DSO 699: Statistics Theory
Special Topics in Data Sciences an Operations

Week 4

Adel Javanmard

USC Marshall
Department of Data Sciences and Operations

September 14, 2023

Last class

We talked about:

- **Global testing:** testing the global null

$$H_0 = \cap_{i=1}^{n} H_{0,i} \,,$$

  which holds if and only if all the individual nulls are true.
- χ^2-test (as an approximation to the Fisher's combination test)
- Power of χ^2 test
- Simes test
  - ✓ Another test for global null
  - ✓ Uniformly more power powerful than Bonferroni
- Test based on empirical cdf of p-values
  - ✓ The Kolmogorov- Smirnov Test
  - ✓ Anderson-Darling Test
  - ✓ Tukey's Higher-Criticism Test

Outline for today

1. Family-wise Error Rate (FWER)
2. Methods for controlling FWER
3. The Closure Principle

Family-wise Error Rate (FWER)

Global Testing versus Multiple Testing

	Healthy (n_0)	Prostate cancer (n_1)
Expression level of gene i	$Y_{ij}^{(0)}, 1 \leq j \leq n_0$	$Y_{ij}^{(1)}, 1 \leq j \leq n_1$

- 20K genes in the human body
- hundreds of healthy patients
- 50-100 patients with prostate cancer ($n_1 \ll n_0$)

$$H_{0,i} : \mathbb{E}[\, Y_{ij}^{(0)}] = \mathbb{E}[\, Y_{ij}^{(1)}]\,,$$

or

$$H_{0,i} :\; Y_{ij}^{(0)} \overset{d}{=} \; Y_{ij}^{(1)}\,.$$

Global Testing versus Multiple Testing

	Healthy (n_0)	Prostate cancer (n_1)
Expression level of gene i	$Y_{ij}^{(0)}, 1 \leq j \leq n_0$	$Y_{ij}^{(1)}, 1 \leq j \leq n_1$

- 20K genes in the human body
- hundreds of healthy patients
- 50-100 patients with prostate cancer ($n_1 \ll n_0$)

$$H_{0,i} : \mathbb{E}[\,Y_{ij}^{(0)}] = \mathbb{E}[\,Y_{ij}^{(1)}]\,,$$

or

$$H_{0,i}:\; Y_{ij}^{(0)} \overset{d}{=} \; Y_{ij}^{(1)}\,.$$

Test statistic for $H_{0,i}$:

$$T = \frac{\bar{Y}_i^{(1)} - \bar{Y}_i^{(0)}}{s_i} \qquad\qquad s_i :\; \text{estimate of SE}$$

Construct p-values: $p_i = \mathbb{P}(|t_{n_0+n_1-2}| > |T_i|)$, (tail of student t-distribution)

Global Testing versus Multiple Testing

- **Global testing:** Testing the global null $H_0 = \cap_i H_{0,i}$
  (asks whether *any* gene is associated with the disease)
- **Multiple testing:** Testing the collection of null $H_{0,i}$ against the alternatives $H_{1,i}$
  (asks whether *this* gene is associate with the disease, for each gene, separately)

Global Testing versus Multiple Testing

- **Global testing:** Testing the global null $H_0 = \cap_i H_{0,i}$
  (asks whether *any* gene is associated with the disease)
- **Multiple testing:** Testing the collection of null $H_{0,i}$ against the alternatives $H_{1,i}$
  (asks whether *this* gene is associate with the disease, for each gene, separately)

Outcomes in multiple testing:

	Fail to reject	Rejected	Total
True null	U	V	n_0
False null	T	S	$n - n_0$
Total	$n - R$	R	n

Quantities of primary interest: V (# false discoveries), R (# total discoveries)

Family-wise Error Rate (FWER)

FWER: probability of making at least one false discovery

$$\mathrm{FWER} = \mathbb{P}(V \geq 1)$$

We aim at controlling FWER less than a target level α, in

- *Strong sense:* under all configurations of true and false hypotheses
- *Weak sense:* under some (specific) configurations of true and false hypotheses (e.g. when all are nulls)

Family-wise Error Rate (FWER)

FWER: probability of making at least one false discovery

$$\mathrm{FWER} = \mathbb{P}(V \geq 1)$$

We aim at controlling FWER less than a target level α, in

- *Strong sense:* under all configurations of true and false hypotheses
- *Weak sense:* under some (specific) configurations of true and false hypotheses (e.g. when all are nulls)

Other variations:

$$k\text{-FWER} = \mathbb{P}(V \geq k)$$

Methods for controlling FWER

Bonferroni's method

Procedure: reject $H_{0,i}$ if and only if $p_i \leq \alpha/n$, for $i = 1, \ldots, n$.

