Analysis of IA data Made Easy

Adapted from Ms. Vonbargen’s website Biologyforlife.com

There are many ways to analyze data. Which statistical test to run depends on your experiment.

There are 4 statistical tests that most of the IB ESS experiments fall under:

* ANOVA
* Pearson’s Correlation
* T-test
* Chi square

**These tests tell you if there is a significant difference between your sets of data.**

So in the discussion of your paper, you need to talk about whether or not your findings were significant or not, thus accepting your hypothesis or rejecting it.

Read the tests below and then determine the one statistical test that best fits your experiment. You must include one of these tests in your IA conclusion. In order to get the full points on this section, you will have had to choose the correct test AND come to the correct conclusion based on your findings.

**T-Test**

The T-test is a test of a **statistical significant difference between two groups.**A "significant difference" means that the results that are seen are most likely not due to chance or sampling error.  In any experiment or observation that involves sampling from a population, there is always the possibility that an observed effect would have occurred due to sampling error alone.  But if result is "significant," then the investigator may conclude that the observed effect actually reflects the characteristics of the population rather than just sampling error or chance.

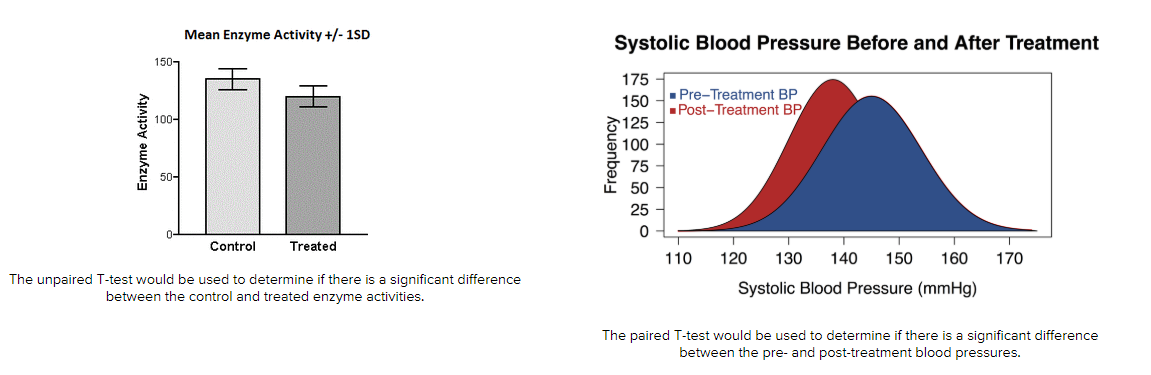
Example of when to use the t-test:

A T-test is used when looking at a numerical value—for example height—and then comparing the averages of two separate populations or groups (such as males and females).

Calculate the averages of your groups and their levels and then put these numbers into the t-test calculator <https://www.socscistatistics.com/tests/studentttest/default2.aspx>. At the bottom of the page the calculator will tell you whether or not your data is significant to the 0.05 for a one tailed hypothesis.

In your IA report both the t value and the p value (A p-value is the probability of concluding there is a significant difference between the groups result when your hypothesis is not supported (meaning, the probability of making the WRONG conclusion).  Since we use a standard “p-value” of 0.05. This means that five times out of a hundred you would find a statistically significant difference between the means even if there were none.

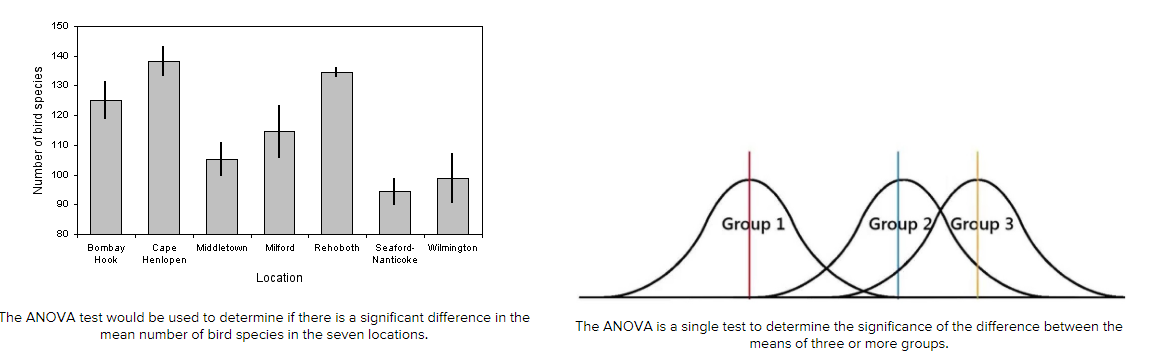
Don’t forget that you also need to graph your data, not just run the numbers in the calculator. Your graph could be a bar graph or a line graph depending on the experiment you are running.

