

LAB 22 – Artificial Selection

Objectives:

- To investigate natural selection as a major mechanism of evolution.
- To investigate how natural selection acts on phenotypic variations in populations.
- To convert a data set from a table of numbers that reflects a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change

Introduction:

Evolution is a process that has existed throughout the history of life on Earth. One of the key driving forces of evolution is natural selection, which is differential reproduction in a population – some organisms in a population may reproduce more than others and leave more viable offspring in the next population or generation. Differential reproduction results in a population with a genetic makeup that is different from that of the previous population. Thus, populations may change over time. This process of change is evolution. With natural selection, environmental factors play a key role in determining which organisms reproduce and how many of their offspring survive. In artificial selection, humans determine which organisms reproduce, allowing some individuals to reproduce more than others.

In *On the Origin of Species*, Charles Darwin used artificial selection – the kind of selection that is used to develop domestic breeds of animals and plants – as a way to understand and explain natural selection. Like natural selection, artificial selection requires variation in the population under selection. For selection to work, the variations must be inheritable. To conduct artificial selection, humans decide on a specific trait of a plant or animal to enhance or diminish and then select which individuals with that desired trait will breed, producing the next generation and the next population.

You and your classmates will perform multiple rounds of artificial selection on a population of Wisconsin Fast Plants®. First, you will identify a few traits that vary in the population and that you can quantify easily. You will then perform artificial selection by cross-pollinating only selected plants. You'll collect the seeds, plant them, and then sample the next-generation population and see if it is different from the previous one.

Procedure:**Preparing the Bottle Growing System**

1. Prepare growing containers. We will follow a modified direction set from the Wisconsin Fast Plant website (www.fastplants.org/grow.lighting.bottle.php). The number of seeds we will use depends on the size container we have. Follow **Table 1** for the bottle size we have.

Table 1: Bottle size vs. number of plants.

Size of Bottle	Recommended Number of Plants	Number of Fertilizer Pellets
16.9 ounces (0.5 L)	4	8
24-33 ounces (~0.75 – 1 L)	6	12
2 L	8	16

2. Some of our growing containers have already been built by previous classes. However, if we do not have enough, we will have to construct new ones. The steps will resemble those found in **Figure 1** and are listed below.

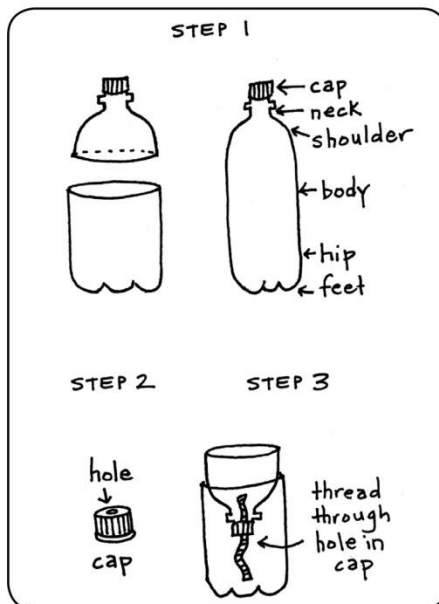


Figure 1: Building a bottle growing system for Wisconsin Fast Plants.

3. For whatever size bottle you are using, remove the label. Cut the bottle *at least* 1 cm below the shoulder (the curve at the top). If you cut it any higher up, the top section will fall into the reservoir of water and ruin the plants.
4. Poke or drill or melt a 1 cm hole in the bottle cap.
5. Thread a thoroughly wet wick strip (nylon strings) through bottle top, invert the top, and set into base. The wicking strip should reach the bottom of the reservoir. To be effective, the

wick should run up into the soil, not be plastered along a side of the bottle and not to extend above the soil.

6. Remove the top section and fill the reservoir with water.
7. Prepare the potting mix by moistening ahead of time. (It should be moist enough to hold a loose ball if compressed.) Add potting mix until the container is $\frac{1}{4}$ full.
8. Scatter the appropriate number of fertilizer pellets (see **Table 1**) for your size container on the surface of the potting mix. Then fill the container with additional potting mix until the bottle is filled with mix.
9. Lightly water the surface of the potting mix. Water should be dripping from the wick. The surface of the potting mix may drop 2 – 4 cm after watering.
10. Place the appropriate number of seeds, evenly spaced, on top of the potting mix in a circle near the edge of the bottle as shown in **Figure 2**.

