Laboratory 2
Data Collection, Probability and Statistics

One of the most important, yet perhaps most neglected aspects of learning science is proper collection of data and learning to accurately and repeatably take measurements. In last week's lab, you performed an experiment to determine whether there was a significant difference between a treatment and a control group of condoms. In today's lab, you will get a better idea of how statistics play a vital role in hypothesis testing.

The exercises you will do today might seem simple, but by the end of this lab, you should have an idea about how to apply statistical tests to data you will be collecting in other labs for the rest of this semester and beyond.

I. Collecting Data

The natural sciences rely heavily on analysis of quantitative data. In such studies, the investigator's goal is to collect biological observations which can be tabulated as numerical facts, also known as **data** (singular = **datum**). Biological research can yield several different types of data, which are tabulated as numerical facts. Important measurements include those which describe the middle values of the population of interest (**mean**, **mode**, and **median**) and those which describe the dispersion around those middle values (**range**, **variance**, and **standard deviation**). Today you will get some experience with all of these. Before you begin, be sure to review Appendix I, and know the differences between attribute and numerical data.

Exercise 1. - Determining the Nature of your Data

You will work in teams of four. For the next ten minutes, cooperate as a group to make a list of five different classes EACH of (1) attribute data, (2) discrete numerical data, and (3) continuous numerical data that you could MEASURE AND COLLECT FROM THE MEMBERS OF YOUR LABORATORY SECTION.

For example, you might list "hair color" as a type of attribute data, "number of hairs per square centimeter of scalp" as discrete numerical data, and "length of hair" as continuous numerical data. Use your imagination, and be sure you are listing each data set in its proper category. List your examples in the spaces provided below.

When all groups have completed their lists, your TA will go around the room and have you read out some of your examples. Your classmates and TA will discuss whether each group's examples are correctly placed in each of the three categories, and why or why not.

Types of data collectible from *Homo sapiens* in BIL 151 laboratory:

Attribute data
1.
2.
3.
4.
5.
Discrete numerical data
1.
2.
3.
4.
5.

Continuous numerical data
1.
2.
3.
4.
5.

II. Mean, Mode and Median

When an investigator collects *numerical* data from a group of subjects, s/he must determine how and with what frequency the data vary. For example, if one wished to study the distribution of shoe size in the human population, one might measure the shoe size of a subset of the human population (say, 30 individuals) and graph the numbers with "shoe size" on the x-axis and "number of individuals" on the y-axis. The resulting figure would show the **frequency distribution** of the data, a representation of how often a particular value occurs in the sample population.

Usually, data measurements are distributed over a range of values. Measures of the tendency of data points to occur near the center of the range include the population **mean** (the average measurement), the **median** (the measurement located at the exact center of the range) and the **mode** (the most common measurement in the range).

Exercise 2. Gathering Data for a Frequency Distribution

Genes do the strangest things. One characteristic under genetic control is the ratio of the length of your index finger to that of your ring finger. (To determine a ratio like this, simply measure the length of the index finger, and divide it by the length of the ring finger.) Each finger should be measured from the top of the knuckle at its base to the very tip of the finger. There are various tools at your lab station that will allow you to measure finger length; you decide which tools will give you the most accurate values.

Work in groups of four. One person should be selected to measure the finger lengths *on the left hand only* of each of the other group members, as well as his/her own finger length. Once you've finished, calculate a mean for each value, as shown in the table provided below.

Once the first member of the group has finished measuring and recording the ratios of each group member, the other members of the group should each take a turn measuring and calculating the ratios. Each group member should measure independently, and not look at the measurements taken by other group members. You may use the same method as your fellow group members, or you may choose a different method. But whatever protocol you select, be sure you use it consistently for all your measurements.

Write down *your own* measured values in the table provided below:

Group member	**Sex (circle)**	**Index finger length (cm)**	**Ring finger length (cm)**	**Index : ring Ratio**
	M F			
	M F			
	M F			
	M F			
Means →	n/a			

Important note! When you calculate a ratio of two similar values, remember that the units of the values--centimeters, in this case--cancel each other out. The ratio is thus a dimensionless quantity.

Does your mean differ from those calculated by your team members? ______________

Explain why or why not: ____

____

____

____

When you have finished the above, select one group member's ratio values and turn them in to your TA. The TA will list ratio values for each class member on the blackboard. From these data, you will calculate the mean ratio for the entire class.

What is the mean index : ring finger ratio for your entire lab section? ______________

Is it different from the one you calculated from only your group members? __________

Explain why or why not: ____

____

____

____

Which mean ratio--your group's or the entire lab's--do you believe is closer to the true parameter for all *Homo sapiens*, and why? ________________________________

