

# STATS300A - Lecture 18

Dominik Rothenhaeusler  
Scribed by Michael Howes

11/29/21

## Contents

<b>1</b>	<b>Announcements</b>	<b>1</b>
<b>2</b>	<b>Multiple testing</b>	<b>1</b>
2.1	Setting . . . . .	1
2.2	Motivation . . . . .	2
2.3	Different goals . . . . .	2
2.4	Comparing error rates . . . . .	3
<b>3</b>	<b>Multiple testing procedures</b>	<b>3</b>
3.1	Bonferroni correction . . . . .	3
3.2	Holm's procedure . . . . .	4
3.3	Hochberg's procedure . . . . .	5

## 1 Announcements

- The final exam is at 3:30pm December 8<sup>th</sup> Wednesday.
- The exam is an online timed assignment.
- The exam is three hours long and has the same rules as the midterm.
- Approximately 1/3 of the exam will be on the first half of the course and approximately 2/3 of the exam will be on the second half.

## 2 Multiple testing

Historically people have worked in a setting where first they will fix a question, then collect data and then perform inference. Today people are more likely to collect a lot of data, then ask data dependent questions and then do inference. This can be viewed as asking many many questions about the data and requires different techniques.

### 2.1 Setting

As before we have data  $X \sim \mathbb{P} \in \mathcal{P}$  where we observe  $X$  but we do not know  $\mathbb{P}$ . We are given  $n$  null hypotheses  $H_{0,i}$  for  $i = 1, \dots, n$ . Rather than thinking of each null as a partition of our parameter space we will work directly with  $\mathcal{P}$ . That is  $\mathcal{P} = H_{0,i} \cup H_{1,i}$  where  $H_{1,i} = H_{0,i}^c$ .

For each null hypothesis we have a p-value  $p_i$  such that under  $H_{0,i}$ ,

$$\mathbb{P}(p_i \leq t) \leq t.$$

That is, under the null  $H_{0,i}$ , the p-value  $p_i$  stochastically dominates the uniform distribution. For simplicity we will in fact assume that  $p_i$  is uniformly distributed under  $H_{0,i}$  so that  $\mathbb{P}(p_i \leq t) = t$  under  $H_{0,i}$ .

## 2.2 Motivation

What is the problem that multiple testing is meant to solve? Suppose that we have  $X_i \sim \mathcal{N}(\theta_i, 1)$   $i = 1, \dots, n$  where  $n = 10,000$  and  $X_i$  are independent. Suppose we want to test  $H_{0,i} : \theta_i = 0$  against  $H_{0,i} : \theta_i < 0$ . We can define our p-values as

$$p_i = \psi(X_i),$$

where  $\psi$  is the CDF of a standard Gaussian distribution. The test  $\phi_i = \mathbf{1}_{p_i \leq \alpha}$  is thus the UMP level- $\alpha$  test for  $H_{0,i}$ . If  $\alpha = 0.05$  and  $\theta_i = 0$  for all  $i$ , then we would expect approximately  $n \times \alpha = 500$  false discoveries.

## 2.3 Different goals

There are different quantities we can work with in multiple testing. For example:

- (a) We can test the *global null*. That is we wish to test the null  $H_0 = \bigcap_{i=1}^n H_{0,i}$  where every null is true. In this setting we wish to find a function  $\Phi_G : [0, 1]^n \rightarrow \{0, 1\}$  where  $\Phi_G$  is a function of our p-values  $p = (p_1, \dots, p_n)$  and  $\Phi_G(p) = 1$  means that for the given p-values  $p$  we reject the global null and  $\Phi_G(p) = 0$  means we do not reject the global null. In this setting we wish to control the *global type I error* of  $\Phi_G$  which is

$$\text{Global type I error}(\Phi_G) = \begin{cases} \mathbb{P}(\Phi_G = 1) & \text{if } H_0 \text{ is true,} \\ 0 & \text{if } H_0 \text{ is false.} \end{cases}$$

- (b) We can also work with the *family wise error rate (FWER)*. This is the probability of making one or more false discoveries. In this case we want a function  $\Phi : [0, 1]^n \rightarrow \{0, 1\}^n$  where each component  $\Phi_i$  is a function of our p-values and  $\Phi_i(p) = 1$  means we reject the null  $H_{0,i}$  and  $\Phi_i(p) = 0$  means we do not reject the null  $H_{0,i}$ . In this setting we can define a random variable  $V$  which counts the  $i$ 's such that  $\Phi_i(p) = 1$  and  $H_{0,i}$  is true. Thus  $V$  is the number of false discoveries. We then define

$$FWER = FWER(\Phi) := \mathbb{P}(V > 1).$$

We wish to find powerful procedures  $\Phi$  such that  $FWER \leq \alpha$ .

