

Multiple Comparisons (Bonferroni vs FDR)

Contents

1 Review: Philosophy I (Familywise) Correction for Error Rate	2
1.1 Simulation of FWER	2
1.2 Results Table of FWER Simulation	2
1.3 Bonferroni Correction	2
1.3.1 Results Table of Bonferroni Simulation	2
1.3.2 Plot of Bonferroni Corrected Pvalues	3
1.3.3 Increase Power, Rinse and Repeat	3
1.3.4 Results Table of Higher Powered Bonferroni Simulation (Power = .5)	3
1.3.5 Results Table of Super Powered Bonferroni Simulation (Power = .95)	4
2 Philosophy II False Rate of Discovery	4
2.1 Benjamini and Hochberg procedure	5
2.2 FDR simulation	5
2.2.1 Results Table of FDR Simulation	5
2.2.2 Plot of FDR Corrected Pvalues	6
2.2.3 Results Table of FDR Simulation (alpha=.10)	6
2.2.4 Notes	6

1 Review: Philosophy I (Familywise) Correction for Error Rate

Dominant Philosophy in Experimental Psychology

Declared	Null is True (no difference)	Null is False (difference)	Total
Sig	Type 1 (V)	True Positives (S)	R
Non-Sig	True Negative (U)	Type II (T)	M-R
Total	Mo	M-Mo	M

Familywise error = V / Mo = Probability of getting at least one test wrong

1.1 Simulation of FWER

Let's imagine that 25% of the brain regions are different from each other at a set *effect size*. We will run a Monte-Carlo Simulation: 50,000 t-tests, $N = 10$ per group, $\alpha_{pc} = .05$, which will yield Power = .50. Our goal is to find those 25 regions and we will assume each test is independent and we sample with replacement. Further, we will know which regions were actually different (**We know the “truth”**).

1.2 Results Table of FWER Simulation

Declared	Null is True (no difference)	Null is False (difference)	Total
Sig	1882	6245	8127
Non-Sig	35618	6255	41873
Total	37500	12500	50000

FWER = $V / Mo = 0.05$

Type II = $T/M-Mo = 0.5$, which is correct since power = .50

If you ran 10K studies 16.25% of them would have been declared significant (R/M), but in reality 25% of the studies had a true effect.

Here is another way to view it: Of the tests we declared significant a large percentage of them were not real ($V / R = 23.16\%$).

1.3 Bonferroni Correction

$$\alpha_{pc} = \frac{\alpha_{EW}}{j} = \frac{.05}{50,000} = 1 \times 10^{-6}$$

We will keep everything else the same from the last simulation.

1.3.1 Results Table of Bonferroni Simulation

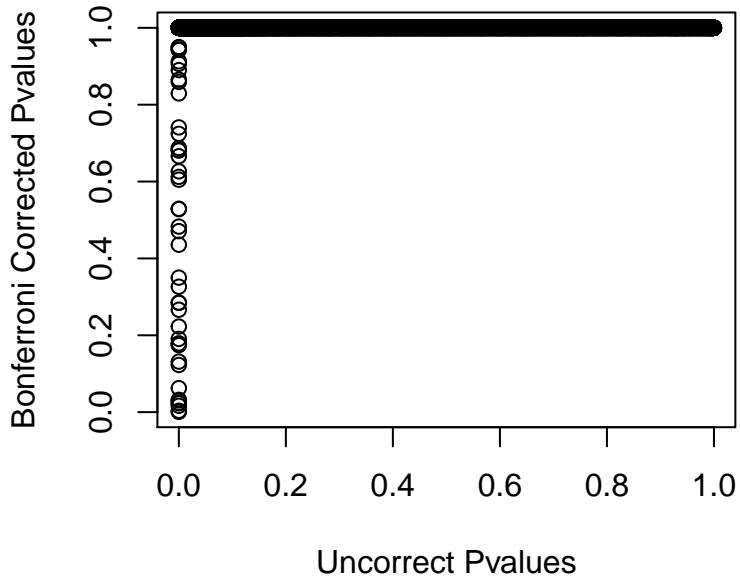
Declared	Null is True (no difference)	Null is False (difference)	Total
Sig	0	7	7
Non-Sig	37500	12493	49993
Total	37500	12500	50000

FWER = V / Mo = 0

Type II = T/M-Mo = 0.999. So we missed everything (power is basically 0)

So we made no mistakes, but we found nothing. *NIH never gives you another penny.*

1.3.2 Plot of Bonferroni Corrected Pvalues



1.3.3 Increase Power, Rinse and Repeat

Given our original $d = 0.926$, we would recalculate the sample size we would need per group (62) given our new alpha $\alpha_{pc} = 5 \times 10^{-6}$. That is a 520% increase in sample size!

Let's try to increase our sample size to 62 per group and try again

1.3.4 Results Table of Higher Powered Bonferroni Simulation (Power = .5)

Declared	Null is True (no difference)	Null is False (difference)	Total
Sig	0	6296	6296
Non-Sig	37500	6204	43704
Total	37500	12500	50000

FWER = V / Mo = 0

Type II = T/M-Mo = 0.496. So we fixed our power problem and made no FWER! BUT:

650 Per scan * N = 10 * 2 groups = 13,000 dollars

650 Per scan * N = 62 * 2 groups = 80,600 dollars

And yet we still miss 50% of the results because we are powered at .50. If we go to power .95, as we don't want to

miss 50% of the results.