____

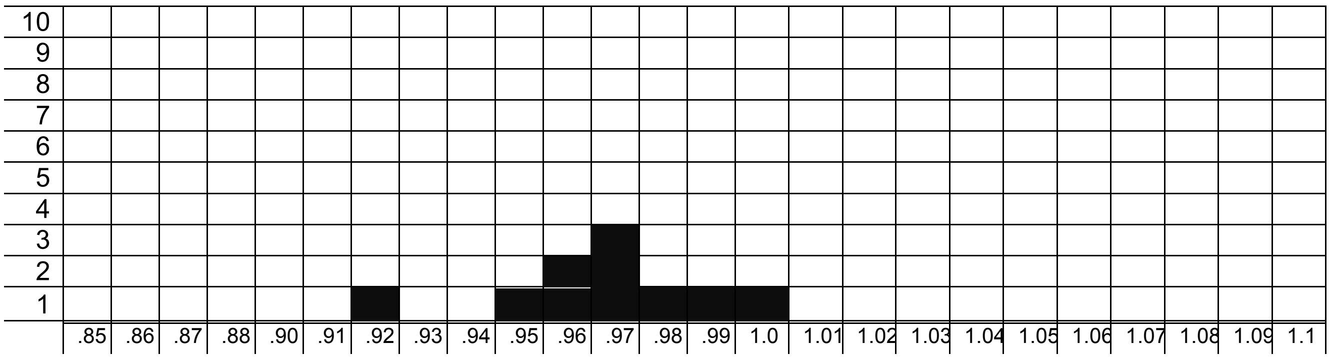
____

____

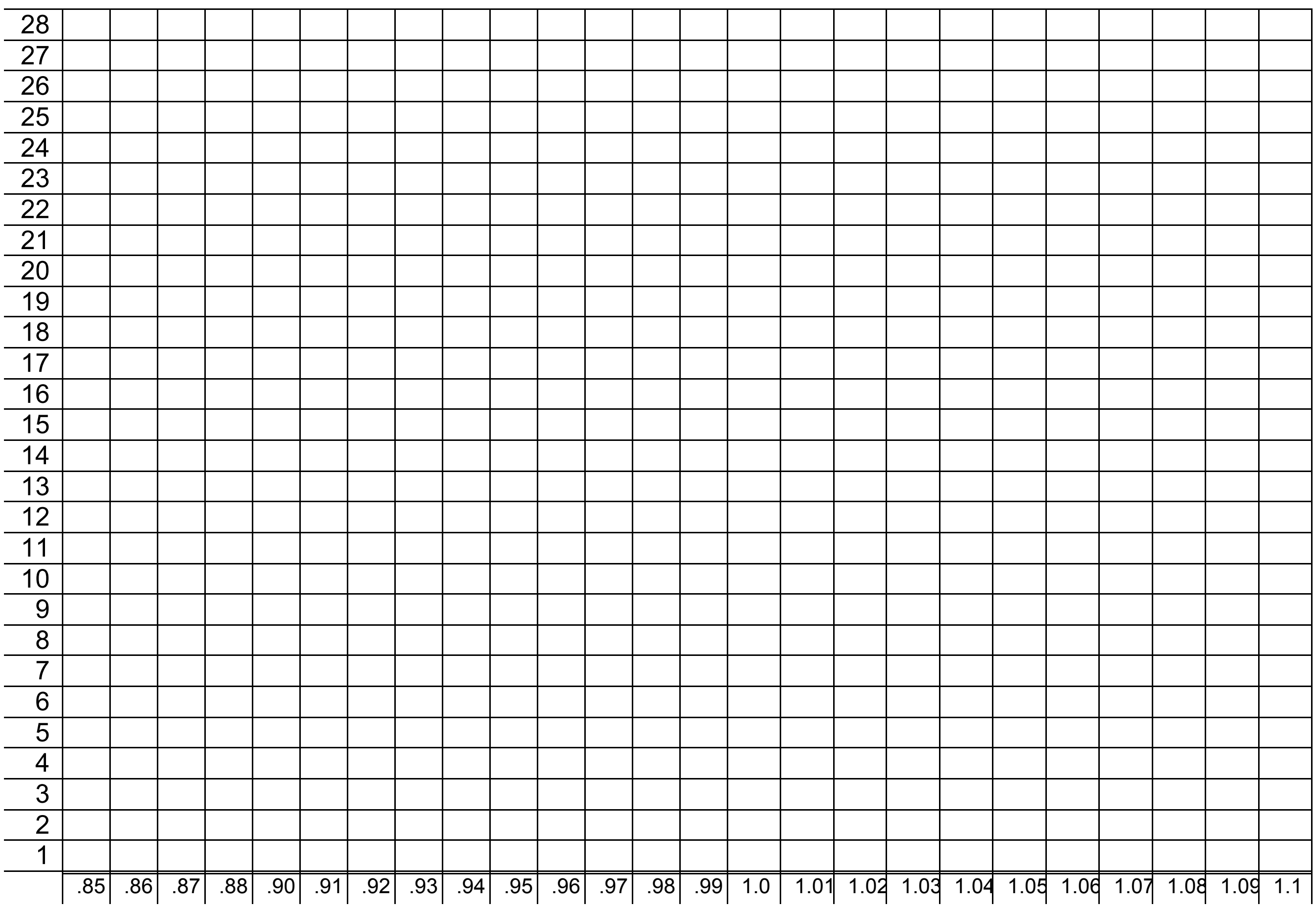
You will complete a histogram of the ratios in the grid provided in Figure 2-1. The x axis shows the range of ratios, and the y-axis is the number of samples with any particular value on the x-axis. For example, if there were 10 students in your lab, and their ratios were as follows:

0.92, 0.95, 0.96, 0.96, 0.98, 0.97, 0.99, 1.01, 0.97, and 0.97

Your histogram would look like this:



In the grid below, enter the index : ring finger ratio <u>for every individual member of your lab section</u>. The number of squares blackened should equal the number of measurements.



