Statistics: An Introduction

In last week's activity, we discussed the role that systematic and random errors play in a scientific investigation. While the presence of systematic errors means that your data from any experimental run will be biased in a particular direction, the presence of random errors means that data from successive runs of the experiment will always vary, even if the exact same procedure is followed. This variability in the outcome results in a distribution of data points if the experiment is run multiple times. The question of how to compare a distribution of data points to the single data point output by a theoretical model is answered by using the mathematical field known as statistics.

Statistics are very commonplace, and we're exposed to them every day. They allow us to summarize large volumes of mathematical information and are thus found wherever there's data to be presented. The most recognizable use of statistics for the typical college student is when exam grades are posted. You want to know two things: your score and the class average. The class average matters because it is an easy way for you to compare your performance with the "average" student in the class, and to get an idea for the potential for the "curving" of exam scores.

If you're a sports fan, statistics can be important when comparing players' performance, determining success or failure in fantasy sports leagues, and in the all-too-frequent debates on the greatest players of all time. They allow you to quickly summarize an entire player's career or performance into a few numbers, and to make

Fig. 1: Average 8th grade science scores (*DOE*)

comparisons between players for the same statistic. Statistics also play a prominent role in politics, where supporters or detractors of legislation often use statistics to back their position and discredit their opponents' point of view. When politicians are proposing tax cuts, opponents of the cut will state the vast majority of the benefit goes to the top 1% of taxpayers, while proponents will state that the average taxpayer stands to have their tax burden reduced by a sizable amount. In many of these cases, both sides are using the same data source and are technically telling the truth, but the impression one gets from the two points of view is radically different. Advertising works in much the same way, where commercials will present statistics from industry ratings (automobile ratings, consumer research studies) that show their products favorably and their competitor's products unfavorably. Curiously enough, their competitor will sometimes run a commercial giving the exact opposite impression while citing the same source.

The gathering and analysis of data forms the backbone of science, and so statistics obviously play an important role in this discipline. Statistics are used to describe numerous parameters (temperature, growth rates, age, etc.) and to make comparisons between experimental groups. In this week's exercise, you will be given an introduction to statistics, with emphasis on the statistical measures you will be using throughout the semester to analyze and present data. We will begin by examining descriptive statistics, which are used to summarize groups of data. We will then cover data presentation, in which the proper methods for displaying data in tables and figures will be described. You will frequently be using the skills
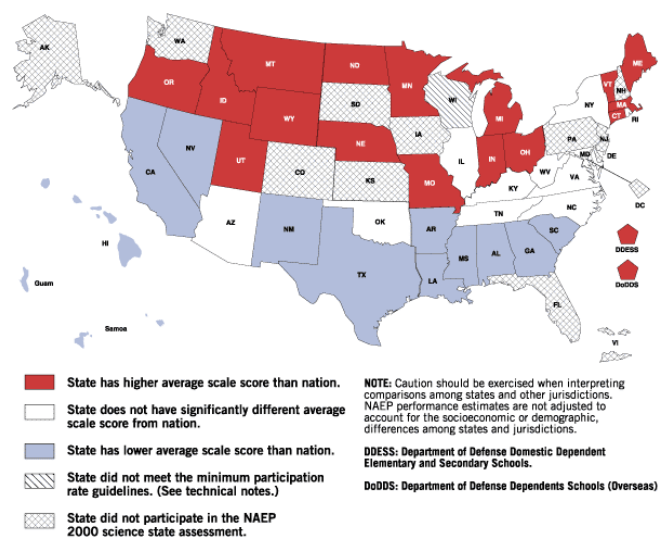
you learn here throughout the semester, so you are strongly encouraged to review the material in this exercise diligently.

Descriptive Statistics: Measures of Central Tendency

When looking at a data set, we often are interested in knowing the "center" of the group of values we're examining. If you are evaluating your exam score, you want to know the "average" score for the class. If you are predicting the winner of a basketball game, you want to know the "average" heights of the two teams. There are three such measures of central tendency: the mode, the median, and the mean. To evaluate these measures, let's examine a set of exam scores. To protect the innocent, we will refer to the different students by a coded letter.

To examine measures of central tendency, it is helpful to arrange the values in descending order.

97, 93, 90, 88, 88, 85, 83, 82, 77, 75, 74, 71, 69, 66, 59

Student	Score (pts.)
a	93
b	88
c	77
d	85
e	74
f	69
g	90
h	66
i	82
j	88
k	59
l	83
m	75
n	97
o	71

Mode

The mode is the value that appears most frequently in your data set. The term "mode" comes from the French word for "fashion", as the most fashionable clothes are the ones seen most commonly on the street. The mode of a data set does not need to be near the center of the data set, it simply has to be the most common. It is important to note that some data sets can have more than one mode if two or more values appear the same number of times. Modes are useful in that they tell you the most common value in the data set, and can shed some light on the data set's tendencies. Modes give you an idea of the tendencies of the data set by showing you which value is the most common, but the mode is not always representative of the value that forms the "center" of the data set. To determine this, we need to examine the median or mean of the values.

PRACTICE QUESTION #1:

Determine the mode of the data set below. Answer provided at the end of this section.

