Name _____ Date _____ Partners _____

LAB 1:

PHYSICAL MEASUREMENTS

"A measurement whose accuracy is unknown has no use whatever. It is therefore necessary to know how to estimate the reliability of experimental data and how to convey this information to others."

—E. Bright Wilson, Jr., *An Introduction to Scientific Research*

OBJECTIVES

- To learn how to use measurements to best estimate the "true" values of physical quantities
- To learn how to estimate how close the measured value is likely to be to the "true" value
- To learn some of the notation that scientists and engineers use to express the results of measurements
- To learn some relevant concepts from the mathematical theory of probability and statistics

OVERVIEW

Our mental picture of a physical quantity is that there exists some unchanging, underlying value. It is through measurements we try to find this value. Experience has shown that the results of measurements deviate from these "true" values.

Accuracy and Precision

According to many dictionaries, "accuracy" and "precision" are synonyms. To scientists, however, they refer to two distinct (yet closely related) concepts. When we say that a measurement is "accurate", we mean that it is very near to the "true" value. When we say that a measurement is "precise", we mean that it is very reproducible. [Of course, we want to make accurate AND precise measurements.] Associated with each of these concepts is a type of error.

Systematic errors are due to problems with the technique or measuring instrument. For example, as many of the rulers found in labs have worn ends, length measurements could be wrong. One can make very precise (reproducible) measurements that are quite inaccurate (far from the true value).

Random errors are caused by fluctuations in the very quantities that we are measuring. You could have a well calibrated pressure gauge, but if the pressure is fluctuating, your reading of the gauge, while perhaps accurate, would be imprecise (not very reproducible).

Through careful design and attention to detail, we can usually eliminate (or correct for) systematic errors. Using the worn ruler example above, we could either replace the ruler or we could carefully determine the "zero offset" and simply add it to our recorded measurements.

Random errors, on the other hand, are less easily eliminated or corrected. We usually have to rely upon the mathematical tools of *probability* and *statistics* to help us determine the "true" value that we seek. Using the fluctuating gauge example above, we could make a series of independent measurements of the pressure and take their average as our best estimate of the true value.

Measurements of physical quantities are expressed in numbers. The numbers we record are called *data*, and numbers we compute from our data are called *statistics*[*].

Finally, we must also mention *careless* errors. These usually manifest themselves by producing clearly wrong results. For example, the miswiring of probes and sensors is an all too common cause of poor results in this lab, so please pay attention.

Probability

Scientists base their treatment of random errors on the theory of probability. We will not delve too deeply into this fundamental subject, but will only touch on some highlights. Probability concerns random events. To some events we can assign a theoretical, or *a priori*, probability. For instance, the probability of a "perfect" coin landing heads (or tails, but not both) is 1/2 (50%) for each of the two possible outcomes; the *a priori* probability of a "perfect" die[†] falling with a particular one of its six sides uppermost is 1/6 (16.7%).

The previous examples illustrate four basic principles about probability:

- The possible outcomes have to be mutually exclusive. If a coin lands heads, it does not land tails, and *vice versa*.
- The list of outcomes has to exhaust all possibilities. In the example of the coin we implicitly assumed that the coin neither landed on its edge, nor could it be evaporated by a lightning bolt while in the air, or any other improbable, but not impossible, potential outcome. (And ditto for the die.)
- Probabilities are always numbers between zero and one, inclusive. A probability of one means the outcome always happens, while a probability of zero means the outcome never happens.
- When all possible outcomes are included, the sum of the probabilities of each exclusive outcome is one. That is, the probability that *something* happens is one. So if we flip a coin, the probability that it lands heads *or* tails is 1/2 + 1/2 = 1. If we toss a die, the probability that it lands with 1, 2, 3, 4, 5, *or* 6 spots showing is 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 = 1.

The mapping of a probability to each possible outcome is called a *probability distribution*. Just as our mental picture of there being a "true" value that we can only estimate, we also envision a "true" probability distribution that we can only estimate through observation. Using the dice toss example to illustrate, if we toss five dice,

---

[*] A statistic is, by definition, a number we can compute from a set of data. An example is the average, or *mean*. Another is the *variance*, which we shall define below.
[†] …one of a pair of dice.

we should not be too surprised to get a five and four sixes (think of the game *Yahtzee*). Our estimate of the probability distribution would then be 1/5 for 5 spots, 4/5 for 6 spots, and and 0 for 1, 2, 3, or 4 spots. We do expect that our estimate would improve as the number of tosses[*] gets "large". In fact, it is only in the limit of an infinite number of tosses that we can expect to approach the theoretical, "true" probability distribution.