<u>Figure 2-1.</u> Frequencies of index : ring finger ratios in male and female students of BIL 151 section ______.

Voila! You have created a **frequency distribution**! This is simply a representation of how often (i.e., how frequently) each value occurs in your sampled group (all your lab colleagues), how the values are distributed across all possible values, and around the mean.

What is the median of the entire lab section's ratios? ______________________

What is the mode? ____

Do these values differ from the median and mode of your four-person team? _________

Explain any difference: ____

____

____

III. Range, Variance and Standard Deviation

As you may recall, when you are studying some measurable aspect of a population (such as the index : ring finger length ratio), the degree of variation around the mean should always be considered. In biological systems, there is almost always a great deal of variation around the mean of any given value. In many biological studies, the estimation of variances is as important, if not more important, than the mean.

Measurements of dispersion around the mean include the **range**, **variance** and **standard deviation**. The simplest of these is the range, which is defined as the highest value minus the lowest value. Unfortunately, the greater the sample size, the greater the range, and because it employs essentially only the two extreme values, a great deal of information about variation between those extremes is lost.

More useful are the variance and standard deviation, which are measures of deviations from the mean. The **variance (s^2)** is calculated as

$$s^2 = \frac{\sum (\bar{x} - x)^2}{n-1}$$

In which $\bar{x}$ is the mean, x is each individual value, and n is the sample size.

The **standard deviation (s)** is the square root of the variance:

$$s = \sqrt{\frac{\sum (\bar{x} - x)^2}{n-1}}$$

Exercise 3. Calculating Range, Variance and Standard Deviation of your Frequency Distribution.

What is the range of index : ring finger ratios for your lab section? _________________

What is the variance? ________________________________

What is the standard deviation? ______________________

(Can you feel your IQ rising? Do not be alarmed. The symptoms will subside as soon as you go home and turn on the TV and watch "American Idol.")

IV. Probability: What is meant by "expected results?"

Probability calculations allow us to define the range of possible results of a series of trials, often in the form of a bell-shaped curve (a frequency distribution) representing the likelihood of each possible result. In other words, these probability calculations allow us to determine **expected results** from a particular series of events, be they a series of coin flips, rolls of the dice, or carefully controlled experimental trials.

When an investigator collects data from a subset of a population, s/he is interested in determining whether some hypothesis about that population is true or false. How does an investigator know whether any "unusual" results are significantly different from what was expected, or whether such variation is due simply to random chance?

Let's use the example of an event that has several possible outcomes, with each of those outcomes having an equal probability of occurring. For example, in any given birth, a single offspring will be genetically either male or female. There is a 1/2 chance that it will be male and a 1/2 chance that it will be female.

The **probability (P)** of any of those possible outcomes can be expressed as:

$$a/n$$

...where

a = the # of occurrences of the event in question and
n = the total number chances a particular result has to occur

For example, because a die (that's the singular of dice) has six sides, there is always a 1/6 chance on any roll that any given number (let's say a "6") will come up.

When the probability of various events are known, scientific investigators can consider their **combined probabilities**: the likelihood of two or more events happening together. Let's illustrate with some actual games of chance.

Exercise 4. Dice-o-Rama!

At your lab station you will find several sets of dice. Work in pairs and use the dice to familiarize yourself with the three general rules of probability that form the basis of the **Sum Rule**, the **Product Rule** and the **Binomial Theorem**.

A. The Sum Rule: Dependent Probabilities

The Sum Rule is used to determine the likelihood that either of two events occurs when one precludes the other. In other words, each time a particular event has a chance to occur, it's an "either/or" choice, such as flipping a coin or rolling a die once.

Every time you flip a coin, it will come up either heads or tails--not both. Every time you roll a die, only one face will show, and no others. Every time a couple has a single child, it will be either genetically male or female--not both (no matter what you might read in the Weekly World News).

In the case of a single roll of a die, the probability of rolling a either a "one" or a "six" on any given roll can be expressed as the *sum* of the probabilities of rolling either a one or a six. For example, the probability of rolling either a one or a six in one roll can be calculated as

$$P = (a/n)_1 + (a/n)_6$$

…in which the subscripts indicate the probability of rolling a "1" or a "6", respectively. Given what you know about the probability of rolling a one or a six with a six-sided die, calculate the likelihood of rolling either a one or a six on a single roll of that die.

Your answer: ________________________________

This means that you *expect* __________ displays of either 1 or 6 in three rolls.

Note:

The probability above is strictly for rolling a 2 followed by a five. If you are not concerned about the order (i.e., "What is the probability of rolling a 2 and a 5 with two rolls of a die?"), then the two rolls are considered separately.

Each of these two outcomes (2 or 5) has a probability of 1/6. You can get a 2 and a 5 by rolling a 2 (1/6 probability) and then a 5 (1/6 probability). Or you can get a 2 and a 5 by rolling a 5 (1/6 probability) and then a 2 (1/6 probability). The total probability for this specific wording would be 2 x (1/6 x 1/6) = 1/18, since each individual event has two chances to happen.

Exercise 4a.

Let's see if it works! Each member of your pair should roll a single die three times in a row and repeat this 5 times. Record results in the grid provided.

