Chapter 2

Error Analysis

Name:________________ **Lab Partner:**____________ **Section:**_____

2.1 Purpose

In this experiment error analysis and propagation will be explored.

2.2 Introduction

Experimental physics is the foundation upon which the theoretical models of physical phenomena are built. Theory is an explanation of what experiment shows us and is, of course, inherently limited by the accuracy, precision, and basic viability of the experiment or experiments. We need to know error and uncertainty to design better, more accurate and precise experiments. These experiments will, hopefully, allow us to test and evaluate more of the subtle features of nature and thus allow for improvement of existing theories. In this lab, we will introduce the concepts of error, accuracy, and precision. You should understand what the various categories of error are, how to evaluate them, and how to express your experimental results with error limits and confidence factors. It should be stated that experimental error is not a quantity that can be eliminated. Indeed, the existence of experimental error is in itself an indication of the correctness of the theory.

In this laboratory course you will be using many measurement devices: meter stick, balance (scales), spark timer, force table, timers and compass, among others. From these you will draw conclusions on what you have observed, and use these numbers to calculate other numbers. Each measurement device has an error associated with it that is in turn, propagated and usually magnified in doing calculations. We will start this experiment with an evaluation of such devices and how you should determine their error. We will then proceed to random and systematic error and how to work with them. We will see that the numbers you list need to have significant digits. Finally, a method of propagation of these numbers through your calculations will allow you to present your experimental results with a reasonable idea of how accurate your result is.

2.2.1 Measurement Devices

The normal method by which uncertainty is determined for such things as analog voltmeters, meter sticks and the like is to assume the error in any given measurement is one-half the smallest scale division of the analog device. Thus, a ruler that is graduated in millimeters has an error of ± 0.5 mm. For example, you have measured how far a cart has traveled before it comes to a complete stop. This distance is then determined with a ruler and is found to be 1.234 meters and a small amount more or less than one millimeter. By convention, you are allowed to round up or down to the nearest half millimeter. So, the measurement appears nearer the middle of the distance between the millimeter marks than either side. You will then express the number as 1.2345 meters. If the distance is nearer to the millimeter marks the distance is then 1.234 or 1.235 depending on the position of the cart's edge. So, to completely state your measured quantity, you can write the following:

1.2345 ± 0.0005 m

As another example, you have measured the time that the cart was in motion. You found it to be 2.25 seconds by means of a digital stopwatch that has digits to the hundredth of a second. This is expressed as:

2.25 ± 0.01 seconds

Note that it is not possible to determine whether or not the time point is nearer the middle of the hundredth or the end marks. The stopwatch only gives whole digits to the hundredth. This is an exception to the $\frac{1}{2}$ smallest scale division rule (ssd).

We will encounter several computer sensors which will have digital displays. When encountering a digital device, the device will display the last significant digit of the display gives the uncertainty of the measurement

2.2.2 Random Uncertainty

In many experiments you will measure a large enough number of data points that it is statistically possible to use an averaging method of determining the uncertainty in your experiment. The average of a number of data points can then be considered the true value of what you are measuring. The ssd rule can then be replaced by statistical considerations. First, the Arithmetic Mean is the average value of "enough" data measurements. The subject of what is "enough" will be reserved for future work. It is, however, more than one or two measurements.

2.2.3 Assumptions in Statistical Uncertainty

Statistical analysis of experimental uncertainty rests upon several assumptions pertaining to the number of measurements, the "distribution" of the measurements about a "mean", and the size of the sample in relation to a total population of some sort. The number of measurements are generally considered to be adequate if they are fifty or more. The

