Invertebrate Zoology
Statistics Assignment

Due Date: ______________________

The goals of this assignment are to:

1. Introduce or reinforce data gathering skills.
2. Introduce or reinforce basic statistical and data analysis skills.
3. Apply critical thinking to the refined data to answer questions based on the evidence collected.

One of the goals of a scientist is to be able to answer questions with the greatest possible reliance on observable facts, and the least reliance on intuition. While intuition has great importance in finding the right questions to ask, and in finding ways of investigation, once the data is gathered the scientist should rely on the facts at hand. Patterns in the data may be revealed through good graphical analysis, and the patterns should then be tested with statistics to see if they are "real" – or simply the result of the scientist looking at the data and "seeing" a preconceived result. This is an example of bias; one shields oneself from bias by using commonly agreed upon statistical tests as impartial arbitrators of what is "real"

Sometimes the results are unambiguous. Every time you drop a penny it falls to the ground. No one needs statistical analysis to prove the existence of gravity. On the other hand, sometimes the penny lies heads up, sometimes heads down. Determining if this is a random event or something influenced by other factors may require the application of statistics; statistics are also useful to draw conclusions about a larger population by sampling a smaller portion of it.

Results in biology are seldom so clear-cut as to eliminate the need for statistics. There are several basic tests and graphical analyses that should be in every biologist's "toolkit". Among the graphing techniques are:

1. The **scatterplot**, which is used to look for correlation between two variables, or to track a variable over time.
2. The **trendline**, which is the superposition of a line drawn from a mathematical model over a scatterplot.
3. The **histogram**, which is used to look for patterns in abundance.

Any pattern that is revealed by the graphical analysis should be examined by statistical tests to see if the pattern is "real". In most cases, this means determining if the pattern is different enough from what might be expected in a random world. For instance, flipping 51 heads out of 100 tosses would not be unexpected; flipping 80 heads out of 100 tosses, or flipping 20 heads in a row might be unexpected and suggest that something else is at work. The statistical tests that will be of the most use to you in testing apparent patterns are:

1. The ***t*-test,** which is used to tell if two averages (the composite of many measurements) differ in a statistically significant way.
2. The **correlation coefficient,** which is used to test the statistical significance of a trendline.
3. The **Chi-square** test, which is used to determine if experimental results differ enough from expected results to suggest "real" difference.
4. The **ANOVA** test, which is kind of a "super" t-test to tell if any of a group of mean values differs from the rest. If the results are positive, you then have to go back with multiple t-test and see which mean or means is different

In this exercise, you will construct a scatterplot with a trendline. You will also generate a statistical summary, conduct a t-test and use a correlation coefficient.

Biological Background (do some research to fill in the blanks):

Mealworms are larvae of beetles in the order ____________________________. They feed on grains, and many are important pests of stored products such as flour. Their life cycle takes several months to complete. They molt ____________ times. The time between molts is called an instar, we refer to 1st instar larvae, 2nd instar larvae, and so on. The second-to-last molt produces the pupa, and the final molt produces the adult.

Your assignment will be to measure and weigh a sampling of mealworms from a culture 2 times, one week apart. You will characterize the data statistically, graph it out, and apply statistical techniques to test for any patterns the graphs reveal. You will then use your critical thinking skills to explain the nature of the culture in the container that you tested.

Step 1: making the measurements.

Ask the instructor where the mealworm culture will be located. Near the culture you will also find the other tools you will need to make the measurements: a scale, a ruler, forceps, paper towels, a sieve, weigh boats (dishes), culture dishes, a refrigerator, dust masks, and a computer linked into the network.

The first step is to use the sieve to sift a *random* sample of the mealworm larvae from the culture. Be sure to sample all areas of the culture and not to bias your results by going after specific mealworms. Be gentle in your sifting and try to avoid putting excess flour dust into the air. You may want to wear a dust mask to avoid breathing the dust.

Place any pupae or adults back into the culture. Place the larvae into the glass culture dish. Be sure you have at *least* 50 larvae. There is no set number to have; just have more than 50. The more you count, the more accurate your results will be, but you don't have the time to measure them all.

