Introduction to Statistics and Random Errors

Content Discussion and Activities

PHYS 104L

1 Goal

The goal of this week's activities is to provide a foundational understanding regarding statistics and random errors. Distinguishing between random and systematic errors, how different errors affect measurements and results differently, how statistical calculations can be performed to estimate uncertainties due to random errors, and practice in the use of Excel to perform these calculations are primary topics to be covered.

2 Introduction

We have already discussed and covered the fundamental importance of measurements in science, units associated with measurements, calculations involving measured quantities, and the ultimate limitation of reading errors when making measurements. Now we will build on this work by focusing on understanding other types of error effects which can affect measurements and any subsequent calculated results. As described in prior weeks, the **reading error** of the measuring device used is the limiting source of uncertainty in a measured value. If one is measuring a length and one is doing so with a measuring device with a reading error of 0.1 cm, then that measured value must have an uncertainty of at least 0.1 cm. The actual uncertainty in our measurement could turn out to be larger than 0.1 cm if there are other sources of uncertainty or error affecting the measurement, but the uncertainty can't be less than the reading error. When making measurements, whenever possible, one should repeat a particular measurement several times, using the same technique and the same equipment/device. If a measurement has been repeated several times and the same value is obtained on each of those measurements, that doesn't mean there is zero uncertainty in the measured value. For example, if the length of some object is measured five times with resulting values of 72.4 cm, 72.4 cm, 72.4 cm, 72.4 cm and 72.4 cm; we cannot conclude that the value for this object's length is exactly 72.4 cm! The value is still limited by the reading error of 0.1 cm. The actual value might well be 72.436792 cm and we would have no way of knowing because of the reading error of our measuring device. If we repeat a measurement several times and get the same value each time, the uncertainty in that measurement still must be at least the reading error. Even though we got the same value for the length on each of the five measurements, we would list our result for the length as 72.4 ± 0.1 cm. While we should always repeat measurements multiple times to check, in some cases it might be obvious that repeating a measurement several times will always result in the same value. In these situations, it is just assumed that one would get the same values and the reading error is assumed to be the uncertainty.

3A Discussion - Random and Systematic Errors

While measuring the same value on multiple, repeated measurements can occur, often when one repeats a measurement of a particular physical quantity multiple times one will get different values even though the quantity should be remaining constant. The variation in these repeated measurements is due to **random errors** affecting each individual measurement. A random error source is one that, as the name implies, causes a measurement to differ from the true value by random amounts, varying from trial to trial. Sometimes the measured value will be too big, but it is just as likely that it will be too small. Sometimes the measured value will be off by a lot and other times it will be off only by a little. How much and in which direction the measured value is off varies from one measurement to the next. These are the hallmark characteristics of random errors. A good example of a random error effect is human reaction time when starting and stopping a stopwatch. We may try our best to synch the "start" and "stop" of pushing the stopwatch button with the "start" and "stop" of the event, however, it is likely we will be off by some amount both pushes. We could be off by a lot and then only off by a little bit the next time. We could get a time that is too long and then the next time get one that is too short. The amount we are off by on any one time interval measurement varies randomly. Random variation in a set of measured values could be caused either by the equipment being used, the technique being used, or even due to actual random variation in the quantity itself that we are trying to measure. The key outcome is that when non-negligible random errors are affecting our measurements, we expect the measured values to vary from trial to trial as we repeat the measurement multiple times. How much the values will vary or scatter from trial to trial will depend on how big the random error effects are. It should make sense that the bigger the random error effect, the larger the variation in measured values will be, and the smaller the random error effect is, the smaller the variation in measured values will be. What would happen if the magnitude of the random error effects were so small that we didn't observe any scatter or variation in our measured values? In this case, we would get the same value for every measurement trial after trial; we would have an uncertainty limited by the reading error of the measuring device. This situation is the case in which we need to know and use the reading error for the measuring device as our uncertainty: when the size of any random error effect is smaller than the reading error. In this case, we say random error effects are negligible. Notice, if one was to acquire a measuring device with a smaller reading error and repeat the measurements, the values might now vary from trial to trial indicating the presence of a random error effect larger than our new smaller reading error.

