Reason 3: Multiple Hypothesis Testing

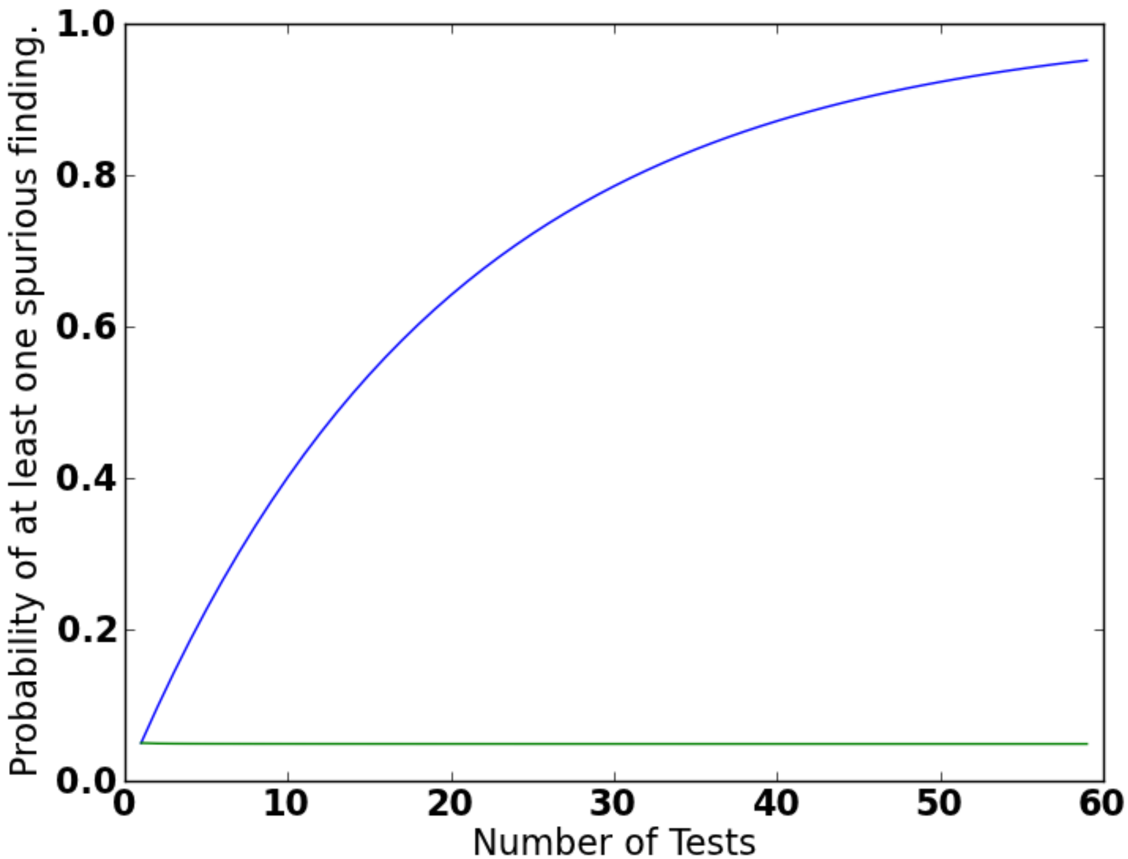
- If you perform experiments over and over, you're bound to find something
- This is a bit different than the publication bias problem: Same sample, different hypotheses
- Significance level must be adjusted down when performing multiple hypothesis tests

P(detecting an effect when there is none) = α = 0.05

P(detecting an effect when it exists) = $1 - \alpha$

P(detecting an effect when it exists on every experiment) = $(1 - \alpha)^k$

P(detecting an effect when there is none on <u>at least one</u> experiment) = $1 - (1 - \alpha)^k$



$\alpha = 0.05$

"Familywise Error Rate"

Familywise Error Rate Corrections

- Bonferroni Correction
  - Just divide by the number of hypotheses

$$\alpha_c = \frac{\alpha}{k}$$

- Šidák Correction
  - Asserts independence

$$\alpha = 1 - (1 - \alpha_c)^k$$

$$\alpha_c = 1 - (1 - \alpha)^{\frac{1}{k}}$$

False Discovery Rate

	Reject H0	Do Not Reject H0	Total
H0 is true	FD	TN	T
H0 is false	TD	FN	F
Total	D	N	$TFDN$

T/F = True/False
D/N = Discovery/Nondiscovery

$$Q = FDR = \frac{FD}{D}$$

FDR (2)

- Bonferroni correction and other FWER corrections tend to wipe out evidence of the most interesting effects; they suffer from low power.
- FDR control offers a way to increase power while maintaining a bound on the ratio of wrong conclusions
- Intuition:
  - 4 false discoveries out of 10 rejected null hypotheses

  is a more serious error than

  - 20 false discoveries out of 100 rejected null hypotheses.