Probability Distributions

The probability distributions we've discussed so far have been for discrete possible outcomes (coin flips and die tosses). When we measure quantities that are not necessarily discrete (such as pressure read from an analog gauge), our probability distributions become more correctly termed *probability density function* (although you often see "probability distribution" used indiscriminately). The defining property of a probability distribution is that its sum (*integral*) over a range of possible measured values tells us the probability of a measurement yielding a value within the range.

The most common probability distribution encountered in the lab is the *Gaussian* distribution. The Gaussian distribution is also known as the *normal* distribution. You may have heard it called the *bell* curve (because it is shaped somewhat like a fancy bell) when applied to grade distributions. The mathematical form of the Gaussian distribution is:
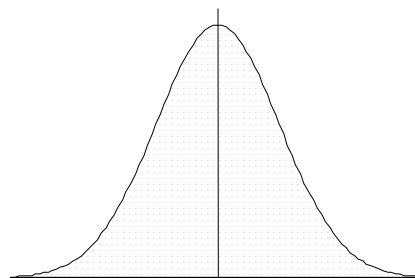

Figure 1 Gaussian Distribution

$$P_G(d) = \frac{1}{\sqrt{2\pi}\sigma} e^{-d^2/2\sigma^2} \qquad (1)$$

The variables will be discussed later. The Gaussian distribution is ubiquitous because it is the end result you get if you have a number of processes, each with their own probability distribution, that "mix together" to yield a final result. We will come back to probability distributions after we've discussed some statistics.

Statistics

Measurements of physical quantities are expressed in numbers. The numbers we record are called *data*, and numbers we compute from our data are called *statistics*. A statistic is, by definition, a number we can compute from a set of data.

Perhaps the single most important statistic is the mean or average. Often we will use a "bar" over a variable (e.g., $\bar{x}$) or "angle brackets" (e.g., $\langle x \rangle$) to indicate that it is an

---

[*] Each toss is, in the language of statistics, called a *trial*. A scientist or engineer would probably say that it is a *measurement* or *observation*.

[‡‡‡] You may wonder where the "5" in the denominator came from. It takes into account the fact that, as we're averaging over five dice, the averages have five possible results per unit range.

average. So, if we have N measurements x_i (i.e., x_1, x_2, ..., x_N), the average is given by:

$$\bar{x} = \langle x \rangle = (x_1 + x_2 + ... + x_N)/N = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (2)$$

The average of a set of measurements is usually our best estimate of the "true" value:

$$\bar{x} \approx x \qquad (3)$$

Note: For these discussions, we will denote the "true" value as a variable without adornment (e.g., x).

In general, a given measurement will differ from the "true" value by some amount. That amount is called a *deviation*. Denoting a deviation by d, we then obtain:

$$d_i = x - x_i \approx \bar{x} - x_i \qquad (4)$$

Clearly, the average deviation is zero, because

$$\bar{d} = \frac{1}{N}\sum_{i=1}^{N}(x - x_i) = \frac{x}{N}\sum_{i=1}^{N} 1 - \frac{1}{N}\sum_{i=1}^{N} x_i = x - \bar{x} \approx 0.$$

Another notable statistic is the *variance*, defined as the mean square deviation:

$$\text{var}(x) \equiv (d_1^2 + d_2^2 + ... + d_N^2)/N = \frac{1}{N}\sum_{i=1}^{N} d_i^2 = \frac{1}{N}\sum_{i=1}^{N}(x - x_i)^2 \qquad (5)$$

The variance is useful because it gives us a measure of the spread or statistical uncertainty in the measurements.

You may have noticed a slight problem with the expression for the variance: We usually don't know the "true" value x; we have only an estimate, $\bar{x}$, from our measurements. It turns out that using $\bar{x}$ instead of x in Equation (5) systematically underestimates the variance. It can be shown that our best estimate of the "true" variance is given by *sample variance*:

$$\text{var}_{sample}(x) \equiv \frac{1}{N-1}\sum_{i=1}^{N}(\bar{x} - x_i)^2 \qquad (6)$$

A related statistic is the *standard deviation* σ_x, which is simply the square root of the variance. This value is appropriate when we have a large, complete population of values.

$$\sigma_x = \sqrt{\text{var}(x)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x - x_i)^2} \qquad (7)$$

If we have a situation where we can make *all possible* measurements, then we should use Equation (7). Equation (7) defines a statistic that, for clarity, is often called the *population standard deviation*.