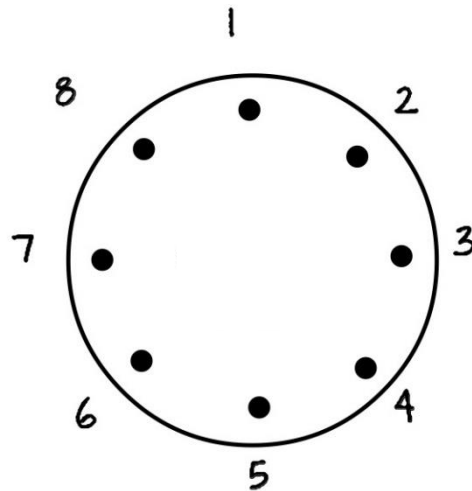


Figure 2: Location of seeds in Bottle Growing System.

11. Cover the seeds and soil with a layer of vermiculite approximately 1 cm deep. Lightly water to moisten the vermiculite.
12. Place the top of the bottle on its water reservoir, and tape the shoulder to the lower compartment to help keep them together.

Observations of the Fast Plants

13. Each day, check your plants and make sure that the reservoirs are full, especially on Fridays. These reservoirs have enough volume to last a three-day weekend for small plants. If algae growth is evident, refresh the water in the reservoir. As your plants grow, record your observations daily. Also try to identify a trait that you could measure or observe reliably. Look for variation in the plants you are growing and describe any you see in your notebook. Observe your classmates' plants as well. Are there also variations in their plants?
14. When the plants are about 7 to 12 days old (see **Figure 3**), we will start to look at a variable trait for artificial selection. We will want to use a trait that varies between plants in a single bottle and also between containers. This trait will not be a "Yes or No" type, but one that can be scored on a continuum. For this we will look at trichomes – these appear to be small hairs on the stems of these plants.



Figure 3: 7 to 12 day old Wisconsin Fast Plants. Trichomes are indicated by the arrow.

15. In your lab notebook, compile a list of all other possible traits your class identified. We might use these in one of our later rounds of selection.
16. Score each of your plants for the number of trichomes. You might need a magnifier to do this accurately. Construct a data table for all of your (and the classes) data.
17. Create a histogram that shows the frequency of trichome length. Calculate appropriate descriptive statistics for the class data for the first generation: sample size, mean, median, range, standard deviation, etc.
18. You are now ready to make selection decisions. Directional selection tends to move the variability of a trait in one direction or the other (increase or decrease the trait in the next population). Find the top (or bottom) 10% of plants with that trait in the entire class's population (e.g., out of a population of 150 plants, the 15 hairiest plants), and mark any that are in your plant bottle container. Using scissors, cut off the tops of the remaining plants in your container (those not in the top 10%).
19. Just as you did in Step 17, construct a new data table, histogram, and calculate descriptive statistics for the selected population of plants. Record the data in your lab notebook. Once you have finished, isolate these selected plants from the rest of the population. Move the bottles of selected plants to another light system so that the plants can finish out their life cycle in isolation. This population will serve as the parents for a new generation.
20. On about day 14–16, when several flowers are present on each of the selected plants, cross-pollinate the selected plants with a single bee stick or pollinating device. Fast Plants are self-incompatible — each plant must be fertilized by pollen from another plant. Collect and distribute pollen from every flower on every plant in the selected population. Reserve this bee stick for only the selected population. Avoid contaminating with the pollen from the remaining Fast Plants. Pollinate flowers in the selected population for the next three days with the same bee stick. Likewise, with separate bee sticks you can pollinate the plants from the larger population, but be careful to keep them separate from the selected population.

21. Maintain the plants through the rest of their life cycle. As the seedpods form be sure to limit each of the plants to 8 to 10 seedpods. Any more will likely result in poor seed quality. Once the seedpods start to turn yellow (about day 28–36), remove the fertilizer water from the reservoirs and allow the plants to dry for several days. After the plants and seedpods have dried (about a week later), harvest the seedpods from the selected population into a small paper bag for further drying.
22. Continue to monitor, pollinate, and maintain your control plants throughout the rest of their life cycle. Just be careful to keep the original population and the selected population separate.
23. You should now have two populations of second-generation seeds: (1) a population that is the offspring of the selected plants from generation one and (2) a population that is the offspring of the remaining plants from generation one. Take seeds from the selected population and plant them to grow the second generation of plants under conditions that are identical to those you used for generation one. Use new bottle containers or, if you choose to use the previous bottle systems, make sure that you thoroughly clean the systems and sterilize with a dilute (10%) bleach solution. Use new wicking cord and new soil. To get your seed, break open the seedpods into a small plastic petri dish lid.
24. When the second-generation plants are about seven to 12 days old, reexamine the plants and score for the trait you selected. Score the plants at the same life history stage using the same method.
25. Compile, analyze, and graph the class data as you did for the first generation.
26. Unless you plan on growing these plants for another generation (maybe another round of selection – which I'd like to do), you do not have to save these plants. You can discard them and clean up your growing equipment at this point. If we are going to another generation, repeat steps 18 – 25.