Data set: 97, 93, 90, 88, 88, 85, 83, 82, 77, 75, 74, 71, 69, 66, 59

Median

The median is the value that occurs in the middle of the data set. To determine the median, examine the values in descending order and find the value in the exact middle of the data set. If the data set contains an odd number of values, the median will be the middle value. For example, in an ordered data set of 19 values, value #10 would be the median as there would be nine values below it and nine values above it. In a data set of 20 values, the median would be the value halfway between the tenth and eleventh values.

PRACTICE QUESTION #2:

Determine the median of the data set below. Answer provided at the end of this section.

Data set: 97, 93, 90, 88, 88, 85, 83, 82, 77, 75, 74, 71, 69, 66, 59

Mean

You are already familiar with the mean of a data set, commonly called the "average". To calculate the mean you sum all of the values and divide by the number of values. While the terms "mean" and "average" are synonymous in common language, we will not refer to the mean as the average, as the median also gives a good representation of the average, and yet is a different statistic.

PRACTICE QUESTION #3:
Determine the mean of the data set below. Answer provided at the end of the exercise.

Data set: 97, 93, 90, 88, 88, 85, 83, 82, 77, 75, 74, 71, 69, 66, 59

Comparing Measures of Central Tendency

So which measure should I use? If I have to describe the central tendency of a data set, which statistic should I use? Well, it depends. Let's compare the three statistics for the exam score data set.

Mode	88
Median	82
Mean	79.8

In most cases, the mode is usually not representative of the central value in a data set, so it is rarely used in this capacity. Values can be the most common but rather unrepresentative of the middle of the data set, as in this case. When deciding whether to use the median or mean, it often falls to a value judgment on your part. Which one do you think is most representative of the central tendency of the data set? In this example, the two values were quite similar, but this is not always the case. Let's take the original data set, make a change, and see how this affects the two statistics. Let's change the lowest exam score from a 59 to a 17, and see what happens.

Old data set: 97, 93, 90, 88, 88, 85, 83, 82, 77, 75, 74, 71, 69, 66, 59
New data set: 97, 93, 90, 88, 88, 85, 83, 82, 77, 75, 74, 71, 69, 66, 17

If we calculate the statistics for the new data set we get:

Mode	88
Median	82
Mean	77.0

Note that the difference between the median and the mean went from 2.1 points to 5 points with a change in only one grade. This shows the influence of "outliers" on means. Outliers are values that fall well outside the range of the other values in the data set. Because means are obtained by adding all of the values together and dividing by the number of values, means will be heavily influenced by outliers. Medians, which fall in the middle of the data set regardless of the magnitude of the largest and smallest values, will be less affected. So medians are the way to go then, right? Not necessarily...

Let's look at a totally different group of exam scores, already ordered for you:

All new data set: 97, 96, 94, 94, 93, 92, 91, 90, 71, 70, 70, 69, 69, 67, 59

If we calculate the statistics for this data set we get:

Mode	94, 70, 69
Median	90
Mean	81.5

As this example illustrates, the median can also be unrepresentative of the central tendencies of a data set. While many students scored in the 90-100% range, there were many students with much lower scores, and the median doesn't give you any indication of that. At the same time, the mean value for this data set is 81.5, and yet not one of the 15 students scored even remotely close to this value on the exam. So what have we decided is the best way to show the central tendency of a given data set? We've seen that modes, medians, and means all have their shortcomings, and the choice of the most appropriate statistic often depends upon the data set. What we really need is a statistic that gives us an indication of the data's "center", and another that describes the "spread" of the data around that value.

Descriptive Statistics: Measures of Dispersion

We saw in the previous sections that having an idea of the spread (dispersion) of data points is necessary to accurately interpret a data set. If you are traveling to a faraway city and are given only the mean daily temperature, you will have no idea of the temperature extremes (high and low temperature for the day), and so would have little idea of the types of clothes to pack. An understanding of data dispersion can therefore be quite important. In this section, we will look at three measures of dispersion: the range, variance, and standard deviation.

Range
The range describes the highest and lowest values in a data set. The range is commonly used in weather reports, where the daily high and low temperatures are reported. The range tells you that all of the values in the set fall within two values, and so gives an indication of the spread of the data.

PRACTICE QUESTION #4:
Determine the range of the data set below. Answer provided at the end of the exercise.

Data set: 97, 93, 90, 88, 88, 85, 83, 82, 77, 75, 74, 71, 69, 66, 59
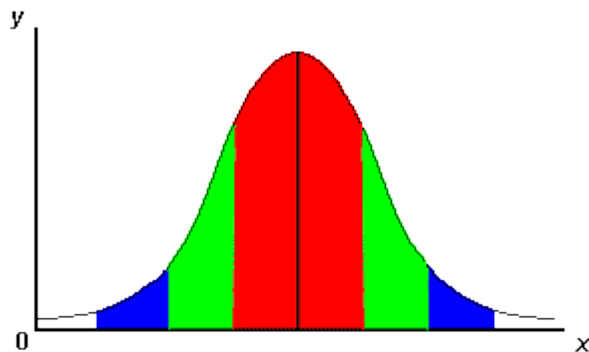
Variance and Standard Deviation
While the range is useful for describing the "borders" of a data set, it tells us almost nothing about the points that fall between the two extremes. In most cases, we are interested in knowing how far each of the data points is from the mean or median of the data set, as this would enable us to see the spread in all of the data points. Since the median doesn't take the magnitude of each data point into account and the mean does, the mean is the better option for a central point. So let's determine the individual point deviations for the original exam score data set by subtracting each score from the mean.