Place the larvae in the glass culture dish into the refrigerator for at least 10 minutes. While you are waiting, plug in the scale, and then turn on the computer (ignore any error messages) and start *Excel* by clicking on its icon on the toolbar at the top of the screen. When *Excel* comes up, make 3 columns as shown in the sample data below. Lay out a paper towel on the desk near the scale and the computer, and place the ruler on it. Place an empty weighing boat on the scale.

Once the worms have cooled for 10 minutes, you are ready to begin taking measurements. Got to the refrigerator with the forceps and a weighing boat. Choose 5 of the mealworms (depending on how quickly you work this may be too many) and place them in the weighing boat. Return to the scale. Working with one larva at a time, in rapid succession:

1. measure the length of the larva in millimeters
2. record the length on the computer
3. hit the Tare button on the scale to reset it (leave the weighing boat in place)
4. place the larva in the weighing boat on the scale and measure the weight
5. record the weight on the computer
6. dump the larva into another culture dish (all of the larvae will end up in this dish)

Repeat this process with the remaining larvae in the weighing boat and then return to the refrigerator to get another 5 larvae. Continue measurements until all of the larvae have been measured. If a larva is curved rather than lying straight then it has warmed up too much and you will need to return it to the refrigerator. Be gentle with the larvae – you will have to measure them again.

When you have measured all the larvae:

1. Weigh them all together to get a “group weight”
2. Record this number on the computer.
3. Return the larvae to the culture.
4. Save the Excel file on the K: drive in the Scratch\Invertebrate Directory
5. Make a backup copy on the floppy disk you were given in class.
6. Clean up everything you got out. Wash the desktop off and wash your hands as well.
7. Turn off and unplug the computer.
8. Unplug the scale.
9. Go to the computer lab to finish the work.

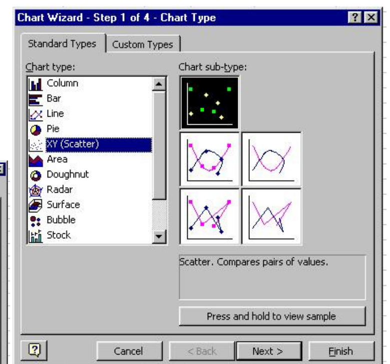
Step 2 – Scatterplot

In the computer lab, you should go to Excel and open up the file you just created. The data should look something like this:

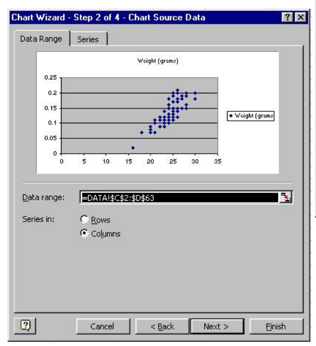
	Length (mm)	Weight (grams)
1	25	0.2
2	25	0.19
3	18	0.07
4	21	0.11
5	22	0.12
6	26	0.18
7	26	0.2
8	25	0.17
9	25	0.12
10	21	0.1
11	24	0.16
12	26	0.21
13	24	0.15
14	28	0.2
15	22	0.12

1. Create a scatterplot. To do this, highlight the data in the length and weight columns and click on the chart wizard button:

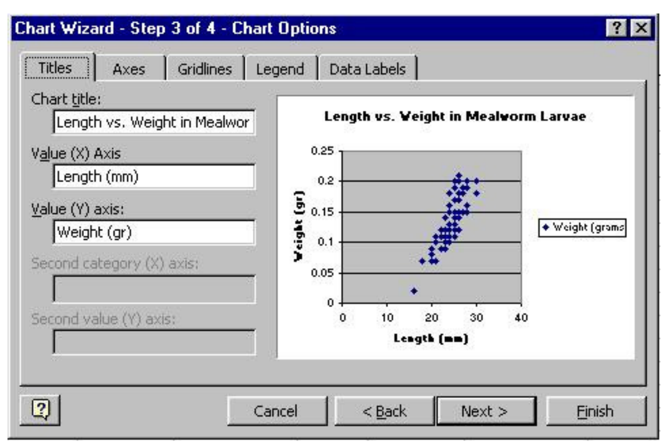
2. In the first wizard, select the XY (Scatter) chart type and the sub type without lines, then click next.