Statistical Analysis of Generational Data

There are many forms of statistical inference for quantitative data. When measuring a single quantitative variable, like birth weight in lambs, calcium concentration in the blood of pregnant women, or migration rate of birds, we often wish to infer the mean value of the population from which we drew the sample. However, the mean of a randomly chosen sample will not necessarily be the same or even close to the population mean. Suppose we wanted to know the average weight of newborn lambs on a particular farm. By chance, we may end up with a random sample that includes an excess of lightweight lambs and therefore a sample mean that is less than the overall mean in the population.

To infer the population mean from the sample data, we can calculate a **confidence interval for the mean**. This is a statistically derived range of values that is centered on the sample mean and is likely to include the population mean. For example, based on the sample of 34 *Abramis brama* weights from Lake Laengelmavesi (**Table 2**), the 95% confidence interval for the mean weight ranges from 556 grams to 696 grams. (This was not worked out for you here, but is listed just to explain this concept.) The true average weight for this species of fish is likely, but not guaranteed, to fall within this range. For this population, the mean is 626, the standard deviation is 207, and the sample size is 34. (More on this later...)

Table 2: Mass of Individual *Abramius brama* from Lake Laengelmavesi.

Weight (g)	Frequency	Relative Frequency
201-300	2	0.06
301-400	3	0.09
401-500	8	0.24
501-600	3	0.09
601-700	8	0.24
701-800	3	0.09
801-900	1	0.03
901-1000	6	0.18
TOTAL	34	1.00

Biologists frequently wish to compare the mean values in two or more groups; for example, newborn lamb weights on several different farms, calcium concentration in women in early and late stages of pregnancy, or migration rates in birds of different species. Based on the means and standard deviations calculated for each of the samples, they infer whether or not the means in the different populations are statistically different from one another. There are several statistical methods for this, and the correct method depends on the number of groups, the experimental design, and the nature of the data.

To illustrate this, we can apply one type of *t*-test to the Lake Laengelmavesi data to assess whether the two fish species *Abramis brama* and *Leusiscus idus* have significantly different mean weights (**Table 3**).

Table 3: Mass of Individual *Leusiscus idus* from Lake Laengelmavesi.

Weight (g)	Frequency	Relative Frequency
201-300	2	0.33
301-400	1	0.17
401-500	0	0.00
501-600	1	0.17
601-700	0	0.00
701-800	1	0.17
801-900	0	0.00
901-1000	1	0.17
TOTAL	6	1.00

We begin by stating our hypotheses and choosing a significance level:

H_0 : *Abramis brama* and *Leusiscus idus* have the same mean weight.

H_A : *Abramis brama* and *Leusiscus idus* have different mean weights.

α = (this is the significance level – we use 0.05 or 5% or 1:20 as our cutoff)

The test statistic is calculated using the means, standard deviations, and sizes of the two samples. Using the t-test function in Excel (again, not shown here – but will be explained soon enough) we find the p -value to be $p = 0.497$. **The p-value is automatically calculated if you enter the data and formulas correctly!** Since p is considerably greater than α , we fail to reject (which means ACCEPT – but said the correct way) the null hypothesis and conclude that our study does not provide evidence that the two species have different mean weights. **Figure 4** below describes the steps of a t -test, a simple method for comparing the means in two different groups.

What is the t -test? A standard method for assessing whether the means of two groups are statistically different from each another.

Step 1: State the null and alternative *hypotheses*:

H_0 : The two populations have the same mean.

H_A : The two populations have different means.

Step 2: Choose a significance level, α , to limit the risk of a Type 1 error.