Value	97	93	90	88	88	85	83	82	77	75	74	71	69	66	59
Deviation	-17.2	-13.2	-10.2	-8.2	-8.2	-5.2	-3.2	-2.2	2.8	4.8	5.8	8.8	10.8	13.8	20.8

If we sum all of the deviations from the mean, we end up with zero (which shouldn't be surprising given how a mean is calculated), and this tells us nothing about the data's dispersion. To get around this, we need to look at the absolute values of the deviations, and one way to do this is to square each of them and add them up. If we divide this sum by one less than the number of values, this will give us the variance of each data point. The variance describes how far each value is from the mean. So our problem is solved, right? Wrong. While the variance gives you an indication of the deviation of each value from the mean, the variance is in "squared" units and it is difficult to relate to the "unsquared" mean. We can undo the squaring by taking the square root of the variance, which will provide us with a measure of data dispersion in the same units as the mean. This value is known as the standard deviation (SD).

Coupled with the mean, the standard deviation is useful because it allows us to examine the dispersion of a data set quickly and easily. Data that have a bell-shaped distribution when plotted, such as those shown in Figure 2, are said to have a "normal" distribution. A normal distribution is symmetrical, with the majority of the points near the center (mean) and fewer as you progress away from the center. If the data set resembles a normal distribution, we can conclude that about 68% of the data points will fall within one SD of the mean (area indicated in red), 95% will fall within two SD's of the mean (red area + green area), and 99% will fall within three SD's of the mean (red area + green area + blue area). By knowing the mean and the spread of points around it, we can get a good indication of the characteristics of a data set.

For the data set above, for example, the SD = 10.8 points, so we can conclude that 68% of the scores fell between 69.0 and 90.6 points.

Figure 2. Normal distribution with standard deviations

NOTE: When listing the mean and standard deviation of a data set, convention dictates that you list the mean first, followed by the standard deviation in parentheses with a plus/minus sign.
The mean and SD for the exam scores example would therefore be written as: mean (SD) = 79.8 (±10.8)

Central Tendency and Dispersion
By this point, you've seen that there are several measures of central tendency and dispersion for data, and that each one can be inappropriate for certain data sets. The point of data analysis is to present the data in a concise but clear manner, so that you have sufficient information to make conclusions. While presenting all of the statistical measures we've discussed for a data set might be informative, the sheer volume of information hinders interpretation. For every data set, you must strive to describe the data in the most effective manner possible. We've seen that the mean, coupled with the standard deviation, is a good way to summarize a data set as it provides the reader with a measure of both the central tendency and the dispersion of the data. You should not, however, always limit yourself to these two statistics. If you determine that the median, range, mode, or other statistic is useful for a particular data set, you should provide it.

Statistical Tools
Calculating descriptive statistics by hand can be a daunting task, particularly when large numbers of data points are involved. To facilitate data analysis in the lab exercises this semester, we suggest you use a statistical analysis tool. Microsoft Excel can calculate the descriptive statistics in this exercise, as do some online tools. Some examples are listed below.

VassarStats	QuickCalcs
Online statistical package	Online statistical package
Vassar College	*GraphPad Software*
http://vassarstats.net/	http://www.graphpad.com/quickcalcs/CImean1.cfm
choose "Miscellanea" then "Basic Sample Stats"	

Presenting Data: Tables
You have seen the value in using descriptive statistics to summarize data sets, and we will now examine the various methods available to visually present data. We will describe how to choose the proper format for presenting data, and then detail the procedures for constructing tables and graphs.

A data table, if properly designed, can be a highly effective way to organize data. For example, if we are conducting an experiment in which we will be monitoring oxygen concentrations in samples of pond water over 45 minutes, we would design a table for our data like the one below. Like all tables, the data would be entered in the appropriate cell, referenced by column and row headings.

Time (min.)	Darkness	Ambient Light
0		
5		
10		
15		
20		
25		
30		
35		
40		
45		

Table 1. Oxygen concentrations (mg per L) over 45 minutes in pond water samples exposed to darkness or ambient light levels.

Note that there are certain formatting characteristics for an effective table:
(a.) Each table is numbered sequentially, beginning with Table 1.
(b.) The title of the table should be descriptive, allowing readers to interpret the table based on the title alone. It should include both the dependent and independent variables when applicable.
(c.) The cells in the table should be roughly equally spaced for an organized look, and the column and row headings should stand out against the data.
(d.) The units for the data should always be identified, either within the table or in the title. We would then enter data into the table during the experiment.

