Experimental Error: Precision and Accuracy in Measurements

1/12/2022

Introduction

Experimental error, more appropriately called *uncertainty*, is really just a phrase to describe the accuracy and/or precision of an experiment. The use of the word "error" is not intended to imply that the person performing the experiment did anything incorrectly, instead it is used to reflect the fact that no matter how carefully we design, set up, and carry out an experiment, there are some fundamental limits on how well we can know the answer. In this experiment we will make several measurements and compare them, as a class, in order to gain insight into the idea of experimental error, accuracy, and precision.

Equipment

- **2-meter stick**
- **marble maze**
- **stopwatch**
- **marble**

Preliminary Work

In your group, discuss what the terms "accuracy" and "precision" mean. Try to come up with a definition for them and write them down in your notebook. Among the things to consider: Are the terms synonymous? If so, why would we need both terms? If they are not synonymous, is it possible to be precise but inaccurate? Can you think of an example? Is it possible to be accurate but imprecise? Again, give a concrete example. Do this *before* continuing on below!

A classic example for precision and accuracy is target shooting. If you were to fire ten arrows at a target, you can make two basic measurements for them. The first is how close each one is to the center, and how close each one is to all the other shots, called "grouping". If all ten shots are very close to the bull's eye, you can say you were both accurate and precise. If your grouping is small, but they are all off, say low and to the left, you could say you were precise, but inaccurate—there would be some kind of *systematic error* which you could then compensate for at some later time such as a misaligned aiming pin, a constant wind, etc. If your grouping was large, but clustered around the bull's eye relatively the same, then you could say you had good accuracy (on average!) but poor precision—in this case you would have some kind of *random error* for which you can't compensate, but you may be able to characterize.

Convince yourself and your partners that the above statements make sense by drawing pictures of the above three cases. Show another version of the last two cases but one with either more or less error. Finally, add a fourth case that is both inaccurate and imprecise with a verbal description what that means. What kind of error(s) are involved? In this case, can you easily characterize and/or compensate for them?

Adjust your definitions from above based on this new information. Note that in a lab notebook, you *never* erase or completely obliterate anything you've written before—simply cross out anything that is no longer useful, and rewrite it either in the same place, or make a note where your correction is newly written.

Procedure

In this lab we will perform three measurements, combine them as a class, and analyze them. Our analysis will include looking for systematic and random errors, discovering a method to find the "best value" for those measurements, and a method for describing how well we have represented that measurement—how accurate and precise it is. We will also make a calculation based on two of these measurements and attempt to understand uncertainty in derived quantities.

Note that there are two sections below—you can do them in any order and since there is only a single Marble Maze for the class, you can skip down to that section whenever the Maze if free to use. This way we can use our lab time more efficiently.

Spatial Measurements

Take a look at your two-meter stick and describe how it is that you can make length measurements with it—what are you *doing* when you make a measurement. What do the marks on it represent? What is the distance between the closest division marks? If you were to make a measurement with it, how precisely do you think you can know that length? To the nearest mark? Or can you estimate some fraction of the distance between those marks? Which of these two methods do you think would give you a more accurate value?

In general it is difficult (impossible?) to know the accuracy of any single measurement—try to explain why this is true through discussion with your group. What we are going to do is try to figure out the length, width, and surface area of "a standard lab table". Have each person in your group measure the length and width of your lab table. Be sure that each person is making an independent, unbiased, measurement. That means, there must be no outside influence from anyone else, including knowledge of the values acquired by anyone else.

When each person had made their two measurements, add them to the appropriate list on the board. When everyone in the room has finished adding to the list, copy down all the measurements. In your groups, discuss the variations in these measurements. Did everyone measure to the same level of precision? How do you know? Do they seem to cluster around a single value? If so, roughly estimate what that number seems to be. Does the clustering seem small or large? Do any of the numbers seem unusually large or small? If so, do you think there is some systematic reason for that? Find the person who made the odd measurement and see if you can figure out if there is some systematic error. Do the numbers appear to be randomly distributed about your estimated cluster center? What do you suppose is the reason for these differences?

