Discovering Patterns: Variation in Biological Systems

Fiddler crabs, pond snails, crayfish and isopods make good subjects for this lab.

Examine the organisms at your lab bench. Note differences in individuals. Measure a trait that might have survival value when the organism is faced with the risk of being eaten. Then proceed to the computer and organize your data in Minitab. (Instructions will be available at the computers).

I. Frequency Histogram: Visualization of the variability in a sample

One of the major goals of science is to discover and explain patterns in the natural world. A first step in looking for patterns is to organize the data into a frequency histogram (Figure 1.) The distribution of data about a mean gives a sense of how much variability there is in the parameter

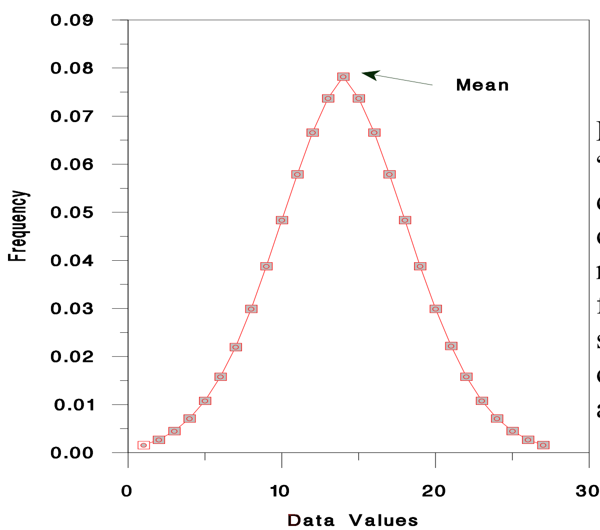

Figure 1. Normally distributed data form a "bell-shaped" curve. In this type of data distribution an equal number of data points lie on each side of the mean. If the data points are not equally distributed about the mean the frequency distribution will take on a skewed shape. If an excess of data points lie to the right of the mean the distribution is positively skewed and if they lie to the left it is negatively skewed.

of interest and the pattern of variation often leads to hypotheses to explain its source.

II. Mean: A measure of the central tendency of a sample

Variation is very important in understanding biological systems. Because it is so important, we need to quantify how much variation exists in a system and the central tendency of a set of measurements. The **mean** is the measurement that is most often used to quantify the central tendency of data. The mean is simply the average of the group, or the sum of all the measurements divided by the number of measurements. Mathematically, the mean ($\bar{x}$) is:

$$\bar{X} = \frac{\sum X_i}{n}$$

where ΣX_i stands for the summation of all the individual, measured values (X_i), n is the number of measurements and $X_1, X_2, \ldots X_n$, are the individual measurements. By convention "n" is used to represent the sample size and "N" is used to represent the number of individuals in the population.

III. Standard Deviation: One way to express variability

The mean is one way of describing a large data set but by itself it tells you nothing about the variability of the data. The **range** (high value - low value) can add information but because it depends on just 2 values, one still does not have a good idea about how the data spread about the mean. The **standard deviation** (*s*) is a measurement that describes, on average, how far each value is away from the mean. However, we can't simply take each value, subtract it from the mean, and take an average of the differences -- that average would always be zero. Why? To avoid this problem mathematicians add the squares of the differences between the mean and each value, and obtain a number that is similar to the average of the squared differences, and then take its square root. The calculation for the standard deviation is:

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

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

The standard deviation is an estimate of variability within a population. If the parameter that you measured is normally distributed, then the standard deviation can be used to make predictions about the numbers of value that fall within a specific range around the mean. For example, in a normally distributed set of measurements, 68% of the values will lie within the mean $\pm$ 1s, 95% of the values will lie within the mean $\pm$ 2s, and 99% of the values lie within the mean $\pm$ 3s.

The method of calculating standard deviation is shown below for a sample of 6 measurements:

	Measurement (heartbeats/min) X_i	Measurement minus mean $(X_i - \bar{X})$	Square of difference $(X_i - \bar{X})^2$
	72	3	9
	66	- 3	9
	52	- 17	289
	85	16	256
	78	9	81
	61	- 8	64
Total:	414	0	708

$\bar{X}$ = 414 / 6 = 69 n = 6

s = √ (708)/(6-1) = 11.9

The mean, standard deviation, and sample size are usually reported as: 69.0 $\pm$ 11.9 (n=6)

We can summarize a whole set of data as just three values: a mean, a measurement of variation such as a standard deviation, and a sample size. This is particularly useful when the sample size is very large. For example, if you measured 1000 individuals it would be much easier to understand the data by seeing the mean and standard deviation than to look at all 1000 measurements. What does a large standard deviation tell you about a population? What does a small standard deviation tell you?