**ANOVA**

The ANOVA test is a statistical test that can be done in place of multiple T-tests when **comparing the means of more than two groups at a time.** The t-test tells us if the variation between**two** groups is "significant".  If you have 5 five levels of a manipulated variable or more in an experiment, you will need to run the ANOVA test.

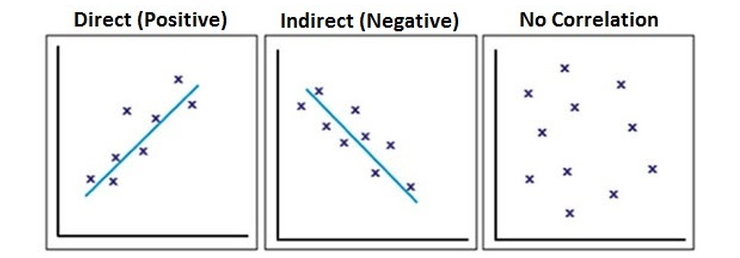
Use the calculator <https://www.socscistatistics.com/tests/anova/default2.aspx> to find out whether there is a significant difference between your manipulations (independent variables). Report the quantitative values in the analysis of your IA. See T-test for more explanation of what these numbers mean.

Don’t forget that you also need to graph your data, not just run the numbers in the calculator. Your graph could be a bar graph or a line graph depending on the experiment you are running.



**Correlation Coefficient**

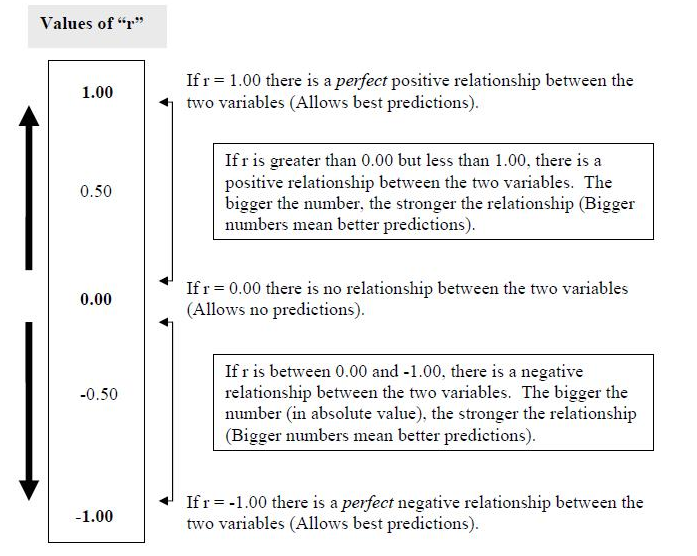
One way to get a general idea about whether or not two variables are related is to plot them on a “scatter plot”. If the dots on the scatter plot tend to go from the lower left to the upper right it means that as one variable goes up the other variable tends to go up also. This is a called a “direct (or  positive) relationship.”  On the other hand, if the dots on the scatter plot tend to go from the upper left corner to the lower right corner of the scatter plot, it means that as values on one  variable go up values on the other variable go down. This is called an “indirect (or negative) relationship."



Use the online calculator to find the correlation <https://www.socscistatistics.com/tests/pearson/default2.aspx>

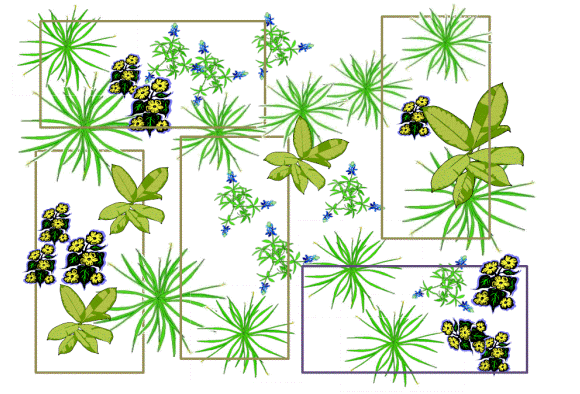
**Important Things Correlation Coefficients Tell You ​**

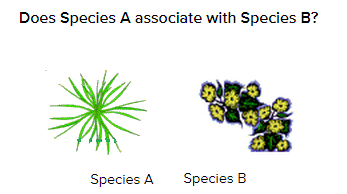
*It Tells You The Direction Of A Relationship:*  
If your correlation coefficient is a negative number you can tell, just by looking  at it, that there is an indirect, negative relationship between the two variables.  As you may recall, a negative relationship means that as values on one variable increase (go up) the values on the other variable tend to decrease (go down) in a predictable manner.   If your     correlation coefficient is a positive number, then you know that you have a direct, positive relationship. This means that as one variable increases (or decreases) the values of the other variable tend to go in the same direction. If one increases, so does the other. If one decreases, so does the other in a predictable manner.  
  