Bonferroni's method

Procedure: reject $H_{0,i}$ if and only if $p_i \le \alpha/n$, for $i = 1, \dots, n$.

Control of FWER

Bonferroni's method controls FWER at level α (in strong sense):

$$\mathrm{FWER} \le \mathbb{E}[V] \le \frac{n_0}{n}\alpha\,.$$

Proof. By Markov's inequality,

$$\begin{aligned}\mathbb{P}(V \ge 1) \le \mathbb{E}[V] &= \mathbb{E}[\sum_i \mathbb{I}(H_{0,i} \text{ is true }, p_i \le \alpha/n)] \\ &= \sum_{i:H_{0,i}\text{true}} \mathbb{P}(p_i \le \frac{\alpha}{n}) \le \frac{n_0}{n}\alpha\,.\end{aligned}$$

Bonferroni's method

Procedure: reject $H_{0,i}$ if and only if $p_i \le \alpha/n$, for $i = 1, \dots, n$.

Control of FWER

Bonferroni's method controls FWER at level α (in strong sense):

$$\mathrm{FWER} \le \mathbb{E}[V] \le \frac{n_0}{n}\alpha\,.$$

Proof. By Markov's inequality,

$$\mathbb{P}(V \ge 1) \le \mathbb{E}[V] = \mathbb{E}[\sum_i \mathbb{I}(H_{0,i} \text{ is true }, p_i \le \alpha/n)]$$

$$= \sum_{i: H_{0,i} \text{true}} \mathbb{P}(p_i \le \frac{\alpha}{n}) \le \frac{n_0}{n}\alpha\,.$$

Rem1. Bonferroni controls the stronger notion $\mathbb{E}[V]$.
Rem2. Bonferroni's method is valid for (arbitrarily) dependent tests.

Šidák Procedure

Procedure: reject $H_{0,i}$ if and only if $p_i \leq 1 - (1-\alpha)^{1/n}$, for $i = 1, \ldots, n$.

- less stringent than the Bonferroni correction, but only slightly!

$$1 - (1-\alpha)^{1/n} = \frac{\alpha}{n} + \frac{1}{2n}(1 - \frac{1}{n})\alpha^2 + \ldots$$

- Only applicable to independent p-values!

Šidák Procedure

Procedure: reject $H_{0,i}$ if and only if $p_i \leq 1-(1-\alpha)^{1/n}$, for $i=1,\dots,n$.

- less stringent than the Bonferroni correction, but only slightly!

$$1-(1-\alpha)^{1/n} = \frac{\alpha}{n} + \frac{1}{2n}(1-\frac{1}{n})\alpha^2 + \dots$$

- Only applicable to independent p-values!

Why it controls FWER?

$$\mathbb{P}(V \geq 1) = 1-\mathbb{P}(V=0) = 1-(1-\alpha_{th})^{n_0} \leq 1-(1-\alpha_{th})^{n}$$

So we want $1-(1-\alpha_{th})^n < \alpha \rightarrow \alpha_{th} \leq 1-(1-\alpha)^{1/n}$.

Weak control of FWER

A two-stage procedure: Consider the following multiple testing procedure (proposed by Fisher):

1. Construct a valid test for global null $H_0 = \cap_{i=1}^{n} H_{0,i}$
2. If H_0 is rejected then reject $H_{0,i}$ if $p_i \leq \alpha$

Weak control of FWER

A two-stage procedure: Consider the following multiple testing procedure (proposed by Fisher):

1. Construct a valid test for global null $H_0 = \cap_{i=1}^{n} H_{0,i}$
2. If H_0 is rejected then reject $H_{0,i}$ if $p_i \leq \alpha$

Lemma

The above two-stage procedure controls FWER under the global null (*Weak* control!)

- *Why?*

Weak control of FWER

A two-stage procedure: Consider the following multiple testing procedure (proposed by Fisher):

1. Construct a valid test for global null $H_0 = \cap_{i=1}^{n} H_{0,i}$
2. If H_0 is rejected then reject $H_{0,i}$ if $p_i \leq \alpha$

Lemma

The above two-stage procedure controls FWER under the global null (*Weak* control!)

- Why?

- Does it control FWER in a strong case?