While **random errors** result in measured values scattering from trial to trial, **systematic errors** are a different type of error with a very different type of effect on measured values. A **systematic error**, as the name also implies, causes a measured value to be off by the **same amount** in the **same direction** time after time. This outcome is very different from that of a random error. A systematic error doesn't cause scatter, rather, it causes the measurements to all be shifted by some constant amount. With just random errors present, the measured values will vary and scatter around the correct value. With just systematic errors present, the measured values will all be the same, but that constant value will be offset from the actual value. With both systematic and random errors present, the measured values will scatter about a constant (but incorrect) value. The source of a systematic error may be the equipment being used, the technique being used, or something else affecting the actual quantity being measured. Here is an example of a systematic error. What if you are measuring the length of an object, being very careful to minimize random errors, but you are using a meterstick with the first 7 cm cut off of it and you don't notice? No matter how careful you are, every measured length is going to be off by 7 cm; a constant offset.

3B Activity – Random and Systematic Errors

1.) On your activity sheet, for several different measurements, a source of potential error is described. Take a few minutes and discuss each situation with your partner or group to determine whether the source of error is a **random**, **systematic**, or **reading** error. Mark it as such.

4A Discussion – Precision and Accuracy

Precision and accuracy are two words often used interchangeably in most situations. However, for a scientist describing measurements, data or results, precision and accuracy mean very different things; a difference which is important for all of us to understand. **Precision** refers to the consistency of a measurement as it is repeated. The way you can assess the precision of your data is by seeing how much it varies from trial to trial. Very precise data would have very little scatter or variation to it. As you might have already guessed, **random errors** are what affect the **precision** of your measured data and calculated results. If there are large random error effects, then the precision of your results will be lower. Recall, our experimental results are always a range, not a single number, rather a Best Estimate ± Uncertainty. For our analysis, random errors determine the size of the uncertainty (unless there is negligible data scatter in which case the reading error will determine the uncertainty). Random errors determine the width of your experimental range. This outcome should make sense. If there are large random error effects, the data will scatter a lot and our calculations will indicate a wide range of possible values for the quantity of interest. On the other hand, if there are very small random error effects, the data will have very little scatter and our calculations will give a very narrow range of possible values for the quantity of interest. A narrow results range corresponds to precise data!

Accuracy, on the other hand, refers to the correctness of your data and results. There is no way to assess the accuracy of your data just by looking at your data or results. The only way to check for accuracy is to compare your results with an accepted, expected, or theoretically predicted result and see if they agree. If your results do not agree with what you were expecting, it is the accuracy of your results that is in question. The reason a result would not be in agreement with what was expected is that the result was offset by so much that it didn't agree. As you again may have guessed, **systematic errors** are what affect the **accuracy** of your measured data and calculated results. If there is a lack of agreement between your experimental and expected results, systematic errors offsetting your results may well be the culprit. Thinking of your experimental result, random errors determine the width of the range while systematic errors would be responsible for shifting the center of that range one way or the other. If the systematic error effects are large enough, the range can get shifted so much that it does not agree with the accepted value. Systematic errors do not affect the width of the range because they do not cause any scatter. A result which is in agreement with an expected value is called an accurate result. Accurate data is not necessarily precise just as precise data is not necessarily accurate. It is important to understand this difference so that "accuracy" and "precision" can be used to correctly describe your data and results and better help indicate the importance of random or systematic errors.

4B Activity – Precision and Accuracy

1.) On your activity sheet are four blank dart boards. You and your partners are tasked with considering a situation where 8 darts are thrown at a board with those throws being subject to different combinations of random and systematic errors. You are to assume the thrower is trying to hit the bullseye. Discuss with your partner and others how you would expect the different combination of errors to result in those darts landing for the following four different cases:

Board A - Assume large random errors but negligible systematic errors.

