DEPARTMENT OF Life Sciences

**Physiology**

**LABORATORY 1-2**

**Graphing and Graph Interpretation (36pts)**

**OBJECTIVES**

After completing this exercise, you should be able to:

1. Identify the type of graphs used to visualize biological data.

2. Evaluate graphs to identify trends in a data set.

3. Obtain descriptive statistical values such as mean, mode, and standard deviation from a data set.

4. Prepare an appropriate graph on excel from a given data set.

5. Draw conclusions about the data set based on your graph(s).

**Why Biologists Graph Data**

Why might a student or scientist wish to convey their data in a graphical form? Most people are

visual learners by nature and representing one’s data graphically is a visual picture of that data set. By representing one’s data in a graph we can quickly see trends in large amounts of data. Is one variable related to another in some way? For example, with an increase in the body mass index (BMI), the concentration of total cholesterol in the blood increases (BMI is a measure of relative size based on the mass and height of an individual). If we were to graph BMI on one axis and total cholesterol level on another axis and plot this for a group of people, we would see that the data points rise as BMI is increasing (Graph 4).

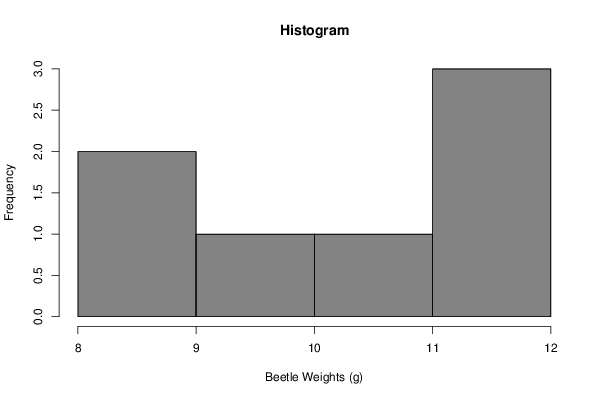
Graphing a set of data can also allow us to see the dispersion of the data: are the data points grouped closely together or are the data points widely distributed? By displaying data graphically we can visualize descriptive statistics such as the mean, mode, median, range and standard deviation of a set of data. The mean, median and mode tell us about the central tendency of the data, or what most of the data looks like; while the range, standard deviation and variance tells us about the spread or dispersion of the data. In this way a graph can complement your data tables. Some graphs can also be used as a tool to evaluate the statistical relationship between the variables. Perhaps you would like to study the relationship between the physical activity and the muscle endurance of your fellow students. Let’s say, you make a hypothesis statement: “If the students exercise regularly then their muscle endurance is greater”. You want to test your hypothesis so you devise a test, recruit volunteers and collect data. Let’s say it appears that the average, or mean, time of endurance of student athletes is greater than that of non-athlete students. How do you show that this difference is real? You could apply an appropriate statistical test with accompanying graphs of the data to support your hypothesis.

**Types of Graphs**

While many types of graphs exist to convey important features of a set of data, or population, there are a few common types of graphs that are typically used in biology. These are bar charts and histograms, x-y or scatter plots and fever plots or line charts. Bar charts use rectangles to represent data and each bar represents a variable. A histogram is a type of bar chart which quantifies the distribution of one variable over the range of values. For example, let’s say you measure the weights of beetles on four islands in the Caribbean.

You want to see the trends in beetle weight for Island I. You can create a histogram with ranges of

beetle weights in grams on the x-axis and number of beetles on the y-axis for Island I.