Note that the standard deviation has the same problem as does the variance in that we

don't know x. Again we find that using $\bar{x}$ instead of x systematically underestimates the standard deviation. We define the ***sample standard deviation*** to be the square root of the sample variance:

$$s_x = \sqrt{\text{var}_{sample}(x)} = \sqrt{\left(\frac{1}{N-1}\right)\sum_{i=1}^{N}(\bar{x} - x_i)^2} \tag{8}$$

The sample standard deviation is our best estimate of the "true" standard deviation. The sample standard deviation gives the best estimate of error in any single measurement x_i.

We also often need the best estimate of error $\sigma_{\bar{x}}$ in determining the mean value $\bar{x}$ of the population. It is given by

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{N}}s_x \tag{9}$$

To illustrate some of these points, consider the following: Suppose we want to know the average height and associated standard deviation of the entering class of students. We could measure every entering student (the entire population) and simply calculate the average. We would then simply calculate x and σ_x directly. Tracking down all of the entering students, however, would be very tedious. We could, instead, measure a representative sample and calculate $\bar{x}$ and s_x as estimates of x and σ_x.

Spreadsheet programs (such as MS Excel™ or Corel Quattro Pro™) as well as some calculators (such as HP and TI) also have built-in statistical functions. For example, **AVERAGE** (Excel), **AVG** (Quattro) and $\bar{x}$ (calculator) calculate the average of a range of cells; whereas **STDEV** (Excel), **STDS** (Quattro) and s_x (calculator) calculate the sample standard deviations. **STDEVP** (Excel), **STD** (Quattro Pro) and σ_x (calculator) calculate the population standard deviation.

Probable Error

We now return to probability distributions. Consider Equation (1), the expression for a Gaussian distribution. You should now have some idea as to why we wrote it in terms of d and σ. Most of the time we find that our measurements (x_i) deviate from the "true" value (x) and that these deviations (d_i) follow a Gaussian distribution with a standard deviation of σ. So, what is the significance of σ? Remember that the integral of a probability distribution over some range gives the probability of obtaining a result within that range. A straightforward calculation shows that the integral of P_G (see Equation (1)) from $-\sigma$ to $+\sigma$ is about 2/3. This means that there is probability of 2/3 for any single measurement being within $\pm\sigma$ of the "true" value. It is in this sense that we introduce the concept of probable error.

Whenever we give a result, we also want to specify a probable error in such a way that we think that there is a 2/3 probability that the "true" value is within the range of values between our result minus the probable error to our result plus the probable

error. In other words, if $\bar{x}$ is our best estimate of the "true" value x and $\sigma_{\bar{x}}$ is our best estimate of the probable error in $\bar{x}$, then there is a 2/3 probability that:

$$\bar{x} - \sigma_{\bar{x}} \le x \le \bar{x} + \sigma_{\bar{x}}$$

When we report results, we use the following notation:

$$\bar{x} \pm \sigma_{\bar{x}}$$

Thus, for example, the electron mass is given in data tables as

$$m_e = (9.109534 \pm 0.000047) \times 10^{-31} \text{ kg.}$$

By this we mean that the electron mass lies between 9.109487×10^{-31} kg and 9.109581×10^{-31} kg, with a probability of roughly 2/3.

Significant Figures

In informal usage the last *significant digit* implies something about the precision of the measurement. For example, if we measure a rod to be 101.3 mm long but consider the result accurate to only ±0.5 mm, we round off and say, "The length is 101 mm". That is, we believe the length lies between 100.5 mm and 101.5 mm, and is closest to 101 mm. The implication, if no error is stated explicitly, is that the uncertainty is ½ of one digit, in the place following the last significant digit.