Weak control of FWER

A two-stage procedure: Consider the following multiple testing procedure (proposed by Fisher):

1. Construct a valid test for global null $H_0 = \cap_{i=1}^{n} H_{0,i}$
2. If H_0 is rejected then reject $H_{0,i}$ if $p_i \leq \alpha$

Lemma

The above two-stage procedure controls FWER under the global null (*Weak control!*)

- *Why?*
- *Does it control FWER in a strong case?*

Suppose we use Bonferroni's method for the first stage. Consider one very strong signal (say p-value = 10^{-10} and $n = 10^5$).
If all other hypotheses are null you could make on average $n\alpha$ false rejections!

Holm's Step-Down Procedure

FWER version of Simes
Simes with stopping

Consider the ordered p-values

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(n)}$$

corresponding to null hypotheses $H_{(1)}, \ldots, H_{(n)}$.

- **Step 1:** If $p_{(1)} \leq \alpha/n$, then reject $H_{(1)}$ and go to Step 2. Otherwise, accept $H_{(1)}$, $H_{(2)}$, ..., $H_{(n)}$ and stop.
- **Step i:** If $p_{(i)} \leq \alpha/(n-i+1)$, then reject $H_{(i)}$ and go to Step $i+1$. Otherwise, accept $H_{(i)}$, $H_{(i+1)}$, ..., $H_{(n)}$ and stop.
- **Step n:** If $p_{(n)} \leq \alpha$, then reject $H_{(n)}$. Otherwise, accept $H_{(n)}$.

Holm's Step-Down Procedure (alternative description)

straingforward rephrase, no trick.

stepdown procedure is easily rephrased as data-dependent threshold

1. Order the p-values: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(n)}$

2. Define

the first time

$$L = \min\left\{ j : \ p_{(j)} > \frac{\alpha}{n + 1 - j} \right\}$$

strict inequality

3. Reject all $H_{0,j}$ for which $p_j < p_{(L)}$.

i.e. data-driven

- The rejection threshold $p_{(L)}$ depends on all p-values. In contrast, Bonferroni's threshold α/n is regardless of p-values.
- Holm's method makes no independence assumptions about the tests! (same as Bonferroni's method)
- Holm's method is less conservative than Bonferroni. (Why?)

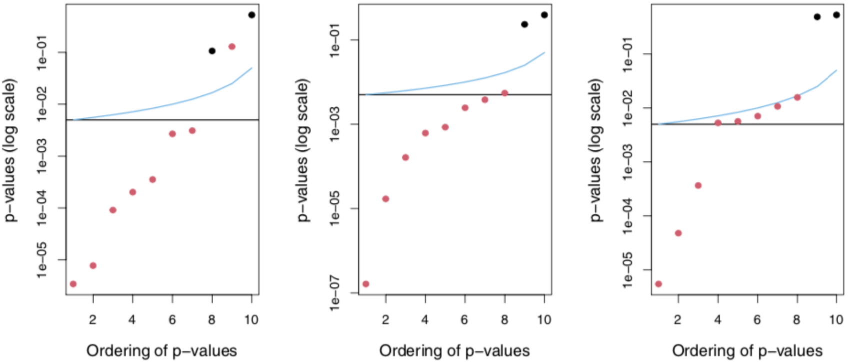
Holm's Procedure (cont'd)

Control of FWER

Holm's procedure controls the FWER at level α in a strong sense.

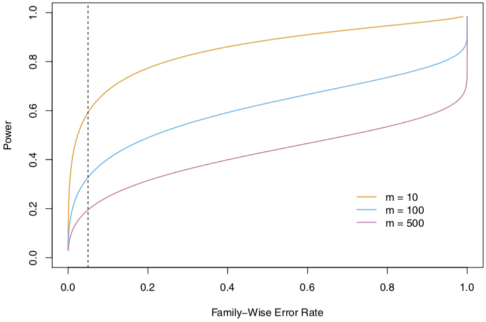
Proof. (Whiteboard)

Examples. (*Comparison of Holm's and Bonferroni's procedures*)



Trade-Off Between the FWER and Power

A simulation illustration:



(m: total number of tests)

- As m grows, controlling FWER results in lower power
- For $m = 1000$ attempting to control the FWER will make it almost impossible to make any rejection!
- We may want to work with a different (more relaxed) measure of error in large scale testing.

The Closure Principle

connection btwen global and multiple testin

An intriguing question ...

Q: Suppose you are given a valid method for global testing (i., type I error is controlled under α for testing global null). Can you use it to develop a procedure to control FWER?