Within your group, try to define a mathematical method to figure out how to find the actual value of the cluster center for these two sets of numbers. Discuss this with other groups and try to come to a consensus on your formula. For example, if you have two measurements, say 3 and 5, what is the center of that cluster? How did you get it? What about, say, 2, 3, 4, 5, and 6? What about 7, 8, 9, and 10? At some level, these calculations cluster centers represents your best possible values for the length and width of a lab table. How do you think you could get an even more accurate number?

The next step is to quantify the level of accuracy—we want to know a value for how big the scatter in the measurements is. For each measurement, write down next to it the difference between it and the cluster center from above. Be sure you do the difference in the same way for each number! That is, subtract the average from each measurement. What do you notice about these difference values? How many of them are negative? How many are positive? Based on what you did above, what do you suppose you could do to calculate a best value of these scatters? Try that and discuss what you get—do you think this value really represents the scatter well?

One thing you may notice is that this number is *very* small relative to all of the individual scatter values. If we want to get some idea of the "average" scatter what do you think we must do to all these numbers? Try to write a mathematical formula to describe the process. Discuss this with your colleagues in class and, again, try to come to some consensus.

Using these best values, calculate the area of the lab table; make a note of the formula you used. What do you suppose the uncertainty is of this value? Should it be more or less than the uncertainties in the length and width? Think about what the relative uncertainty is in the length and width—that is, what is the ratio of the errors to the values themselves? Are they the same? What do you suppose should be the relative error of your calculated value? In class, come to some consensus on what this relative error in the area calculation "should be". Now, knowing it, and the value of the area itself, calculate the error in the area.

Temporal Measurements

Sometimes in the universe there are events over which we have very little direct control, but we wish to understand them. We will model this idea with the Marble Maze—a board with nails randomly placed on it down which we will roll a marble. The primary question is: How long does it take for the marble to reach the bottom?

At some point in the lab period, each individual will use the stopwatch to time how long it takes the marble to complete the marble maze. Start the stopwatch as you release the marble at the top of the maze and stop it when the marble has left the maze. Make a note of your time on the board and when everyone in class has completed the experiment, copy the entire list into your lab notebook.

Look at these time values and note any observations you can make about them. Are they all the same? Do they appear to cluster around any particular value? Do any of the numbers seem extraordinarily large or small? Do you think these values are due to random or systematic error? Again, find the person who made that particular measurement and make your determination. Make a hypothesis about the nature of (of) the scatter in these values—do you think the reasons for it are the same or different than the reason for the scatter in the first half?

Try to determine a mathematical value for the center of the cluster of these time measurements. What do you suppose this value physically represents? That is, can we "really" answer the question at the beginning of this section? If not, what question *are* we answering with this value we just calculated?

Again we would like to characterize the accuracy of this value. In your notebook, again note the difference between each measurement and the calculated average and analyze them as you did the length and width differences above: Do they seem to cluster about some value? How many are negative and how many are positive? Try to calculate a best value for this scatter—come to some class consensus on a formula for this and calculate it.

Additional Thoughts

1. Take your three data sets and enter them into columns A, B, and C in your favorite spreadsheet program (Microsoft Excel is on our computers)—be sure to enter a text header in the first row of each column.

2. Create a *histogram* of the data. If you don't know how to do this, use the program Help function or your colleagues to learn how. You can also discuss this further on the WebCT bulletin board.

3. What is the general shape of the graph? From the graph, try to estimate the cluster center and the best value for the scatter. How do these compare to your above calculate values?

4. Use the built in AVERAGE and STDEV functions on each of the columns of your data and compare/contrast to your graphical and above calculated results.

5. Go to the web and try do some research on *average*, *mean*, *median*, *mode*, and *standard deviation*. Be sure to note which sites you used and try to assess the sites you use for how well you can trust the information they are give you—note anything about the sites(s) used that give you confidence in the information they present.