Explaining Patterns: Analysis of Simple Experiments

Preparation for Laboratory: Web tutorial 1 (http://biology.kenyon.edu - go to Biology 109 Resources)

Last week you examined the variability in a group of organisms and learned about ways to summarize that variability using the mean, standard deviation and sample size. You also generated a hypothesis based on an observation, made a prediction and collected data that would allow you to test your prediction. This week we will examine your experiments and apply statistical tests to help you determine if the data you collected supports or disproves your hypothesis.

I. Exploring the Variation in a Data Set

Organize the Data.

During this laboratory you will be working with data collected using two basic types of design. If you used a before/after design, get data from a group that used a design involving two distinct groups and go to the computer. Open the file containing your data. Enter the data from the other group into two new minitab columns. Label each column so you know the source of the data. Use the Calc menu to calculate the mean and standard deviation for each column.

Reorganize the data. Combine the data from your experiment into a single column. Then combine the data from the other experiment into a single column. Instructions for doing this are detailed below:

> **Manip** →**Stack/Unstack** → **Stack Columns** →**Stack the following columns** (*enter the two columns of data by LDC on them)* →**Store the stacked data in** (*enter a name for the column of stacked data*)→**Store the subscripts in** (*enter a name for the code column*). This subscript column will have two numbers one will code for the data that came from the first column you added to the stack and the other number will code for the data that was in the second column. Check to make sure you are comfortable with the way the data were stacked.

Measures of Variability

There are two types of error that contribute to the variation that we see in a set of measurements. These are (1) **measurement** or **experimental error** and (2) **natural variation** that exists in any living or non-living system. Measurement error can be affected by the consistency in the way a measurement was taken, the care taken by the investigator in obtaining those measurements, the amount of experience that the investigator has, and equipment design and condition. For the most part, measurement error can be controlled by the investigator. When planning experiments, you should think through your experimental design very carefully in order to reduce as much of the measurement error as possible. The second type of error, natural variation, is something that is not under the control of the investigator. The fact that individual variation is referred to as a source of error does not imply that natural variation is "bad". It is simply a property of living organisms. Individuals in any biological population are not identical.

We will be using three different measurements to quantify the variation or the spread around the mean. These are the **standard deviation**, the **standard error**, and the **95% confidence interval**.

Standard Deviation: estimates the variation in a larger population based on a sample

You are already familiar with the standard deviation. But to review the calculation for the standard deviation is:

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

where $X_i - \bar{X}$ is the difference of each value from the mean and n the total number of measurements.

The standard deviation is an estimate of variability within a population. If the parameter that you measured was normally distributed, then the standard deviation can be used to make predictions about the numbers of value that fall within a specific range around the mean.

Standard Error of the Mean: index of the variation in the sample's estimate of the population mean.

When we take measurements from a small sample of a larger population and calculate an average value from those measurements, it is often useful to know how confident we are that the mean that we calculate is representative of the whole population. Obviously, as we sample more individuals from the population, we become more confident that the mean that we calculate is close to the true mean for the population. The **standard error** is a measurement of our confidence in our mean value that reflects the amount of variation in the population (as measured by the standard deviation) and the number of measurements that were taken. To calculate the standard error, divide the standard deviation (*s*) by the square root of the sample size (n):

$$\mathbf{S.E.} = \frac{\mathbf{s}}{\sqrt{\mathbf{n}}}$$

95% Confidence Interval: range in which the mean should lie 95% of the time with repeated sampling.

Another useful measurement of variation is the 95% confidence interval. This interval is a range of numbers in which the true mean for the entire population is likely to fall. In other words, we can say with a certainty of 95% that the true mean of the population that we sampled falls between the upper and lower values specified by the 95% confidence interval. To calculate the 95% confidence interval, multiply the standard error by the critical value shown in the table below. Note that the critical value is dependent upon your sample size. In this case, degrees of freedom, df = sample size -1.