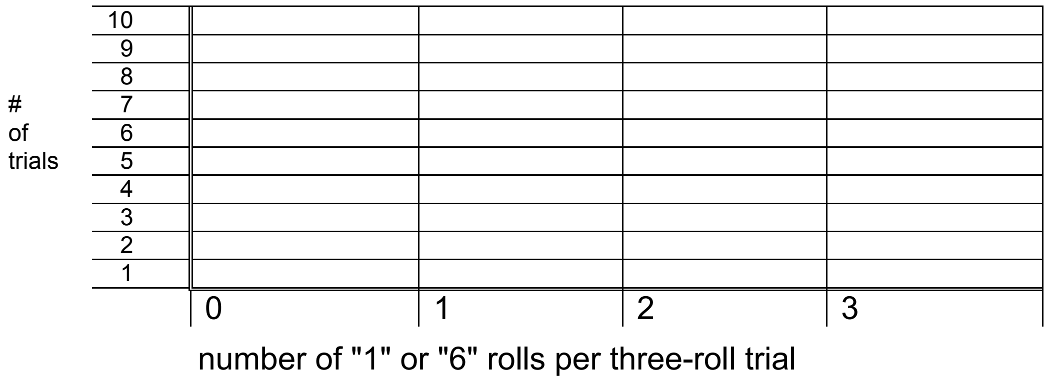
Grid for raw data: Each of the sets below (Set #1 - 5) represents one set of three rolls. To enter your raw data, roll the die three times and circle the number of each roll for each set. For example, if you roll (in order): 3-5-6, then in the lines provided for set #1, you will circle "3" in the first row, "5" in the second row, and "6" in the third row.

Do a total of 5 sets of three rolls per person. At the end, you should have a total of 10 sets of three rolls between you and your lab partner.

You can liken each set of three rolls to a single experimental trial in a series of experiments. Each set of data you get for each trial (3-5-**6** or 4-4-2 or **6-1-6**, etc.) is considered one data point in your set of experimental trials. (In the three examples just given, the boldface type indicates ONE roll of 1 or 6 in the first set, ZERO rolls of 1 or 6 in the second set, and THREE rolls of 1 or 6 in the third set.)

Set #	Partner #1: Circle the result below						Partner #2: Circle the result below					
1	1	2	3	4	5	6	1	2	3	4	5	6
	1	2	3	4	5	6	1	2	3	4	5	6
	1	2	3	4	5	6	1	2	3	4	5	6
2	1	2	3	4	5	6	1	2	3	4	5	6
	1	2	3	4	5	6	1	2	3	4	5	6
	1	2	3	4	5	6	1	2	3	4	5	6
3	1	2	3	4	5	6	1	2	3	4	5	6
	1	2	3	4	5	6	1	2	3	4	5	6
	1	2	3	4	5	6	1	2	3	4	5	6
4	1	2	3	4	5	6	1	2	3	4	5	6
	1	2	3	4	5	6	1	2	3	4	5	6
	1	2	3	4	5	6	1	2	3	4	5	6
5	1	2	3	4	5	6	1	2	3	4	5	6
	1	2	3	4	5	6	1	2	3	4	5	6
	1	2	3	4	5	6	1	2	3	4	5	6

In Figure 2-1 below, create a frequency distribution of the results obtained by you and your lab partner. (Do this the same way you did it for the histogram of finger ratio measurements: darken the squares corresponding to the number of "1" or "6" for as many rolls as you obtained them for each of your ten sets (trials) of three rolls.



<u>Figure 2-2.</u> Frequency distribution of rolls of "1" or "6" in trials of three rolls.

Which of the four columns do you *expect to* have the highest frequency of "1" or "6" rolls? ______________________________

Why? ______________________________

____

Did your actual results confirm what you expected, given the known probability? Explain why or why not: ______________________________

____

When every pair of students has finished filling in a frequency distribution, your TA will ask each team to display its graph. Do all the graphs reflect the same general results? Are there any that differ from the most probable outcome? Explain.

____

____

____

Even though the expected result was one "1" or one "6" each time you rolled the die three times, you probably got a distribution of all possible results in your trials. As you can see, any set of three rolls could yield 0, 1, 2 or 3 incidences of a "1" or a "6". This demonstrates that although a roll of "1" or "6" is <u>*probable*</u> in three rolls, it will not happen every time. Variation is part of nature, and an important aspect of any scientific study.

The larger the sample size, the more closely you will approximate the true population parameter. If you were to pool all the rolls of your entire lab section, what do you think the frequency distribution would look like?

When a subset of a population fails to show the expected result <u>*only because of small sample size*</u> (as you may have seen in your die-rolling trials), the deviation from the expected is said to be due to **random sampling error**. Of greater interest to the scientist, of course, are deviations from the expected that occur <u>*despite*</u> an adequate sample size.

<u>B. The Product Rule: Independent Probabilities</u>

The Product Rule is used to determine the likelihood of two events occurring when the two possible results are independent of each other. Unlike the Sum Rule, which is used to determine "either/or" probabilities, the Product Rule is used to determine "and" probabilities, such as the results of flipping a coin, rolling a single die two times in a row, or having two female children in two births.