Time (min.)	Darkness	Ambient Light
0	2.1	2.1
5	2.1	2.1
10	2.1	2.2
15	2.0	2.2
20	2.0	2.2
25	2.0	2.3
30	1.9	2.3
35	1.9	2.3
40	1.9	2.4
45	1.9	2.4

Tables can be an important tool for organizing and recording data during an experiment, but also can serve role in presenting data. For example, let's say that we wish to present the mode, median, and mean exam scores for two groups of students taught with different educational approaches. One group of students examined smog formation with a hands-on laboratory activity, while the other used a computer

simulation. Both groups were given the same exam afterwards and the scores compared in the two groups. Compare the two ways to present this information - first as a table, and then in paragraph form.

	Laboratory activity	
	Hands-on	Simulation
Mode	76	79
Median	83.3	84.0
Mean	84.1	82.6

Table 2. Exam scores (percent correct) for students using hands-on lab activity and computer simulation.

"The mode score for the hands-on group was 76%, and 79% for the simulation group. The median score for the hands-on group was 83.3%, and 84.0% for the hands-on group. The mean score for the hands-on group was 84.1% and 82.6% for the simulation group".

When comparing the two approaches, notice that although they both present the same exact information, the data is more easily interpreted in the table. Because of this ability to organize information and present values, tables are widely used in scientific reports and articles to present data. You will similarly find many instances throughout the semester in which this method of presentation will prove useful. While tables can be an effective way to record data and present certain types of information, there are times when a figure (graph) is more appropriate.

Presenting Data: Figures
Figures include illustrations, diagrams, maps, and graphs, and can be extremely useful in scientific writing. Humans are a visual species (hence, "a picture is worth a thousand words"), and figures are therefore beneficial in that they depict information visually. Let's look at the different types of figures often used in scientific writing.

Presenting Data: Illustrations, Diagrams, and Maps
The ultimate goal of science writing is to present information to the reader concisely, but completely. The use of illustrations, diagrams, and maps can be a great way to efficiently present large amounts of information, and should be used whenever appropriate. Illustrations can be used to identify structures in an organism, or to show a new species or experimental apparatus. Diagrams are often used to show the relationship between variables in complex systems, or to visually depict a chemical reaction or other sequence of events. Maps are most commonly used to show the locations of sampling sites when experiments are conducted in the field. While using these types of figures can be helpful, their overuse can be detrimental. You would not, for example, include an illustration of a rainbow trout in a report on the study of the effects of low pH water on this species. Everyone reading your work should know what a rainbow trout looks like, and the inclusion of the illustration is therefore simply wasted space. If an illustration clarifies a concept or provides valuable information, include it in your report. If it simply adds "visual appeal", omit it. Don't be afraid to use illustrations, diagrams, and maps, just use them wisely.

Formatting characteristics for illustrations, diagrams, and maps:
(a.) Each illustration/diagram/map is numbered sequentially, beginning with Figure 1.
(b.) The title of the illustration/diagram/map should be descriptive, allowing readers to interpret the table based on the title alone.
(c.) The illustration/diagram/map should be of high quality and appropriate size.

Graphs
The type of figure you will use most frequently is a graph. Like tables, graphs present data to the reader, but do so in a visual manner that it often superior to tables. Graphs can be an effective way to present

information, but it is important that they are properly constructed. Let's look at a sample graph to demonstrate the basic characteristics of a graph. Note that the independent variable is graphed on the x-axis and the dependent variable on the y-axis, the x-axis (horizontal) and y-axis (vertical) are labeled, units are listed where appropriate, and the scale of the y-axis is tailored to the data set. If "time" is a factor in your graph, it should be graphed on the x-axis, as shown in Figure 4). Recall that graphs, illustrations, diagrams, and maps are all called "Figures", and should be numbered in the order in which they appear in the report. Graphs should also be of high quality and appropriate size.
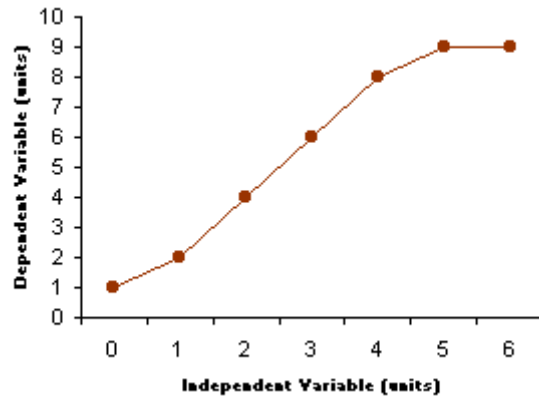
Fig. 3: Sample graph

