Statistical Testing for Dummies!!!
Your Idiotproof Guide to Choosing the Right Statistical Test for the Job!

Hey, there, fellow Statistical Dummies! Guess what! There are just five major statistical tests that you will want to be familiar with in your two years of Marine & Environmental Science at CBGS:

1. Standard t-test
2. Paired t-test
3. One-way ANOVA (Analysis of Variance)
4. Two-way ANOVA
5. Linear Regression

Don't let the weird names scare you! It's all just typical statistician mumbo-jumbo. But the tests themselves are powerful, valuable devices to help you in your scientific research. And a computer can do all the icky, gnarly mathematical computations for you. All you have to do is pick the right test for your particular lab experiment or field study. The statistical test that you select will depend upon your experimental design, especially the sorts of **Groups** (Control and/or Experimental), **Variables** (Independent and Response), and **Treatment Levels** that you are working with.

Keep in mind that a statistical test is always a test on your **Null Hypothesis**.

More specifically, it tests the **Probability** that your Null Hypothesis is valid.

More to the point, it tests the probability that the two (or more) **Estimated Means** from your study (and expressed in your null hypothesis) belong to the <u>same</u> "Distribution of the Means," or whether they probably belong to two or more <u>different</u> Distributions of the Mean.

In short, each of these five tests is a statistical comparison of two (or more) MEANS, the averages that you get from each separate GROUP in your experiment or field study.

1. <u>**Standard t-test**</u> – The most basic type of statistical test, for use when you are comparing the means from exactly **TWO** Groups, such as the Control Group versus the Experimental Group.

   (ex) Your experiment is studying the effect of a new herbicide on the growth of the invasive grass *Phragmites*. You have TWO groups of plants: an Experimental Group that has been sprayed with the poison and a Control Group that has not been sprayed. After you have calculated the average growth for each of the two groups, run a t-test to see if you have detected a "statistically significant" difference in their growth. The t-test returns a p-value that expresses the probability that this null hypothesis is wrong:

   $$H_o:\ G_C = G_E$$

   …where G stands for mean growth (control vs. experimental)

2. **<u>Paired t-test</u>** – An extremely powerful test for detecting differences (it is, in fact, the most "sensitive" of all our five tests). It is usually used for **"Before vs. After"** type experiments, where the ***same*** individuals are measured before and after the application of some sort of treatment. It can also be used for **"Left vs. Right"** experiments, where two sides of an individual are given two different treatments.
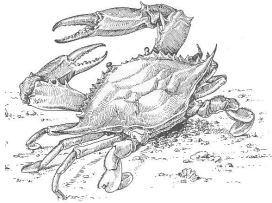
(ex) You are curious to know if depth perception is important for blue crabs in capturing small fish as prey. One day you provide a dozen hungry crabs, each in his own aquarium, with a meal of ten minnows that you release freely into the water. For each crab you record their "hit:miss" ratio, the number of times they successfully lunge with a pincher at a passing fish versus the number of times they miss the fish. You let the crabs go hungry again for several days, then repeat the procedure on <u>the same 12</u> crabs, except this time you fit each crab with a little eye patch that blocks its vision in one eye and thus ruins its 3-D depth perception. You organize your results in "before-after" <u>pairs</u>, <u>crab-by-crab</u>, and run a Paired t-test on them.   The Paired t-test returns a p-value on the validity of this null hypothesis:

$$H_o:\ S_B = S_A$$

…where S stands for mean rate of successful attacks (before vs. after)

(ex) Red snapper, a favorite target of recreational fishermen in the Gulf of Mexico, often get caught in the trawl nets of shrimp boats. Recreational fishermen don't like this! You have invented a new kind of trawl net bearing a trapdoor that is intended to let red snapper escape from the net, but without letting the shrimp escape. Shrimp boats can tow <u>two</u> nets simultaneously, one on each side. You rig a single shrimp boat with a traditional shrimp trawl on the starboard side and your new experimental trawl on the other side. You do a series of successive tows, counting the number of red snapper captured by each net on each tow. Afterwards you organize your data in left-right <u>pairs</u>, <u>tow-by-tow</u>, and run a Paired t-test on the following null hypothesis:
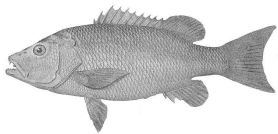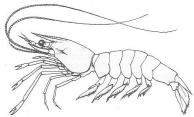
$$H_o:\ C_P = C_S$$

…where C stands for mean number of red snapper captured
(port vs. starboard)

3. **<u>One-Way ANOVA</u>** – Similar to a t-test, except that this test can be used to compare the means from **THREE OR MORE** groups (t-tests can only compare TWO groups at a time, and for statistical reasons it is generally considered "illegal" to use t-tests over and over again on different groups from a single experiment).

