannot prove that the pendulum was allowed to swing an extra time. The correct answer is to report the datum, put it on your graph, and explicitly say that you are not including it in your analysis.

There is a slippery slope in outlier analysis. If you really think you know what the answer is supposed to be, you can be tempted to reject as outliers the *good* measurements and keep the *bad* measurements. I have seen this done in the real literature. You can also keep points that are really flaky, that lead you to wrong conclusions, for example you keep them because they are your only measurements of pendulum period using a really long string. Feynman won his Nobel prize in significant part by saying 'My V-A theory is correct. The data are wrong. The experiment that disproved V-A relies entirely on the two points where the particle accelerator was seriously overloaded to get to way high energies, and those points are no good.'