Lecture 30: Multiple testing considerations

Motivation: differential gene expression

Suppose a biologist is interested in identifying genes which are *differentially expressed* under different biological treatments. The biologist observes 10 subjects under treatment A, and 10 subjects under treatment B. Gene expression measurements $X_{i,j}$ (treatment A) and $Y_{i,j}$ (treatment B) are recorded for 1000 different genes ($i = 1, \ldots, 1000, j = 1, \ldots, 10$).

For each gene i, the biologist tests $H_0 : \mu_{i,A} = \mu_{i,B}$, rejecting when the p-value is below a threshold α.

If H_0 is actually true for all 1000 genes, how many false positives do we expect?

$\alpha \cdot 1000$

Motivation: multiple testing

In what other settings might we test many hypotheses?

Outcomes for multiple hypothesis tests

Suppose we test m null hypotheses $H_{0,1}, \ldots, H_{0,m}$ of which m_0 are truly null (H_0 is true)

Ideally, we reject for H_0 the $m - m_0$ true alternatives, and fail to reject the m_0 true nulls. Possible outcomes:

	H_0 true	H_0 false	
Reject	V		
Fail to reject			
	m_0	$m - m_0$	m

We want to control V, the # of false positives (# of type I errors).

$$FWER = P(V > 0)$$ (probability of at least one type I error)

Family-wise error rate

Definition: Suppose we test m null hypotheses $H_{0,1}, \ldots, H_{0,m}$. The *family-wise error rate* is the probability of making *at least one* type I error:

$$FWER = P\left(\bigcup_{i:H_{0,i} \text{ is true}} \{\text{reject } H_{0,i}\} \right)$$

Suppose all m tests are independent, and $H_{0,i}$ is true for all tests. For each test, we reject if the corresponding p-value $p_i < \alpha$. What is the FWER?

$$FWER = P\left(\bigcup_{i:H_{0,i} \text{ true}} \{\text{reject } H_{0,i}\} \right) = P\left(\bigcup_{i=1}^{m} \text{reject } H_{0,i} \right)$$

$$= 1 - P(\text{fail to reject all } H_{0,i})$$

$$1 - (1-\alpha)^m \qquad \text{(independence)}$$

The Sidak correction

$$FWER = P\left(\bigcup_{i:H_{0,i} \text{ is true}} \{\text{reject } H_{0,i}\} \right)$$

If all m hypotheses are independent, at what threshold α^* should we reject each test, such FWER $\leq \alpha$?

$FWER \leq 1 - (1-\alpha^*)^m$ (= exactly when all m are truly null)

$\Rightarrow 1-(1-\alpha^*)^m = \alpha \Rightarrow \alpha^* = 1-(1-\alpha)^{\frac{1}{m}}$

Example: $m=100$, $\alpha = 0.05$ $\alpha^* \approx 0.000\,51$

<u>Sidak correction</u>: Reject $H_{0,i}$ if $p_i < 1-(1-\alpha)^{\frac{1}{m}}$

The Bonferroni correction

union bound:

$P(\cup_i A_i) \leq \sum_i P(A_i)$

$$FWER = P\left(\bigcup_{i:H_{0,i} \text{ is true}} \{\text{reject } H_{0,i}\} \right)$$

Suppose we reject each $H_{0,i}$ if $p_i < \alpha^*$

$$P\left(\bigcup_{i:H_{0,i} \text{ is true}} \text{reject } H_{0,i} \right) \leq \sum_{i:H_{0,i} \text{ true}} P(\text{reject } H_{0,i}) \leq \sum_{i:H_{0,i} \text{ is true}} \alpha^* \leq m\alpha^*$$

<u>Bonferroni correction</u>: Reject when $p_i < \frac{\alpha}{m}$

$\Rightarrow$ $FWER \leq m\left(\frac{\alpha}{m}\right) = \alpha$

(in fact, $FWER \leq m_0\left(\frac{\alpha}{m}\right)$ but don't usually know m_0)

$\frac{\alpha}{m} < 1-(1-\alpha)^{\frac{1}{m}}$ so Bonferroni is more conservative than Sidak

Holm's procedure

Suppose we test 5 hypotheses, and observe p-values 0.4, 0.01, 0, 0, 0. Does it still seem reasonable to use the Bonferroni cutoff $\alpha/5$ for each test?

No – want to use other p-values to make a decision for each hypothesis

Idea: order p-values $p_{(1)} \le p_{(2)} \le \dots \le p_{(m)}$

First test: if $p_{(1)} < \frac{\alpha}{m}$ (Bonferroni threshold), reject $H_{0(1)}$

If we reject $H_{0(1)}$, consider $p_{(2)}$. There are now $m-1$ tests left, so reject $H_{0(2)}$ if $p_{(2)} < \frac{\alpha}{m-1}$

⋮

Continue procedure: reject $H_{0(i)}$ if $p_{(i)} < \frac{\alpha}{m-i+1}$

As soon as $p_{(i)} > \frac{\alpha}{m-i+1}$ for some i, stop the procedure (fail to reject $H_{0(i)}, \dots$)

Holm's procedure

FWER $\le P\left(\min_{i \in I_0} p_{(i)} < \frac{\alpha}{m_0}\right)$

(union bound) $\le m_0 \left(\frac{\alpha}{m_0}\right) = \alpha$ //

Suppose we test m null hypotheses $H_{0,1}, \ldots, H_{0,m}$. Let p_i be the corresponding p-value for test i.

- Order the p-values $p_{(1)} \le p_{(2)} \le \cdots \le p_{(m)}$
- Let $i^* = \min \left\{ i : p_{(i)} > \frac{\alpha}{m-i+1} \right\}$
- Reject $H_{0,(i)}$ for all $i < i^*$

Claim: Holm's procedure controls FWER at level α

Proof: Let $I_0 = \{i : H_{0(i)}$ is true$\}$. Let $m_0 = |I_0|$ (# of the nulls)

Let $j = \min(I_0)$ ($\forall i < j$, $H_{0(i)}$ is false)

Holm's procedure compares $p_{(j)}$ to $\frac{\alpha}{m-j+1}$. If $p_{(j)} > \frac{\alpha}{m-j+1}$, fail to reject all the nulls. And since $j = \min(I_0)$ and there are m_0 elements in I_0, $m-j+1 \ge m_0$

$\Rightarrow$ fail to reject all the nulls if $p_{(j)} > \frac{\alpha}{m_0}$