Board B - Assume small random errors but a large systematic error.

Board C - Assume small random errors and negligible systematic errors.

Board D - Assume large random errors and a large systematic error.

2.) Under each board, identify whether you would describe those darts as precise or accurate by indicating Yes or No in the appropriate blanks.

3.) After discussing the results of steps 1 and 2, the instructor will set up a pendulum at the front of the lab room. The instructor will start the pendulum swinging back and forth. The period of a pendulum is the time it takes to complete one full cycle of motion (starting at one extreme side, the time it takes to swing over and then back to where it started). Each student in lab is to use a hand stopwatch to directly measure the period of the pendulum. Make 5 independent measurements of this period and record them in the top row of the appropriate data sheet.

4.) Once you have made your measurements and recorded them on your activity sheet, the instructor will have a place for everyone in the class to record their measured times. As each student enters their measured values, copy those onto your activity sheet such that you have the five measured times for each student recorded. The instructor will give you an expected value.

5.) For each set of 5 measurements, record the minimum time and the maximum time in the appropriate columns. As a rule, do these "ranges" tend to overlap with each other or not? Do several overlap with just a couple that don't or do the "ranges" not really overlap with each other at all? Explain these observations in terms of how you think random and systematic errors were affecting the measured values.

5A Discussion – Statistical Quantities

In the prior activity, we just used the largest and smallest values measured to define a range. These values are very dependent on the specific data we happened to collect. If you had collected six measurements, the sixth one might have been larger or smaller and substantially changed what we would have listed as a range. We also can't really associate much meaning to that range other than it is the highest to lowest values. We would like something better. In other words, let's assume we have made several measurements for a particular value. We expected all of those values to be the same in the absence of any random errors. However, it turns out we do have non-negligible random errors affecting each measurement so the measured values have varied from trial to trial. Just as in your last activity, we'll assume 5 measurements for a particular time interval. Our goal is to use those measured values to determine our **Best Estimate** for the quantity and an **Uncertainty** to associate with that Best Estimate. Well, there just happen to be some statistical methods that we can use to get a Best Estimate and a meaningful value for the Uncertainty in our Best Estimate using the measured values.

We know our measured values were subject to random errors. Knowing this, how should we calculate our Best Estimate for the value of what we were trying to measure? Random errors mean that each measurement was just as likely to be smaller than the actual value as they were to be larger than the actual value. Since these errors vary randomly around the correct value, it would make sense to take the **average** (or mean) of them to try to get Best Estimate since the random variations should tend to cancel out. It should also make sense that if we were to increase the number of trials, then the average should increasingly approach the actual value as more measurements make it even more likely that the individual variations due to random errors will cancel out.

Now that we know how to get the Best Estimate, how do we get this Uncertainty in our Best Estimate? We use some statistics, which we've already done by calculating an average. The first thing we do is compute the **standard deviation**, and from that we compute two additional items: **standard error** and **associated error**. The associated error is what we will use as our Uncertainty. The standard deviation is a measure of the typical (standard) difference (deviation) between an individual measured value and the actual value we were trying to measure. The standard deviation turns out to be the uncertainty in any one of our individual measured values due to all the random errors affecting the measurement. We expect the uncertainty in any one measured value to be larger than the Uncertainty in the Best Estimate. This is because in averaging to get the Best Estimate, some of those random variations will have canceled out. The standard error **is** an uncertainty in the Best Estimate. However, if we used the standard error for our Uncertainty in the Best Estimate, we would have a range with which we would expect agreement 67% of the time, and so we would not be getting agreement 33% of the time just due to bad luck and random errors not canceling out well. We believe this is too high of a chance for bad luck to prevent agreement with our results. To increase the probability of expected agreement, we need to make the Uncertainty larger (so the result range is wider). To address this, we will calculate the associated error and use that for our Uncertainty. The associated error will give a range in which we expect agreement 95% of the time. A 5% chance of having bad luck is small enough for us to ignore. Calculating these 4 statistical quantities; average, standard deviation, standard error, and associated error, then allows us to list a range in which we are 95% confident the actual value we were trying to determine should be within. This statistical analysis is the best way we have for using data to calculate an experimental range by effectively estimating the Uncertainty due to random error effects affecting our measurements and data.