The scaling of the axes is an area that tends to give students trouble, and is deserving of further discussion. The scale on an axis should cover the range of values in your data set, and does not have to start at zero. If you are graphing a set of values that range from 70 to 85, the scale on the axis should run from about 65 to 90 (start slightly below lowest value and slightly above greatest value). The scale should also cover the entire height or width of the axis, so you should begin the scale at one end of the graph, end it at the other end, and divide the increments between equally. While this scaling is often automatically done by software applications with graphing capability, you should nonetheless be diligent in checking them to ensure they are correct. As shown in Figure 4, scaling can make a big difference in data interpretation. The graph on the left seems to suggest no difference in pH level in the two treatments, but with proper scaling the differences in the groups become apparent.
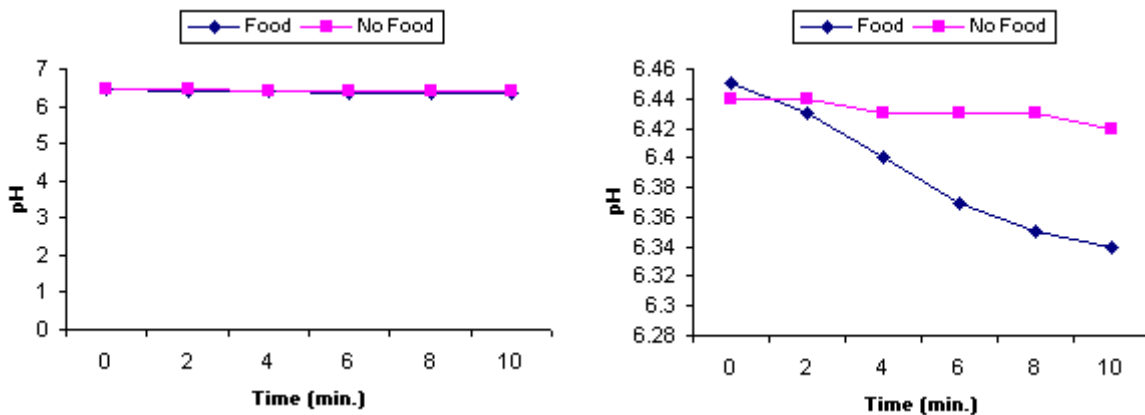

Figure 4. Effects of axis scale on data presentation

There are exceptions to this rule, though. If you are comparing percentages (e.g., exam scores), it is conventional that you make the axis run from 0% to 100% even if the values graphed are both near the top of the scale. When in doubt, use your best judgment.

Presenting Data: Types of Graphs
Although a variety of graphs exist, we will focus here on the two types that you will use most often in this course: bar graphs and line graphs. Examples of each type of graph are shown below.
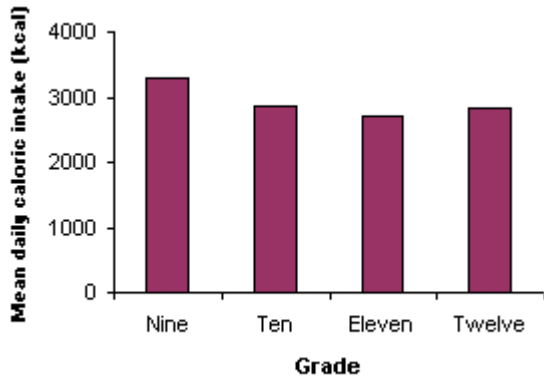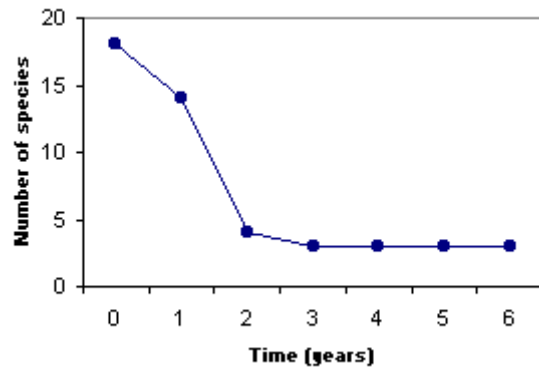
Figure 5. Sample bar graph

Figure 6. Sample line graph

Choosing the appropriate type of graph is often difficult for students, even if provided with only two choices. Let's therefore outline a few criteria regarding graph choice.

When Bar Graphs are Useful

If you can break your data into discrete groups like those shown here, then a bar graph is appropriate. Bar graphs are useful for graphing non-continuous data, such as data from different experimental groups. Note that in each case, the dependent variable is on the y-axis and the independent variable on the x-axis, and all of the formatting issues listed above are addressed.
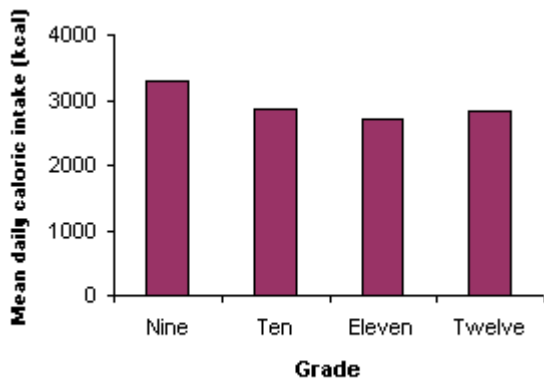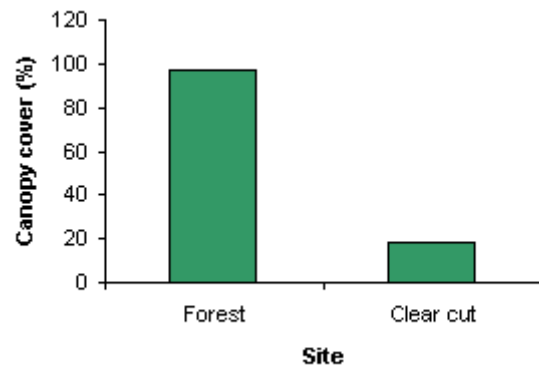

