Experiment IV: Computer Analysis of Experiments II and III

Goals

- Learn the basic concepts in error analysis
- Use a computer spreadsheet program, Excel, to analyze data from Experiments II and III

Introduction and Background

Experimental Errors**:** Experimental error or uncertainty is inherent in any experimental result at some level, however small. It is set by a combination of the design of the experiment, the quality of the apparatus, and the care and skill of the experimenter. Generally, it is possible to separate the sources of experimental errors into two categories: **random** and **systematic**.

Random Errors**:** The existence of random errors in a measurement can be inferred if repetition of the measurement does not give the same result each time. By definition, random errors are those that tend to average out upon repetition of the measurement. Hence for random errors, the more repetitions of a measurement, the less uncertainty there is in the resulting average. From the mathematics of statistics, it can be shown that the uncertainty due to random error in the average of N measurements decreases as $\sqrt{N}$ when N is large. For example, 400 measurements should give an average with half the uncertainty due to random errors as compared to 100 measurements.

The first step in treated the random error in a large number N of repeated measurements is to calculate the average:

$$\bar{y} = \frac{1}{N}(y_1 + y_2 + y_3 + ... + y_N) \tag{4-1}$$

which is the desired result and sometimes called the mean in statistics.

To establish the uncertainty in the average, $\bar{y}$, the usual procedure is to first calculate what is called the standard deviation, σ, of the measurements. If no one measurement is more accurate than any other, then σ is defined by

$$\sigma = \sqrt{\frac{1}{N-1}[(\bar{y}-y_1)^2 + (\bar{y}-y_2)^2 + ... + (\bar{y}-y_N)^2} \tag{4-2}$$

The physical significance is that any one additional y measurement has about a 2 in 3 chance of falling between $\pm\sigma$ of the average $\bar{y}$. Statistical analysis further shows that the average $\bar{y}$ has a 2 in 3 chance of falling within $\pm\sigma$ of the true value. Thus it is common to append $\pm\sigma$ as a measure of the uncertainty due to the randomness in the measurements. The standard deviation for the average (mean) is

$$\sigma_m = \frac{\sigma}{\sqrt{N}} \tag{4-3}$$

and the final result is often expressed as

$$\bar{y} \pm \frac{\sigma}{\sqrt{N}} \tag{4-4}$$

Systematic Errors**:** A systematic error is one which tends to repeat and thus create a shift in the average from the true value. A systematic error cannot be averaged out by repeated measurements. Systematic errors may result from the experimenter, the apparatus, or the poor design of the experiment. Because they are not revealed by repeated measurements, care must be taken to investigate and account for all possible sources of errors. This can be very difficult to do. Such errors sometimes remain unknown until other experimenters with other apparatus obtain convincing evidence that a previous result is off from the true value by more than the originally reported uncertainty.

Treatment of Random Errors**:** If the quantity we want to determine is the quantity we measure, the treatment of random errors is straightforward: We simply repeat the same measurement many times and calculate the average and the standard deviation. However, in most cases, it is difficult to directly measure the quantity we want to determine. Rather, we use what we measure to calculate what we want to determine. For example, in the "Determination of *g*" lab, what we measured was velocity and we used that to infer the acceleration *a*. In this case there are two ways to treat the random errors (we use the "Determination of *g*" lab as an example):

i) Calculate the quantity we want to determine (*a*) from every pair of data we measure (*v*), thus creating a whole set of values for *a* which we would average;
ii) Graphical method: the data is used to create a linear plot (*v* versus *t*) and the slope of the best-fit straight line gives the average of the quantity we want to determine (*a*).

The first method is cumbersome and requires a lot of calculation. The second method is the only one we will use in this class. In this method, the random errors in the data points result in an uncertainty in the slope (Did you agonize over how to draw the best-fit line?). There is always a certain degree of arbitrariness associated with a hand drawn best-fit line, while a computer can determine quantitatively what the best-fit line should be. In computing the best-fit straight line, the spreadsheet program (Excel) chooses a linear function of the form: $y = mx + c$, where *m* is the slope and *c* the intercept on the *y*-axis. *m* and *c* are the so-called adjustable fitting parameters. Excel would choose some starting values for *m* and *c* and gradually adjust them until the following sum is minimum:

$$[y_1(\text{measured}) - y_1(\text{best-fit})]^2 + [y_2(\text{measured}) - y_2(\text{best-fit})]^2 + \ldots + [y_N(\text{measured}) - y_N(\text{best-fit})]^2.$$

This sum is an indicator of the deviation between the actual data and the best-fit values, thus must be minimized. The method is called, for evident reasons, "the method of least squares fitting". Since we will only deal with fitting of linear functions (straight lines), we will use the "**Linear Regression**" function in Excel, which does least squares fitting for linear functions. In a linear regression, along with best-fit values of *m* and *c*, the associated uncertainties are also given.