Zeros to the left of the first non-zero digit do not count in the tally of significant figures. If we say $U = 0.001325$ Volts, the zero to the left of the decimal point, and the two zeros between the decimal point and the digits 1,325 merely locate the decimal point; they do not indicate precision. (The zero to the left of the decimal point is included because decimal points are small and hard to see. It is just a visual clue—and it is a good idea to provide this clue when you write down numerical results in a laboratory!) The voltage U has thus been stated to four, not seven, significant figures. When we write it this way, we say we know its value to about ½ part in 1,000 (strictly, ½ part in 1,325 or one part in 2,650). We could bring this out more clearly by writing either $U = 1.325 \times 10^{-3}$ V, or $U = 1.325$ mV.

When reporting a result with an explicit error estimate, keep enough digits so that your probable error is given to two significant digits (as we did in the previous section).

Important: NEVER round off "intermediate results" when performing a chain of calculations. The associated round-off errors can quickly "propagate" (see next section) and cause your final result to be unnecessarily inaccurate.

Propagation of Errors

More often than not, we want to use our measured quantities in further calculations. The question that then arises is: How do the errors "propagate"? In other words: What is the probable error in a particular calculated quantity given the probable er-

rors in the input values?

Before we answer this question, we want to introduce three new terms:

The *relative error* of a quantity Q is simply its probable error, σ_Q, divided by the absolute value of Q. For example, if a length is known to 49±4 cm, we say it has a relative error of 4/49 = 0.082.

It is often useful to express such fractions in *percent*. In this case we would say that we had a relative error of 8.2%.

When we say that quantities *add in quadrature*, we mean that you first square the individual quantities, then sum squared quantities, and then take the square root of the sum of the squared quantities.

We will simply give the results for propagating errors rather than derive the formulas.

1. If the functional form of the derived quantity (f) is simply the product of a constant (C) times a quantity with known probable error (x and σ_x), then the probable error in the derived quantity is the product of the absolute value of the constant and the probable error in the quantity:

$$f(x) = Cx \rightarrow \sigma_f = |C|\sigma_x \qquad (10)$$

2. If the functional form of the derived quantity (f) is simply the sum or difference of two quantities with known probable error (x and σ_x and y and σ_x), then the probable error in the derived quantity is the quadrature sum of the errors:

$$f(x, y) = x + y \ \text{ or } \ f(x, y) = x - y \ \rightarrow \sigma_f = \sqrt{\sigma_x^2 + \sigma_y^2} \qquad (11)$$

3. If the functional form of the derived quantity (f) is simply the product or ratio of two quantities with known probable error (x and σ_x and y and σ_y), then the *relative* probable error in the derived quantity is the quadrature sum of the *relative* errors:

$$f(x, y) = x \times y \ \text{ or } \ f(x, y) = x/y \rightarrow \sigma_f /|f| = \sqrt{(\sigma_x / x)^2 + (\sigma_y / y)^2} \qquad (12)$$

4. If the functional form of the derived quantity (f) is a quantity with known probable error (x and σ_x) raised to some constant power (a), then the *relative* probable error in the derived quantity is the product of the absolute value of the constant and the *relative* probable error in the quantity:

$$f(x) = x^a \rightarrow \sigma_f /|f| = |a|\sigma_x /|x| \qquad (13)$$

5. If the functional form of the derived quantity (f) is the log of a quantity with known probable error (x and σ_x), then the probable error in the derived quantity is the relative probable error in the quantity:

$$f(x) = \ln(x) \rightarrow \sigma_f = \sigma_x /|x| \qquad (14)$$

6. If the functional form of the derived quantity (f) is the exponential of a

quantity with known probable error (x and σ_x), then the relative probable error in the derived quantity is the probable error in the quantity:

$$f(x) = e^x \rightarrow \sigma_f / |f| = \sigma_x \tag{15}$$

And, finally, we give the general form (you are not expected to know or use this equation; it is only given for "completeness"):

$$f(x, y, ...) \rightarrow \sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + ...$$

Application: Probable Error in the Mean

Suppose that we make two independent measurements of some quantity: x_1 and x_2. Our best estimate of x, the "true" value, is given by the mean, $\bar{x} = (x_1 + x_2)/2$, and our best estimate of the probable error in x_1 and in x_2 is given by the <u>sample standard deviation</u>:

$$\sigma_{x_1} = \sigma_{x_2} = s_x = \sqrt{\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2\right]/(2-1)} \ .$$