Figure 7. Sample bar graph

Figure 8. Sample bar graph

When Line Graphs are Useful

If your data are continuous (each point is directly related to the next and can be connected by an infinite number of intermediate points), then a line graph is the way to go. Line graphs are commonly used in scientific studies to present data when time (a continuous variable) is one of the variables involved. As before, note the dependent variable is on the y-axis, "time" is on the x-axis, and all of the formatting issues are addressed.
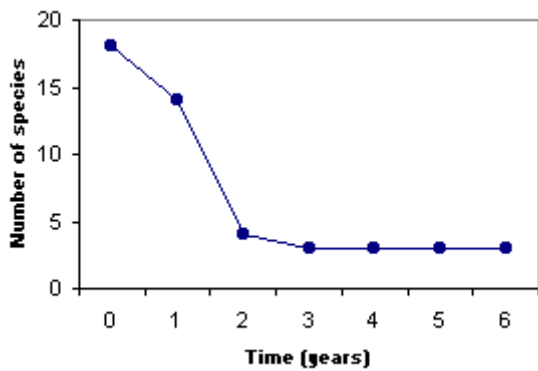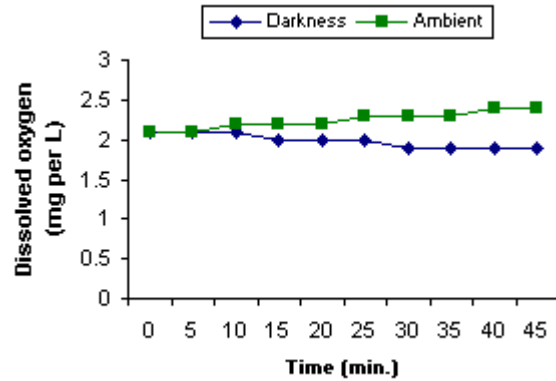
Figure 9. Sample line graph

Figure 10. Sample line graph

Scatter Plots

While bar and line graphs are used commonly, there is another type of graph worth mentioning - the scatter plot. In some cases you may need to graph two variables against one another to determine their relationship. You graph one variable on the X-axis and the other on the Y-axis, and then graph your points accordingly. By doing this, you can see if one variable increases as the other increases (positive relationship), one increases as the other decreases (negative relationship), or if the two variables show no relationship to one another. For example, assume that you are given the data below and you want to find the relationship between the two variables.

Variable 1	5.5	8.1	2.7	3.0	7.0	3.5	2.0	1.5	7.4	4.2	2.4	2.9	5.6	6.4	1.9	6.1
Variable 2	11.0	17.6	5.1	5.0	14.0	7.5	3.8	4.1	14.0	8.4	6.8	5.1	12.9	10.6	12.8	3.8

If you place variable 1 on the X-axis and variable 2 on the Y-axis and graph the points, you get a graph that looks like this.
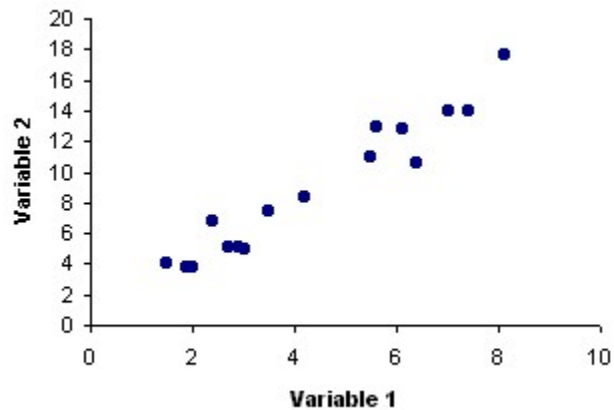

Figure 11. Sample scatter plot

Note that the points form an upward-sloping line, which indicates the two variables have a positive relationship to one another. Had the line sloped downward, this would suggest a negative relationship. If the points were scattered about and no clear relationship was visible, this would mean the two variables are unrelated to one another. The relationship can be indicated by drawing a "best-fit line" through the

points. The line should go through all of the points in such a way that it minimizes the distance between each point and the line.
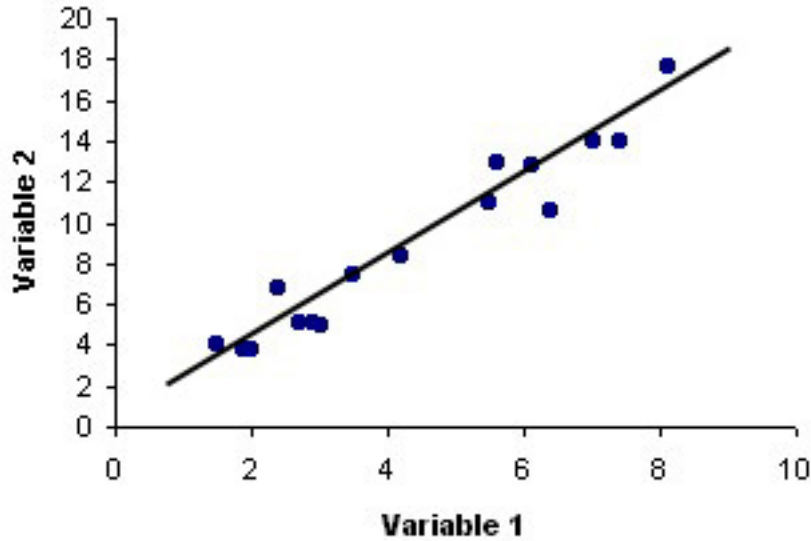

Figure 12. Scatter plot with best-fit line

Presenting Data: Choosing the Values to Graph
Students can also have difficulty with graphing because they are unsure exactly what data to graph. Let's examine an example to show this.

