9. Data Analysis

Your students will need to interpret the data they collect in order to characterize the sand crab population at their beach. Organizing the raw data will allow students to identify any existing patterns, and to understand the distribution and abundance of sand crabs on a larger scale. To better visualize the data, students should begin by graphing the data. As they do, they may develop questions that could be tested using the information in the database. Skill level and the amount of time available for the project will dictate the ways by which students describe the data.

A. Graphing

Following data entry, students can generate graphs to display their results. The online system does not require a username and password to view the results. In the results section, students can select graphs for crab distribution along the beach, size fequency, and sex frequency. Students can compare their results to those of other schools monitoring beaches along the coast of California.

If necessary, spend some time reviewing line graphs and histograms (bar graphs) with your students.

B. Statistics

A critical part of ecological investigation is statistical analysis. Statistical tests evaluate significant differences or similarities between data. They are used to draw conclusions about the patterns in the data with a higher level of confidence than using simple visual analysis. Explain to your students that one can measure a difference in nature, but the difference can be due to chance variation, and not representative of actual conditons. To reinforce this idea with your students, have them all flip a coin 10 times. Simply by chance, their results can differ from 5 heads and 5 tails. When sampling a sand crab population, you could not be 100% certain of conditions unless you were to census every crab. However, statisitics allows us to estimate with 95% certainty ($P = 0.05$) what is happening in the natural environment. A common factor in statistical tests is the null hypothesis, which assumes there is no effect or difference between treatments. The researcher is looking for a difference. Your students can explore the following questions using Chi Square analysis. Sample tables and examples are included for clarification.

For a Chi Square Tutorial, visit http://www.georgetown.edu/faculty/ballc/webtools/web_chi.html

<u>B1. Is the sand crab sex ratio even or skewed?</u>

Students can use the **Chi Square (χ^2) Test for Goodness of Fit** to compare the frequency distribution of their data with a theoretical expected distribution. The Chi Square Test examines the differences between distributions of discrete data, not percentages. When testing for a skewed sex ratio, the expected frequency distribution is 1:1, or 50 percent males : 50 percent females.

Null hypothesis:

The **observed** frequency distribution (ratio) is equal to the **expected** frequency distribution (ratio).

Formula: $\chi^2 = \sum (o - e)^2 / e$

For the different distributions to be compared, square the difference between the observed value (*o*) and the expected value (*e*), and divide this square by the expected value. The figures calculated for each category are then summed to obtain χ^2.

Calculation Table:

	MALES	*FEMALES*	
observed (obs.)			
expected (exp.)			
obs. – exp.			
(obs. – exp.)^2			
(obs. – exp.)^2 / exp.		+	=χ2

Degrees of freedom:
The degrees of freedom (d.f.) is equal to the number of categories (columns) minus 1.

Interpretation:
You will now compare the calculated χ^2 to a critical value from the χ^2 Table. χ^2 tables can be found at the following web addresses. A probablity level of 0.05 should be used.

- http://www.ento.vt.edu/~sharov/PopEcol/tables/chisq.html
- http://bmj.bmjjournals.com/collections/statsbk/apptabc.shtml
- http://www.richland.cc.il.us/james/lecture/m170/tbl-chi.html

If the value calculated for χ^2 is **equal to or greater than** the critical value given in the χ^2 Table, for the appropriate degrees of freedom, the null hypothesis can be rejected. If the calculated χ^2 value is **less than** the critical value, the null hypothesis cannot be rejected, indicating there is no statistically significant difference between the ratios.

Example:
You want to know if there are more male or female sand crabs at your study beach on a given date. You randomly sample the beach, and record 52 males and 64 females. Your null hypothesis is that the sand crab sex ratio is 1 : 1 (the number of males is proportionately the same as the number of females). To obtain the theoretically expected values, sum the two observed values (52 + 64 = 116) and divide by two (116 / 2 = 58).

There are two categories (males and females), so the degrees of freedom is 1. The χ^2 Table at this d.f. and the probability level of 0.05 shows the critical value is 3.84. Since the calculated χ^2 value (1.24) is less than the critical value (3.84), you do not reject the null hypothesis. Although you collected more females than males, you cannot conclude that a significant difference exists between the number of male and female sand crabs on the sampling day. It is likely that the observed difference is the result of random variation in your collection.

<u>B2. Do sand crab sex ratios vary significantly by season or location?</u>
To test for a seasonal difference or a difference between beaches, students can use the **Chi Square (χ^2) Test of Independence**. This test can determine whether two or more frequency distributions are the same, and requires a contigency table to calculate the expected values.

Arrange the calculations in a table like this:

	MALES	*FEMALES*	
observed (obs.)	52	64	
expected (exp.)	58	58	
obs. – exp.	6	6	
$(\text{obs.} - \text{exp.})^2$	36	36	
$(\text{obs.} - \text{exp.})^2$ / exp.	0.62 +	0.62	1.24=$\chi 2$

Null hypothesis:
The frequency distribution of Sample A is equal to the frequency distribution of Sample B.

Contigency Table:

		males	*females*	Row Totals
Fall	observed			
	expected			
Spring	observed			
	expected			
Column Totals				

Degrees of freedom:
The d.f. is equal to (the number of columns minus 1) multiplied by (the number of rows minus 1).

Example:
You want to know if the distribution of sand crabs at your beach differs between Fall and Spring. The null hypothesis is # Males : # Females in the Fall is equal to the # Males : # Females in the Spring. You observe 72 males and 45 females in the Fall, and 15 males and 88 females in the Spring, and record this observed distribution in a contigency table as follows.

In this case, the expected distribution is not obvious, but must be calculated from the data. You calculate each expected value by multiplying its row total by its column total, and dividing by the grand total. For example, to determine the expected value for the Fall for males, (117 x 87) / 220 = 46.3. This calculation must be done for each row and column category.

		males	*females*	Row Totals
Fall	observed	72	45	117
	expected	46.3	70.7	
Spring	observed	15	88	103
	expected	40.7	62.3	
Column Totals		87	133	220

Now, you must calculate $(obs. - exp.)^2$ / exp. for each set of values and then total them to determine χ^2.

	FALL		SPRING		
	males	*females*	*males*	*females*	
obs. - exp.	25.7	-25.7	-25.7	25.7	
$(obs. - exp.)^2$	660.5	660.5	660.5	660.5	
$(obs. - exp.)^2$ / exp.	14.27	+ 9.34	+ 16.23	+ 10.6	50.44=χ2

The sum of these calculations is χ^2, or 50.44.
The degrees of freedom for the test is (2-1) x (2-1) = 1. Consulting a χ^2 Table at the 0.05 level and 1 d.f. results in a critical value of 3.84, which is much smaller than the calculated χ^2 of 50.44, so you can reject the null hypothesis that the ratios are equal. There are different sex ratios in Fall and Spring; there are significantly more females in the Spring.

Other Comparisons

The same test could be used to investigate differences between two or more beaches. In addition to comparing the ratio of females to males, students could also compare the ratio of adults to juveniles.