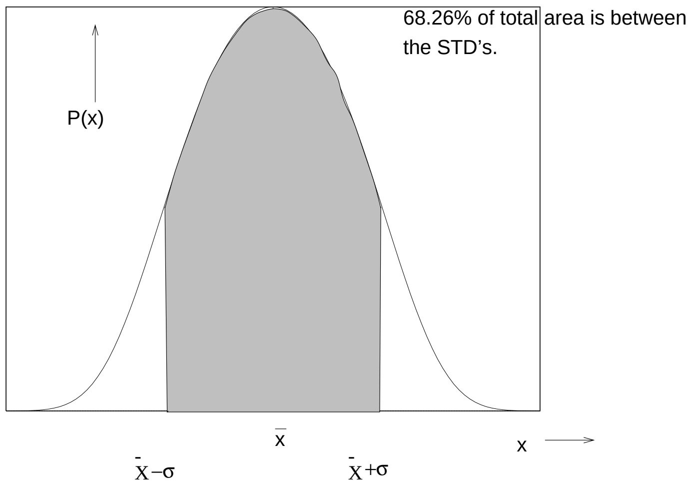

Figure 2.1: The Gaussian distribution is symmetric about the mean value ($\bar{x}$), asymptotically approaches zero far from $\bar{x}$, and determines the probability that a measurement x will be between x_1 and x_2 (shaded area).

"distribution" of the data points is termed "normal" if they follow a Gaussian distribution function (see Figures 2.1 and 2.2):

The function in Figure 2.1 is also termed the Normal Distribution since the total area under the curve is "normalized" to one. This is a probability function in that the probability of a given measurement value being between two other values is the area under the curve between the values. This is the shaded area seen in the figure. Indeed, this is usually given as a probability formula:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

where $P(x)$ is the probability that a given value x will be observed for some quantity when measured, $\bar{x}$ is the arithmetic mean, and σ is the Standard Deviation defined below. A very important aspect of the standard deviation is the area enclosed under the normal curve between $\bar{x} - \sigma$ and $\bar{x} + \sigma$ is 68.26% of the total area under the Normal Distribution Curve. Keep this in mind because we will discuss "Confidence Factors" shortly.

2.2.4 Basic Statistical Quantities

The arithmetic mean is the average of several (i) values:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{2.1}$$

where $\bar{X}$ is the arithmetic mean, n is the number of data points, and X_i are the values for the individual points.

The next quantity we must deal with is the deviation. It is the difference between the Mean and the individual points that were measured. We need this to arrive at an expression for the uncertainty in the Mean. The deviation is defined by:

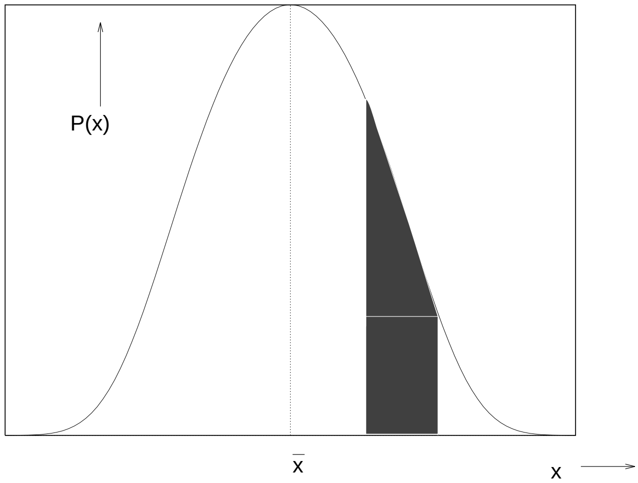

Figure 2.2: The Gaussian distribution

$$D_i = X_i - \bar{X}$$

.

where D_i is the deviation, X_i are the values measured, and $\bar{X}$ is again, the arithmetic mean. There is a deviation value for each data point and it is either positive or negative in sign. Sometimes the term 'residual' is used instead of 'deviation'.