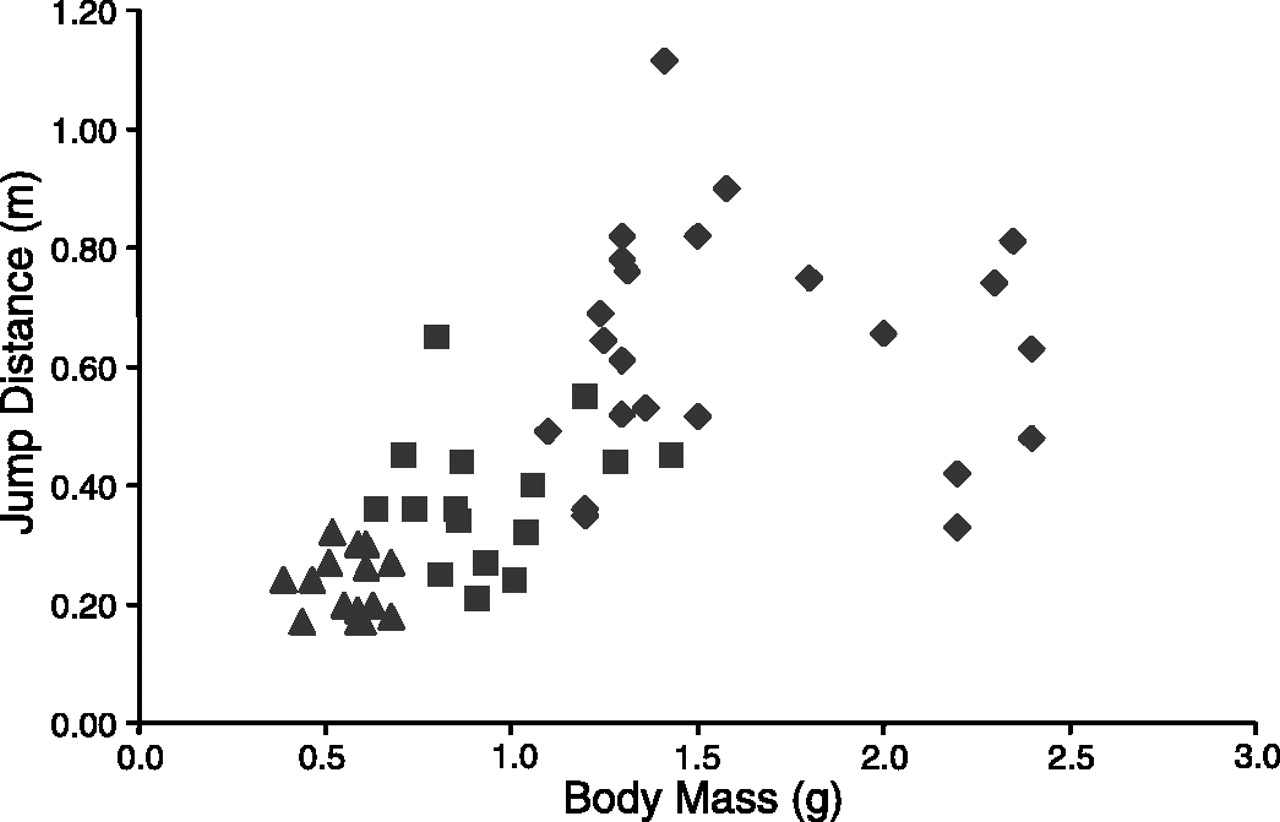


**Graph 1**. Histogram of beetle weights in grams for Island I.

**Q1**: What conclusions can you draw from this graph? (2pts)

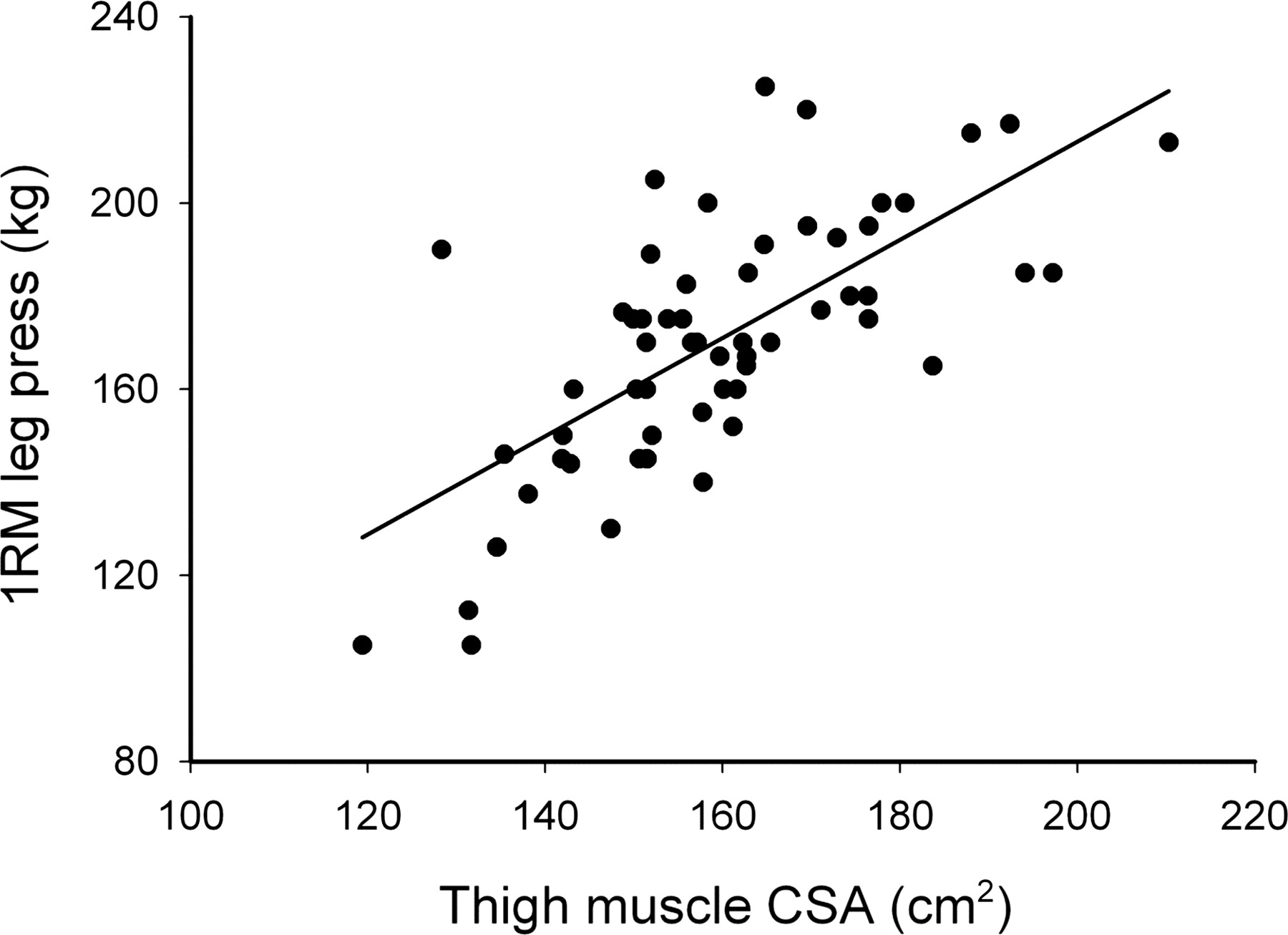
An x-y or scatter plot displays one variable in relation to another in the form of points or dots on a grid

as shown in Graph 2.



**Graph 2.** Scatter plot. This graph shows the distance that locusts can jump, relative to their body mass.

**Q 2**: What conclusions can you draw from this graph? (2pts)



**Graph 3** shows an example of a scatter plot combined with a line plot to mathematically construct a best-fit line through the data points. With the formula that created this line, you could predict how much weight someone could lift in a leg press using the cross-sectional area of their thigh muscle.

**Q 3**: What can you tell from this graph? What does this tell you that the previous graph did not? (2pts)

**How to interpret information from graphs**

When gleaning information from a graph there a few basic questions to ask yourself. *First, discern the type of graph that displays the data.* Is the graph a bar chart, histogram, pie chart, scatter plot, line graph or curve? Different types of graphs display different information about the data. A bar chart can compare averages (means) between different samples (eg. average popcorn kernels popped per bag in different brands of microwave popcorn). A histogram shows the trends in a single set of data (Graph 1). A scatter plot and/or line graph conveys a relationship between variables. That relationship may be positive, both variables increase together (Graph 2 & 3) or negative, as one variable increases another decreases. The relationship between variables may not be linear, but a relationship may still exist (Graphs 4).

*Once you have established the type of graph, you can look more closely at the variables*. Look at the title of the graph, graph description, and x and y axes. Decide what variables were measured, the measurements used and if one variable is dependent on the other. Typically, the dependent variable is displayed on the y-axis, but there may be no dependence between the variables. Some graphs may try to establish that one variable is dependent on another. A single graph is not enough evidence to conclusively demonstrate dependence between variables, but can be used to support the hypothesis that one variable is dependent on another.