Here's an example of how to do this. Imagine that the times listed in the table below are those for timing an object moving through some distance:

Trial	Time (s)
1	2.34
2	2.31
3	2.28
4	2.39
5	2.32

We first compute the **average**: $t_{avg} = (2.34s + 2.31s + 2.28s + 2.39s + 2.32s)/5 = 2.328s$. Note how the units are in the equation and also in the answer, so there is no question as to what they are. To find the **standard deviation**, we need to compute the difference between this average and each of the values of our data set. We then square these differences and add them together, and finally we divide by one less than the total number of trials and square root that:

$$\sigma = \sqrt{\frac{(2.34s - 2.328s)^2 + (2.31s - 2.328s)^2 + (2.28s - 2.328s)^2 + (2.39s - 2.328s)^2 + (2.32s - 2.328s)^2}{4}}$$

so that

$$\sigma = 0.040865633 \text{ s}$$

where again the units appear in the equation and are correct in the result. To compute the **standard error**, we take our standard deviation and divide it by the square root of the number of trials:

$$SE = 0.040865633\,s/\sqrt{5} = 0.018275667s$$

again with proper units. Finally, the **associated error** is obtained by multiplying this value by 1.96:

AE=1.96*0.018275667 s =0.035820307 s

so that we could then state our result as: $t_{avg} = 2.328 \pm 0.035820307s$. We would expect the actual time, for the event we were measuring, to be within this range 95% of the time. Your instructor will demonstrate how you can use an Excel spreadsheet to do these statistics calculations fairly quickly and easily rather than doing them by hand.

Writing down this result brings up another item we need to address, and that is how to handle all of the digits in these long numbers that calculators and spreadsheets give us. Or, in other words, how many significant figures should there be? Our approach to this is fairly straight forward. First, when you obtain a number corresponding to a measurement from an instrument, that number will have a certain number of digits which is based on the reading error for the instrument. So, for a stopwatch, you will have numbers like 1.27s, 12.38s, etc. You may have three or four digits, depending on the amount of time you measure, but in all cases the stopwatch gives two decimal places, and so all numbers in your data table should be given to this number of decimal places (including trailing 0s, which spreadsheet programs like to truncate). For a ruler or meter stick, you can read to one millimeter, so all numbers should have one decimal place if expressed in cm. Next, for numbers that are results that you want to show us as your final result to something we ask you for in a lab handout (like the example above's last value for t_{avg}) we want you to find the first non-zero number in the associated error (AE) value, regardless of where it is relative to the decimal point. You will keep this digit as the Uncertainty value, and you should round it up one if the digit to its right is 5 or greater. Then, depending on where this digit is relative to the decimal point, you keep that same amount of precision in the Best Estimate. Considering the above result as an example, if we look at the Uncertainty, the first non-zero digit is the 3, but since it is followed by a 5, we will round it up to a 4. That gives us 0.04s as our Uncertainty. Since the four is in the hundredths place, we will express our value for Best Estimate out to this decimal place, rounding if necessary. So, then, our result for the average time in our example, rounded to the correct number of digits, would be: $t_{avg} = 2.33 \pm 0.04s$.

5B Activity – Statistical Quantities

1.) You will be using the stopwatch at your table. Reset it and then try to start and stop it so that you stop it as close to 2.00 seconds as possible. Record this time on your activity sheet. Repeat this 11 more times so that you have a total of 12 attempts at stopping the watch at exactly 2.00 seconds. Don't ignore any attempts unless the buttons don't work or there is some kind of equipment malfunction.

2.) Calculate and record the average, standard deviation, standard error, and associated error using a data set comprised of your first four attempts. Use these values to record a properly rounded result and range.

3.) Repeat step 2 again, this time using all 12 measured values. Check to see if this result is in agreement with the result from step 2.