N = 106

650 Per scan * N = 106 * 2 groups = 137,800 dollars

1.3.5 Results Table of Super Powered Bonferroni Simulation (Power = .95)

Declared	Null is True (no difference)	Null is False (difference)	Total
Sig	0	11869	11869
Non-Sig	37500	631	38131
Total	37500	12500	50000

FWER = V / Mo = 0

Type II = T/M-Mo = 0.05.

All of the low low price of 137,800 dollars for 1 study!

2 Philosophy II False Rate of Discovery

FWER is appropriate when you want NO false positives

Dominant Philosophy in Neuroscience, genomics, etc. Any field where they are not as worried about false positives but interested in making discoveries.

Declared	Null is True (no difference)	Null is False (difference)	Total
Sig	Type 1 (V)	True Positives (S)	R
Non-Sig	True Negative (U)	Type II (T)	M-R
Total	Mo	M-Mo	M

FDR error = V / R

2.1 Benjamini and Hochberg procedure

1. Sort all p-values from all tests you did (smallest to largest)
2. Assign ranks to the p-values (lowest=1, etc.).
3. Calculate each individual p-value's Benjamini-Hochberg critical value, using the formula (rank/total number of tests)*q, where: q = the false discovery rate (a percentage, chosen by you or via algorithm which is beyond our scope)
4. Compare your original p-values to the B-H critical value; find the largest pvalue that is smaller than the critical value.

For example 5 t-tests only:

Test	Pvalue	Rank	BH Critical Value (Rank/m)*q [.1]	Significant?
Test 5	.001	1	0.02	Yes
Test 1	.030	2	0.04	Yes
Test 3	.045	3	0.06	No
Test 2	.056	4	0.08	No
Test 4	.230	4	0.1	No

2.2 FDR simulation

Repeat our original simulation but correct the pvalues using FDR correction. Note we will set FDR level to be about .05 to compare to Bonferroni, but people could change their alpha to be higher (.1 or .25 as in the Dead Salmon Paper).

2.2.1 Results Table of FDR Simulation

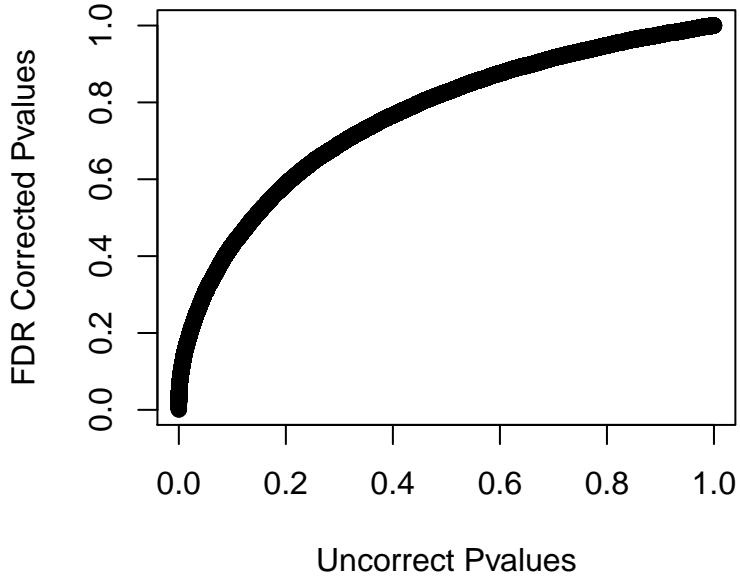
Declared	Null is True (no difference)	Null is False (difference)	Total
Sig	24	614	638
Non-Sig	37476	11886	49362
Total	37500	12500	50000

We can compare this to our Bonferroni Correction Table.

$$FDR = V / R = 0.0376$$

$$Type\ II = T/M-Mo = 0.951$$

2.2.2 Plot of FDR Corrected Pvalues



2.2.3 Results Table of FDR Simulation ($\alpha=.10$)

Declared	Null is True (no difference)	Null is False (difference)	Total
Sig	165	2051	2216
Non-Sig	37335	10449	47784
Total	37500	12500	50000

We can compare this to our Bonferroni Correction Table.

$$FDR = V / R = 0.0745$$

$$\text{Type II} = T/M-Mo = 0.836$$

2.2.4 Notes

- FDR q values are controlled via an algorithm in R, and we will not go into the details.
- FDR (BH procedure) is only one of methods, there are others, but we will not go into them.
- FDR correction is often applied to correction matrices because it's more logical as you are doing exploratory work. Simply pass your pvalues into this function: `Correct.Pvalues<-p.adjust(Pvector, method = 'fdr')`. You can decide on what to call significant (alpha .05 or .10 depending on how exploratory you want to be).
- To avoid these issues, FMRI hypothesis can be constrained to be not “whole brain” or voxel by voxel comparisons.