Note that s_x is *not* our best estimate of $\sigma_{\bar{x}}$, the probable error in $\bar{x}$. We must use the propagation of errors formulas to get $\sigma_{\bar{x}}$. Now, $\bar{x}$ is not exactly in one of the simple forms where we have a propagation of errors formula. However, we can see that it is of the form of a constant, (½), times something else, $(x_1 + x_2)$, and so:

$$\sigma_{\bar{x}} = \left|\frac{1}{2}\right| \sigma_{x_1 + x_2}$$

The "something else" is a simple sum of two quantities with known probable errors (s_x) and we do have a formula for that:

$$\sigma_{x_1 + x_2} = \sqrt{\sigma_{x_1}^2 + \sigma_{x2}^2} = \sqrt{s_x^2 + s_x^2} = \sqrt{2} s_x$$

So we obtain the desired result for two measurements:

$$\sigma_{\bar{x}} = \tfrac{1}{\sqrt{2}} s_x$$

By taking a second measurement, we have reduced our probable error by a factor of $1/\sqrt{2}$. You can probably see now how you would go about showing that adding third, x_3, changes this factor to $1/\sqrt{3}$. The general result (for N measurements) for the probable error in the mean is:

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{N}} s_x \tag{16}$$

INVESTIGATION 1: PROBABILITY AND STATISTICS

In this Investigation we will use dice to explore some aspects of statistics and prob-

ability theory.
You will need the following:
- five dice
- Styrofoam cup

ACTIVITY 1-1: SINGLE DIE

If each face of a die is equally likely to come up on top, it is clear that the probability distribution will be flat and that the *average* number of spots will be $(1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$ spots. It is perhaps not as clear (albeit straightforward to show) that the standard deviation (from this average) is

$$\sqrt{\left((1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2\right)/6} = 1.7078251$$

We will now test these expectations.

1. Poll the members of your group as to Excel expertise. Assign the task of "operating the computer" to the least experienced group member. The most experienced member should take on the role of "mentor".

2. Open **L01.A1-1 Single Die.xls** (either from within Excel or by "double clicking" on the file in Windows Explorer). Make sure that the tab at the bottom of the page reads "DieToss" (click that tab, if necessary).

3. Roll one die six times, each time entering the number of spots on the top face into sequential rows of column A.

NOTE: Hitting "Enter" after you key in a value will advance to the next row. "Tab" will advance to the next column.

Now that we have some data, we can calculate some statistics.

4. Move to cell B1 and enter the Excel formula "=COUNT(A:A)" (enter the equals sign, but not the quotes!) and hit "Enter". Excel will count the cells in column A that contain numbers and put the result into the cell.

5. Move to cell C1 and enter "Tosses". To pretty things up a bit, highlight column C (by clicking on the column label) and click the **B** button on the toolbar so that entries in this column are rendered in boldface font.

6. Enter the formula "=AVERAGE(A:A)" into B2 and "**Average**" into C2.

NOTE: Before answering any questions or predictions, discuss the issues among your group members and try to come to consensus. However, the written response should be in your own words.

Question 1-1: Discuss the agreement of the experimental average with the theoretical value of 3.5.

7. Enter the formula "=STDEV(A:A)" in B3 and "**StDev**" in C3.

Question 1-2: Discuss the agreement of the experimental standard deviation with the theoretical value of 1.7078251 spots. (Remember that the unit in this case is "spots". The standard deviation will also have units.) Do you expect better or worse agreement as the number of rolls increases? Explain your reasoning.

Now we'll look at the *probability distribution* by creating and plotting a histogram.

8. Enter "**Spots**" into E1 and "**Count**" into F1. Fill in the numbers 1 through 6 into cells E2 through E7.

HINT: To fill in these values, enter "1" into E2 and "2" into E3 and then select these two cells and "click and drag" the lower right hand corner of the selection to automatically fill in the rest of the values.

9. Enter the formula "=COUNTIF(A:A,E2)" into F2. Excel will count all of the cells in column A that contain numbers equal to the number in cell E2.

10. Select the cell F2 and "click and drag" the lower right hand corner of the selection to automatically fill in the rest of the cells (through F7).

NOTE: Excel nicely modifies the cell references in the formula when you "click and drag" copy. Sometimes we don't want this to happen. In such cases, precede the column and/or row label with a "$".