Step 3: Calculate the *test statistic*: $t_s = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

Notation: \bar{y}_1 and \bar{y}_2 are the sample means; s_1 and s_2 are the sample standard deviations; and n_1 and n_2 are the sample sizes.

Step 4: Use the test statistic to assess whether the data are consistent with the null hypothesis:

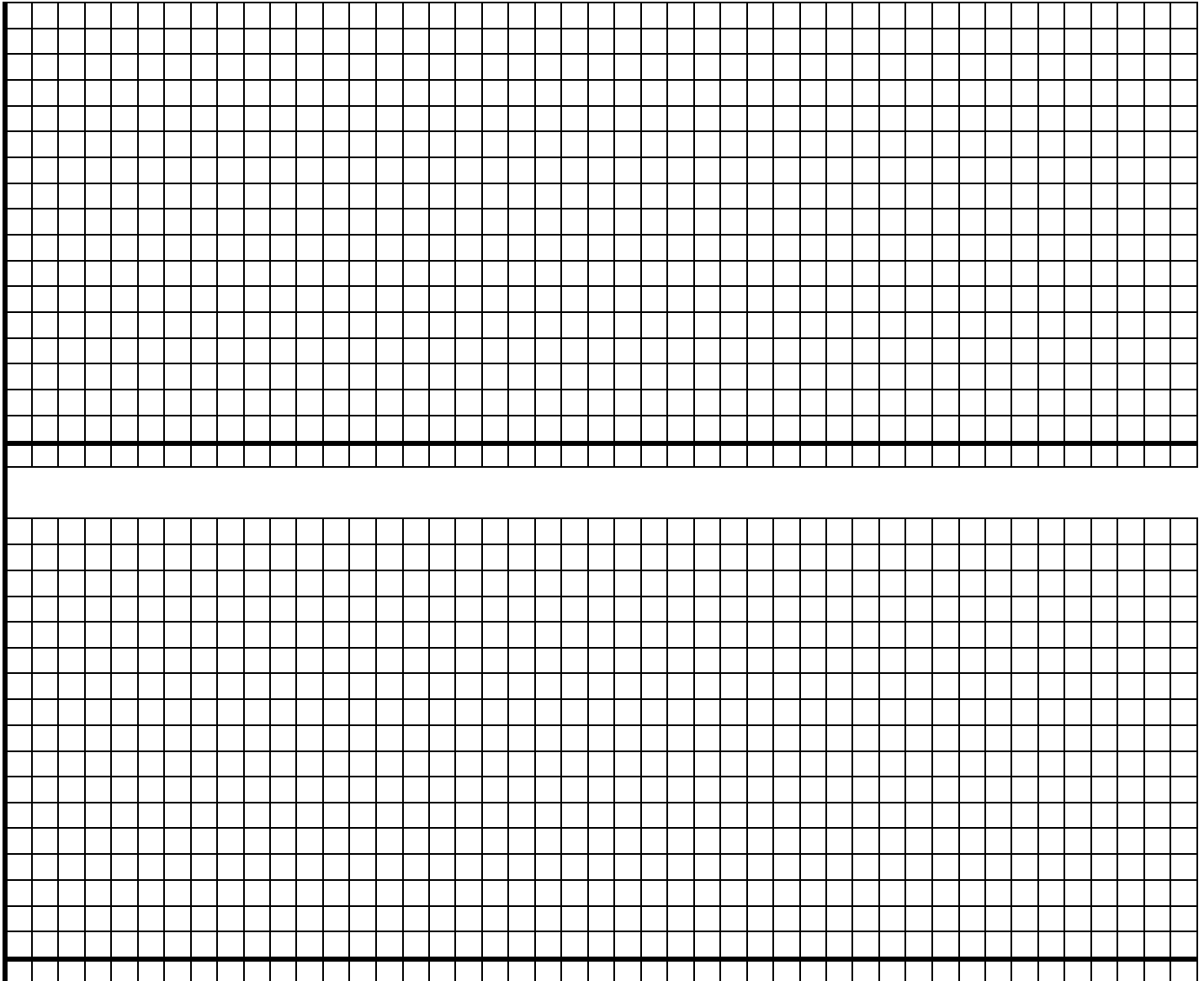
Calculate the p -value (p) using statistical software or by hand using statistical tables.

Step 5: *Draw conclusions* from the test:

If $p \leq \alpha$, then reject H_0 , and conclude that the two means are significantly different.

If $p > \alpha$, then we do not have sufficient evidence to conclude that the means differ.

