Method Sheet 111

Applying correction for multiple testing in statistical tests

Overview

This method sheet explains how to apply corrections to p-values when multiple statistical tests have been performed sequentially on the same dataset. This is necessary to account for the danger of discovering false positive “significant” results through the simple process of calculating multiple p-values. We must recognise that even if we were looking at completely random data, with no pattern at all, the chances of obtaining a p-value of less than 0.05 is one in twenty just by chance alone. This means that if, for example, you compared 20 different drug treatments to a control using 20 individual t-tests, it is likely that at least one of the tests would give a false positive “significant” p-value, even if none of the drugs have any real activity. The method of accounting for this effect is called ‘correction for multiple testing’.

The Bonferroni method of correction for multiple testing

1) The Bonferroni method of correction for multiple testing is very easy to apply and widely accepted as one of the most robust ways of doing so.
2) First calculate your p-values in the way you normally would (for example, with Student’s t-tests, or Pearson’s correlation analyses).
3) Count how many individual p-values you have calculated for that particular experimental endpoint (e.g. cell growth, enzyme activity, absorbance, etc.).
4) Do not count the p-values from other tests you may have done that relate to separate experiments - focus only on those p-values that relate to the same experiment and/or type of measurement.
5) Now that you have this method, there are two ways to apply the Bonferroni correction - both are equally valid, so feel free to choose whichever you prefer.

Option 1: Multiply the p-values by the ‘n’

1) The first option is to simply multiply all of the p-values for a particular set of comparisons by the number of tests that were performed.
2) For example, if you have a p-value of 0.001 for one correlation within a series of 7 related correlation tests, multiply the p-value by 7 to get 0.007.
3) When using this option, keep the p-value threshold for significance the same, at 0.05.
4) In this example, 0.007 is below 0.05, so the result remains significant after correction for multiple testing using the Bonferroni method.

Option 2: Divide the threshold for significance by the 'n'

1) The second option is to simply divide the threshold for significance by the number of tests that were performed, but leave the p-values unchanged.
2) For example, if you have a p-value of 0.001 for one correlation within a series of 7 related correlation tests, you would set the threshold for significance to be no longer 0.05 as we normally would for one test, but instead set it at 0.05 / 7 or 0.0071.
3) When using this option, do not adjust the p-values themselves.
4) In this example, 0.001 is below 0.0071, so the result remains significant after correction for multiple testing using the Bonferroni method.

What to report in your dissertation

1) If you choose to apply option 1, it can be helpful to report your p-values in two separate columns of a table, one showing the uncorrected p-values and the other showing the p-values after correction.
2) Remember to be clear in the Methods section of your dissertation which specific analyses you have chosen to apply Bonferroni correction to.
3) Remember also to explain clearly the use of Bonferroni correction in your figure and table legends in the Results section of your report when you use it.

Notes

- When you use post-hoc tests, such as Tukey's or Dunnett's tests with ANOVA, there is no requirement to perform an additional Bonferroni correction on the p-values arising from these analyses, since these post-hoc tests already incorporate a correction for multiple testing.
- If you only perform only one t-test, or one Pearson's correlation analysis, to a particular dataset, you do not have to perform any form of correction for multiple testing.