3. In the second wizard, the graph should look like the one below. Click next.



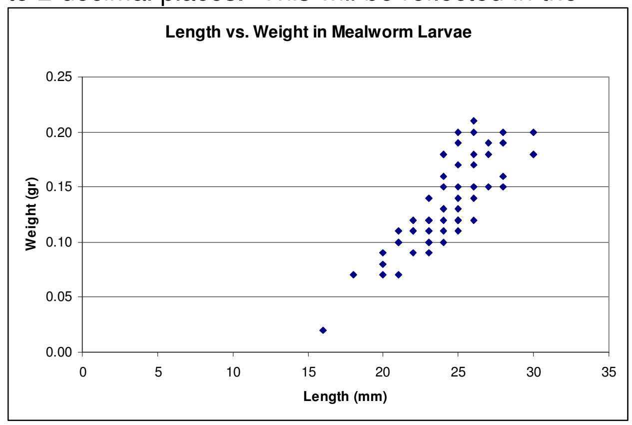
4. In the third wizard, fill in the title and the axis labels. Click next.
5. In the 4th wizard (not pictured), simply click on next to paste the chart onto the data sheet. Once the chart is inserted, drag it to a good location and size it as you see fit. Several remaining things need to be done to properly format the graph.
6. Right-click on the gray area in the center of the graph (not on a data point or grid line). Select “none” for the fill; this will get rid of the gray background.
7. Left-click once on the legend (which says “Weight”) then hit the delete key to remove the legend, which is redundant.



8. Highlight the weight column (just the numbers, not the label at the top). Click on the increase decimals button : 

This will make all of the values in the column 2 decimal places long. If you go too far, use the decrease decimals button to get all the numbers to 2 decimal places. This will be reflected in the *y* axis on the chart.

9. Your graph should now look something like this:

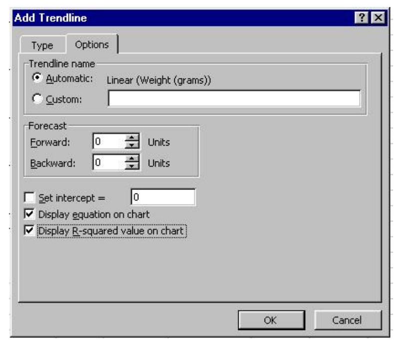


Step 3 - Trendline

1. Right-click on one of the data points. This should open up a menu with several choices including "Add trendline". Choose that option.

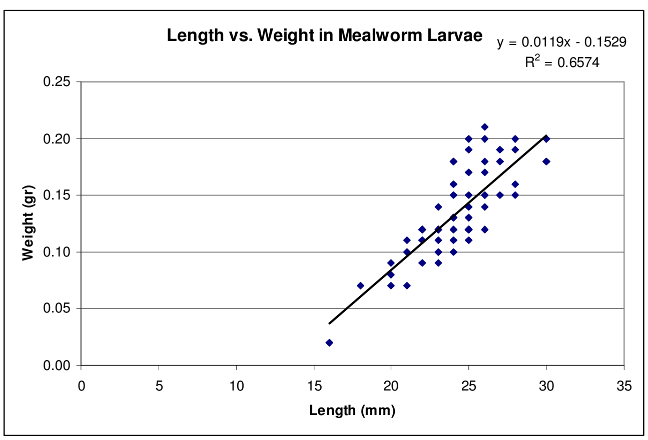
2. On the box that opens choose linear, then click the options tab.