Figure 4: How to calculate a p -value via a t -test.



3. On each of the above histograms, mark the means with solid vertical lines labeled " \bar{x} "(use a ruler!). Mark ± 1 , ± 2 , and ± 3 s.d.'s with dotted vertical lines labeled "sd=1", "sd=-1" etc.).
4. From the above graphs, can you say for certain that the experimental group is different than the control group with respect to the independent variable?

After you complete the above calculations and histograms, you will next determine whether or not you are able to say that the two groups are different.

Artificial Selection: Inferential Statistics and Summary Graph

Let's start by defining our hypothesis (H_A) and null hypothesis (H_0) for our investigation:

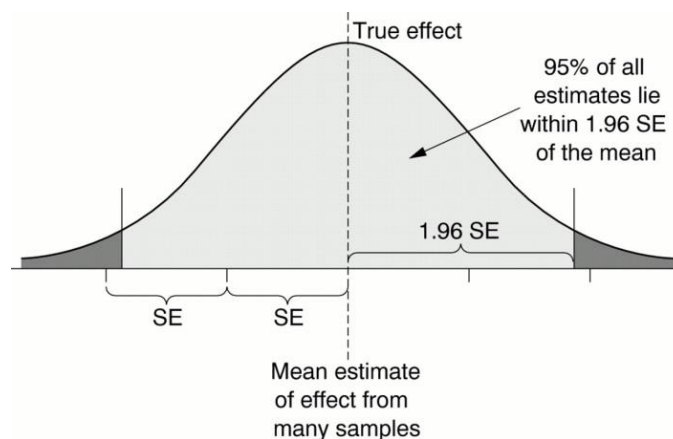
H_A : _____

H_0 : _____

To reject the null hypothesis (H_0), we need to demonstrate that the two normally distributed means (\bar{x}_{control} and $\bar{x}_{\text{experimental}}$) are not equal. More precisely, we need to show that the experimental and control groups are not just one big population. We can accomplish this by applying a statistical test like *chi-square test for goodness of fit* (you'll have to know this test – and only this test – for the AP test) or *Student's T test* (this is more appropriate for our data but you will not have to know how to do it). ***The result of a statistical test is the probability (p-value) of obtaining the collected set of data through chance alone.***

We will use a t-test generated with Excel to analyze our data. The **probability** of obtaining your data assuming H_0 is the **p-value**. If $p < 0.05$ it means that, if the null hypothesis is 'true', there is less than a 5% chance of obtaining data as extreme as ours; so we can reject the null hypothesis (H_0). If $p > 0.05$, we "fail to reject the null hypothesis". Note the language in quotes – we cannot say that we "accept the alternative hypothesis".

If you were to repeat this experiment a thousand times, you would probably get slightly different means each time. In fact, a histogram of these hypothetical means would form a bell-shaped curve (below). Standard error is the standard deviation of the distribution of these hypothetical means. We know that on a normal (bell-shaped) curve, 95% of the points fall between ± 2 standard deviations (actually, ± 1.96 standard deviations) of the mean. So, there is a 95% chance that the true mean is within 1.96 standard errors of the sample mean. So, although you cannot say with confidence that your *measured mean* is exactly the *true mean* of an infinite population, you can say that there is a 95% chance that the *true mean* is within 1.96 standard errors of your *measured mean*.



Calculating Standard Error of a Mean (SEM)

$$\text{standard error of the mean} = \frac{s.d.}{\sqrt{N}}$$

Notice that as the sample size (N) gets bigger, standard error gets smaller. The larger your sample size, the more confident you can be that your measured mean is close to the true mean.

The 95% confidence interval around the mean is calculated by multiplying standard error by ± 1.96 . There is a 95% chance that the "true mean" falls between $\bar{x} - (1.96)(S.E.)$ and $\bar{x} + (1.96)(S.E.)$. (assuming a normal distribution)

Calculate standard error and the 95% confidence interval below. Show your work.

CONTROL**EXPERIMENTAL**

 $\bar{x}_{\text{control}} =$ $\bar{x}_{\text{experimental}} =$ $\sigma_{\text{control}} =$ $\sigma_{\text{experimental}} =$ $SE_{\bar{x}_{\text{control}}} =$ $SE_{\bar{x}_{\text{experimental}}} =$

Using Excel to Carry Out a Student *t*-test

The following will return the probability (the p-value from our Inferential Statistics section in the Biology Skills Lab from last September) associated with a Student's *t*-test. Enter your data from the two separate groups into two properly labeled columns in a new Excel sheet. (I recommend using columns A and B.)