We are next faced with the task of determining the uncertainty in the mean itself. Note that this is completely statistical in its treatment and does not account for device error directly. We will discuss this further. For now, the definition of the uncertainty in the arithmetic mean is called the "Standard Deviation". This is the calculated interval in which the absolute value of any measured data point in the experiment, has a 68% chance of being found.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(D_i)^2}{n-1}} \tag{2.2}$$

where D_i is the ith deviation and n is the number of measurements. The purpose of squaring the Deviations is to ensure that an accurate portrayal of the uncertainty of the mean is presented. The inclusion of arithmetic sign would cause the under estimation of the uncertainty. The assumption behind this is that any data measurement has an equal likelihood of being higher or lower than the mean. (n - 1) is used in the denominator instead of n to adhere to the theory of errors. One valuable consequence of this is that a single measurement of a quantity does not allow us to estimate the spread in values. The percent probability of the absolute value of data points falling within integral multiples of σ are as follows: 68% - σ, 95.4% - 2σ, 99.7% - 3σ.

2.2.5 Significant Figures

The subject of significant Figures in data reporting and calculation has been ignored until now so that the meaning of this can be viewed in light of the statistical treatment of experimental data. Let's begin by stating the accepted method of presenting the arithmetic mean and its associated standard deviation.

$$X = \bar{X} \pm \sigma.$$

$\bar{X}$ is stated as, for example, the meter stick measurement previously discussed. It will have as many significant figures in the number with the last digit as $\frac{1}{2}$ ssd. The σ value will then be listed as whatever you have calculated for the SD to one significant digit that has the same magnitude as the least significant digit of $\bar{X}$. NOTE: if for some reason the σ has a greater magnitude than the least significant digit of $\bar{X}$, then the least significant digit of $\bar{X}$ will be on the same magnitude as σ.

Example: Distance = 1.2345 ± 0.001 meters will need to be changed to 1.234 ± 0.001 meters. You may add a zero after the σ *if it is justified by the calculation*, i.e. 1.234 ± 0.0010 *not* 1.234 ± 0.0013. The last digit in the significant figure is unjustified because it is not in the Arithmetic Mean.

You should, by now be familiar with scientific notation. To state your values in this manner the following convention is used:

measured charge $= (2.71 \pm 0.05) \times 10^{-19}$ Coulombs.

This is easier to read and understand than measured charge $= 2.17 \times 10^{-19} \pm 5 \times 10^{-21}$ Coulombs.

2.2.6 Propagation of Uncertainty

When you do calculations with measured numbers to arrive at another number, the error in the calculated number is determined by what calculations are done. If you add two numbers the error in the sum is different from the error in the two numbers you multiply together. The following section contains formulas for propagating uncertainties through your calculations. Using these basic equations, you can calculate the uncertainty is complicated expressions. You will probably have to refer to this section through out the course when doing uncertainty propagation.

For a function F = F(x, y, z, ...) and constant C, some basic equations for propagating errors are:

1. $F = x + y$, the expression of the error is then $(\Delta F)^2 = (\Delta x)^2 + (\Delta y)^2$.
2. $F = x \cdot y$, the error is given by $\frac{\Delta F}{F} = \sqrt{(\frac{\Delta x}{x})^2 + (\frac{\Delta y}{y})^2}$.
3. $F = x^2$, the error is given by $\Delta F = 2x(\Delta x)$.
4. $F = Cx^n y^m z^k$, the error is given by $\frac{\Delta F}{F} = \sqrt{(\frac{n\Delta x}{x})^2 + (\frac{m\Delta y}{y})^2 + (\frac{k\Delta z}{z})^2}$.

5. $F = e^x$, the error is given by $\frac{\Delta F}{F} = \Delta x$.

6. $F = C \cdot ln(x)$ the error is given by $\Delta F = A\frac{\Delta x}{x}$

The above examples will enable you to express the error equations of most of the calculations you will encounter. Normally the Δx, Δy and Δz values are the standard deviations (σ) from previous calculations or the ssd rule. Note that these expressions can be substituted back to arrive at expressions for the more complicated equations.

You will have to perform this analysis several times this semester, so try to become proficient at writing down error equations. When ever you write down a result, you should use the correct number of significant digits and include an uncertainty (error).