95% C.I. = S.E. * critical value (where df = n-1)

Data can be summarized as means $\pm$ 95% C.I. with the sample size in parentheses, e.g. 69.0 $\pm$ 12.6 (n=6)

These numbers tell you that you are 95% certain that if you were able to measure every individual in the population the true mean would fall between 56.4 and 81.6 (69 - 12.6 = 56.4 and 69 + 12.6 = 81.6).

Note that as the sample size increases the critical value decreases. Large samples increase the confidence we have in our estimate of the true population mean. When the sample size is 61 (df=60) the 95% CI is two times the standard error of the mean.

Table 1. Critical values for calculating 95% confidence intervals. Values are from Rohlf and Sokal (1981). **In this table if the calculated value is > the critical value then $p < 0.05$ (p = the probability a statistic of the calculated magnitude could be due to chance).**

df	Crit. Value	df	Crit. Value	df	Crit. Value
1	12.706	12	2.179	23	2.069
2	4.303	13	2.160	24	2.064
3	3.182	14	2.145	25	2.060
4	2.776	15	2.131	26	2.056
5	2.571	16	2.120	27	2.052
6	2.447	17	2.110	28	2.048
7	2.365	18	2.101	29	2.045
8	2.306	19	2.093	30	2.042
9	2.262	20	2.086	40	2.021
10	2.228	21	2.080	**60	2.000
11	2.201	22	2.074	120	1.980
				∞	1.960

Use Minitab to calculate the mean, standard deviation, standard error, and 95% confidence interval for the unstacked data (refer to the statistics guide on page 27). You will be working with the four columns of unstacked data. Then following the instructions given in the statistics guide make an interval plot of the data set for each experiment. Examine the error bars and note the amount of overlap. Do you think the two data sets are significantly different?

II. Examination of experimental design

Exchange your methods sheet with someone who was not in your group. Is the question/hypothesis clearly stated? Read the description of what they did and add your comments. Was past tense used? Was the level of detail appropriate? Could you repeat the experiment? If you have questions about their design, please note them.

Choose a member of your group to present your experimental design and the hypothesis you were testing.

III. Using Statistics to Evaluate Data

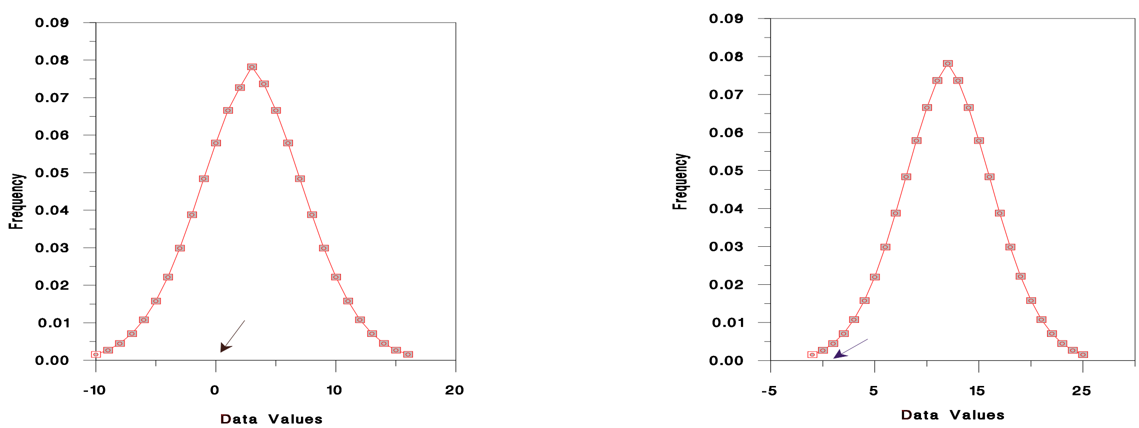
A. Comparing a mean value to an expected value

One way to examine the effect of a treatment is to determine if the treatment causes a change from a baseline condition. For example you might expect that exercise would cause an increase in heartrate. Furthermore, we would expect that the increase would be higher than the small fluctuations we would see in the resting heartrates from one moment to the next. But how much higher do they have to be in order for us to be confident that we are seeing an effect of exercise? 1 beat/min. or 25 beats/min? Would an average increase of 10 beats/min. be enough to convincingly say that exercise is having an effect?

