Opinionated Lessons in Statistics

by Bill Press

#16 Multiple Hypotheses

Let's talk about multiple hypothesis testing.

The "Bonferroni correction" is widely used.

Carlo Emilio Bonferroni
(1892 – 1960)

It is very conservative, hence usually not the most powerful test.

α = prob. that <u>one or more</u> of N tests will accidentally fall in their critical regions α'

$$\alpha = 1 - (1 - \alpha')^N \approx N\alpha'$$

This assumes that the N tests are all independent. That's rarely true.

The opposite limit would be to repeat the <u>same test</u> N times on the same data (N non-communicating graduate students open the same statistics book).

$$\alpha = \alpha'$$

The truth is always somewhere in-between.

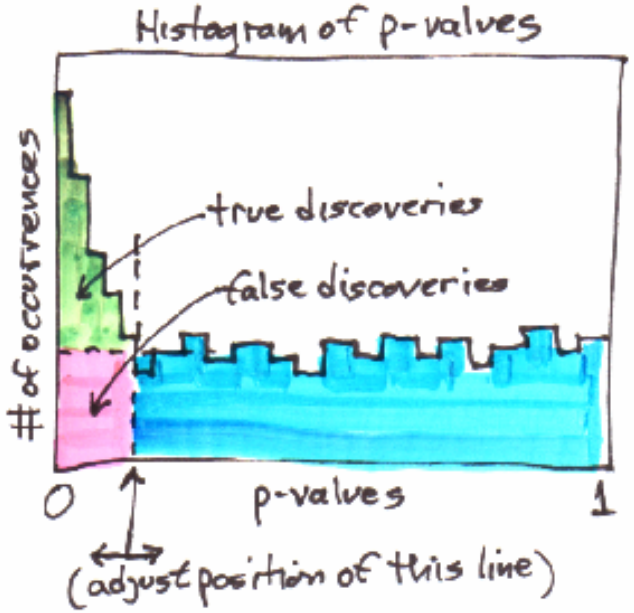
Slavish adherence to Bonferroni is a curse on biomedical research, but it is better than the alternative of having a literature full of wrong results!

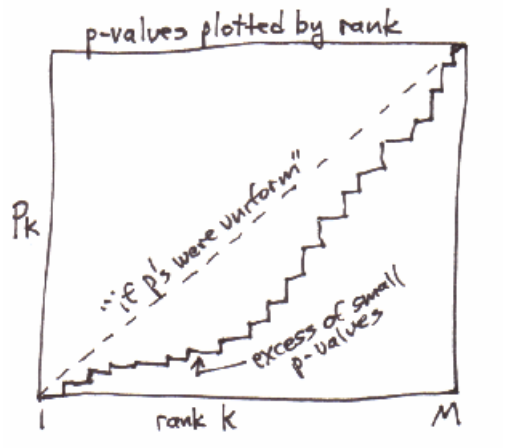
For large-scale screens can use False Discovery Rate (FDR) instead.

False Discovery Rate (Benjamini & Hochberg)

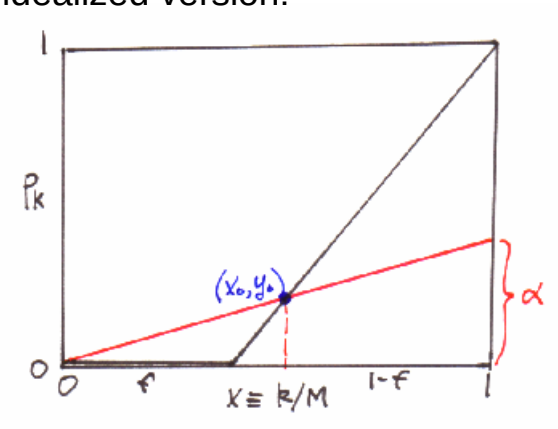
This is often a good alternative to Bonferroni, when the latter is too conservative.

- You have a lot of p-values
  - e.g., one per drug for 1000 drugs
  - or, one per gene for 10000 genes
- They are not uniform
  - there is an excess at small values
  - so some must be "causal"
- How do you set p to control α, the fraction of discovery calls that are false?
  - say, α = 5%





idealized version:



Prescription: call as discoveries all $p_k < \frac{k}{M}\alpha$

Proof:

$$\alpha x_0 = \frac{x_0 - f}{1 - f} \;\Rightarrow\; x_0 = \frac{f}{1 - \alpha(1-f)}$$

$$\Rightarrow \mathrm{FDR} = \frac{x_0 - f}{x_0} = \alpha(1-f) < \alpha$$

(There are fancier proofs for the nonidealized version.)

OK, enough p-values for now.

Keep in mind that we are still "closet Bayesians", however…

Bayesians have much less difficulty with multiple hypotheses in the happy case that they are EME.

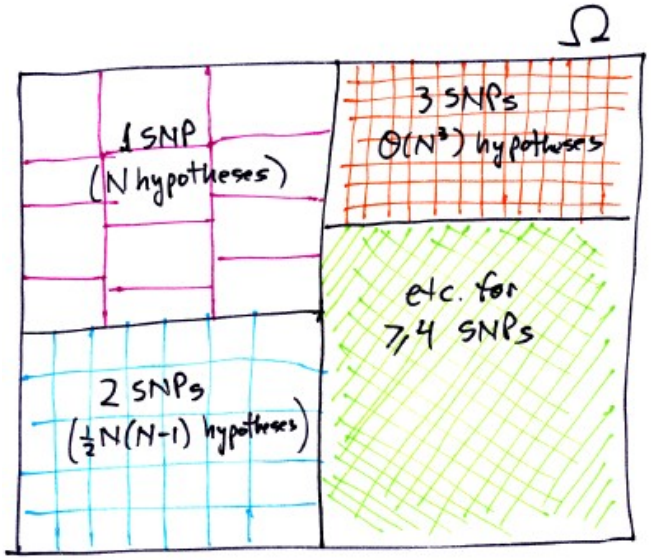
Example: We have a model where one, or a combination of, single nucleotide polymorphisms (SNPs) causes a particular kind of cancer.

We genotype patients and controls for N SNPs, each with 2 alleles.

So there are 2^N measurable hypotheses, and under each we can compute $P(\text{data}|H_i)$

p-value with Bonferroni would make a statistically significant finding impossible!

We are saved by the prior $P(H_i)$ which must have $\sum P(H_i) = 1$



Quite typically, our prior for models with "one main factor" (here, one SNP) will be larger than with "two main factors" (2 SNPs) and so on.

Now, do the "Bayes thing" and see if the evidence factor increases any individual model to high posterior probability.

Look, Ma, no multiple hypothesis correction!