- (c) We can also work with the *false discover rate (FDR)*. Let  $R$  be the total number of rejections and define

$$FDR = FDR(\Phi) := \mathbb{E} \left[ \frac{V}{\max\{R, 1\}} \right].$$

Again we are interested in powerful procedures  $\Phi$  such that  $FDR \leq \alpha$ .

## 2.4 Comparing error rates

Given a procedure  $\Phi : [0, 1]^n \rightarrow \{0, 1\}$  for  $H_{0,i}$ ,  $i = 1, \dots, n$ , we can define a global procedure for  $H_0 = \bigcap_{i=1}^n H_{0,i}$  by

$$\Phi_G(p) = \max\{\Phi_1(p), \dots, \Phi_n(p)\}.$$

Thus  $\Phi_G$  rejects the global null  $H_0$  if and only if for some  $i$ ,  $\Phi_i$  rejects the null  $H_{0,i}$ . This procedure is natural in settings where we expect the false nulls to be sparse. For this choice of  $\Phi_G$ , we have the following comparison between the different quantities we want to control:

$$\text{Global null type I error}(\Phi_G) \leq FDR(\Phi) \leq FWER(\Phi).$$

*Proof.* If the global null  $H_0$  is false, then  $\text{Global null type I error}(\Phi_G) = 0$  so we automatically have  $\text{Global null type I error}(\Phi_G) \leq FDR$ . If  $H_0$  is true, then all of nulls  $H_{0,i}$  are true and so every rejection is a false rejection. This implies that  $V = R$  and so

$$FDR = \mathbb{P}(V > 0) = \mathbb{P}(\Phi_G = 1) = \text{Global null type I error}(\Phi_G).$$

For the second inequality we have  $V \leq R$  and so

$$\frac{V}{\max\{R, 1\}} \leq \frac{V}{\max\{V, 1\}} = \mathbf{1}_{V>0}.$$

Thus

$$FDR = \mathbb{E} \left[ \frac{V}{\max\{R, 1\}} \right] \leq \mathbb{E}[\mathbf{1}_{V>0}] = \mathbb{P}(V > 0) = FWER. \quad \square$$

The different error criteria have different uses.

- Testing the global null is for “detecting.”
- Controlling the FWER or FDR is for “locating.”

Multiple testing is an active area of research and if you are interested, you should consider attending the [International seminar on selective inference](#).

## 3 Multiple testing procedures

We will now consider a number of methods that can be used when doing multiple testing.

### 3.1 Bonferroni correction

Define  $\Phi_i = \mathbf{1}_{p_i \leq \frac{\alpha}{n}}$ . We will show that this procedure control FWER at  $\alpha$ . That is,  $FWER \leq \alpha$ . Note that

$$\begin{aligned} \mathbb{P}(V > 0) &= \mathbb{P}(\Phi_i = 1 \text{ for some } i \text{ such that } H_{0,i} \text{ is true}) \\ &\leq \sum_{i, H_{0,i} \text{ is true}} \mathbb{P}(\Phi_i = 1) \\ &= \sum_{i, H_{0,i} \text{ is true}} \mathbb{P}(p_i \leq \alpha/n) \\ &= \frac{n_0}{n} \alpha \\ &\leq \alpha, \end{aligned}$$

where  $n_0 \leq n$  is the number of true nulls. Note that we did not put any independence assumptions on our p-values. The optimality of Bonferroni depends on the correlation between our p-values. Suppose that  $p_i$  are all independent and uniform and consider a test of the form  $\Phi_i(p) = \mathbf{1}_{p_i \leq t}$  for some value of  $t$  that does not depend on  $i$ . Suppose that the global null is true. Under this assumption, we have

$$\begin{aligned} FWER &= 1 - \mathbb{P}(V = 0) \\ &= 1 - \mathbb{P}(p_i > t, \text{ for all } i) \\ &= 1 - (1 - t)^n. \end{aligned}$$