If exercise is having no effect on heartrates, then some individuals heartrates might go up and some might go down, but on average, there should be little change. Another way of thinking about this is to say that if exercise has no effect on heartrate, then the average differences in the heartrates before and after exercise will be zero. Of course, it is very unlikely that the average will actually be zero if exercise has no effect. Even if we took two measurements of the resting heartrates for the class, we would not expect the mean to be exactly the same. They would change a little bit over even a short period of time. If exercise has no effect on heartrates then the average difference before and after exercise would likely be close to zero, but not exactly zero. What do we mean by "close to zero"? How close is close enough?

One way to address this question is to ask whether the "expected value", zero, is within the "normal" range of numbers spanned by the heartrate changes (Figure 1). Just as we could look at individual heartrates to see if they fall within the normal heartbeat range, we can see if the value "zero" falls within the normal range of changes in heartrates before and after exercise. If zero is within that range, we can say the differences that we see are not significantly different from zero. If zero falls outside of the normal range of heartrate changes, then we can be very sure (although not 100% sure) that the mean change in heartrates before and after exercise is very likely to be different from zero. But again, we face the question, what is the "normal" range of values? When we look at individual heartrates we question whether "normal" meant the top and bottom 5% of the population or the top and bottom 1%? How do we decide which value to use? **Scientists have decided that values falling in the outer 5% of a distribution should be considered extreme values**. Five percent is an arbitrary value. There are no statistical or biological reasons for picking 5%. It is simply a convention.

In deciding whether a mean is different from an expected value, we should also take the sample size into account. Clearly the more individuals we sample, the more confidence we have that our measured mean is close to the actual mean for a population. Because we want to account for sample size, we will be using standard error rather than a standard deviation to test our hypothesis.



A. B.

Figure 1. A one sample t-test or paired t-test is used to evaluate the question: Is the expected value (in this case 0) within the "normal" range of values for a group of numbers (A), or does it fall outside the "normal" range. In this case, we are defining the "normal" range as the 95% of the values that would be closest to the mean value.

Now, let's formulate a hypothesis and test it with data from one of the before/after experiments. The hypothesis that we will test is called our **null hypothesis**:

H_O: The change in heartrate due to exercise equals zero.
(or exercise has no effect on heartrate)

If we find our null hypothesis to be false, then we are forced to accept the **alternative hypothesis**:

H_A: The change in heartrate due to exercise is not equal to zero.
(or exercise affects heartrates)

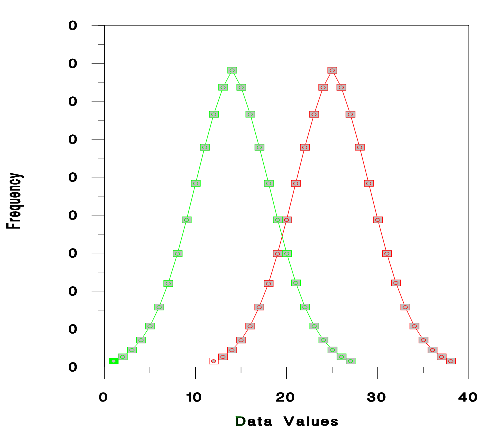
To test the hypothesis, we must first calculate the t-statistic:

$$t = \frac{|\bar{X} - \mu|}{SE}$$ **where μ = the expected value** and **df = n -1**

