http://www.theseashore.org.uk/theseashore/Stats%20for%20twits/Hyptesting%20stats%20format.html

The general structure of hypothesis testing statistics

Even if you think you know nothing about statistics, it's almost certain that you do. You will probably have heard of terms like the "average" value or maybe the "range" of some data or possibly its "standard deviation" or "variance". All of these things tell you something about a set of data, they are known as descriptive statistics.

The statistics we are concerned with here are called hypothesis-testing statistics. For the most part you will be using them to compare one set of data with another set of data. The format you follow is similar for most of these tests:

Invent a null hypothesis

Using your devastating powers of observation, you have noticed that adult humans seem on the whole to be taller than baby ones. You could easily investigate this by measuring some of each kind and comparing the average values of the two sets of data.

An hypothesis is simply a statement which offers an explanation of your observations. In this case our experimental hypothesis might be that all adult humans have had special cosmetic surgery to lengthen their legs and make them taller than babies. Alternatively, we might suggest that adults have been around for longer and therefore have grown bigger. Both of these would be experimental hypotheses, the latter being the more reasonable one.

A null hypothesis is a special sort of hypothesis which you invent purely for the purpose of doing the statistical test. It does not have to agree with your experimental hypothesis. The word null means a condition of nothingness or lacking any distinction. A null hypothesis is sometimes called a hypothesis of no difference. It is always stated as though there were no difference between the two things you are comparing. If we were doing a test that compared our two average (or mean) heights a suitable null hypothesis would be:

There is no significant difference between the means of the two sets of data

Remember, it might be obvious that there is a difference but you state it like this anyway. Having done the statistical test you will end up either accepting or rejecting this statement.

Calculate the value of the test statistic

All the tests do something different but the general pattern of what you do is the same. The next thing you do is use your data to calculate a value of the test statistic you are using (this will have a name, usually a letter like "**t**", "**U**", "rs"). You calculate a value that is **specific for your data.**

Find the critical value of the test statistic

Statisticians have spent a long time working out what are known as critical values of test statistics for all combinations of circumstances and sets of data. You must extract from one of their tables of critical values the value that applies to your combination of circumstances. What the value is depends on the number of items of data in each data set and the degree of precision you want to use in either accepting or rejecting your null hypothesis. This is the real value of these techniques, they allow you to say how certain you are when you either accept or reject the null hypothesis. You get to choose how certain you want to be.

Here is part of a table of critical values for a statistic called Spearman's rank correlation coefficient:

Significance level:				
Number of pairs of data (n):	10%	5%	2%	1%
5	0.9	1	1	
6	0.829	0.886	0.943	1
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.6	0.683	0.783	0.833
10	0.564	0.648	0.746	0.794
12	0.506	0.591	0.712	0.777
14	0.456	0.544	0.645	0.715
16	0.425	0.506	0.601	0.665
18	0.399	0.475	0.564	0.625
20	0.377	0.45	0.534	0.591
22	0.359	0.428	0.508	0.562
24	0.343	0.409	0.485	0.537
26	0.329	0.392	0.465	0.515
28	0.317	0.377	0.448	0.496
30	0.306	0.364	0.432	0.478

This is where you get to pick the degree of precision you want in either accepting or rejecting your null hypothesis. Lets us say that you wanted to be as certain as you could be (using our table above) that you would be correct in accepting or rejecting your null hypothesis. Enter the table at the 1% significance column and find the appropriate critical value by going along the appropriate row. For the sake of argument, let's say we have 10 pairs of data. As you can see the critical value is 0.794.

Draw a conclusion

With this particular test, if the value you've calculated for your own data is the same or bigger than this, you reject the null hypothesis. If the value for your data is smaller than the critical value, you accept the null hypothesis. In accepting or rejecting it at the 1% significance level you are saying: "If I did this test a very large number of times I would expect to be correct in accepting or rejecting my null hypothesis 99% of the time. I would expect a different result due to chance only 1% of the time". Put simply (and not quite accurately but hopefully you know what I mean): "I'm 99% certain that I'm right in accepting or rejecting my null hypothesis".

If you are not so concerned with being near certain you can pick a bigger % significance level. If you picked the 5% level the critical value would be 0.648. This is smaller than the critical value for 1% significance and it will be easier for your value (calculated from your own data) to beat it – and thus easier reject the hypothesis of no difference. However if you do it at this level you would expect different results due to chance 5% of the time. In other words 95 times out of a 100 you'd expect to be correct in accepting or rejecting your null hypothesis. 5 times out of 100 you'd expect a different result due to chance.

There is no law about what level of significance you choose but given the inherent variability of biological systems, it has become generally accepted that a level of 5% is acceptable for field data.

http://www.theseashore.org.uk/theseashore/Stats%20for%20twits/What%20different%20tests%20do.html

What some of the tests do:

T test

A T test will tell you if the means of two sets of normally distributed, unmatched, continuous data, with interval level measurements are significantly different to one another. (If you have a big sample, e.g. 25+ data points, you can use it for count data as well.) Some people call it a z test when you have a big sample but we use the same formulae. For any T test you do the null hypothesis will be:
There is no significant difference between the means of the 2 sets of data

Spearman's rank correlation coefficient

Spearman's rank correlation coefficient will tell you whether 2 variables are correlated, i.e. Does one variable change as the other one changes? It will tell you whether the relationship is positive (both go up together) or negative (one goes up as the other goes down) and the strength of any correlation. It assumes that any relationship is roughly a straight line one. For any Spearman's rank correlation coefficient you do, the null hypothesis will be:
There is no correlation between the 2 variables

Chi-squared test

A chi-squared test does a lot of things, but for the most part we use it in a simple way to see if an observed set of data (which has to be counts of things in categories) differs significantly from what we might expect, given our null hypothesis. For any chi-squared test you do the null hypothesis will be:
There is no significant difference between the observed and the expected frequencies

Mann-Whitney U test

This test tells you whether the median values of two sets of data are significantly different from one another. It has the advantage that the data does not have to be normally distributed and you can use it on smallish quantities of count data. For any Mann-Whitney U test you do, the null hypothesis will always be:
There is no significant difference between the medians of the two sets of data