**Q4**: What does it mean for one variable to be dependent on another? (2pts)

Look at the axes of the graph. Are the divisions between measurements equal? Does the axis start at zero (graphs do not need to start at zero). What is the range of measurements for each variable (the highest and lowest value)? What were the measurements used for each variable? Is the scale logarithmic?

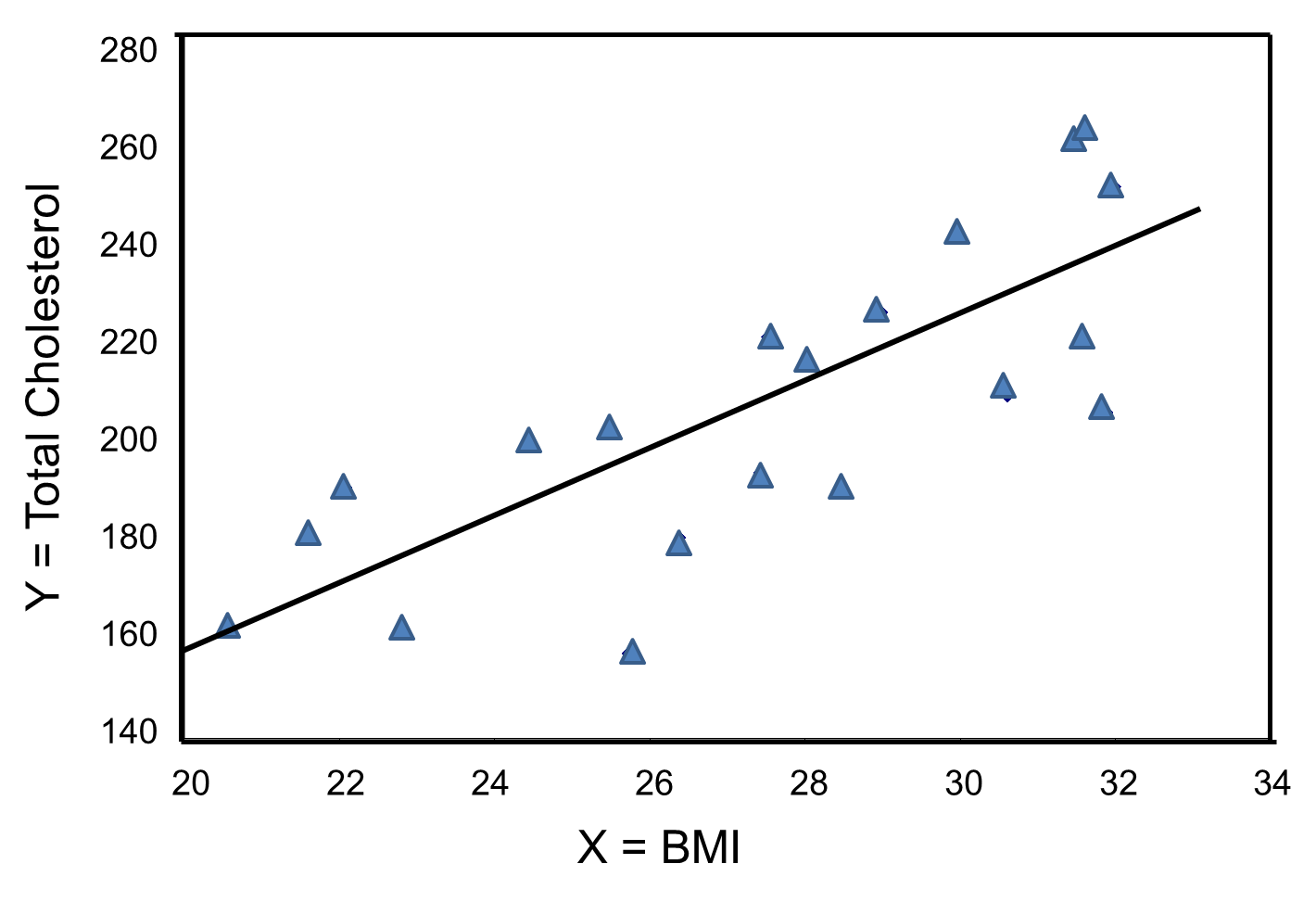
*Now look at the variation of the data*. This gives an indication of how dependable the measurements are. How many times did the author repeat the measurements? Are these measurements precise? Are the measurements close together or widely dispersed? If the same conditions exist one might expect that the values would be close to each other (precise) if the experiment is repeated. Wide variation in measurements indicates a problem with the experiment.

