**Using Excel to do Basic Statistical Tests**

Before you do statistical tests, it’s important to understand what they are. This help sheet assumes that you have this basic knowledge and some general idea about how to use Excel.

Descriptive statistics

There are two major summary statistics that one usually wants. One is a measure of the central tendency of the numbers, like a mean (average), median, or mode. The second is a measure of how different the numbers are, which we measure using standard deviation, standard error, or variance.

* Enter your data in columns.
* Click on the empty cell after your last number in a column.
* Hit the = sign in the bar at the top of the spreadsheet.
* Hit the down arrow to the left of the = sign. Now some options will appear.
* Click on "average"
* If it is not on the list, click "more functions", choose "statistical", then "average".
* Do the same thing (using the next empty cell) to get "STDEV". This gives you the standard deviation.

T tests

T tests can be used to compare two groups or treatments.

* Click on any empty cell.
* Hit the = sign in the bar at the top of the spreadsheet.
* Hit the down arrow to the left of the = sign. Now some options will appear.
* If "TTEST" is not on the list, click "more functions", choose "statistical", then "TTEST".
* A dialog box will appear. Click in the box next to "Array 1".
* Drag the dialog box out of the way, then highlight your first column of numbers.
* Click in the box next to "Array 2" and highlight your second column of numbers.

To answer the "tails" question, remember your prediction about the direction of the difference between the groups. If you predicted group A would be lower than group B, pick 1 tail. If you predicted group B would lower than group A, pick one tail. If you didn’t predict which would be higher, use 2 tails. You can’t change your mind after the data are gathered.

There are three types of T test you can use on Excel. Let’s say you wanted to test whether heart rate increased after drinking a cup of hot sauce (don’t actually try this!) or whether plant growth would increase after adding fertilizer to pots of soil. In these cases you would be comparing the heart rate of the same people, or the growth of the same pot of plants before and after the treatment. This would require a "paired" or "dependent" T test. Excel calls this a "type 1" test.

Let’s look at another situation. Say you want to know whether nursing students consume more coffee than do biology students. You would then have two groups of test subjects rather than taking 2 measurements on each person. Now you would use an "unpaired" or "independent" T-test. Excel calls these "type 2" or "type 3" tests. Now the tricky part is to decide which of these to use. Are the standard deviations about the same for both groups, or are they different? You can test this statistically, but let’s just work with how they seem. If in doubt, go with "type 3" for unequal variances.

Now hit "OK" and see what the number is. This is your P-value. Remember that a P-value below 0.05 is generally considered statistically significant, while one of 0.05 or greater indicates no difference between the groups. If your number looks like this: 2.03188E-7, Excel is giving you the number in its version of scientific notation. This number is actually 2.03 X 10 -7, or 0.000000203.

Correlations

Use correlations to look at whether changes in one variable are accompanied by changes in another variable. For example, let’s say you wanted to know whether the number of credits taken by students correlates with the number of hours they sleep per day.

* Enter your data in two columns.
* Click on any empty cell.
* Hit the = sign in the bar at the top of the spreadsheet.
* Hit the down arrow to the left of the = sign. Now some options will appear.
* If "CORREL" is not on the list, click "more functions", choose "statistical", then "CORREL".
* A dialog box will appear. Click in the box next to "Array 1".
* Drag the dialog box out of the way, then highlight your first column of numbers.
* Click in the box next to "Array 2" and highlight your second column of numbers.
* Hit OK and read the number. This is the correlation coefficient.

A "1" would mean complete correlation. A "0" would mean none. A negative number means that when one variable goes up, the other goes down. A positive number means they rise and fall together. This program does not return a P-value for correlations. Note that correlations work well for "straight line" relationships, but don’t work as well with most curves. You can add a line that estimates the trend by clicking on "Chart " and "Add Trendline ".

Chi Square Test

Use the chi (pronounced KY as in sky) square test to look at whether actual data differ from a random distribution. For example, say you want to find out whether students prefer particular T-shirt colors. Assume there are five different colors and each student could get one free at registration (there are enough so that everyone could choose the same color). If people chose at random, the proportion of each color chosen would be equal (about 20% of the total shirts chosen would be in each category). You might not be surprised to find 19% of the shirts chosen were red and 21% were black, but when do you have enough evidence to say people are choosing them non-randomly? This test will tell you.

Label one row "Actual" and the other " Expected". Now make each column a separate category (such as red, black, white, etc.). In the "Actual" row, enter the numbers (not percent) you actually observed. In the " Expected" row, enter the number you would expect to find if the selection were random. In our example, we had five colors, so we expect to find 20% of our total chosen shirts in each color if people are choosing randomly. Multiply the total number of chosen shirts by 20% to get the expected number for each color. Note: the test assumes that each choice is independent, so for this test to be valid, you would have to make sure each person couldn't tell what others had taken.

* Click on any empty cell.
* Hit the = sign in the bar at the top of the spreadsheet.
* Hit the down arrow to the left of the = sign. Now some options will appear.
* If "CHITEST" is not on the list, click "more functions", choose "statistical", then "CHITEST".
* A dialog box will appear. Click in the box next to "Actual\_range".
* Drag the dialog box out of the way, then highlight your first row of numbers.
* Click in the box next to "Expected\_range" and highlight your second row of numbers.

The program will return a P-value. In our case, that is the probability that people are choosing shirts randomly. The smaller the number, the greater the likelihood that they are choosing by color. Remember that a P-value below 0.05 is generally considered statistically significant, while one of 0.05 or greater indicates the distribution is random. If your number looks like this: 2.03188E-7, Excel is giving you the number in its version of scientific notation. This number is actually 2.03 X 10 -7, or 0.000000203.

Reminders:

* Graph and look at your data to see whether your results make sense.
* Hypotheses can be rejected or supported, but not proven.

*Last update: 4/18/02 by* [*Rebecca Burton*](http://depts.alverno.edu/nsmt/burton.htm)*,* [*Dept. of Biology*](http://depts.alverno.edu/nsmt/Indexbio.htm)*,* [*Alverno College*](http://www.alverno.edu/)