Let's assume that a researcher is attempting to determine the effects of pesticide pollution on the hatching success of fish eggs. He has noticed that fish eggs in streams near agricultural fields fare poorly, while those in undisturbed areas hatch successfully. The researcher sets up a lab experiment in which ten groups of fish eggs are allowed to develop in unmanipulated stream water, and ten groups of eggs develop in stream water with the addition of pesticide. The proportion of eggs hatching in each group is tallied (Table 3), and descriptive statistics for the two groups compared.

Group	Proportion eggs hatching	
	No Pesticide	Pesticide
1	0.80	0.50
2	0.76	0.45
3	0.81	0.68
4	0.90	0.77
5	0.95	0.64
6	0.84	0.60
7	0.88	0.54
8	0.99	0.57
9	0.86	0.62
10	0.93	0.57

| Statistic	No Pesticide	Pesticide
Mean proportion eggs hatching	0.87	0.59
Standard deviation	0.07	0.09

Table 3. Effects of pesticide pollution on hatching success of fish eggs.

Now that we have the data, how do we graph it? First, we determine that a bar graph is most appropriate, as the independent variable is the two groups, and this is categorical, not continuous. Second, we must determine what to graph. If we graph all of the values, we get a graph like the one below.
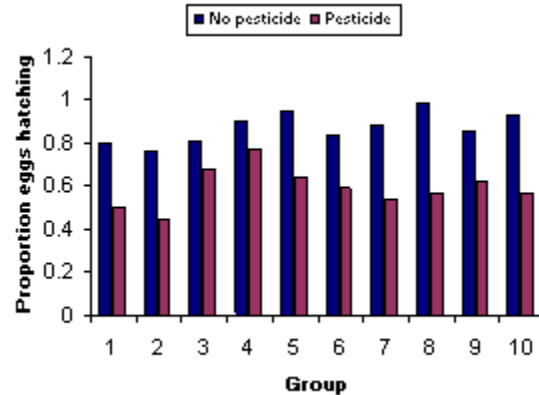
Figure 13. Effects of pesticide pollution on hatching success of fish eggs.

But what does this graph tell you? There are 20 bars, 2 categories, and lots of different group numbers of mix things up. If we take a step back and look at the study, what we're really interested in doing is comparing the hatching success in groups exposed to pesticide and the groups not exposed to pesticide. So what we really want to do is to graph the mean of the two groups and compare them to one another. If we do that, we get a graph that looks like:
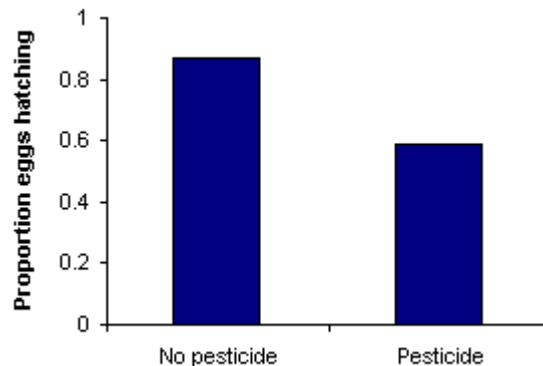
Figure 14. Effects of pesticide pollution on hatching success of fish eggs.

Now while this graph is useful in that it readily allows us to compare the means of the two groups, it doesn't provide us with any indication of the spread of the data in each group. Knowing the spread of the data can be important in making conclusions, so you need to provide the reader with this information. This is done by adding "error bars" to data points or bars on a graph that indicate the standard deviation of the data. In this example, we would add the standard deviation error bar to the bar for each group. In

bar and line charts, error bars are drawn both above and below the mean to a distance equal to their value.
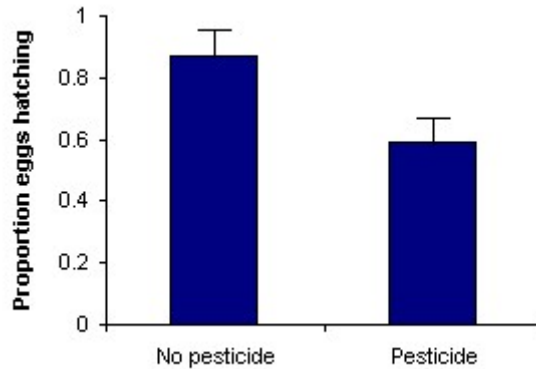
Figure 15. Effects of pesticide pollution on hatching success of fish eggs.