IV. Hypotheses: Ideas suggested to explain variation.

Variation is characteristic of all living things. By examining the pattern of variation it is sometimes possible to suggest hypotheses to explain its source. Often more than one hypothesis can be suggested to explain a pattern. You were asked to select a variable that you thought might have survival value. If it is strongly selected for you might expect in future generations the variability in this trait would decrease as animals at one end of the trait range would be selected against (fail to reproduce). This might cause a shift in the mean and a decrease in variability. You could test this hypothesis by studying the trait through several generations. Another possible explanation for the variation is that it is related to the sex of the organism. While taking your measurements you might have noticed that males tended to be at the upper end of the frequency distribution and females at the lower end.

Our understanding of patterns advances through the process of experimentation and rigorous statistical evaluation of the data. This approach may allow us to disprove one of the hypotheses. Through this process of elimination of hypotheses we gradually build support for the hypothesis that we have been unable to disprove. This is why science can be an exciting field.

V. Experimental Design: A way of testing hypotheses about patterns

Scientists collect information to help them clarify patterns that occur in nature. High levels of natural variation can make finding differences between groups more difficult to detect. This is why it is important when designing experiments or collecting data to eliminate bias and reduce errors due to the manner in which the data are collected. This type of variation is known as **measurement error** and care must be taken to reduce it if the biologically important variation in the parameter under study is to be accurately represented.

As you decide on a question to investigate for the second half of this lab **keep it simple**. Try to state your hypothesis as an expectation or in the form of an "if...then" statement. This will help clarify the type of data you will want to collect. Data collection should be designed to eliminate an incorrect hypothesis. Make sure you all agree on how the data will be collected. Several types of organisms will be available in the lab.

Decide on a question using one of the two designs suggested below:

- **compare two groups** (eg. males vs. females, large vs. small) or use two conditions with half the individuals tested in each condition (eg. land vs. water)
- **compare the same individuals** in a before/after situation (eg. before/after exercise) or the same individuals measured in two different conditions.

Tools for measuring dependent variables include the following: rulers, calipers, balances, stop watches.

Equipment for conducting experiments: racetracks, plastic containers, aquaria

Possible Independent Variables:

- **substrate** - sand, gravel, rocks
- **habitat** - dry vs wet
- **slope of ramp** - flat vs angled
- **sex**
- **light level** - shade vs. sun
- **exercise** - standardize

Define your data collection protocol:

- what to measure
- how to measure
- unit of measurement
- elimination of bias

Design a table to collect data. When designing the table allow enough room that the data collected by your lab partners can be added to it. When the data are compiled each group will have a total of 15-20 observations for each situation. If you decide to compare two groups you will measure 15-20 individuals from each group. If you are doing a before/after design you will use only 15-20 individuals, but have 2 measurements taken on each. Record the 2 measurements so the pairs of measurement are kept together. Your table should be given a table number (eg. Table 1.), followed by a caption that describes what type of data are in the tables. Columns of data should be carefully labeled and include the unit used for measurement.

Return the organisms you studied to the container you got them from.

Assignment week 1

Make sure that you have a table in your lab notebook that contains all of the group data. Before coming to class the group data should be entered into a Minitab file and saved. Do not do any analysis at this point.

Summarize what you have done using the headings below:

- Title:
- Introduction: State the question/hypothesis and your expectation(s). One or two sentences should suffice.
- Methods: Write a thorough methods section that describes the experimental design and the manner in which data were collected. You can probably do this in less than a page.

Refer to how to write a scientific paper in your lab manual or use the information on writing papers provided online at http://biology.kenyon.edu -follow the link to Biology 109 Resources.

When you get to class next week

Discuss with the members of your group the manner in which the data were collected. Did the manner in which you collected the data limit measurement error? Did you all use the same unit of measure? Later during the period one member of each group will be asked to present the experimental design for your group. Make sure that you all agree on how the data were collected.