The result of each roll of a die is independent of the next. The probability of rolling two particular numbers back-to-back in a specified order can be calculated as the *product* of each number's individual probability.

Let's use the dice again. This time we'll calculate the probability that in rolling one die twice in a row, you will roll a two, followed by a five. The probability is calculated as

$$\mathbf{P = (a/n)_2 \ x \ (a/n)_5}$$

…where the subscripts indicate the probability of the chosen number.

Given what you know about the probability of rolling a two or a five, calculate the likelihood of rolling a two followed by a five in two back-to-back rolls of one die.

Your answer: ________________________________.

You expect to get _______ rolls of 2 followed by 5 in _______rolls of a single die.

NOTE:

The formula above is strictly for calculating the probability of rolling a 2 followed by a five. If you wanted to calculate the probability of simply rolling a 2 and a 5 in two rolls of the die, but do not want to specify the order, then each roll is considered one separate event. You can get a combination of 2 and 5 by rolling first a 2, then a 5. Or you can get the same combination by rolling first a 5, then a 2. Either outcome (2 or 5) has a 1/6 probability, so combining them will yield 1/6 x 1/6 = 1/36. But since there are two ways this can happen, this probability must be doubled, making the total probability 2 x (1/6 x 1/6) = 1/18. Have we lost you yet? You'll thank us next time you're in Vegas.

Exercise 4b.

Work in pairs. Each partner should roll two sets while the other partner records results, then switch roles and repeat. When you're done, you and your partner will have a total of four sets of 36 rolls between you.

On the following two pages, grids to record your number pairings are provided. Note that each of your 36 rolls is paired with the roll *immediately after* it. For example, if your set of rolls is:

1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

Your pairings should be read as:

1/2 2/3 3/4 4/5 5/6 6/1 1/2… and so forth.

On the following four pages are grids for four sets of 36 rolls. Each page has a table with 36 squares for you to enter the number that comes up on each individual roll. When you have finished your four sets, note how many times you got a sequence of 2 followed by 5.

Set #1: Enter the number rolled in each of your 36 attempts:

Number of times you got 2 followed by 5 in set #1: ________________________

Set #2: Enter the number rolled in each of your 36 attempts:

Number of times you got 2 followed by 5 in set #2: ________________________

Set #3: Enter the number rolled in each of your 36 attempts:

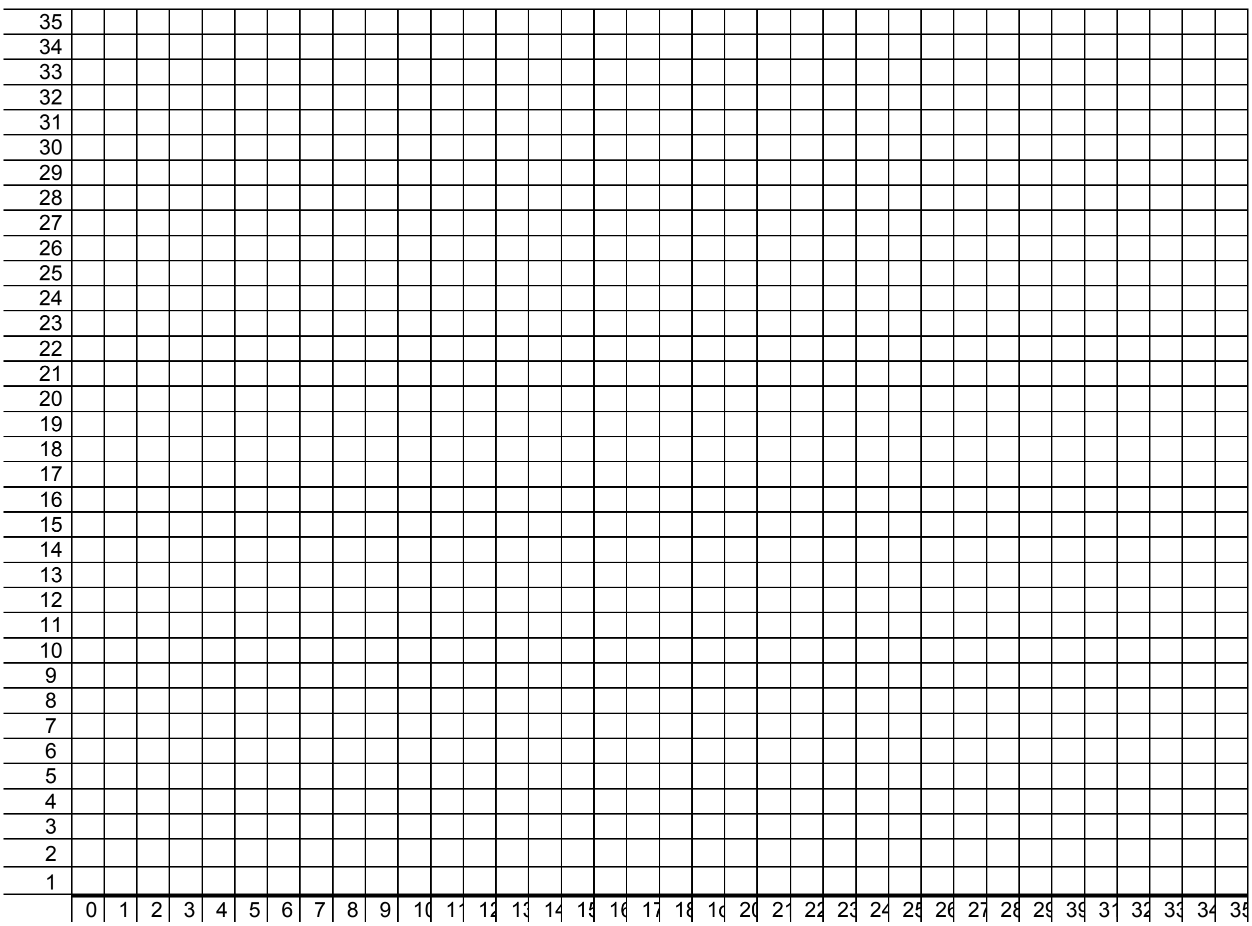
Number of times you rolled 2 followed by 5 in set #3: ______________________

