False positives and the q-value.

A false positive finding in an experiment occurs when you conclude that something has an effect on a variable when in fact it does not. There is a risk of this in all experiments, but they can be a particular problem when tests are done on lots of variables (100's or 1000's) such as happens in proteomics, transcriptomics or metabolomics. False discovery rates (FDRs) and q-values are a way of quantifying the problem.

To see how they arise, we first consider an experiment in which there is no treatment effect on any of the variables measured. In this case, any positive findings must be false. For each of the many variables, we will have calculated a p-value. What will these p-values look like? It will be something like the plot below: they will be evenly spread between 0 and 1. 5% of them will be less than p=0.05. These are the 5% false positives present in every experiment, and they cannot be reduced by a bigger or better experiment or a more careful analysis.

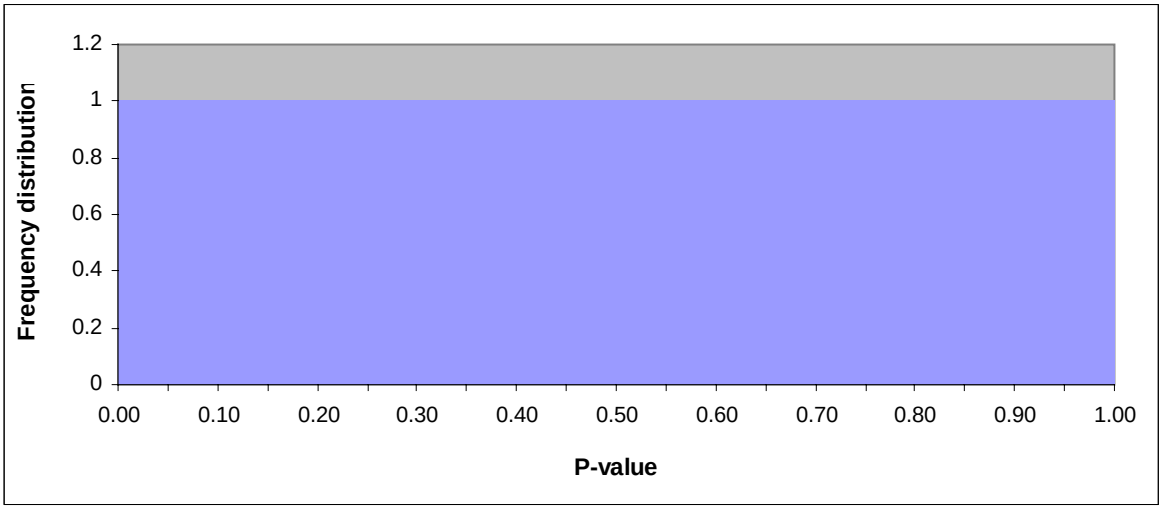

Fig 1. Distribution of p-values when there is no effect

Now suppose that we have an experiment where the treatments do have an effect on some of the variables. Let us suppose that 20% of them are affected (although we do not know this in advance). For the 80% which are unaffected, their p-values will be evenly spread between 0 and 1. For the 20% which are affected, the p-values will have a distribution concentrated towards the lower end of the 0-1 range. This means that the total distribution of p-values looks like the diagram below:

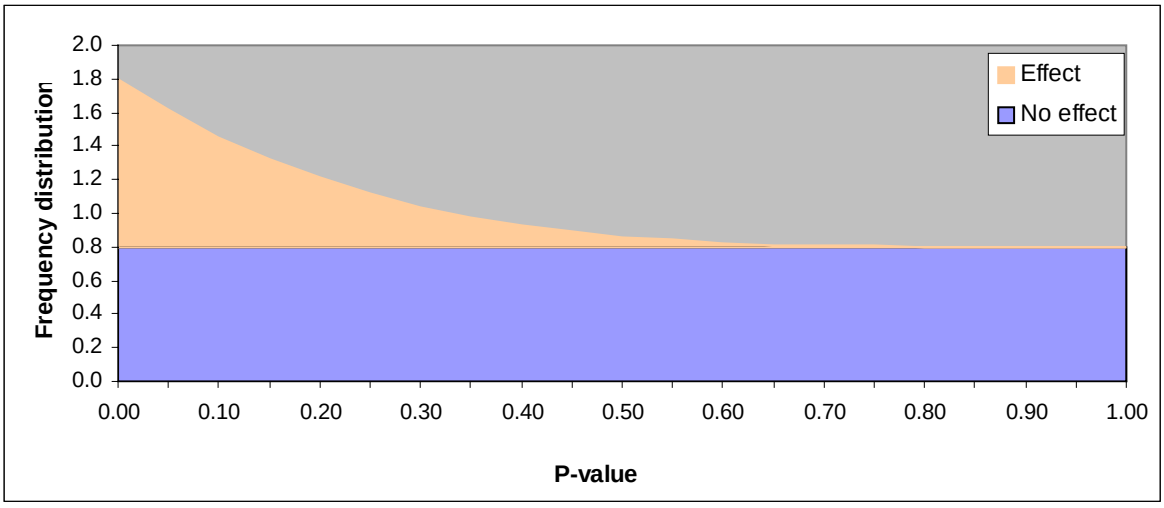

Fig 2. Distribution of p-values when some of the variables are affected.

By looking at the shape of this distribution, we can estimate how much of it is in the 'No effect' (rectangular shaped) part, and how much is in 'Effect' part. Now suppose we choose a p-value cutoff in order to conclude which variables are affected. This will divide the variables into 4 types (Fig 3). The **q-value** is the proportion of variables chosen as positive which are false positives. In Fig 3, it is about 50% for the cutoff shown.