Given a family of hypotheses $\{H_i\}_{i=1}^n$ with corresponding p-values $\{p_i\}_{i=1}^n$. Define the closure of the family as

$$H_I := \cap_{i \in I} H_i \quad \text{for all } I \subseteq \{1, \dots, n\}$$

index subset

You are given a method that for each I, gives a test ϕ_I for testing H_I so that

$$\mathbb{P}(\phi_I = 1 | H_I \text{ is true }) \leq \alpha \, .$$

(Note that H_I is the "global null" for the set I).
(e.g., ϕ_I can be constructed using Bonferroni, Simes, Fisher's test, etc)

black box global testing

Closure Principle

A two-pass procedures:

- First test each H_I using the α-test ϕ_I
- Reject H_i if and only if for all $I \supseteq \{i\}$, H_I is rejected.
  i.e. all related subset_global_test is rejected

In other words, (recall that $\phi_I = 1$, if rejection), let

$$T_i = \min_{I \supseteq \{i\}} \phi_I$$

and reject if $T_i = 1$

An example

Consider $n = 4$ hypotheses

not tree

$$\underline{H_{1234}}$$

$$\underline{H_{123}} \quad \underline{H_{124}} \qquad \underline{H_{134}} \quad H_{234}$$

$$\underline{H_{12}} \quad \underline{H_{13}} \quad \underline{H_{14}} \qquad H_{23} \quad H_{24} \quad H_{34}$$

$$\underline{H_1} \quad \underline{H_2} \qquad H_3 \quad H_4$$

these must be tested too

start from single global testing

Suppose that the underlined are the rejected ones at level α based on ϕ_I.
Which of the individual tests should be rejected by the closure procedure?

Properties of the closure procedure

Theorem (FWER control)

The closure principle controls the FWER, in a strong sense.

Proof. (Whiteboard)

Properties of the closure procedure

Theorem (FWER control)

The closure principle controls the FWER, in a strong sense.

Proof. (Whiteboard)

Computational complexity. A drawback of the closure procedure is the computational complexity (need to test $2^n - 1$ test for a collection of n tests.)

A reduction? Sometimes! (for specific test ϕ_I)

depends on what global test we use i.e. BF?

Closing Bonferroni

Suppose we use Bonferroni's method to construct ϕ_I:

$$\phi_I = 1 \Leftrightarrow \min_{i \in I} p_i \leq \frac{\alpha}{|I|}$$

Reduction: Sort the p-values $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(n)}$
Let $S_j := \{(j), (j+1), \ldots, (n)\}$ (the index set for the top $n - j + 1$ p-values)
Consider now any subset I so that $(j) \in I \subseteq S_j$. We claim that

fix j.

$$\phi_{S_j} = 1 \Longrightarrow \phi_I = 1$$

If prove this claim, then we can only test $\{H_{S_j}\}_{j=1}^{n}$.

Closing Bonferroni (cont'd)

claim

For any subset $(j) \in I \subseteq S_j$,

$$\phi_{S_j} = 1 \Longrightarrow \phi_I = 1$$

Proof.

$$\phi_{S_j} = 1 \to p_{(j)} = \min\{p_i : i \in S_j\} \le \frac{\alpha}{|S_j|} \le \frac{\alpha}{|I|},$$

since $I \subseteq j$. But $p_{(j)} = \min\{p_i : i \in I\}$. So $\phi_I = 1$.

2 **Summary.** We only need to test $\{H_{S_j}\}_{j=1}^{n}$, so:

$$\begin{aligned} H_{(j)} \text{ is rejected by closed test } &\leftrightarrow \phi_{S_1} = 1, \ldots, \phi_{S_j} = 1 \\ &\leftrightarrow p_{(1)} \le \frac{\alpha}{|S_1|}, \ldots, p_{(j)} \le \frac{\alpha}{|S_j|} \\ &\leftrightarrow p_{(1)} \le \frac{\alpha}{n}, \ldots, p_{(j)} \le \frac{\alpha}{n-j+1} \end{aligned}$$

Closing Bonferroni (cont'd)

Does it remind you of anything?

Closing Bonferroni (cont'd)

Does it remind you of anything?
It is exactly the Holm's procedure!

Holm's procedure:

1. Define

$$L = \min\left\{j : \; p_{(j)} > \frac{\alpha}{n+1-j}\right\}$$

2. Reject all $H_{0,j}$ for which $p_j < p_{(L)}$.

Recap

We talked about

- Family-wise Error Rate (FWER): $\mathbb{P}(V \geq 1)$
  - ✓ Strong control
  - ✓ Weak control
- Methods for controlling FWER
  - ✓ Bonferroni's method
  - ✓ Šidák's method
  - ✓ Holm's step-down procedure
- The closure principle (global testing $\rightarrow$ FWER control)

FIGHT ON!