*Sample size*. How large was the population that was measured? A small sample size can bias the data interpretation and may indicate trends that do not exist when a larger sample size is measured. In clinical medical trials it is important to look at gender, age and even culture as there are differences in responses to treatments depending on these variables. A medication that works well for Caucasian adult males may not be the best option for young Asian females. Another consideration in variation is the treatment of the variables. When the test is repeated, are the variables all treated the same in all the experiments? If conditions are changed, it can be difficult to draw conclusions about the data. This is often the case in medical studies when different trials are compared. The trials often focus on different variables and measure variables based on different tests. This can make comparisons between studies difficult.

*Next try to figure out what the author’s main point about the data may be*. What does the graph tell you about the data? How do the treatments compare to the controls? The treatments are the subjects in which one variable has been changed while the controls are the normal conditions. In drug trials the control patients are given a placebo drug. A placebo is a pill that looks just like the drug, but does not contain the ingredient that is being tested. It is sometimes referred to as a sugar pill. So neither the doctors, nurses or patients know who is given the test drug and who is given the placebo. This is called a double-blind test since neither the testers or patients know who is in the control and who is in the treatment group. This helps to prevent any bias of data interpretation. Once you decide the main point of the graph, read what the author says about his or her interpretation of the data. Do you agree? Why do you think the author’s point is valid or invalid?

**Lab activities:**

Use the graph below to answer the following questions:

**Graph 4.** The total cholesterol of 20 adult men, shown relative to their BMI. (5pts)

1. What type of graph is this? What do you know about the data from the type of graph?
2. What did the researchers measure?
3. Which is the independent variable? Which is the dependent variable?
4. Do these two variables have a relationship? If so, describe it.
5. Describe the variation in the data. What might explain the variation that you see here?

**Making your own graphs- Histograms (10pts)**

For this portion of the lab you will use your data from the Measurement Lab 1 to construct your own graphs in Microsoft excel. The instructions for creating these graphs are for windows. If you are using a mac, these instructions may not apply.

**1. Enter your data.**

Open Microsoft excel on your laptop and enter the height data for the entire class from Lab 1 into a column with the first cell labeled “Height (cm)”. Do not enter the mean or standard deviation. Later instructions will allow excel to calculate these parameters. In another column labeled “Wrist Circumference (cm)” enter the circumference of the wrists of your class members.

*Save your data by clicking the MS icon in the upper left corner of the screen*. Choose the “save as” option from the pull down menu and save the file in a format that will be compatible with your home system.

**2. Calculate the mean for your data**. Click on the open cell just below your height data. Click on the “formulas” tab of the top of the menu bar. Click “Insert formula” on the left of the menu options. In the pop-up menu box choose “AVERAGE” for the function. If this does not appear as a choice, type “average” in the search box and click OK. Once you have selected “AVERAGE” be sure that only the height values are selected for calculation. The cells selected in the box “number one” should appear as A2:A23 or whatever cell of height data in column A is the last. Now click OK and the mean should appear in the box. Type Mean next to your average. Don’t forget to correct for significant digits. Excel will not do this for you. Now calculate the mean for the wrist data in the same way and enter the wrist mean circumference in the empty cell below the wrist data. Don’t forget to label the mean and correct for significant digits. *Save your file*.

How do the excel calculated means compare to your calculated mean using your calculator? Are they the same or different? Why do you suppose that is?

**3. Calculate the standard deviation for the height and wrist circumference.** Click on the open cell just below your average. Click on the “formulas” tab of the top of the menu bar. Click “Insert formula” on the left of the menu options. In the pop-up menu box choose “STDEV” for the function. If this does not appear as a choice, type “standard deviation” in the search box and click OK. Once you have selected “STDEV” be sure that only the height values are selected for calculation. Do not include the mean as a value to be included. Now click OK and the standard deviation should appear in the box. Type “Standard Deviation” next to your standard deviation. Don’t forget to correct for significant digits. Now calculate the standard deviation for the wrist data in the same way and enter the wrist circumference standard deviation in the empty cell below the average. *Save your file.*

How do the excel calculated standard deviations compare to your calculated standard deviations using your calculator? Are they the same or different? Why do you suppose that is?