3. Click the bottom two options on the page, as shown at the right:



4. Click OK.

5. Using the left mouse button, click and drag to move the inserted equation off the graph to a position where it can be read. The graph should now look like this:



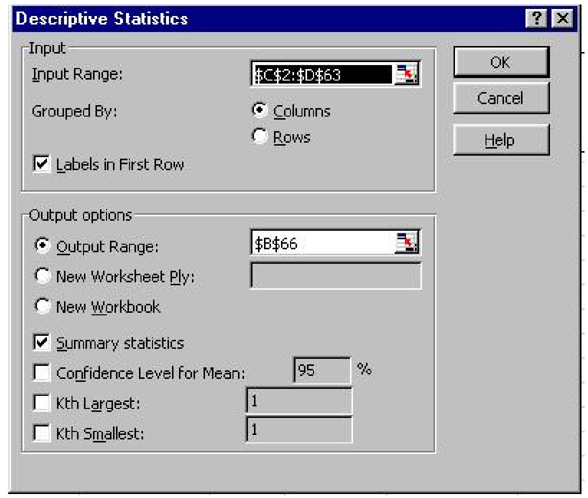
You now have a scatterplot with a trendline and can answer the following questions:

Are length and weight correlated in mealworms?
If so, what is the mathematical relationship?
How strong is the correlation?

The fact that you got a straight line shows that there is a linear relationship between the variables. The nature of the relationship is in the equation, where y is the weight and x is the length. In this case, the weight is 0.0119 times the length – 0.1529. The fact that the line slopes up to the right indicates that as length increases, so does weight. Also, the R^2 value tells you how good the correlation is. The closer the R^2 value is to one, the better the correlation.

Step 4 – Statistical summary.

1. Select Tools:Data Analysis from the menu.
2. Select Descriptive Statistics from the list that is presented, and click OK.
3. The Descriptive Statistics Wizard will come up. Fill it out in a similar way to the one presented here:
4. In the Input Range enter the cells where your data can be found. You can click on the small box to the right to go to the spreadsheet and highlight your data. Include the column headings.
5. Be sure to click the Labels in First Row box
6. For the Output Range, select an area of your sheet with nothing to the right or below.
7. Click the summary statistics box.
8. Click OK. If you get a message about overwriting data, click cancel and try again with a different output range.
9. Your results should look something like this:



Length (mm)		*Weight (grams)*	
Mean	24.24590164	Mean	0.134590164
Standard Error	0.351154021	Standard Error	0.005135188
Median	24	Median	0.12
Mode	25	Mode	0.12
Standard Deviation	2.742600577	Standard Deviation	0.040107097
Sample Variance	7.521857923	Sample Variance	0.001608579
Kurtosis	0.765780336	Kurtosis	-0.173701118
Skewness	-0.291385066	Skewness	-0.083054008
Range	14	Range	0.19
Minimum	16	Minimum	0.02
Maximum	30	Maximum	0.21
Sum	1479	Sum	8.21
Count	61	Count	61

There is a lot of data here; this isn't a statistics course and we won't go over it all. The mean is the average of all the lengths or weights. Median and mode are also measures of where the "center" of the data points lies. Standard error, sample variance, and standard deviation are all measurements of how close the data points are to each other. Kurtosis and skewness help determine if the population is distributed normally; minimum, maximum and range tell you the high and low points and how far apart they are; the sum is calculated by adding up all the data points, and the count is the number of data points. Divide the sum by the count and you get the mean, which is where we started.

It is often interesting to compare the mean weight that Excel generates from your data to the mean weight of the mealworms as calculated by their total weight divided by the number of mealworms.

Step 5 –* t*-test.

In order to conduct a *t*-test you need two sets of data. The *t*-test helps you answer the question "Are the means of these two data sets the same or not?" Or, to be more precise, the *t*-test allows you to reject the hypothesis that the two data sets have the same mean with a certain chance of making a mistake. The possibility of making a mistake comes about because of the variation within natural populations. If you wanted to compare the heights of people in two different cities, you might watch 100 people pass though a doorway with the heights marked on it. If, by chance, in one city you did your measurements while an elementary school went on a field trip, and in the other city you caught the athletes at the city basketball tournament, you would conclude (incorrectly) that the two cities had different average heights. To protect against making this type of mistake you set a benchmark – the alpha (α) value at a high level. If you set it at 5%, that means there is only a 5% chance that you might erroneously conclude that the means are different when in fact you just had bad luck in sampling.