*Correlation Coefficients Always Fall Between -1.00 and     +1.00:*  
A correlation coefficient of -1.00 tells  you that there is a perfect negative relationship between the two variables. This means that as values on one variable increase there is a perfectly predictable decrease in values on the other variable. In other words, as one variable goes up, the other goes in the opposite direction (it goes down).  A correlation coefficient of +1.00 tells you that there is a perfect positive relationship between the two variables. This means that as values on one variable increase there is a perfectly predictable increase in values on the other variable. In other words, as one variable goes up so does the other.  A correlation coefficient of 0.00 tells you that there is a zero correlation, or no relationship, between the two variables. In other words, as one variable changes (goes up or down) you can’t really say anything about what happens to the other variable.  
  
*Larger Correlation Coefficients Mean Stronger Relationships*  
Most correlation coefficients (assuming there really is a relationship between the two variables you are examining) tend to be somewhat lower than plus or minus 1.00 (meaning that they are not perfect relationships) but are somewhat above 0.00.    Remember that a correlation coefficient of 0.00 means that there is no relationship between your two variables based on the data  you are looking at. The closer a correlation coefficient is to 0.00, the weaker the relationship is and the less able you are to tell exactly what happens to one variable based on knowledge of the other variable. The closer a correlation coefficient approaches plus or minus 1.00 the stronger the relationship is and the more accurately you are able to predict what happens to one variable based on the knowledge you have of the other variable.



**Chi-square test for Independence**

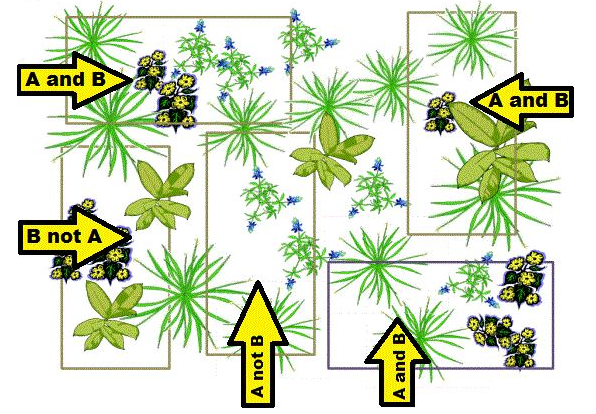
Chi-square Test for Independence is a statistical test commonly used to determine if there is a significant association between two variables.  For example, a biologist might want to determine if two species of organisms associate (are found together) in a community.



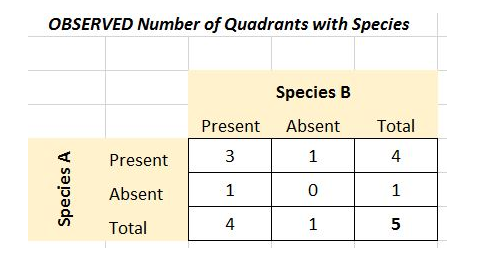


**How to Calculate a Chi-Square Test of Independence**

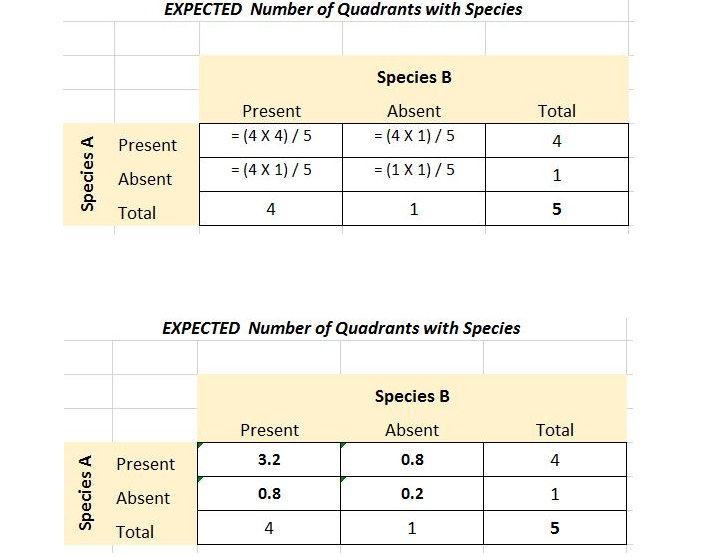
The first step is to collect raw data for the occurrence of each variable.  This is often done via random sampling using a quadrant.  In our example, there are five quadrats.  Determine:

* The number of quadrats with both species present
* The number of quadrats with Species A but not Species B
* The number of quadrats with Species B but not Species A
* The number of quadrats with neither species

Then create a "contingency table" to display your results.  In the Chi-Square test, these are your OBSERVED values.

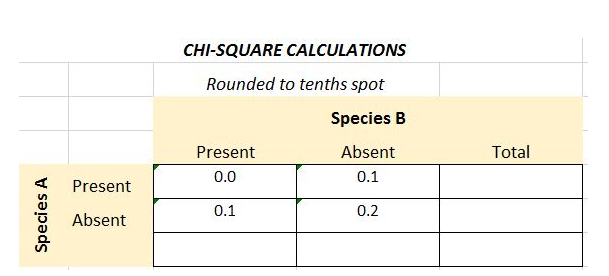


Next you need to determine what would be EXPECTED assuming the species are randomly distributed with respect to each other.   **Expected frequencies = (row total X column total) / grand total**



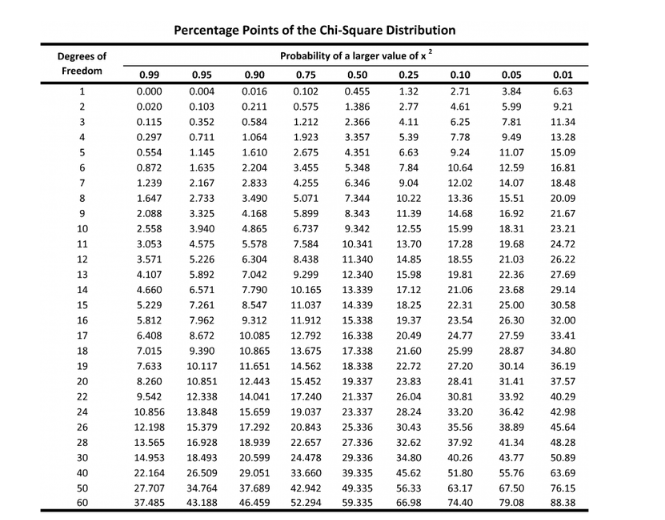
Now that you have OBSERVED and EXPECTED values, apply the Chi-Square formula in each part of the contingency table by determining **(O-E)2 / E for each** box.





The final calculated chi-square value is determined by summing the values:

* X2 = 0.0 + 0.1 = 0.1 + 0.2 = **0.4**



* If the calculated value is lower than the 0.05 level of significance, accept the null hypothesis and conclude that there is NO significant association between the variables.

* If the calculated value is higher than the 0.05 level of significance, reject the null hypothesis and conclude that there IS a significant association between the variables.  ​

​  
For example, with a DF=1, a value greater than 3.841 is required to be considered statistically significant (at p = 0.05). Since the X2 we calculated (0.4) is less than 3.841, there is NOT a significant association between Species A and Species B.  The location of Species A has no significant effect on the location of Species B, any association between species is likely due to chance and sampling error.

**Practice Scenarios**

Read the experiments below and determine the appropriate statistical test. Write the name of the test in the space.

#1 You have a group of individuals randomly split into smaller groups and completing different tasks. For example, might be studying effects of tea on weight loss and form 3 different groups: green tea, black tea, no tea.

#2 You have a group of individuals based on the attributes they possess. For example, you may be studying leg strength of people according to their weight. You split participants into weight categories (obese, overweight, normal) and measure their leg strength on a weight machine.

#3 You are trying to see if there is a correlation between income AND gender for anxiety level at job interviews. You group people into the following categories: male with low income, male with medium income, male with high income AND female with low income, female with medium income, and female with high income.

#4 You want to find out if people living in New York spend more or less money on movies than people living in Kansas.

#5 A sample of students were given a pretest before studying a particular concept. The teacher teaches the concept and then gives the test again at the end of the unit.

#6 You want to see if there is a relationship between age and blood pressure.

#7 You want to see if there is a relationship between student’s achievement motivation and GPA.

#8 You want to if chickadee birds and junco birds hang out together in the same habitat.