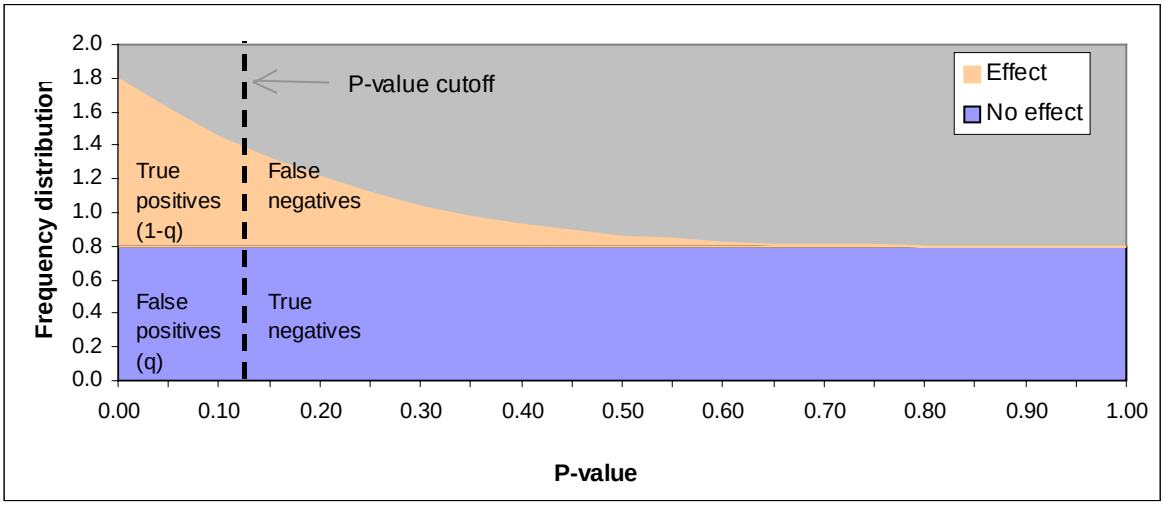

Fig 3. Using a p-value cutoff to decide which variables are affected by treatment.

What do we use it for? We can either use it to quantify the false positive problem by estimating what percentage of the variables declared to be affected are not. Or by choosing a q-value which we consider acceptable, use that to determine the p-value cutoff to use - it doesn't have to be 5%.

It may well be asked at this point that having quantified the false positives, can we not say which ones they are? Unfortunately this can't be done. It is like trying to design a fire alarm which makes a different sound depending on whether it's a real fire or a false alarm: a nice idea but it won't work. The only way to find out is to do some more investigation: search the building or in the case of the science, do some more research.

So, how do we calculate q-values? All we need is a collection of p-values from the same sort of test (e.g. t-test) on the same sort of variable (e.g. gel spots) in a single experiment. Then we can quantify the shape of the distribution by fitting some mathematical form to the curve in Fig 2. There is Genstat code to do this, and an Excel macro which uses a different method. There is also some specialised software to calculate q-values.

But what about the false negatives? These are also unsatisfactory. We can also quantify how many of these we have, but (like the false positives) can't say which they are. The issue of false negatives is related to the power of the experiment. By increasing this, such as by doing a bigger experiment with more replicates, we can push the orange part of the distribution towards the left. An example of what this might look like for the experiment in Fig 3 is shown below: there are no more affected variables, but we are more likely to detect them. We have reduced the q-value to about 38% for the same cutoff.

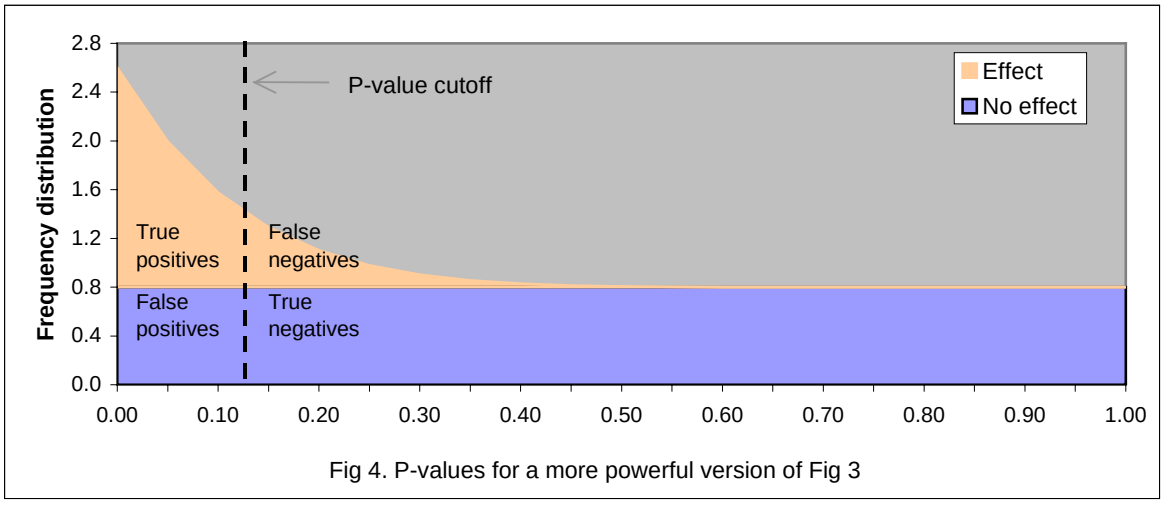

Fig 4. P-values for a more powerful version of Fig 3