11. Select cells F1 through F7 and press the "Chart Wizard" button on the tool bar (it looks like ▐▐). Select Chart Type "Column" (the default), then press "next". Select the 'Series' tab (at the top) and press the button to the right of the box next to "Category (X) axis labels". Select cells E2 through E7 and press the right hand side button again to get back to the wizard. Press "Finish" and you should see a plot of the histogram.

Question 1-3: As each of the six faces is equally likely, we might naively expect that, with six tosses, to have each face turn up exactly once. Discuss how well your histogram agrees with this expectation. Do you expect better or worse agreement as the number of rolls increases? Explain your reasoning.

Rolling the die many, many, times and recording the results would get rather tedious, so we will resort to computer simulations to get a feel for how things change as the number of rolls increases.

12. Select the "SimDieToss" sheet (click on the tab at the bottom of the screen). You should see a sheet very much like the one you just made, with the notable addition of a button labeled "Toss Dice". Try clicking on the button a few times to see what happens. There is also a column labeled **P** on the Excel sheet. The formula in these cells calculates the ratio of the number of times that the corresponding number of spots appears to the total number of tosses. This represents our "experimental" estimate of the probability that that face will turn up.

13. By changing the value of cell B1 (initially six), you control the number of simulated tosses. Try different values (10, 100, 1,000, 10,000, up to 2^{16}-1 = 65,535, Excel's limit on the number of rows). [Note that you have to "change focus" from the cell (by hitting "enter" or clicking on another cell) to actually change the value before clicking the button.] Click the "Toss Dice" button a few times for each value and observe how the statistics (average and standard deviation) and the histogram change.

> **NOTE:** Excel is configured to automatically recalculate if any cells change. If you ever suspect that Excel has not done so, you can click the F9 key to tell Excel to "Calculate Now". You can also go to Tools→Options menu item, select the Calculation tab, and click the "Automatic" button to restore the automatic calculation feature.

Question 1-4: Qualitatively describe how the statistics and the histogram change as the number of tosses increases. Do the statistics converge to the expected values?

ACTIVITY 1-2: YAHTZEE

We will now explore the assertion that you get a Gaussian probability distribution if we "mix together" a number of random processes. If we toss dice five at a time, each die will still have the flat probability distribution that we observed in 0. Our assertion implies that if we calculate the average of the five dice for each roll, then these averages will have a probability distribution that is (approximately) Gaussian.

It is fairly clear that the average (of the averages) will still be 3.5, but what about the standard deviation? A direct enumeration of all 7,776 possible combinations of five dice (the entire population) yields a standard deviation of 0.763762616 (to a ridiculous nine significant digits). [Not surprisingly, the average did turn out to be 3.5.]

Alternatively, we could use the "Propagation of Errors" formalism. As each die's value is independent of the others, we can see from "Probable Error in the Mean" discussion that the standard deviation of the means will be $1/\sqrt{5}$ times the standard deviation for a single die (see Equation (16)). Not surprisingly, we get the same result.

8. Open **L01.A1-2 Yahtzee.xls**. Make sure that the tab at the bottom of the page reads "Dice". To save time, this sheet already has some labels, formulas, and plots pre-configured. [Note that Excel may be complaining that we are asking it to divide by zero. That is OK; we'll soon "feed" it some data!]

9. Take some time to explore and understand the sheet. Click on the various cells and make sure that you understand what it is configured to do.

    On the left-hand side, columns A through E are labeled **D1** through **D5**. Each row in this set of columns will correspond to a roll of the five dice.

    Column G is labeled **Average**. Each row in this column contains a formula to calculate the average of the corresponding **D1** through **D5** values.

    There are two groupings of statistics and histograms. The first, **Individual Dice**, are calculated without any grouping (in other words, columns A through E). These reproduce the 0 results.

The second, **Five at a Time**, are calculated using the row averages (column G). Here each roll of five dice are treated as a single "measurement".

10. Roll five dice six times and enter the results into the labeled cells (one row per roll). Observe how the statistics and histograms change as the number of rolls increase.

Question 1-5: Discuss qualitatively how the statistics and histogram agree with expectations.

Again, we wish to avoid tedium and will resort to computer simulations.

11. Select the "SimDice" sheet.
     Again you should see a sheet much
     like the one you just used. This time,
     however, in addition to a "Toss
     Dice" button, we have calculated a
     "theoretical" histogram based upon
     our   assumed   probability
     distributions.   The   experimental
     histograms are plotted in a reddish
     color and the theoretical histograms
     are plotted in a pale blue.

