

Multiple Testing

It was mentioned in the p -value limitations unit that running more than one test at a time, on the same data is an example of p -value hacking. This is because there is an expected false positive rate for running one test, and if we run multiple tests say using different combinations of features this expected rate should be higher.

Bonferroni Correction

The **Bonferroni Correction** is a simple way to rectify the over testing issue.

Suppose we want to test a combined hypothesis as a threshold of α . The Bonferroni correction procedure then tests each of the individual hypotheses at a threshold of:

$$\alpha_{\text{Bonferroni}} = \frac{\alpha}{\text{\# of Hypotheses in Combined Hypothesis}}$$

This type of adjustment for dealing with multiple testing was popular for some time, but today when the number of tests can scale into the hundreds or thousands the method has become less widely used when compared to other methods.

Other methods to correct for multiple tests

The Bonferroni Correction is probably the simplest way to avoid increasing your false positive rate when running several tests at once. You simply divide your threshold by the number of tests you are running. This conservative method does penalize potentially significant results, so there is an increase in false negatives. This may necessitate considering alternative ways of dealing with multiple tests.

Note: If we have a data set with 1000 observations and we repeat a t-test for each observation then (with $\alpha=0.05$) we can expect 50 observation to be found significant purely by random chance.

A bonferroni Correction in the above example is likely too conservative. Fortunately, there are a number of other methods that have been developed.

The [statsmodels](#) library in Python has the following methods in the [multipletests submodule](#):

- *bonferroni* : one-step correction
- *sidak* : one-step correction
- *holm-sidak* : step down method using Sidak adjustments
- *holm* : step-down method using Bonferroni adjustments
- *simes-hochberg* : step-up method (independent)

- *hommel* : closed method based on Simes tests (non-negative)
- *fdr_bh* : Benjamini/Hochberg (non-negative)
- *fdr_by* : Benjamini/Yekutieli (negative)
- *fdr_tsbh* : two stage fdr correction (non-negative)
- *fdr_tsbky* : two stage fdr correction (non-negative)

The *fdr_bh* method is a commonly used method. The *p*-values are ranked, multiplied by the number of features, and divided by their corresponding rank. It is another form of *p*-value adjustment.

These methods are reasonable for many types of data, but they still have some limitations. One major assumption is that the *p*-values follow a certain distribution, but in practice this is known to be invalid for certain types of data like *genomics*. There is a whole body of literature dedicated to permutation tests as a form of assessing significance in these types of situations. Recall that we use a permutation test in the A/B testing example as well. Here is an example of several methods.

```

1 import numpy as np
2 from statsmodels.stats.multitest import multipletests
3
4 pvals = np.random.uniform(0.001, 0.06, 12)
5 _results = multipletests(pvals, alpha=0.05, method='bonferroni', is_sorted=False
6 , returnsorted=False)
7 rejected_bonferroni, adjusted_bonferroni = _results[0], _results[1]
8 # adjusted_bonferroni is the adjusted p_value
9
10 _results = multipletests(pvals, alpha=0.05, method='fdr_bh', is_sorted=False,
11 returnsorted=False)
12 rejected_bh, adjusted_bh = _results[0], _results[1]
13 # adjusted_bh is the adjusted alpha
14
15 for p,pval in enumerate(pvals):
16     print(round(pval,3), round(adjusted_bonferroni[p],3), round(adjusted_bh[p]
17 ,3))
18

```

```

1 0.015 0.185 0.05
2 0.036 0.438 0.05
3 0.052 0.629 0.052
4 0.045 0.535 0.05
5 0.038 0.458 0.05
6 0.02 0.244 0.05
7 0.04 0.482 0.05
8 0.008 0.095 0.05
9 0.02 0.239 0.05
10 0.046 0.548 0.05
11 0.021 0.256 0.05
12 0.038 0.461 0.05
13

```

Note that another way to think of the Bonferroni correction is just the `pvals * num_tests`.

Additional Resources

- [Influential paper by Benjamini and Hochberg](#)
- [Paper describing the use of permutation tests in multiple hypothesis testing](#)