If we wish to have  $FWER = \alpha$  we get  $t = 1 - (1 - \alpha)^{1/n} \approx \alpha/n$ . So that Bonferroni is approximately optimal for small  $\alpha$ , large  $n$  and independent p-values. If instead the p-values have positive dependence, then Bonferroni is sub-optimal. Suppose in an extreme case that  $p_1 = \dots = p_n$ . Then

$$FWER = 1 - \mathbb{P}(V = 0) = 1 - \mathbb{P}(p_1 > t) = t.$$

So the optimal choice of  $t$  is  $\alpha$  which is much larger than  $\alpha/n$ . In the case of negative dependence, it can be shown that Bonferroni is optimal.

### 3.2 Holm's procedure

How can we improve Bonferroni? Note that after we reject  $H_{0,i}$  we have two possibilities. Either we have made a false discovery or we have made a true discovery and the remaining hypotheses become a multiple testing problem with  $n - 1$  null hypotheses. Thus after making one rejection we can “relax” the rejection criteria. More formally, we first order the p values

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)},$$

and relabel the corresponding null hypotheses  $H_{0,(1)}, H_{0,(2)}, \dots, H_{0,(n)}$ . Let

$$j = \min \left\{ i : p_{(i+1)} > \frac{\alpha}{n-i}, i = 0, 1, \dots, n-1, \right\}.$$

We then reject the nulls  $H_{0,(1)}, \dots, H_{0,(j)}$ . Thus  $p_{(i)} \leq \frac{\alpha}{n-i+1}$  for rejected  $H_{0,(i)}$ .

**Proposition 1.** *Holm's procedure controls FWER at level  $\alpha$ .*

*Proof.* Let  $i_0$  be the first index  $i$  for which  $H_{0,(i)}$  is true. The quantity  $i_0$  is a random variable since it depends on the ordering of the random variables  $p_i$ . Let  $n_0$  be the number of true nulls, we thus have  $i_0 \leq n - n_0 + 1$  and so  $n_0 \leq n - i_0 + 1$  and  $\frac{\alpha}{n_0} \geq \frac{\alpha}{n-i_0+1}$ . Now note that

$$\begin{aligned} FWER &= \mathbb{P}(V > 0) \\ &= \mathbb{P} \left( p_{(1)} \leq \frac{\alpha}{n}, p_{(2)} \leq \frac{\alpha}{n-1}, \dots, p_{(i_0)} \leq \frac{\alpha}{n-i_0+1} \right) \\ &\leq \mathbb{P} \left( p_{(i_0)} \leq \frac{\alpha}{n-i_0+1} \right) \\ &\leq \mathbb{P} \left( p_{(i_0)} \leq \frac{\alpha}{n_0} \right) \\ &= \mathbb{P} \left( p_i \leq \frac{\alpha}{n_0}, \text{ for some } i \text{ such that } H_{0,i} \text{ is true} \right) \\ &\leq \sum_{i, H_{0,i} \text{ is true}} \mathbb{P} \left( p_i \leq \frac{\alpha}{n_0} \right) \\ &= n_0 \cdot \frac{\alpha}{n_0} \\ &= \alpha. \end{aligned}$$

□

### 3.3 Hochberg's procedure

As before order the p-value and null hypotheses so that  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$  and  $p_{(i)}$  corresponds to  $H_{0,(i)}$ . Define

$$j = \max \left\{ i : p_{(i)} \leq \frac{\alpha}{n - i + 1}, i = 1, \dots, n \right\},$$

where we define  $\max \emptyset = 0$ . We then reject  $H_{0,(1)}, \dots, H_{0,(j)}$ . If the p-values are independent, then this procedure also has level  $\alpha$  FWER control. This procedure is more powerful than Holm's procedure in the sense that it rejects more often.