     Calculating the **FLAT** distribution is
     easy, we just assign one-sixth of the
     total number of **D**'s to each die face.

The **GAUSSIAN** distribution is a bit
harder and requires a messier formula.
We have to plug the theoretical mean
and standard deviation as well as the
total number of tosses into Equation (1)
‡‡‡.

Figure 2: Free Fall Apparatus

12. Change the value of cell J1 (initially ten) to vary the number of tosses. Try dif-
     ferent values (10, 100, 1,000, 10,000, etc.). Click the "Toss Dice" button a few
     times for each value and observe how the statistics and the histogram change.

Question 1-6: Qualitatively describe how the statistics and the histogram change as
the number of tosses increases. Do the statistics converge to the expected values?
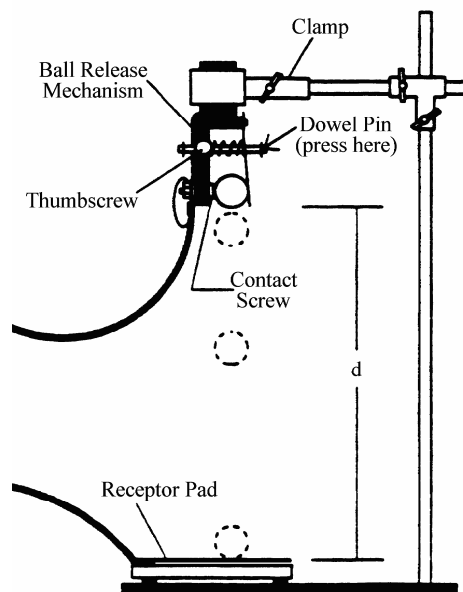

INVESTIGATION 2:   ERRORS

ACTIVITY 2-1: PROPAGATION OF ERRORS

In this investigation we will explore the use of the propagation of errors formulae.
We will measure the time it takes a ball to drop a measured distance. From these
data we will calculate gravitational acceleration *g* and compare it to the accepted
value. To do so in a meaningful way, we must estimate the probable error in our de-

termination.

The distance of fall, *D*, time of fall, *t*, and *g* are related by $D = \frac{1}{2}gt^2$. Hence, we can calculate *g* by:

$$g = 2D/t^2 .\qquad\qquad (17)$$

In the basic free fall experiment, shown in Figure 2, a steel ball is clamped into a spring loaded release mechanism. When the ball is released, it starts the timer, which will keep counting until the ball hits a receptor pad at the bottom of its travel. When the ball strikes the pad, the top plate of the pad is forced against its metal base, and the consequent electrical contact has the effect of stopping the timer. The timer display will then automatically record the time it took for the ball to drop from the release mechanism to the pad.

You will need the following:
- PASCO *Science Workshop* interface
- free fall apparatus
- steel ball
- meter stick

1. Be sure that the interface is connected to the computer. Notify your TA if it is not. Start *Data Studio* and open the activity **L01.A2-1 Ball Drop.ds** [we won't show the extension, "**.ds**", in the future].

2. Take a few minutes to study the display. At the top on the left hand side you will see a window showing possible data sources. Below that there is a window showing possible displays. Under "Table" you will see that there is an entry labeled "Ball Drop". The associated window should also be visible.

   Look at the "Ball Drop" window. You will see a one column table with various statistics (mean, standard deviation, etc.) below it. This is where your data will be displayed.

3. Press the **Setup** button (⚌ Setup) on the toolbar to display how the sensor is to be connected. Make sure that the ball drop apparatus is plugged in to the correct input. Close the Experiment Setup window.

4. Position the clamping mechanism at the top of the stand (so as to maximize the distance the ball falls).

5. Insert the steel ball into the release mechanism, pressing in the dowel pin so the ball is clamped between the contact screw and the hole in the release plate. Tighten the thumbscrew to lock the ball into place. Ask your TA if you have any questions.

6. Loosen the thumbscrew to release the ball. It should hit in the center of the recep-

tor pad. If not, reposition the pad and try again. Drop the ball a few times to get familiar with the process.