4.) Compare each of the four calculated statistics values from step 2 (using 4 values) with those of step 3 (using all 12 values). Are the averages substantially different when using 3 times as much data? What about the standard deviation? What about the standard error and the associated error? Are these answers what you would expect?

5.) If somebody asked you what the uncertainty in any one of your individual measured times was, what value would you give? Explain.

6.) We should expect about 95% of our measured values to be within our result range. Count how many of your 12 measured values are within this range and list this value on your activity sheet. How many would you expect to be within the range? List this value as the expected number. Were your results reasonably consistent with what you expect or not? Explain what this tells you about your measured values.

7.) Return to your data from activity 4B. For each set of student pendulum period data, calculate and list the average, standard deviation, standard error and associated error. Using Excel will make these calculations quicker. Also list a properly rounded result and range. Check to see if each result is in agreement with the given expected value for the theoretical period of that pendulum.

8.) Identify which data sets had the smallest and largest random errors affecting them. Explain how you determined these where the ones.

9.) Identify which data sets had non-negligible systematic errors affecting them. Explain how you determined these where the ones.

Introduction to Statistics and Random Errors

Activity Data Sheet

PHYS 104L

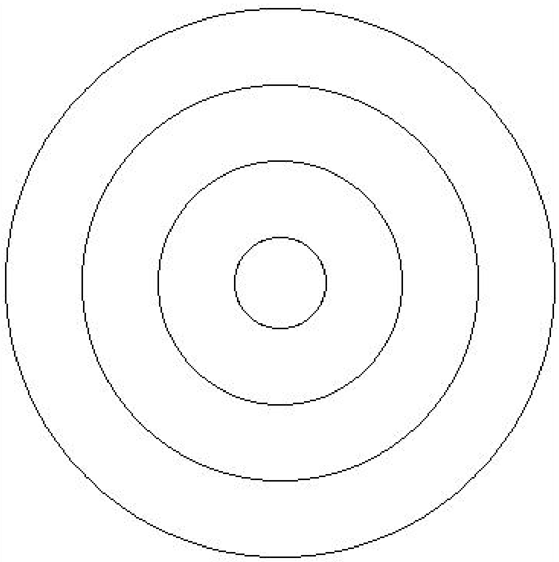
Activity 3B

Measurement	Issue	Error Type
Length of a stick	Lining up the ruler with the stick's end	________________
Mass of an object	Balance only measures to nearest 0.1 gram	________________
Get 1 cup of water	Filling the measuring cup exactly to the line	________________
Time worked	The clock used runs slow	________________
Height of a step	Measured 6.75 inches five straight times	________________
How much fluid in container	Spilled some fluid when pouring it out	________________
Distance ball fell	Measured from top of ball to the floor	________________
Mass of a leaf	Different leaves give different masses	________________

