

Chapter 15

Bonferroni t-Test

The t-test can be used to test the hypothesis that two group means are not different (Chap. 3). When the experimental design involves multiple groups, and, thus, multiple tests, we increase our chance of finding a difference. This is, simply, due to the play of chance rather than a real effect. Multiple testing without any adjustment for this increased chance is called data dredging, and is the source of multiple type I errors (chances of finding a difference where there is none). The Bonferroni t-test (and many other methods) are appropriate for the purpose of adjusting the increased risk of type I errors.

Bonferroni t-Test

The underneath example studies three groups of patients treated with different hemoglobin improving compounds. The mean increases of hemoglobin are given.

	Sample size	Mean hemoglobin (mmol / l)	Standard deviation (mmol / l)
Group 1	16	8.725	0.8445
Group 2	10	10.6300	1.2841
Group 3	15	12.3000	0.9419

An overall analysis of variance test produced a p-value of < 0.01 . The conclusion is that we have a significant difference in the data, but we will need additional testing to find out where exactly the difference is, between group 1 and 2, between group 1 and 3, or between group 2 and 3. The easiest approach is to calculate the t-test for each comparison. It produces a highly significant difference at $p < 0.01$ between group 1 versus 3 with no significant differences between the other comparisons. This highly significant result is, however, unadjusted for multiple comparisons. If one analyzes a set of data with three t-tests, each using a 5% critical value for concluding that there is a significant difference, then there is about $3 \times 5 = 15\%$ chance of finding it. This mechanism is called the Bonferroni inequality.

Bonferroni recommended a solution for the inequality, and proposed to follow in case of three t-tests to use a smaller critical level for concluding that there is a significant difference:

With 1 t-test: critical level = 5%

With 3 t-tests: critical level = $5/3 = 1.6\%$.

The above equations lead rapidly to very small critical values, otherwise called p-values, and is, therefore, considered to be over-conservative. A somewhat less conservative version of the above equation was also developed by Bonferroni., and it is called the Bonferroni t-test.

In case of three comparisons the rejection p-value will be $0.05 \times \frac{2}{3(3-1)} = 0.0166$.

In the given example a p-value of 0.0166 is still larger than 0.01, and, so, the difference observed remained statistically significant, but using a cut-off p-value of 0.0166, instead of 0.05, means that the difference is not *highly* significant anymore.