7. Reposition the ball in the clamp and carefully measure the distance from the bottom of the ball to the top of the pad. This is your estimate of the "true" distance *D*. Record this value. [Pay attention to the units!]

$$D_{exp} : \underline{\hspace{5cm}} \text{ mm}$$

8. Estimate how well you can read your ruler. This is your estimate of σ_D, the "probable error" in *D*, and should indicate that there is roughly a 2/3 probability that D_{exp} is within $\pm\sigma_D$ of *D* (and that there is a 1/3 probability that it is not).

$$\sigma_D: \underline{\hspace{4cm}} \text{ mm}$$

9. Press the **Start** button (▶ Start) on the *Data Studio* toolbar. The button will change into two buttons, a **Keep** button (✔ Keep) and an adjacent red **Stop** button (■).

10. Drop the ball and press **Keep** to record the fall time. Do this for at least ten drops. If you accidentally record a bad time, ignore it for now and continue taking data. We can fix it afterwards.

11. After you have successfully recorded at least ten drops, press the **Stop** button.

12. If you have any "bad data", click on the "Edit Data" button (🖉) on the "Ball Drop" window toolbar. This will create an "editable" copy of the data. You can then delete bad data by selecting the corresponding row and clicking the "Delete Rows" toolbar button (📧).

13. Record the mean, $\bar{t}$, and *sample* standard deviation, s_t. [Again, pay attention to the units.]

$$\bar{t} : \underline{\hspace{3cm}} \text{ ms} \qquad\qquad s_t: \underline{\hspace{3.5cm}} \text{ ms}$$

$\bar{t}$ is your estimate of *t*, the "true" time, and s_t is your estimate of the standard deviation of all possible measurements of *t*.

14. Record *N*, the number of "good" drops (it should be the "Count" in your statistics).

$$N: \underline{\hspace{4cm}}$$

Question 2-1: Determine your best estimate of *g*. Show your work.

NOTE: Don't round off intermediate results when making calculations. Keep full "machine precision" to minimize the effects of round-off errors. Only round off final results and use your error estimate to guide you as to how many digits to keep.

As was the case for the standard error of the mean, our expression for g is not precisely in one of the simple propagation of errors forms and so we must look at it piecemeal. This time we will not work it all out algebraically, but will instead substitute numbers as soon as we can so that we can take a look at their effects on the final probable error.

We see that g is of the form of a constant, 2, times something else, D/t^2, and so we can use Equation (10):

$$\sigma_g = |2|\sigma_{D/t^2}$$

D/t^2 is of the form of a simple product of two quantities (D and t^2) and so from Equation (12) we see:

$$\sigma_{D/t^2}/|D/t^2| = \sqrt{(\sigma_D/D)^2 + (\sigma_{t^2}/t^2)^2}$$

Now we are getting somewhere as we have estimates for σ_D and D. We need only find σ_{t^2}/t^2.

Question 2-2: Calculate σ_D, the error in D, and $\sigma_D/|D|$, the relative error in D. Use appropriate units for σ_D and write $\sigma_D/|D|$ as a percentage. Show your work.

The quantity t^2 is of the form of a quantity raised to a constant power and so we use Equation (13):

$$\sigma_{t^2}/t^2 = |2|\sigma_t/|t|$$

Now we're just about done as we have $\overline{t}$, our estimate of t, We also have what we need to estimate the probable error in our determination of t.

NOTE: We have to be careful with interpreting the notation. Here "σ_t" is the probable error in our estimate of t. It is NOT to be interpreted as the standard deviation of possible such determinations.

So, what is our estimate of σ_t? Since $\overline{t}$ is our estimate of t, then it is the probable error in $\overline{t}$ that we seek. This we know how to do ("probable error in the mean"):

$$\sigma_t \approx \sigma_{\bar{t}} = s_t / \sqrt{N} \qquad\qquad (18)$$

Question 2-3: Calculate your best estimates of σ_t and $\sigma_t/|t|$. Use appropriate units for σ_t and write $\sigma_t/|t|$ as a percentage. Show your work.

Question 2-4: Calculate your best estimates of σ_g and $\sigma_g/|g|$. Use appropriate units for σ_g and write $\sigma_g/|g|$ as a percentage. Show your work.

Question 2-5: Write your experimental determination of g in standard form ($g \pm \sigma_g$). Keep an appropriate number of significant digits (no more than two for σ_g).

Question 2-6: Is your result consistent with the accepted value, 9.809 m/s^2, within experimental errors? Explain and show your work.