Multiple testing issues

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?

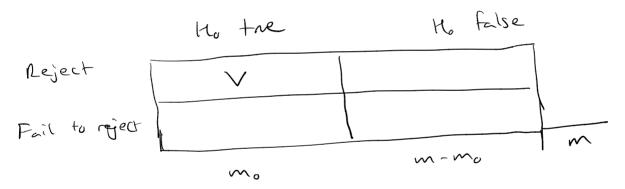
Motivation: multiple testing

In what other settings might we test many hypotheses?

Outcomes for multiple hypothesis tests

Suppose we test m null hypotheses Hari, ..., Harm, of which mo are true. Ideally we would fail to reject the mo true nulls, and reject the m-mo others.

Possible autromes:



we want to control V, the # of false positives (type I errors). One option is the family-wise error rate (FWER):

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:

$$FWER = P\left(igcup_{i:H_{0,i} ext{ is true}} \{ ext{reject } H_{0,i}\}
ight)$$

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?

The Sidak correction

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:

$$FWER = P\left(igcup_{i:H_{0,i} ext{ is true}} \{ ext{reject } H_{0,i}\}
ight)$$

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

FWER
$$= 1 - (1 - \alpha^*)^m$$
 (where band advised when all Ho, i the)
 $= 7 \ 1 - (1 - \alpha^*)^m = \alpha$ $= 7 \ \alpha^* = 1 - (1 - \alpha)^{\frac{1}{m}}$
Sioul correction; Reject Ho, i if $pi \ 1 - (1 - \alpha)^{\frac{1}{m}}$

The Bonferroni correction

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:

$$FWER = P\left(igcup_{i:H_{0,i} ext{ is true}} \{ ext{reject } H_{0,i}\}
ight)$$

Suppose we reject each Ma, if pi 2 was

Bonferroni correction: Reject when PiZ in $= \infty$

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?

Holm's procedure

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

- lacktriangledown Order the p-values $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$
- lacktriangle Let $i^* = \min\left\{i: p_{(i)} > rac{lpha}{m-i+1}
 ight\}$
- lacktriangledown Reject $H_{0,(i)}$ for all $i < i^*$

Claim: Holm's procedure controls FWER at level α Proof: Let $I_0 = \{i : H_{0lis} \text{ is tre} \}$. Let $m_0 = \#\{i : H_{0lis} \text{ is tre} \}$ Let $j = \min(I_0)$. Holm's procedure compares P(i) to $\frac{d}{m-j+1}$ If $P(i) > \frac{d}{m-j+1}$, fail to reject all the nulls

Since $j = \min(I_0)$ and there are mo elements in I_0 , $m-j+1 \ge m_0 \implies fail$ to reject all the nulls if $P(i) > \frac{d}{m_0}$ $\implies FWER = P(\min_{i \in I_0} P_{ii}) \land \frac{d}{m_0} \implies m_0 \pmod{m_0} = \alpha \pmod{m_0}$ 9/10

Class activity

https://sta711-s23.github.io/class_activities/ca_lecture_28.html

for benferrani

```
Side K is slightly more pareful then benferrani, but requires independence.

Holm's procedure is more powerful then benferrani, and requires no additional assumptions why? Benferrani rejects if pi L \frac{d}{m}

For Molm, each p-value P(i) is compared to \frac{d}{m-i+1} = \frac{d}{m}

So the threshold is less steingent for Molm than
```