Multiple testing

Lieven Clement

Statistical Genomics: Master of Science in Bioinformatics

Problem of multiple hypothesis testing

- Consider testing DE for all $m = 16000$ genes simultaneously
- What if we assess each individual test at level α?

$\rightarrow$ Probability to have a false positive among all m simultatenous test $>>> \ \alpha = 0.05$

Table of Outcomes

	H_0 accepted	H_0 rejected	
not DE	TN	FP	m_0
DE	FN	TP	m_1
Total	NR	R	m

- TN: number of true negatives
- FP: number of false positives
- FN: number of false negatives
- TP: number of true positives
- NR: number of non-rejections, R: number of rejections

Table of Outcomes

	H_0 accepted	H_0 rejected	
not DE	TN	FP	m_0
DE	FN	TP	m_1
Total	NR	R	m

Random Variables

Table of Outcomes

		Called Bad	Called Correct	
Unobservable	Bad hit	TN	FP	m_0
	Correct hit	FN	TP	m_1
Observable	Total	NR	R	m

Familywise Error Rate: FWER

- Traditional statistical multiplicity correction
- FWER $= P(FP > 0)$
- FWER: probability of making at least one false positive decision or
  probability to declare at least one gene differentially expressed which is truly non differentially expressed

FWER: single step: e.g. Bonferroni

- Simple method
- m tests are performed at the level α/m
- FWER$\leq \sum\limits_{g=1}^{m} P(reject H_{0g} | H_0 \text{ is true}) = m\alpha/m = \alpha$
- Provides strong control
- Bonferroni is very conservative
- Works for dependent tests
- Adjusted p-value: $\tilde{p}_g = \min(m\ p_g, 1)$
- Example: test 5 genes simultaneously, $\alpha = 0.05$
  $p_1 = 0.002, p_2 = 0.03, p_3 = 0.012, p_4 = 0.7, p_5 = 0.015$

FWER: single step: e.g. Bonferroni

- Simple method
- m tests are performed at the level α/m
- FWER$\leq \sum_{g=1}^{m} P(reject H_{0g}|H_0 \text{ is true}) = m\alpha/m = \alpha$
- Provides strong control
- Bonferroni is very conservative
- Works for dependent tests
- Adjusted p-value: $\tilde{p}_g = \min(m\ p_g, 1)$
- Example: test 5 genes simultaneously, $\alpha = 0.05$
  $p_1 = 0.002, p_2 = 0.03, p_3 = 0.012, p_4 = 0.7, p_5 = 0.015$
  1. c(α)=0.05/5=0.01
  2. Reject H_{01}
  3. $\tilde{p}_1 = 0.01, \tilde{p}_2 = 0.15, \tilde{p}_3 = 0.060, \tilde{p}_4 = 1, \tilde{p}_5 = 0.075$

FWER: step down method of Holm

- order p-values: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
- $\hat{k} = \min\{k : p_{(k)} > \frac{\alpha}{m-k+1}\}$
- if $\hat{k}$ exists then reject all $H_{(0i)}$, $i = 1, \ldots, k-1$
- $\tilde{p}_{(i)} = \min(p_{(i)}(m-i+1), 1)$

FWER: step down method of Holm

- order p-values: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
- $\hat{k} = \min\{k : p_{(k)} > \frac{\alpha}{m-k+1}\}$
- if $\hat{k}$ exists then reject all $H_{(0i)}$, $i = 1, \ldots, k-1$
- $\tilde{p}_{(i)} = \min(p_{(i)}(m-i+1), 1)$
- Suppose 2 tests: $p_{(1)} = 0.001$, $p_{(2)} = 0.0015$
  $\rightarrow \tilde{p}_{(1)} = 0.002$, $\tilde{p}_{(2)} = 0.0015$

FWER: step down method of Holm

- order p-values: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
- $\hat{k} = \min\{k : p_{(k)} > \frac{\alpha}{m-k+1}\}$
- if $\hat{k}$ exists then reject all $H_{(0i)}$, $i = 1, \ldots, k-1$
- $\tilde{p}_{(i)} = \max\limits_{h=1,\ldots,i} \min(p_{(h)}(m-h+1), 1)$
- Suppose 2 tests: $p_{(1)} = 0.001$, $p_{(2)} = 0.0015$
  $\rightarrow \tilde{p}_{(1)} = 0.002$, $\tilde{p}_{(2)} = 0.002$

FWER: step down method of Holm

- order p-values: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
- $\hat{k} = \min\{k : p_{(k)} > \frac{\alpha}{m-k+1}\}$
- if $\hat{k}$ exists then reject all $H_{(0i)}$, $i = 1, \ldots, k-1$
- $\tilde{p}_{(i)} = \max\limits_{h=1,\ldots,i} \min(p_{(h)}(m-h+1), 1)$
- Suppose 2 tests: $p_{(1)} = 0.001$, $p_{(2)} = 0.0015$
  $\rightarrow \tilde{p}_{(1)} = 0.002$, $\tilde{p}_{(2)} = 0.002$
- Strong control of FWER