(ex) Sunlight is a composite of all the colors of the rainbow. Seawater absorbs some of these colors more quickly than others. Red light can penetrate
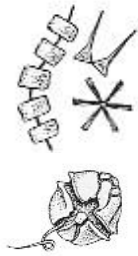
only a few meters into the ocean, yellow light a bit further, and blue deepest of all. Knowing this, you decide to see how this might affect the growth of single-celled phytoplankton ("algae") at different depths in the water column. You set up 30 beakers, each with its own culture of phytoplankton. You grow 10 cultures under a red fluorescent light, 10 under yellow light, and 10 under blue. Since oxygen is a byproduct of photosynthesis, you measure the amount of dissolved oxygen as an indicator of phytoplankton growth. After 24 hours you collect the results and run a One-way ANOVA on the following null hypothesis:

$$H_o: \ O_R = O_Y = O_B$$

…where O stands for mean oxygen output (red vs. yellow vs. blue)

4. **<u>Two-Way ANOVA</u>** – A very useful statistical test, because it's the only one that allows you to compare the means of **TWO OR MORE** groups in response to **TWO DIFFERENT INDEPENDENT VARIABLES**. With this test available, you can set up an experiment in which each member of your sample is exposed to a varying level of *two different treatments!!!* In a field study, this test allows you to compare a mean Response Variable relative to *two different* environmental conditions.

(ex) Periwinkles are snails that crawl up and down the stems of saltmarsh cordgrass to remain above the rising and falling tide. Only at low tide will they venture down onto the exposed mudflats to forage for detritus. At low tide, higher elevations of the marsh are exposed for longer periods than lower elevations. With this in mind, you wonder if the density of periwinkles in any particular portion of the marsh is more of a response to the <u>amount of cordgrass stems</u> there (= independent variable #1), or more of a response to the <u>elevation</u> there (= independent variable #2), or a partial response to <u>both</u>. To find out, you mark off 20 quadrats in a local saltmarsh, half at a very <u>Low</u> elevation near the water's edge and half at a <u>Higher</u> elevation. In 5 of the 10 Low elevation quadrats, you use scissors to trim away two-thirds of the cordgrass stems. You do the same in 5 of the 10 Higher elevation quadrats. The next day at high tide you come back and count the number of periwinkle snails clinging to the stems of the cordgrass in each quadrat. Then you run a Two-way ANOVA to test the following <u>two</u> null hypotheses:

$$H_o: \ P_H = P_L$$
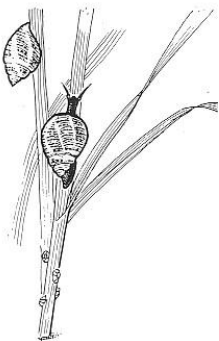
$$H_o: \ P_N = P_T$$

…where P stands for mean periwinkle density
(high elevation vs. low elevation AND normal grass vs. trimmed grass)

5. **Linear Regression** – One of the most common and useful statistical tests. This is for comparing the means of Groups along a **continuum** of **THREE OR MORE** treatment levels, such as a gradually increasing depth of water. It can also be used to compare response means under **THREE OR MORE** treatments set at **regular intervals**, such as a sequence of salinities increasing at intervals of 10 ppt. And it can even be used for a sequence of treatments at **irregular** intervals!   One of the products of a Linear Regression is a graph with a perfect **"Best Fit" Line** that passes as close as possible ("closest to the mostest") to all the data points.

(ex) Young juvenile fish often reside close to the shoreline as a safe haven from larger fish that are unable to swim into shallow water. Based on this, you wonder if juvenile perch in the Bay tend to concentrate in the shallowest waters along local sandy beaches. You and a friend use a seine net to sample the population of young perch at five different depths near shore: 1 foot, 2 feet, 3 feet, 4 feet, and 5 feet. You repeat this process at a dozen different positions along the beach, for a total of 60 tows with the seine. You count the number of juvenile perch each time. Then you run a Linear Regression on the results, to test this null hypothesis:
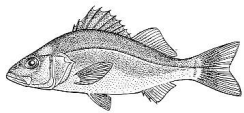
$$H_o: J_1 = J_2 = J_3 = J_4 = J_5$$

…where J stands for the mean number of juvenile perch
caught at each depth

In cases of Linear Regression, an equivalent way to state the null hypothesis is to say that the Best Fit Line will be <u>horizontal</u>, with a <u>slope</u> (M) of <u>zero</u>:

$$H_o: M = 0$$

…where M is the slope of the best fit line

(in accord with **y = m x + b**, the traditional equation for a line)

(ex) To determine whether bloodworms are osmoconformers or osmoregulators, you subject different worms to a series of different salinities: 5, 15, 25, 35, 45, 55, and 65 ppt. You bathe nine worms in each salinity and you weigh each one before and after a twenty-minute bath.   Finally you run a Regression on the percent weight changes.

$$H_o: W_5 = W_{15} = W_{25} = W_{35} = W_{45} = W_{55} = W_{65}$$

…where W stands for the mean weight change at each salinity

OR

$$H_o: M = 0$$

…where M is the slope of the best fit line