Realize that error bars are only appropriate when graphing a group's mean. If you are graphing individual data points, you cannot add error bars as there is only one data point per point on the graph (hence, no mean or standard deviation). Error bars are useful in that they allow you to see the overlap in the data points from two groups. Remembering that all of the data points in a distribution are within two standard deviations of the mean, you can visually estimate the data's distribution by doubling the height of the error bar (because it's only drawn to one standard deviation) for each group and seeing the degree to which their distributions overlap. As shown in the examples below, including error bars can make a big difference in your interpretation of a graph. Although the means of the two groups are identical in each of the graphs, the two groups in the first figure appear to be more different from one another than the two in the second figure due to the smaller error bars and lesser degree of overlap.
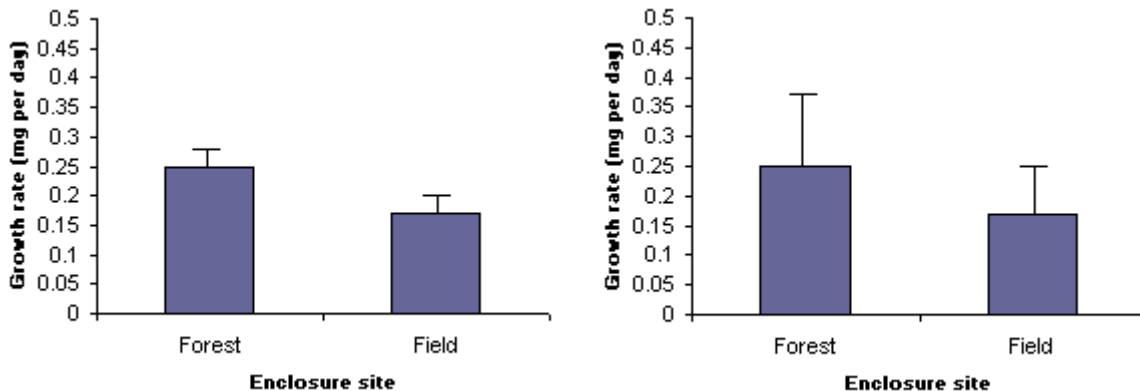
Figure 16. The role of error bars in estimating differences in treatment groups

Creating Graphs
You may choose to make your graphs by hand or with the assistance of a computer software application. In either case, you need to make sure your choice of graphs, the data to graph, and the formatting are all

appropriate. There are several software applications that enable you to create graphs (e.g., Microsoft Excel) and they are often available in campus computer labs. There are also some web sites that allow you to create graphs online, and they are listed below. Providing directions for graph preparation in each would be prohibitively long, so you will need to become familiar with their operation on your own if you opt to use them.

Create a Graph
Online graphing application
National Center for Education Statistics
http://www.nces.ed.gov/nceskids/graphing/

Practice Question Answers

PRACTICE QUESTION #1: Mode = 88

PRACTICE QUESTION #2: Median = 82

PRACTICE QUESTION #3: Mean = 79.8

PRACTICE QUESTION #4: Range = 97 – 59

Activity
In this week's activity, we are going to measure the amount of electricity that is consumed by the major appliances in your home for five days. The amount of electrical energy that you use every day is variable and depends upon a great many factors. To see how much variance there is in that data, we are going to repeat the same measurement of electrical usage over a 5-day time period. The procedure for doing this is simple if you have a few basic facts about each appliance.

To calculate the amount of electrical energy used by your appliances, you need to know 1) the power of the device and 2) the amount of time that the device was on each day. To get the power of the device, you should look for the label on the appliance that lists the pertinent information about the appliance, such as model number, serial number, etc. This label should either list the power in watts, or it should list the amount of current used by the device. If all it lists is the current, you can calculate the power by multiplying the current by 110 V if it has a normal plug or by 220 V if it has the specialized four-prong large plug of a big appliance (dryer, air conditioner, etc.).

Either of these should give you the power in watts. If you convert the amount of time used to hours, then multiplying the power by the time will give you the electrical energy in watt-hours. You can convert this to kilowatt-hours by dividing your answer by 1000. As an example, imagine that you find that your toaster oven has a label that says it has 5 A of current. Since the toaster has a 110 V plug, this means that the toast uses

P = (5 A)(110 V) = 550 W

If you use this for 30 minutes a day, then you use

E = (550 W)(.5 hr) = 225 Whr

This is converted to kilowatt-hours by dividing by 1000, which gives

E = .225 kWhr

Once you made these measurements, you are then going to calculate the median, mean, and standard deviation of the data. The median will require no calculation, actually, as it will just be the middle data point when you order the energy usage from smallest to largest. The mean is easily calculated by adding up each daily total and dividing by 5. In order to calculate the standard deviation, you can either do it the straightforward way (subtract the mean from each daily total, square the differences, add them up, divide by 4, and then take the square root), use one of the online calculators above, or put the data into Excel and use its equation function.

Once you have this data, add up the energy for each day and make a bar graph of the data. Compare this to the national average of about 18 kWhr of electrical energy usage by appliances in a household.

© Matt Laposata and John M. Pratte

Table 1: Electrical Usage by Appliances Per Day

Appliance	Power	Energy Used Each Day (kWhr)					Median	Mean	Std. Dev.
		Day 1	Day 2	Day 3	Day 4	Day			

Data presentation: Plot a bar graph of the total energy for each day

How does your energy usage by appliances compare to the national average?