It would be trivial to compare the weights and heights. Of *course* they are different. Just for fun, I'll do it here so you can see how the *t*-test works:

t-Test: Two-Sample Assuming Unequal Variances

	Length (mm)	*Weight (grams)*
Mean	24.24590164	0.134590164
Variance	7.521857923	0.001608579
Observations	61	61
Hypothesized Mean Difference	0	
df	60	
t Stat	68.6557246	
P(T<=t) one-tail	4.91831E-59	
t Critical one-tail	1.670648544	
P(T<=t) two-tail	9.83662E-59	
t Critical two-tail	2.000297172	

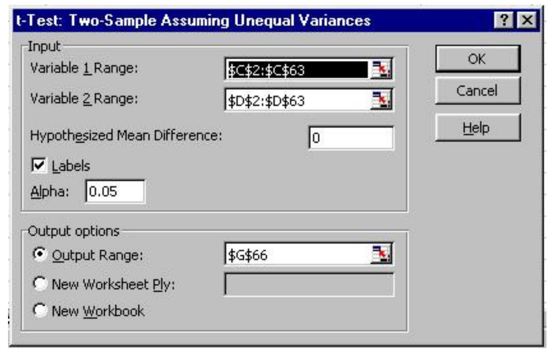
The *t*-test works by mathematically comparing the variances within the two samples with the difference in their means. The number that results from this is compared to a table of values computed for each possible alpha value. Of course, the computer doesn't have a table to go to, the program generates the value on the fly. In Excel, you get a printout like the one above. The important numbers to look at is the t Stat, the P values, and the t Critical values. The t-stat is the number generated by the computer based on your data. The bigger it is, the greater the significance of the difference between the means. The P values tell you the chance of erroneously saying the means are different. The smaller the number the better; you want it at least to be smaller than your alpha value. The t critical numbers are from the table generated by the computer. If your t Stat is greater than the t critical value then you can assume that the means are different with a chance of being wrong due to unlucky sampling of less than the alpha value you selected. The P values give you the exact chance of making that type of mistake; in the example above it is 9.8×10^{-59} (not much of a chance). In this case, we reject the hypothesis that the means are the same, and we're pretty confident that the difference is real, not due to chance.

What about the 1 vs. 2 tails? To put it in a nutshell, use the 1 tail test when you can predict the direction of the difference between the means. If you have been feeding one group of mealworms twice as much as another group, you would expect the group being fed to be heavier, and you would use a 1-tail test. On the other hand, if you were just comparing 2 populations of mealworms and knew nothing about their living conditions, you would have no way of knowing which population was eating better and therefore would be heavier. You would use the 2-tailed test.

In this study the two means that you will be comparing will come from two sets of measurements. That means that you will have to return between 7 and 10 days from when you take your first set of measurements and repeat the measurement process.

To do a *t*-test:

1. Copy the two sets of data to adjacent columns on the same worksheet.
2. Select Tools:Data Analysis from the menu
3. Choose t-test: Two Sample Assuming Equal Variances (if the variances are equal, otherwise choose unequal variances)
4. Fill out the wizard as shown at the right. Your two columns of data (with labels) should be selected in the first 2 boxes.
5. The mean difference should be 0.
6. Check the labels box.
7. Set the Alpha at 0.05
8. Set the output range to an open area on the worksheet.



Step 6 – Putting it all together.

All of this data and analysis are useless if you don't do something with it. The data and analysis are used to *help* you reach *conclusions* and to *support* your arguments as to why your conclusions are right. The data itself is useless.

Your assignment is to write a short paper to answer these questions:

1. Is there a correlation between length and weight in mealworm larvae?
2. What is the nature of this relationship?
3. Do mealworms grow measurably over a period of one week?
4. If so, how much?

The text of your paper should only be a page or two; but since you will be pasting in tables and graphs from *Excel*, the number of pages might be longer. There should also be a paragraph about mealworms and their normal growth patterns (background); this paragraph should be appropriately referenced. Each of the answers to the questions should be backed with data and analysis as shown by material pasted in from *Excel*.