Set #4: Enter the number rolled in each of your 36 attempts:

Number of times you rolled 2 followed by 5 in set #4: ______________________

Once everyone is done, your TA will ask each pair to tell the rest of the class out loud how many rolls of 2 followed by 5 they got in each set. Label the axes of Figure 2-3 appropriately with independent (number of 2/5 sequences per set of 36 rolls) and dependent (frequency of number of 2/5 sequences per set of 36 rolls) variables.

Record these values in Figure 2-3 by blackening the appropriate square as your classmates call out their results. When finished, the figure will be a frequency distribution of how often a roll of 2 followed by 5 occurred in all class trials.



Number of rolls of [2 followed by 5] in sets of 36 individual rolls of a die

Figure 2-3. Frequency distribution of 2/5 sequences in sets of 36 rolls.

Do the pooled data from the entire class approximate what was expected? Explain.

____

____

C. The Binomial Theorem: Multiple Independent Probabilities

The Binomial Theorem is used to generate the probability of two events occurring together over multiple trials when each event has an independent probability of occurring. Let's say we have two alternative events, **X** and **Y**.

The probability of **X** is **p**.
The probability of **Y** is **q**.
n = the number of trials in which either X or Y can occur.
s = the number of times event X occurs in your trials
t = the number of times event Y occurs in your trials

(Recall that, by definition, p + q = 1.0 and s + t = n.)
Let's say that you plan to perform a series of trials in which either result X or result Y can occur. The probability that X will occur s times and Y will occur t times can be calculated with the following equation

$$P = \left(\frac{n!}{s!\ t!}\right) p^s q^t$$

(recall: ! is the symbol for factorial. For example, 10! = 1x2x3x4x5x6x7x8x9x10)

Exercise 4c.

Work in pairs. At your station you will find a container full of colored, flat marbles. Each container is an egg case laid by a female member of a rare species known as *Silicus vitreosus*. (Humor us.) The eggs, which bear a striking resemblance to flat marbles, come in two colors. The blue eggs will hatch out as males, and the pink eggs will hatch out as females. (PLEASE DON'T LOSE ANY MARBLES OR ADD ANY FROM OTHER CONTAINERS!)

Because sex in *Silicus vitreosus* is determined by X and Y chromosomes, with XX being female and XY being male, there is a 50% probability (all other things being the same) that any given egg will be male or female.

Let's consider the laying of a female egg as event **X**, and the laying of a male egg as event **Y**. This means that in any given clutch of eggs, the number of females is s, and the number of males is t. There is a 0.5 chance that an egg will be female, and a 0.5 chance that it will be male.

Let's hatch some eggs! While one partner holds the container, the other should close his/her eyes and randomly draw out 10 marbles. Okay, it's safe to open your eyes. How many of each color did you get? What are the sexes of the babies?

\# of females (Event **X**): ________________________(this is your value for **s**)
\# of males (Event **Y**): ________________________(this is your value for **t**)

Plug your observed values into the binomial equation and calculate P.

What was the probability of your observed result? ________________________

Which combination of colors in a 10-egg clutch do you think would have the highest probability of occurring, and what IS that probability? ______________________________
____

As before, there is a particular probability associated with every combination of colors in any given clutch. Also as before, the occurrence of each color combination could be plotted along a frequency distribution similar to the ones you did for the finger ratios and dice rolls. We will not create another frequency distribution for this example. However, you should feel confident that you could create one, should the need arise.
If so, what would be the appropriate units of the x axis? ___________________________

What about the y-axis? ____

V. Hypotheses and Statistical Tests

The Scientific Process begins with three important steps:

1. The observation of some phenomenon that elicits a question/poses a problem.
2. The formulation of competing, testable hypotheses about that phenomenon
3. The prediction of all possible outcomes of experiments designed to test each hypotheses.

For example, as you wander in a field of wildflowers, you notice that individuals of a species of yellow poppy have two distinct morphologies, one with spines and one without. You also noticed that the individuals with spines were far more numerous than the smooth individuals. You might wonder whether there was a reason for the difference in their numbers, and ask whether the spiny individuals were better able to deter herbivores with their spines than the smooth individuals. There's your question.

The next step is to formulate a testable null hypothesis, such as, "There is no difference in herbivore damage between the spiny and smooth individuals." Your alternative hypothesis would be that there *is* a difference. In this case, the latter is your prediction as well as your hypothesis of interest.