Activity 4B Accuracy and Precision

Board A

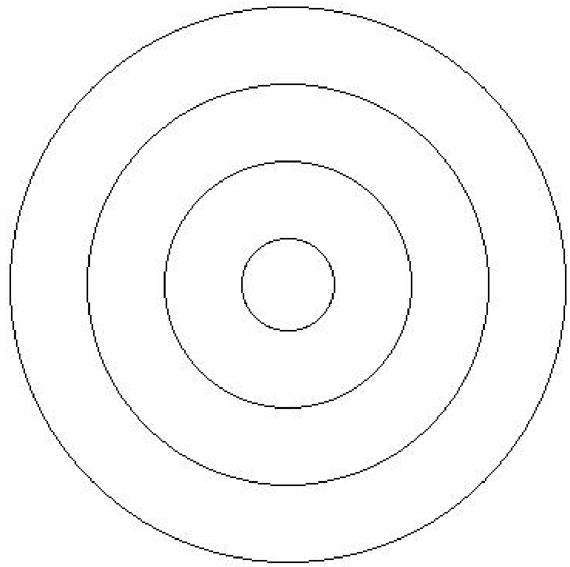
Large Random, Negligible Systematic



Precise? ______ Accurate? _______

Board B

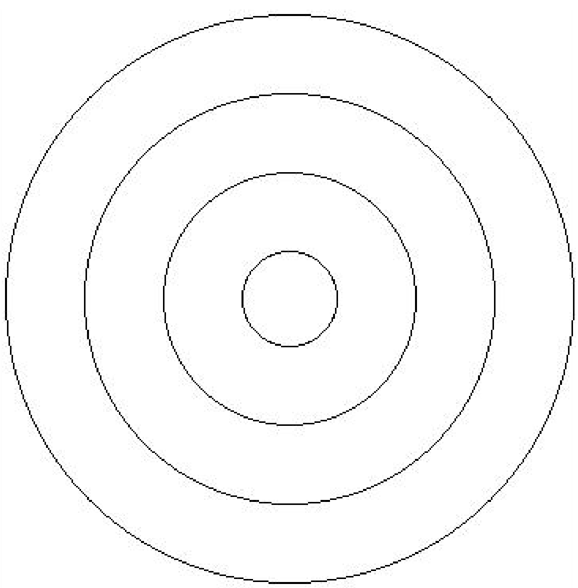
Small Random, Large Systematic



Precise? _______ Accurate? ______

Board C

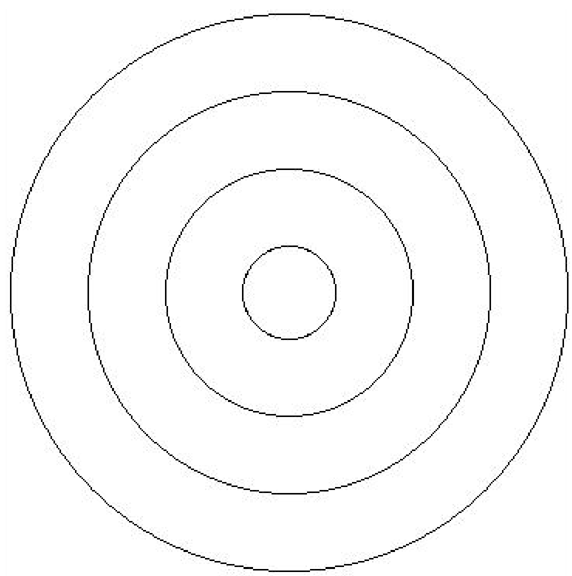
Small Random, Negligible Systematic



Precise? ______ Accurate? _______

Board D

Large Random, Large Systematic



Precise? _______ Accurate? ______

Name	Trial 1 (Sec)	Trial 2 (Sec)	Trial 3 (Sec)	Trial 4 (Sec)	Trial 5 (Sec)	Min (Sec)	Max (Sec)

Expected Value ______________________

Discuss the Overlap of the different"Ranges"?

Which data set, or sets, had the smallest random error effects? Which had the largest? Explain.

Which data set, or sets, would you guess are most likely to have had non-negligible systematic errors affecting the measurements? Why?

Activity 5B   Statistical Quantities

Stopwatch Reading (Seconds)

Statistics Calculation Values

	Average (Seconds)	Stan. Dev. (Seconds)	Stan. Err. (Seconds)	Assoc. Err (Seconds)	Result (Seconds)	Range (Seconds)
Use 1st 4						
All 12						

Are the two results in agreement with each other?

How do the statistical values from the two data sets compare with each other?

What is the best estimate for the numerical uncertainty, due to random errors, in any one of your individual measurements? Explain.

Which of your result ranges were in agreement with the target value of 2.00 seconds? What does that tell you about any systematic errors affecting your measurements?

Name	Average (Sec)	Stan Dev (Sec)	Stan Err (Sec)	Assoc Err (Sec)	Result (Sec)	Range (Sec)	Agree? Yes/No

Which data set, or sets, had the smallest random error effects? Which had the largest? Explain.

Which data set, or sets, would you guess are most likely to have had non-negligible systematic errors affecting the measurements? Why?