FWER Holm: example

- Example: test 5 genes simultaneously, $\alpha = 0.05$
  $p_1 = 0.002, p_2 = 0.03, p_3 = 0.012, p_4 = 0.7, p_5 = 0.015$

$p_{(i)}$	0.002	0.012	0.015	0.03	0.7

$$\longrightarrow$$

$\frac{\alpha}{m-i+1}$	$0.05/5$	$0.05/4$	$0.05/3$	$0.05/2$	0.05
	0.01	0.0125	0.0167	0.025	0.05
reject	yes	yes	yes	no	no
$\tilde{p}_{(i)}$	0.010	0.048	0.045	0.060	0.700

FWER Holm: example

- Example: test 5 genes simultaneously, $\alpha = 0.05$
  $p_1 = 0.002, p_2 = 0.03, p_3 = 0.012, p_4 = 0.7, p_5 = 0.015$

$p_{(i)}$	0.002	0.012	0.015	0.03	0.7

$$\longrightarrow$$

$\frac{\alpha}{m-i+1}$	0.05/5	0.05/4	0.05/3	0.05/2	0.05
	0.01	0.0125	0.0167	0.025	0.05
reject	yes	yes	yes	no	no
$\tilde{p}_{(i)}$	0.010	0.048	0.048	0.060	0.700

- reject $H_{01}, \ H_{03}, \ H_{05}$
- $\tilde{p}_1 = 0.01, \tilde{p}_2 = 0.06, \tilde{p}_3 = 0.048, \tilde{p}_4 = 0.07, \tilde{p}_5 = 0.048$

Method of Shaffer

- Improved FWER correction: account for logical relations that exist between hypotheses
- Perform test at $\alpha/t(j)$
  with $t(j)$ the maximum number of H_0 that can still be true, given the rejection of $j-1$ hypotheses.
- $t(j) \leq n - j + 1$, thus Shaffer uniformly outperforms Holm's method.
- Also useful for post hoc tests

FDR: False discovery rate

- FDR: Expected proportion of false positives on the total number of positives you return.
- An FDR of 1% means that on average we expect 1% false positive proteins in the list of proteins that are called significant.

- Defined by Benjamini and Hochberg in 1995

$$\text{FDR}(|t_\text{thres}|) = \text{E}\left[\frac{FP}{FP+TP}\right] = \frac{\pi_0 Pr(|T| \geq t_\text{thres}|H_0)}{Pr(|T| \geq t_\text{thres})}$$

$$\text{FDR}_\text{BH}(|t_\text{thres}|) = \frac{1 \times p_{t_\text{thres}}}{\frac{\#|t_i| \geq t_\text{thres}}{m}} = \frac{p_{t_\text{thres}} \times m}{\#|t_i| \geq t_\text{thres}}$$

FDR: step-up method of Benjamini-Hochberg (BH-FDR)

- order p-values: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$
- $\hat{k} = \max\{k : p_{(k)} \leq \frac{k\alpha}{m}\}$
- If $\hat{k}$ exists, reject all null hypotheses corresponding to $p_{(1)}, \ldots, p_{(k)}$.
- If no such $\hat{k}$ exists, accept all null hypotheses.
- $\tilde{p}_{(i)} = \min\limits_{j=i,\ldots,m} min(\frac{m}{j} p_{(j)}, 1)$

BH-FDR: example

- Example: test 5 genes simultaneously, $\alpha = 0.05$
  $p_1 = 0.002, p_2 = 0.03, p_3 = 0.012, p_4 = 0.7, p_5 = 0.015$

$p_{(i)}$	0.002	0.012	0.015	0.03	0.7

$$\longleftarrow$$

$\frac{i\alpha}{m}$	$0.05\ \frac{1}{5}$	$0.05\ \frac{2}{5}$	$0.05\ \frac{3}{5}$	$0.05\ \frac{4}{5}$	$0.05\ \frac{5}{5}$
	0.01	0.02	0.03	0.04	0.05
reject	yes	yes	yes	yes	no
$\tilde{p}_{(i)}$	0.01	0.030	0.025	0.0375	0.7

BH-FDR: example

- Example: test 5 genes simultaneously, $\alpha = 0.05$
  $p_1 = 0.002, p_2 = 0.03, p_3 = 0.012, p_4 = 0.7, p_5 = 0.015$

$p_{(i)}$	0.002	0.012	0.015	0.03	0.7

$\longleftarrow$

$\frac{i\alpha}{m}$	$0.05\ \frac{1}{5}$	$0.05\ \frac{2}{5}$	$0.05\ \frac{3}{5}$	$0.05\ \frac{4}{5}$	$0.05\ \frac{5}{5}$
	0.01	0.02	0.03	0.04	0.05
reject	yes	yes	yes	yes	no
$\tilde{p}_{(i)}$	0.01	0.025	0.025	0.0375	0.7

- reject $H_{01}, \ H_{02}, \ H_{03}, \ H_{05}$
- $\tilde{p}_1 = 0.01, \tilde{p}_2 = 0.0375, \tilde{p}_3 = 0.025, \tilde{p}_4 = 0.7, \tilde{p}_5 = 0.025$

Comparison Bonferroni, Holm and BH-FDR