Before you continue, review Appendix I and be confident you know the meanings of **inductive** and **deductive reasoning, null** and **alternative hypotheses, one-tailed** and **two-tailed hypotheses,** and the precise definitions of **hypothesis** and **theory.**

Statistical Tests and Probability Distributions

Probability calculations similar to those you did with the dice and marbles form the basis of one of the scientist's most important tools: the **statistical test**. Once data have been collected, it's not enough to merely "eyeball" them and say, "Eeyup. This is different from what we expected! Something weird is going on here!"

Instead, investigators use statistics calculated from their data sets to determine the likelihood that their results are significantly different from the expected results. Over the decades, many different **probability distributions** have been devised by mathematicians, each one appropriate for different types of data.

Enough statistical tests and their associated probability distributions have been invented to fill many textbooks. Some of these, such as the Chi-square test, the Student t-test, the Analysis of Variance (ANOVA), the Mann-Whitney U test and the Fisher's exact test may sound familiar to you. The specific probability distribution and statistical test appropriate in a given situation depend upon the type of data collected.

One oft-utilized probability distribution is the **t-distribution**, used to determine whether there is a significant difference between the (continuous numerical) means of two groups under study. To make a very long and complex story short, an investigator can use the mean, variance, and standard deviation of his/her data sets to calculate a **t-statistic**. Every possible t-statistic is linked to a certain probability that the data used to calculate it differ because of some factor other than random chance.

Exercise 5. The Student t-test: A tool for determining whether there is a significant difference between two means

The Student t-test can be used to determine whether a difference between two means is significant. These means may calculated from observations that are either **paired** (as when individuals in a single group are subjected to "before and after" measurements, and data points are paired for each tested individual) or **independent** (as when individuals in two *similar* sample populations are measured, but each individual in each sample population is measured only once). Slightly different calculations of the t-statistic must be used in each case. In last week's lab, you used the paired sample t-test. Today, you will learn a slightly different formula, the independent sample t-test.

Work in groups of four. Within each group, one pair of students will measure 30 marbles and calculate their mean volume with one method, and the other pair will measure *the same 30 marbles*, but use a different method. There is more than one way to calculate the volume of solid objects, and because we know you're tired, we'll clue you in on a couple. But feel free to devise your own method if you're full of caffeine and feeling creative!

Method 1. Geometric calculation.

Glass marbles are ostensibly spherical. If you can measure the diameter of the marble, you can calculate its volume with the old standard formula

$$\mathbf{4/3\ (\pi\ r^3)}$$

…where π = 3.14 and **r** is the radius of the sphere. Millimeter rulers and calipers are available for you to use, and you may decide which you prefer. Report your results in cubic centimeters (cm^3, or cc).

Method 2. Volume displacement.

Glassware appropriate for measuring volume also is available at each lab table. By filling a graduated cylinder with a known volume of water, adding a marble and then determining the change in volume…well, it doesn't get much simpler. There are several sizes and types of glassware. We'll leave it up to you to decide the details of how to obtain the most accurate measurements. Report your results in cm^3. (Note that one milliliter (ml) is equal to one cubic centimeter (cc).)

Before you begin your measurements, answer the following:

1. What is your null hypothesis (there may be more than one reasonable one.)?

____

____

2. What is your alternative hypothesis?

____

____

3. Do you predict that the two different measurement methods will yield the same volume for each of your 30 marbles? ___________________________.

If you were measuring two sets of 30 similar marbles with the two methods, the means of your two sample populations would be independent of one another, as a single measurement is made on each individual. However, you will essentially be performing "before" and "after" measurements of a single marble with two different methods. This means that the two values you attain will not be independent of one another, and are said to be **paired**.

In a paired sample t-test, the separate means of two sample populations is not measured. Instead, you calculate the difference between your first measurement and your second measurement on the same individual, and use this to calculate your t-statistic. The difference between your first and second measurements is represented

as **d**, and the mean of all your individual differences as $\bar{\mathbf{d}}$.

The standard deviation and variance are calculated as you did them before, but substituting d and d for x and x in the equations you used in the first exercise. Thus, to calculate the **variance** of your paired samples, you will use the following equation:

$$s_d^2 = \frac{\sum(\bar{d} - d)^2}{n-1}$$

To calculate the **standard deviation**, take the square root of the variance.

Measure the marbles one at a time with Method #1, enter the volume in Column 2 of the grid provided, and then pass the same marble to your partner pair, who will measure it with Method #2, and enter the volume in Column 3.

1. Calculate the **mean** difference in volume **($\bar{d}$)** from the data in Column 4: __________
2. Calculate the **variance** of the differences in volume (s_d^2). ______________________
3. Calculate the **standard deviation** (square root of (s_d^2)): _________________

You're now almost ready to plug these values into the t-test for paired samples. The last thing you must do is to determine the number of independent quantities in your system, or **degrees of freedom**.

4. The degrees of freedom determine the significance level tied to every possible value of a statistic (such as the t-statistic). The degrees of freedom is the number of data points that are free to vary without changing the test statistic, and this changes depending on the type of statistic you are calculating. For the paired sample t-test,

$$\mathbf{df = (n - 1)}$$

...where **n** is the number of independent quantities in your system. For *paired* samples, n = the number of pairs in your system. In this case, n = 30.