- Click on any empty cell (I recommend using one of the top cells in column D – for spacing purposes) and type “=TTEST”
- Then hit tab – NOT ENTER!
- A dialog box will appear. It will say (array 1, array2, tails, type)
- Array 1 equals all the data in your first column. Highlight these cells, then type in a comma.
- For Array 2, highlight the second column of data. It is *not* essential that there is the same sample size for each group.
- To answer the "tails" question, remember your prediction about the direction of the difference between the groups. If you predicted group A would be lower than group B, pick 1 tail. If you predicted group B would lower than group A, pick one tail. If you didn't predict which would be higher, use 2 tails. You can't change your mind after the data are gathered. **We are not going to predict which direction of our difference, so we'll use 2 tails.**
- Answering the type question:
 - There are three types of T test you can use on Excel. Let's say you wanted to test whether heart rate increased after drinking a cup of hot sauce (don't actually try this!) or whether plant growth would increase after adding fertilizer to pots of soil. In these cases you would be comparing the heart rate of the same people, or the growth of the same pot of plants before and after the treatment. This would require a "paired" or "dependent" T test. Excel calls this a "type 1" test.
 - Let's look at another situation. Say you want to know whether nursing students consume more coffee than do biology students. You would then have two groups of test subjects rather than taking 2 measurements on each person. Now you would use an "unpaired" or "independent" T-test. Excel calls these "type 2" or "type 3" tests. Now the tricky part is to decide which of these to use. Are the standard deviations about the same for both groups, or are they different? You can test this statistically, but let's just work with how they seem. If in doubt, go with "type 3" for unequal variances. **We will use the independent test, so type 3.**
- Now hit "Enter" and see what the number is. This is your p-value. Remember that a p-value below 0.05 is generally considered statistically significant, while one of 0.05 or greater indicates no difference between the groups. If your number looks like this: 2.03188E-7, Excel is giving you the number in its version of scientific notation. This number is actually 2.03×10^{-7} , or 0.000000203. You can set the format of the number under right click > Format Cells... to read Number with 2 or 3 decimal places.

For Your “Lab Report”:

Wrapping it up... For this lab, you should have collected a lot of data. Now you need to put it all together and give me your conclusion. Essentially, you will turn in a modified **Results** and **Discussion** section for this lab.

- Before the results, clearly state the null and alternate hypotheses.
- Open with complete **Descriptive Statistics** should be fully articulated, which for each population includes:
 - sample size
 - mean
 - median
 - range
 - standard deviation
 - standard error of the mean
- **A Grouped Frequency Distribution Data Table/Histogram.** You can use the same groupings that we used in this worksheet, as you will generate a histogram of this table as well in Excel.
- Your *t*-test carried out between the control and experimental populations. This must be done properly – so set up your spreadsheet correctly, and come up with your modified discussion section/conclusion. Answer the question in terms of STATISTICAL SIGNIFICANCE! You might have to review Lab 02 with the terminology...

EVALUATION - Lab Report: 22

Writer:

Section Points		Poor					Good					Excellent					Section Scores
		0%	25%	60%	80%	100%	0%	25%	60%	80%	100%	0%	25%	60%	80%	100%	
2	Title																0.0
2	Concise explains the purpose of the investigation (e.g., the effect of additional nitrogen fertilizer on the growth rate of corn)																0.0
5	Abstract																0.0
5	Conveys a brief synopsis of the full report concisely and effectively: objective, brief methodology, condensed results; less than 100 words																0.0
40	Results																0.0
15	<u>Opens</u> with effective statement of overall findings; not interpreted results!																0.0
20	Presents visuals (graphs and data table) clearly and accurately; visuals formatted properly; calculations completed/statistical analysis via Excel																0.0
5	Successfully integrates verbal and visual representations (CAPTIONS)																0.0
25	Discussion																0.0
15	<u>Opens</u> with effective statement of support (or lack thereof) of null/alternate hypothesis; use TTEST results!																0.0
10	Backs up statement with reference to appropriate findings and main concept as described in the introduction - YOUR CONCLUSION!																0.0

72

Points Earned	0.00
Total Possible Points	72
Percentage	0%
Late Penalty	
Final Score	0%