Treatment of Systematic Errors**:** Since systematic errors cannot be minimized by repeated measurements, you will be expected to discuss for each experiment possible sources of systematic errors. In general, this means giving plausible reasons as to why your results might differ from the expected results by more than the uncertainty due to random errors.

Computer Analysis of Experiment II

Experimental Procedure

1. Start the program Microsoft Excel and open the template named "Gravity". Make sure you choose the option "Enable Macros".
   Type the required data into the shaded cells, then click the button "Analyze". The template will calculate the average velocity for each interval and perform the linear regression. You will be asked for the "independent" and "dependent" variables for the regression in a dialog box.

   Click the "Print" button to obtain printouts for the results.

   On the screen or the printout you will see a column called "best fit *v*". This is the value of your measured quantity need to be if it were to fall exactly on the best-fit straight line determined by the computer. Hence the difference between the measured *v* and the best-fit *v* is how far the data point is above or below the line. This is illustrated in Figure 4-1. The column "best fit *v* – average *v*" represents the difference between the best-fit value and your measured value. Looking at this column is a good way to see quickly whether you have a bad data point or an input error: If you have a bad point, this difference will be much larger than the adjacent values.

2. Generate a summary of the results of your analysis by hand from two weeks before and the computer analysis you just performed, in the following format:

	Run 1	Run 2	Run 3
g by hand			
g by computer			
Uncertainty Δg			

Discussions and Questions

1. By how much does each of your values for g (from computer analysis) differ from 980 cm/s^2?
   How does each difference compare with the corresponding uncertainty Δg due to random errors in the data?
   Does it appear that there is a systematic error in your experiment?

2. If you did the lab carefully and the equipment worked properly, you may see the effect of a small systematic error inherent in the design of this experiment.
   When you were measuring the friction force, you tried to get the cart move at a constant velocity. However, when you were measuring the acceleration, the cart was accelerating. In order to get the cart wheels to spin faster and faster, the table has to exert a static friction force on the wheel rim in a direction opposite to the cart motion, in addition to the kinetic friction you have measured and accounted for. This friction force is not present when the

cart was moving with constant velocity, so it has not been accounted for. A calculation of this friction force based on the experimental setup shows that *it should cause the measured g value to come out about 2% lower.*

Now what result would you obtain if you take into account of this systematic error and increase each of your *g* values by 2%?
Are your data accurate enough (with low enough random error) to see this effect?
Do you still have other systematic errors present?

3. List a few other possible sources of systematic errors. (Think about some of the assumptions we have made about the equipment that may or may not be valid).

4. Compare the slopes you obtained two weeks ago with those generated by the computer analysis. Did you do a good job by hand?

This concludes the analysis for Experiment II. You may save your work on a 3.5" floppy disk provided. **Do not save your work on the hard drive!**

Computer Analysis of Experiment III

Experimental Procedure

1. With the program Microsoft Excel already started, open the template named "Centripetal". Make sure you choose the option "Enable Macros".
   Type the required data into the shaded cells, then click the button "Analyze". The template will calculate the average speed and perform the linear regression. You will be asked for the "independent" and "dependent" variables for the regression in a dialog box.
   Click the "Print" button to obtain printouts for the results.

   Do this for all three parts.

Discussions and Questions

1. In most cases, a power law in physics involves exponents that are either integers or simple fractions. Assume this is the case here, use the three values (a, b, c) from Parts A, B, and C you have obtained (from computer analysis) to infer the true values of the exponents. Do you have any ambiguity in any of these exponents (i.e., is it difficult to decide which integer or fraction a or b or c is closet to)?

2. The complete relationship between V, F, M, and R can be written as
   $$V = C \times F^{a} \times M^{b} \times R^{c} \qquad (4\text{-}5)$$
   You have determined what the exponents a, b, and c should be. Now to get an idea what the constant C should be, find the data point in each of the three analyzes that has the best agreement between the measured value and best-fit value. Write down the values of F, R, M, and V corresponding to each of these points. Put these values and the **inferred** values of a, b, and c into Equation 4-5 and calculate three values of C. Average the three values.

Does your result support a simple value for *C* within a few percent? What is this value?

3. Since the centripetal force (*F*) is usually expressed as a function of *M*, *R*, and *V* when dealing with circular motion, rewrite Equation 4-5 in this form with inferred values of *a*, *b*, *c* and *C*.

4. Describe some systematic errors that might have affected your results.

5. How well did you do in determining the slopes of your hand-drawn graphs?

Conclusions

Briefly discuss whether you have accomplished the goals listed at the beginning.