**4. Obtain the mean and standard deviation using the data analysis function.** Now, click the “data”

tab at the top of the menu bar. On the right side of the bar an “analysis” menu option with “data analysis” tab should appear. If this is not an option on your computer, you need to add it. To add this option click the MS icon in the upper left corner of the excel window. Click “excel options” at the bottom right of the pop-up menu. Click “add ins” on the left column of the excel options menu. At the bottom of the menu click Go next to “Manage excel add-ins”. Now click “Analysis Toolpack- VBA” and click OK. Now the data analysis menu option should appear on the data tab menu.

1. Click the “data analysis” option in the data menu. A pop-up menu will appear. Choose “descriptive statistics” and click OK. In the “input range” box select or enter the height data (A2:A23 as in 2 above). Do not include the mean and standard deviation. “Grouped by columns” should be selected. In the output range, new worksheet Ply box type “Heights”. Select summary statistics and click OK. A new work sheet will appear with a table showing the descriptive statistics with your height data. *Save you file*.

How do the mean and standard deviation compare with the formula calculated mean and standard deviation? Are they the same or different? Why do you suppose that is?

1. Do the same for your wrist circumference data. *Save your file*

**5. Calculate the coefficient of variance for the both data sets.** The coefficient of variance (c.v. or C.V.), states the standard deviation as a percentage of the mean:

**C.V. = std. dev./ave. x 100%**

This allows you to compare the spread of the data between two populations with different measurements: for example you could compare tree heights to root area. The c.v. helps us evaluate the precision of our data recording techniques. When measuring samples of the same type you should expect the c.v. to be 10% or less. This indicates that the measurements are within an acceptable range. Larger c.v. percentages may indicate that there was a problem in some aspect of the measuring. A large c.v. may also just reflect the variability of the population measured. For example, if you measured the mass of dogs at a veterinarian’s office, you might

expect that you would have a wide range of masses and high c.v. due to the great variability in size of the different dog breeds.

Now use excel or your calculator to calculate the coefficient of variance for height and wrist circumference. Add the values to the bottom of your height and wrist data columns and label the value as c.v.

**Q5**: Is one c.v. larger than the other are they approximately the same? If one is larger, which one is larger? Why do you think that is? What do you think it means about human variation?

**6. Graph your data for both height and wrist circumference as a histogram.** Go back to your data sheet. Click the data analysis tab in the data menu. Choose “Histogram” from the menu and click OK. In the input range box select or enter the cells that contain the height data. Do not include the mean or standard deviation. Click the “labels” box. In the new worksheet Ply box type “Height histogram”. Click chart output and click OK. A new worksheet will appear with a histogram of your height data.

a. Click “bin”, delete the letters and type “Height (cm)”

b. Often, excel chooses bins or ranges that do not best convey the central tendencies and variation of the data well. The central tendencies of a data population are the best represented by the mean, mode and median. The mean you have calculated already. The median is the middle number of a data set and the mode is the most represented value of a data set. These values were provided by the descriptive statistics function of the data analysis in step 4. The variation and spread of the data is best represented by the range, standard deviation and variance, also provided in step 4. Let’s create our own ranges for the height data. Go back to your data work sheet. Copy and paste your height data to a new worksheet and name it Height histogram by clicking on the tab at the bottom of the sheet. To add a new sheet if needed click the icon with the sunburst next to a new sheet on the bottom right of the excel window.

c. Using your calculator take the range of the mass data (found from the descriptive statistics table) and divide it by 8. This will allow you to create eight bars or bin ranges for your histogram. For example, if your height range divided by 8 equals 5 cm then each bar or bin will be 5cm each. Start a new column and name it height.

d. Now start at the lowest value for the heights or use a convenient starting bin value. I like to have multiples of 5 or 10. If your lowest value is 152 cm you may want to start at 150cm. In the next bin type the next value based on your bin range. Going back to the example above, if my bin ranges are 5cm, then in my next cell, I will enter 155 since it is 5 greater than my lowest value of 150. Continue to add the bin ranges to the cells until you have included your highest value.