adapted from a slide by Christopher Genovese

Benjamini-Hochberg Procedure

- Compute the p-value of m hypotheses
- Order them in increasing order of p-value
  - That is, most likely hypotheses are first

rank order

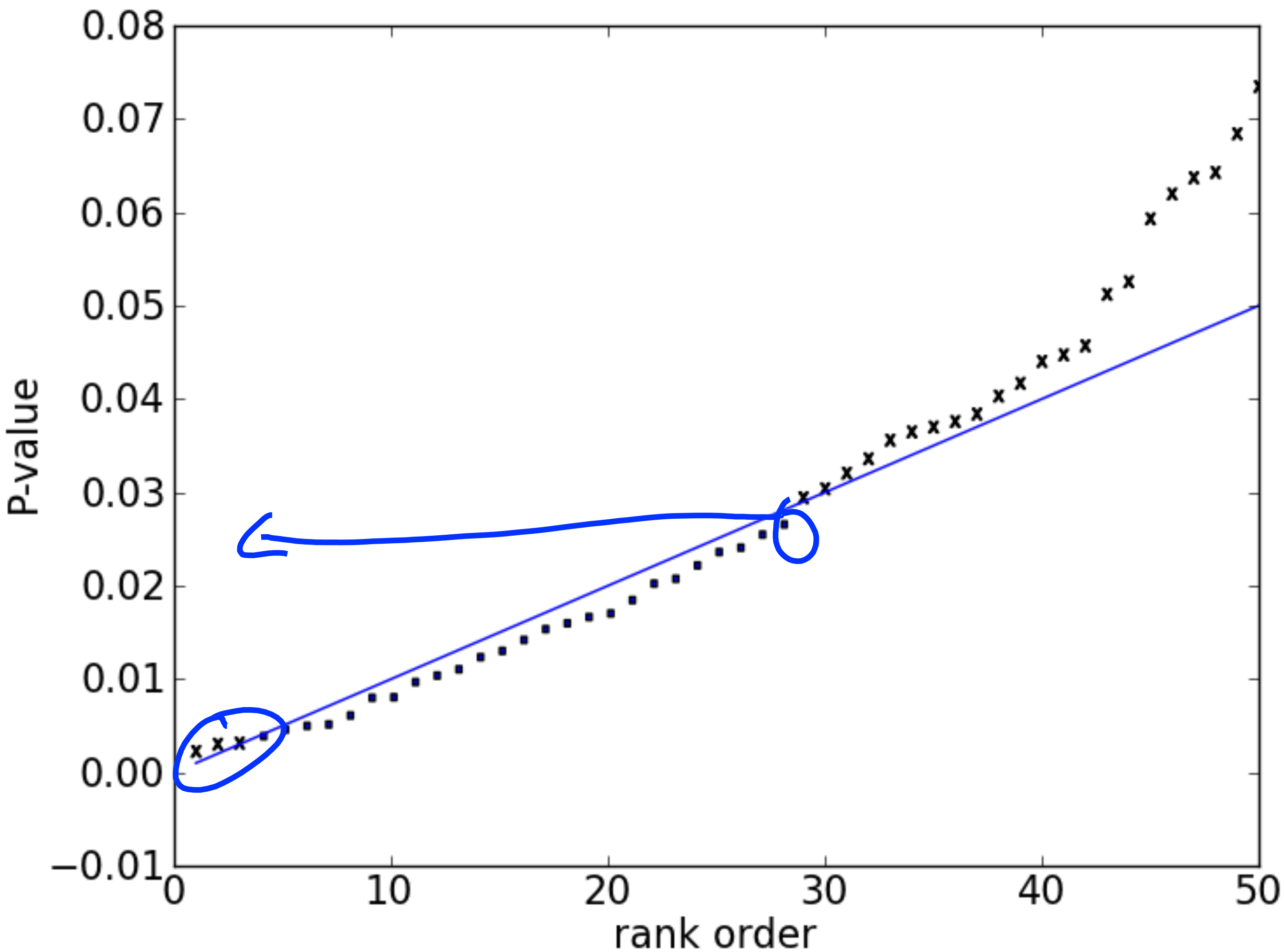
$$P_i \leq \frac{i}{m}\alpha$$

\# of hypotheses

$$FDR \leq \frac{T}{m}\alpha$$

i	$\frac{i}{50}\alpha$ 0.05
1	0.001
2	0.002
3	0.003
…	
20	0.020
…	

Benjamini-Hochberg Procedure



$m = 50$
$\alpha = 0.05$