Your team of four should now calculate a t-statistic for the difference between your two means, and use this to determine whether the difference between them is significant. Use the following equation for the paired sample t-test:

$$t = \frac{(\bar{d} - 0)\sqrt{n}}{s_d}$$

...where **n** is the number of pairs.

What is the value of your t-statistic? ___________________________________

Marble #	Volume in cc (Method I)	Volume in cc (Method 2)	Difference in Volume (d) |column 2 - column 3|
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
TOTAL			

VI. Probability and significance

The term "significant" is often used in every day conversation, yet few people know the statistical meaning of the word. In scientific endeavors, **significance** has a highly specific and important definition. Every time you read the word "significant" in this book, know that we refer to the following scientifically accepted standard:

The difference between an observed and expected result is said to be **statistically *significant*** if and only if:

> Under the assumption that there is no true difference, the probability that the observed difference would be <u>at least as large</u> as that actually seen is less than or equal to 5% (0.05).
>
> Conversely, under the assumption that there is no true difference, the probability that the observed difference would be <u>smaller</u> than that actually seen is greater than 95% (0.95).

Once an investigator has calculated a t-statistic, s/he must be able to draw conclusions from it. How does one determine whether deviations from the expected (null hypothesis) are significant?

As mentioned previously, depending upon the degrees of freedom, there is a specific probability value linked to every possible value of the t-statistic. Thus, *every possible t statistic indicates a certain probability that the data used to calculate it vary significantly from the expected.* Mathematical computation of probability distributions for each statistical test is extremely complex. Fortunately, tables of probability (P) values at various degrees of freedom for each statistical test have been created for us by noble mathematicians (or actually, noble computer jockeys typing in the forumlae the mathematicians have devised).

<u>Exercise 6. Determining the significance level of your t-statistic</u>

Now that you have calculated a t-statistic for your two sets of marble measurements, you must try to interpret what this statistic tells you about the difference between the two methods. Is the difference significant, suggesting that there is something other than chance causing the variation in volume yielded by each method? The answer lies in the table of critical values for the t-statistic, part of which is illustrated in Table 2-2.

1. Locate the appropriate degrees of freedom in the far left column.
2. Look across the df row to find a t value closest to the one you obtained.
3. If the exact value does not appear on the table, note the two t values which most closely border your value, or the single value that is either greater or smaller than your value.
4. Find the P value(s) that correspond to your bordering values, and enter them in the appropriate space(s) below:

__________ > P > _________

If the P value associated with your t-statistic is less than (or equal to) 0.05, there is less than 5% probability that chance deviations could have resulted in a mean volume difference as great or greater than that actually observed. Thus there is a 95% probability that chance deviations could have resulted in a mean volume difference smaller than that actually observed. If this is the case, you must reject your null hypothesis and accept your alternative hypothesis.

(Alternatively, for a "quick and dirty" P value, just find the t-statistic associated with P = 0.05 at your df. If the calculated t-statistic is larger than the one from the table, reject your null-hypothesis.)

Judging from your P value, should you accept or reject your null hypothesis? _________

Briefly explain WHY you believe your statistic indicated that you should reject or fail to reject your null hypothesis.

Do you believe your results to be an accurate reflection of true population values? Support your contention, either way.

Table 2-2. Table of critical values for the two-sample t-test. The P levels (0.05) indicating rejection of the null hypothesis are shown in bold for both one-tailed and two-tailed hypotheses. (From Pearson and Hartley in *Statistics in Medicine* by T. Colton, 1974. Little, Brown and Co., Inc. publishers.)

2-tail -->	0.10	**0.05**	0.02	0.01	0.001
1-tail -->	**0.05**	0.02	0.01	0.005	0.0005
df					
1	6.314	12.706	31.821	63.657	636.619
2	2.920	4.303	6.965	9.925	31.598
3	2.353	3.182	4.541	5.841	12.941
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.859
6	1.934	2.447	3.143	3.707	5.959
7	1.895	2.365	2.998	3.499	5.405
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.539	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.767
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.052	2.473	2.771	3.690
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.659
30	1.697	2.042	2.457	2.750	3.646

VII. Erroneous Conclusions

There is always the secret dread in the heart of any scientist that his/her data were not collected properly, that the sample size was too small, or there was some other confounding problem that caused the data to yield an answer that does not reflect reality. In such cases, the P value associated with your calculated statistic might wrongly lead you to reject the null hypothesis if it is actually true, or perhaps fail to reject the null hypothesis if it is false.

Type I Error: The mistaken rejection of a null hypothesis that is actually true.
Type II error: The mistaken failure to reject a null hypothesis that is false.

As you might intuitively know, a type I error can be more damaging than a type II error. For example, if a pharmaceutical company mistakenly is led to believe that a new drug causes a significant improvement in a medical condition, when it actually does not, it could cost not only tremendous amounts of money, but even lives.

Type II errors are most common when the sample size is too small, and are thus more easily corrected than type I errors.

An experiment should never be considered 100% foolproof. Therefore, In science--as in all endeavors--honesty is of paramount importance. Experimental results do not always confirm predictions. Data manipulation or failure to accurately report results constitute intellectual dishonesty. Science is the pursuit of Truth and Reality, not the pursuit of supporting pet hypotheses that just might be wrong.