

Multiple testing

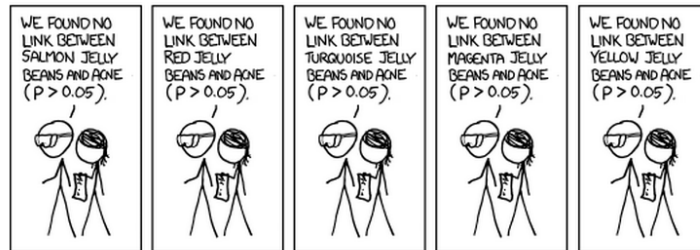
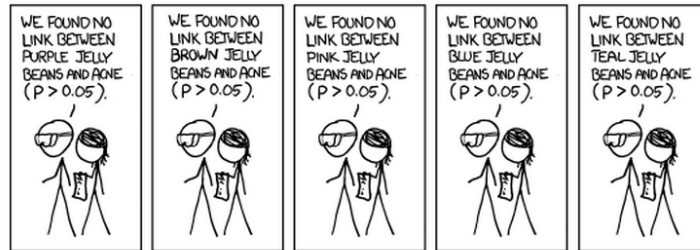
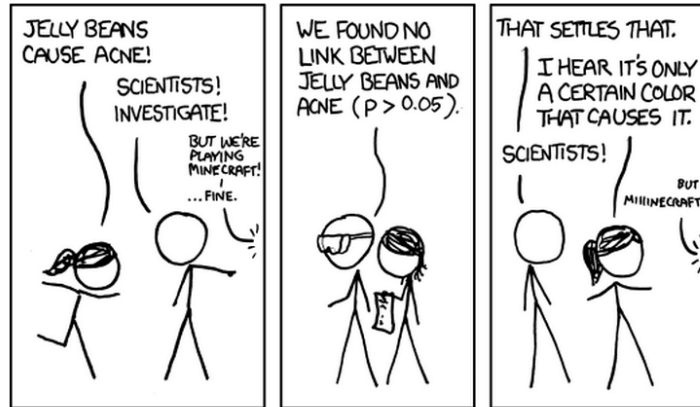
Jeff Leek

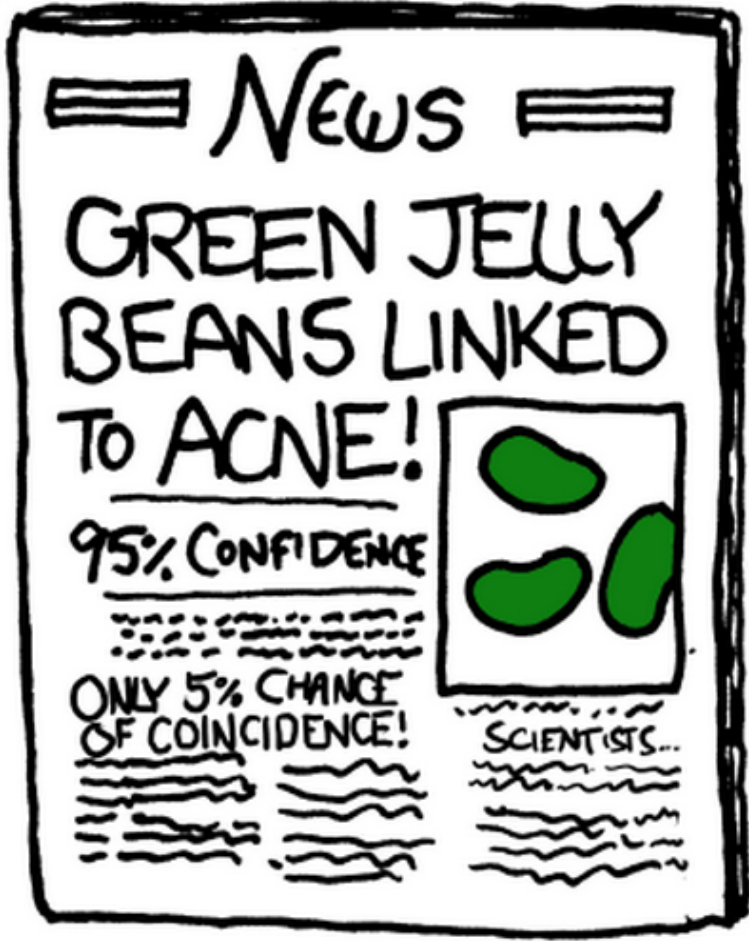
@jtleek


www.jtleek.com

P-values/hypothesis testing are
designed for one







- 
- Measure 10,000 genes
 - Calculate 10,000 p-values
 - Call genes “significant” if p-value < 0.05
 - Expected Number of False Positives:

$$10,000 \times 0.05 = 500 \text{ False Positives}$$

Multiple comparison error rates

- Family wise error rate:

$$\Pr(\# \text{ False Positives} \geq 1)$$

- False discovery rate:

$$E \left[\frac{\# \text{ False Positives}}{\# \text{ Of Discoveries}_-} \right]$$