e. After you table is complete, select the “data analysis” menu option and select “histogram”. Enter the height data values in the input range including the label in the first cell. Click the labels box. For the bin range select the column you just created for the ranges. For the output options click output range and type H1 into the box (or another cell that does not contain data). Click chart output and then click OK. A histogram should appear in the same sheet as the data. *Save your file*.

f. Compare the excel generated histogram to your histogram in which you selected the bins. How are they the same? How are they different? Does one provide more insight in to the central tendencies of the heights then the other?

**Evaluate the histogram. (5pts)**

1. What is the shape of the height distribution? Is it evenly distributed? Is each bin range represented equally? This would produce a flat plateau and is referred to as uniform distribution. Does the distribution look like a bell? This is called a normal distribution. Is the curve skewed to one side so that the sides of the bell are not equal? Is the distribution an upside down bell like in Graph 1? This is a type of bi-model distribution.
2. What can you observe about the central tendency? Can you predict the mean of the data by looking at the histogram? Is the peak of the curve sharp indicating that the central tendency is narrow? This usually indicates that the mean, mode and median are close together. Is the peak spread out or difficult to distinguish? This usually indicates that the central tendency is not strong or that the mean, mode and median are not close together.
3. What can you observe about the variation or spread of the data? The standard deviation of a bell curve or normal distribution is the point of inflection in the curve. Is the point of inflection the same as your calculated standard deviation from Lab 1? Does the graph form a sharp peak or a rolling hill? A sharp peak would indicate that the heights are closely clustered and not widely dispersed. A rolling hill would indicate that the heights are spread widely in relation to the mean and rather dispersed.

Now produce a histogram for your wrist circumferences. Create your own bin ranges and answer the posed above for the wrist data.

**Compare the histograms of the heights and wrist circumferences**. (6pts)

1. Which data are more normally distributed? How did you come to this conclusion?
2. Which data are more widely distributed? How did you decide this?
3. Looking at the two histograms, what conclusions can you draw about heights and wrist circumferences? Consider the fact that these results came from the same people.

**Making your own graphs- Scatter plots**

Now you will have the opportunity to see the relationship between height and wrist circumference.

On your data sheet, click the insert tab on the top of the menu bar and select the scatter plot with only markers. Now you should see the design menu pop up. First select your data. In the top menu click on “select data”. A pop-up menu will appear. Click the “add” tab in the left box. For the series name type Height vs. Wrist Circumference. For the series X values enter the height data (do not include the top cell with the column name). You can type the cells manually or click the icon in the box and click and drag the mouse to select the height data. Click the icon when all the height data is selected. Now enter the wrist circumference cells for the “series Y values”. When you are finished click OK and a chart will appear on the worksheet.

On the chart layout menu click the first chart to add axis titles. For the Y-axis click on “axis title”,

backspace the letters and type “Wrist Circumference (cm)”. On the x-axis type “Height (cm)”. *Save your work.*

Now when you look at your chart, you will notice that excel set the axes to start at zero. We do not need all the wasted space so you can adjust the axes. To do this, right click on the y-axis and select “format axis” at the bottom of the pop-up menu. Click the “fixed” option for “minimum” at the top of the menu and type a multiple of 5 that is lower than your lowest data point. When you click “close” your y-axis should now have 5 mm as the lowest value. Do the same for your x-axis.

To add a best fit line right click on the data points in the chart and select “add trendline” from the pop-up menu. Select “linear” from the “trendline options”. Also select display equation and display R-squared value on chart. When you click “close” a line with the line formula and R2 value appear on the chart. Click and drag the formula and R2 value off the data points and drag it next to the chart title. The R2 value tells you how close the data points fall on the line. The closer to the value “1” the R2 value is, the better the data points fit the line. The closer to zero the R2 value is, the less the points fall on the line. *Save your chart.*

**Analysis of Height and Wrist Circumference**. (4pts)

1. What conclusions can you draw about height and wrist circumference? Is there a correlation? If there is a correlation, is it a positive or negative correlation? How did you decide this?
2. Do you think a line best describes the relationship? Why or why not?