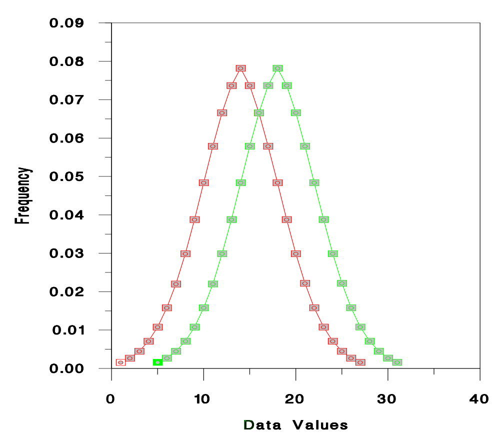
You can use Minitab to do a one sample t-test with the expected value equal to zero, however you must first calculate the difference between the two data sets. In Minitab a paired t-test will automatically subtract one paired value from the other, calculate the mean difference, and calculate the t statistic in a single step. **Use Minitab to do a Paired t-test** to examine the results of the before/after experiment. Examine the printout in the **session window.** What is the value for t? The help you understand the significance of the t statistic. Consider that it is based on a frequency distribution. When the value of t lies near the fringe of the distribution it is unlikely that that value could be obtained by chance alone. Compare the value for t to the critical values in Table 1, where df = n-1. If t is greater than the critical value then we can say that the mean change is "statistically different from zero (t-test: $p<0.05$)". The treatment had a significant effect If we find that t is less than the critical value then we say that the change "is not significantly different from zero (t-test: $p>0.05$)". The p value indicates the level of confidence that we have in making this statement. In this case, we are saying that there is a 5% chance we are making a false statement.

B. Comparing two independent means

Another method for evaluating the outcome of an experiment is to compare **two independent means** with each other. Assume we want to know whether male and female organisms respond differently to a particular treatment. Again, we must deal with the question of how much of a difference is needed to convince ourselves that males and females respond differently (Figure 2).



A.



B.

Figure 2. A two-sample t-test addresses the question: Do the distributions of two samples overlap so little that we are convinced that there are real differences between the two sets of data (A), or is the overlap large enough to have been caused by random chance alone (B)?

The **two-sample t-test** and a related group of **analysis of variance tests** examine the amount of variation in two or more groups to see if the separate groups are less variable than the combined group. Suppose you want to know if males are on average taller than females. You measure the height of two equal-sized groups of males and females. You find that there is a small difference in the average heights. You also determine that there is less variation in the male heights alone and female heights alone than there is if you combine males and females into a single group. Does this mean that sex has a significant effect on height? Not necessarily. It is possible that the differences you see are due to random chance. In order words, if you randomly split the population in half, you could, just by chance, have one group that has a few more tall individuals and another group that has a few more short individuals. These differences could have nothing to do with sex. In order to decide if sex is affecting height, you need to know the probability that the difference in height was due to random chance. Suppose you figure out there was a 75% probability that the differences were due to random chance. Could you be confident that the difference in height was due to sex and not chance? What if the probability was only 50%? or 10%? or 1%? At what point can you confidently say the gender affects height? Again, scientists have decided that the cut-off point is 5%. If the differences in two or more groups are large enough that there is <5% probability that the differences are due to random chance, then the differences are said to be "statistically significant". In other words, if you say that two means are significantly different, there is a 5% or lower chance that the difference in those means is due to random chance, so we conclude that the difference is real.

To determine the statistical significance of a difference between two means, calculate a t-statistic:

$$t = \frac{|(\bar{X}_1 - \bar{X}_2)|}{\sqrt{(s_1^2/n_1 + s_2^2/n_2)}}$$

variance = s^2

average variance = s^2/n

where: $\bar{X}_1$ = mean of sample 1; $\bar{X}_2$ = mean of sample 2

s_1 = standard deviation of sample 1; s_2 = standard deviation of sample 2

n_1 = sample size of sample 1; n_2 = sample size of sample 2

Use Minitab to do a 2 Sample t-test on the independent data you collected. Examine the result in the session window and make sure you are able to interpret it. What is the p value? Is the difference in the means significant? If the means are different, you must see if the difference is in the direction that you expected and look at how large the difference is.

In class assignment: (*This may be modified by your instructor*)

Use the two data sets you analyzed to write a results section. Include at least one figure and a brief text. In both the figure and text support your statements by parenthetically including the results of the statistical tests that you performed. **Submit your results before leaving class.**

Refer to "Introduction to Excel Graphs" for help making the figure.

See the following page for your homework assignment! Do not put off doing it as being rained out is not an option.