Suppose 50 out of 10,000 genes are significant at 0.05 level

No Correction

Expect $0.05 * 10,000 = 500$ false positives

False Discovery Rate

Expect $0.05 * 50 = 2.5$ false positives

Family Wise Error Rate

The probability of at least 1 false positive ≤ 0.05

Controlling error rates

Bonferroni Correction

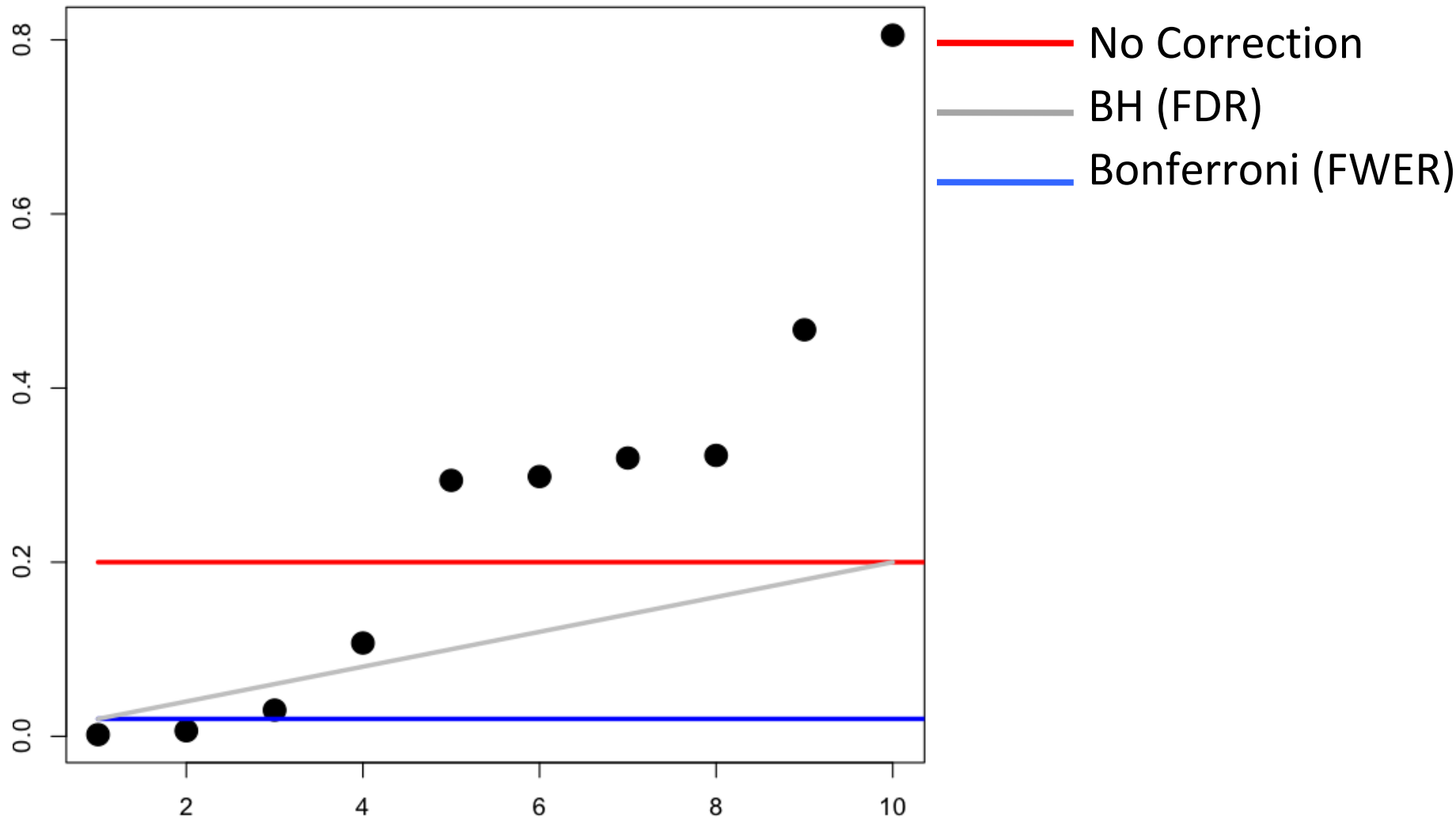
P-values less than α/m are significant

Benjamini-Hochberg Correction

Order the p-values: $p_{(1)}, \dots, p_{(m)}$

If $p_{(i)} \leq \alpha \times i/m$ then it is significant

Example with 10 p-values



Notes and further reading

- Type I errors, family wise error rate, and false discovery rate do not measure the same thing
- These all rely on the p-values being “correct”
 - Things that can go wrong: bad model, batch effects...
- This is a great first read:
 - <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC170937/>