2.3 Procedure

2.3.1 Density measurement of an object

The density of an object, ρ, is the mass, m, of the object divided by the volume. The units are kg/m^3. We will consider a cylinder with diameter, d, and length L. The volume of a cylinder is V = $\frac{\pi d^2 L}{4}$ so the density of th cylinder is:

$$\rho = \frac{4m}{\pi d^2 L} = (\frac{4}{\pi}) m d^{-2} L^{-1} \tag{2.3}$$

Using the fourth error propagation formula, the expression for the error ($\Delta\rho$) is given by:

$$\frac{\Delta\rho}{\rho} = \{(\frac{\Delta m}{m})^2 + 4(\frac{\Delta d}{d})^2 + (\frac{\Delta L}{L})^2\}^{\frac{1}{2}} \tag{2.4}$$

- Weigh the aluminum cylinder on the three different scales. Recorded each of the three measurements with the correct number of significant digits and an uncertainty using the ssd rule. Calculate the mean ($\bar{m}$), squared deviations (D$_i$)^2 and standard deviation (σ) for the three values. When calculating the error in the density measurement, use the mean value for the mass ($\bar{m}$) and the standard deviation for the uncertainty in the mass (Δ m).

Trial	Mass	(Deviation)^2
1		
2		
3		
Sum		

Mean of 3 trials ($\bar{m}$) ____________
Standard Deviation ($\sigma = \Delta m$) ____________

- Measure and record the length of the cylinder (L) and diameter (d) of the aluminum cylinder. Use the ssd value for the uncertainties (ΔL and Δd) in these measurements.

diameter (d) ____________ Δd ____________
Length (L) ____________ ΔL ____________

- Calculate the density and record the value with the correct number of significant digits.

Density (ρ) ____________

Calculate the error in the density using equation 2.4.

error ($\Delta\rho$) ____________

- The standard value for the density of aluminum is 2700 kg/m^3. What is the percentage error for your value? (The Math Appendix at the end of the manual defines the percentage error.) Record the number of errors ($\Delta\rho$) your measured value is from the standard values.

Percentage error ____________
Number of errors ($\Delta\rho$) your value is from the standard value ____________

2.3.2 Gaussian Distribution

Rolling dice provides a convenient simulation of a random Gaussian instrumental errors. In this simulation, two dice will be rolled. The sum of the face values of the two dice provide the number of interest. For example, with two dice there is only one way to roll a total of 2 (1 and 1), but there are three ways to roll a total of 7 (1 and 6; 2 and 5; 3 and 4). Because of this, rolls of two dice will have more sevens than twos. A collection of rolls of two dice forms a Gaussian distribution centered about 7.

- Using the two dice, record the sum of the face values for 40 rolls of the dice. Fill in the table of the number of times each value (2 through 12) occurred.
- Determine the mean value and standard deviation for the 40 dice rolls.

Mean of 40 trials ____________
Standard Deviation (σ) ____________

- Plot a histogram (bar graph) of the frequency distribution of the number of times a given value (2 through 12) has been rolled. Plot on the x axis the possible values going from 2 to 12. On the y axis plot the number of times the corresponding x value has been rolled out of the 40 rolls. Include the properly labeled and referenced graph with your report. Graph paper is provided at the end of this chapter.
- Is the plot approximately a Gaussian shaped curve?

Dice Value	Number
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

2.3.3 Questions

1. In the list of error propagation formulas, no formula is given for F = x - y. The error propagation formula for F = x - y is the same as the formula for F = x + y in the list. Why should this be so?

2. In the list of error propagation formulas, no formula is given for F = $\frac{x}{y}$. Find the correct error propagation formula for division. Hint: use error propagation formula number 4 with $\frac{1}{y}$ = y^{-1}.

3. You can find the mean of the Gaussian distribution by finding a vertical line that divides the area under the histogram in half. In the dice rolling simulation, one half of the area is 20. Count 'events' from the left or right until you have found $\frac{1}{2}$ the area. Is this location near seven on the horizontal scale? Why?

2.4 Conclusion

Write a detailed conclusion about what you have learned. Include all relevant numbers you have measure with errors. Sources of error should also be included.