lecture 18 :

"The entire set of m tests is called a "family". A "family" is "any logical collection of inferences for which it is meaningful to take into account some combined measure of error" on a set of tests where you wish to prevent "data dredging" (e.g. the spurious correlations in 342) or to " ensure a correct "overall" decision in the collection of tests"

We will discuss 2 error properties/ metrics for a family of tests.

The first ~~the~~ is called " familywise error rate " (FWER) defined as:

$FWER := P(V > 0) \leq FWER_0$ ← This is the level of control that I chose e.g 5%

You can show that $FWER \leq FWER_0$ for any $m_0 \leq m$ subset of the m tests, this is called "strong FWER control". We won't study it.

If you can show that $FWER \leq FWER_0$ for $m_0 = m$ then this called "weak FWER control" which we will study. If $m_0 = m$..

		Decision		
		Retain H_0	Reject H_0	
Truth	H_0	U	V	m_0
	H_a	0	0	0
		F	R	m

$$\Rightarrow V = R \Rightarrow FWER = P(R > 0)$$

Our goal is weak FWER control under the most general settings.

$R_1 = 1$ if H_0 is rejected, $R_1 = 0$ if H_0 is retained.

$R_2 = 1$ if H_{0_2} is rejected, $R_2 = 0$ if H_{0_2} is retained

$\vdots$

$R_m = 1$ " " H_{0_m} " , $R_m = 0$, " H_{0_m} " "

$$FWER = P(R > 0) = P(R_1 = 1 \cup R_2 = 1 \cup \cdots \cup (R_m = 1))$$
$$\leq \sum_{i=1}^{m} P(R_i = 1) = m\alpha$$

Recall from Math 241;

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

the Principle of inclusion-exclusion:

$$P(A_1 \cup A_2 \cup \cdots \cup A_m) = \sum P(A_i) - \sum P(A_i - A_j)$$
$$+ \sum P(A_i \cap A_j \cap A_k) - \cdots + - + \cdots$$

and from here you can prove "Boole's inequality:"

$$P(A_1 \cup A_2 \cup \cdots \cup A_m) \leq \sum P(A_i) \ldots$$

$\Rightarrow FWER \leq FWER_0 \Rightarrow m\alpha = FWER_0$

$$\Rightarrow \alpha = \frac{FWER}{m}.$$

This is called the Bonferroni Collection (1936).

$\Rightarrow$ a pval for an individual test must be less than $FWER_0/m$.

Equivalently, you can multyply the Pvalues by $m/FWER_0$ and compare each to $\alpha = 5\%$

$$P_{val} \leq FWER_0/m = \alpha$$

$$\Rightarrow \underbrace{\frac{m}{FWER_0} P_{val}}_{\text{Adjusted p-values}} \leq \alpha.$$

e.g. if $m = 30$, $FWER_0 = 5\% \Rightarrow \alpha = FWER_0/m$
$= 0.00167$.

The obvious problem with this correction is.. it gives you really bad power! Because it is ultra-conservative.

We can do a bit better if we assume the tests are independent. Then,

$$R_1, R_2, \ldots R_m \overset{iid}{\sim} \text{Bern}(\alpha)$$

$$\Rightarrow R \sim \text{Bin}(m, \alpha).$$

$$FWER = P(R>0) = 1 - P(R=0) = 1-(1-\alpha)^m \leq FWER_0$$

$$\Rightarrow 1 - FWER_0 = (1-\alpha)^m$$

$$\Rightarrow 1-\alpha = (1-FWER_0)^{1/m}$$

$$\Rightarrow \alpha = 1 - (1-FWER_0)^{1/m}.$$

This is known as "Dunn- Sidak correction" (1967)

e.g. if $m = 30$, $FWER_0 = 5\% \Rightarrow \alpha = 1 - (95\%)^{1/30}$
$= 0.00171 >$
$.00167$
(the Bonferroni)

Thus you get slightly higher power with Sidak.

$1 - (1-x)^{1/c} \approx x/c$ (First order tylor Series)

There are other methods e.g the "Holm-step-down" Procedure (1979) but we won't study it b/c it is similar to the Simes procedure (1986) which we talk about now. Bonferroni and Sidak never looked at the Pvalue and there's a lot of information there. Remember, Fisher created the p-value to gauge the "strength" of a rejection. Rejecting with a p-value of 0.00001 is much stronger than rejecting with a p-value of 0.01.
Holm and Simes used this for the m-tests,

You get p-values $P_1, P_2, \dots P_m$ but don't retain / reject anything yet!!!

Order them from smallest to largest:

$P_{(1)} \leq P_{(2)}, \dots P \leq P_{(m)}$ (order statistics)

$\uparrow$ min pvalue $\quad\quad\quad\quad$ $\downarrow$ max pvalue

Then locate the following:

"linear step-up"

$$a_* := \max\left\{a : P_{(a)} \leq \frac{a}{m} \text{FWER}_0\right\}$$

$\in \{\text{Bonferroni}, \dots, \alpha \text{ (naive)}\}$

or let $a_* = 0$ if max doesn't exist.

Then set $\alpha = \frac{a_*}{m} \text{FWER}_0$

You can prove that this gives you weak FWER control: it is more that

This is not more powerful than Bonferroni / Sidak

By Construction, you reject all tests upto the a^* th test (if the tests are in order of P value). Then you retain all the other $m-a^*$ tests.

The Problem with FWER in general is maybe it's too conservative: what if you want to trade some false rejections for more power? Let's consider another metric of familywise control (not FWER), called "False discovery Rate" (FDR). First define the "False discovery Proportion (FDP)

$$FDP := \begin{cases} \frac{V}{R} & \text{if } R>0 \\ 0 & \text{if } R=0 \end{cases}$$

the random proportion of rejections that any Type 1 errors

FDR := $E[FDP]$, the expected proportion of rejections that are Type I errors.

Now we wish to control FDR. so we want:

$FDR \leq FDR_0$, a constant you set.

For example if $FDR_0 = 5\%$ and I run m tests and get 100 rejections, then I except ≤ 5 of the rejections to be type I errors, ≥ 95 of the rejections to be ~~do~~ real discoveries.

Note: if $m = m_0$ the FWER = FDR.

Proof:

$$m = m_0 \Rightarrow V = R \Rightarrow FDP = \begin{cases} 1 & \text{if } R > 0 \\ 0 & \text{if } R = 0 \end{cases} = \text{Bern}(P(R>0))$$

Bern$(P(R>0))$

$\Rightarrow$ FDR $=$ E [FDP] $\leq$ P(R > 0) = FWER.

not test :

note, the FDR procedure is more powerful than the FWER procedure.

$\mathbb{1}_{V \geq 1} \geq V/R$ if $V = 0 \Rightarrow 0 \geq 0$

$V = 1 \Rightarrow 1 \geq 1$ or $1/2$ or $1/3$ ... $\forall R$

$V \geq 1 \Rightarrow 1 \geq 1$ or $\frac{V}{V+1}$ ... $\forall R$

$\Rightarrow E[\mathbb{1}_{V \geq 1}] \geq E[V/R]$

$P(V \geq 1) \geq$ FDR

$\Downarrow$

FWER $\geq$ FDR

Benjamini and Hochberg (1995) proved the Simes-procedure controls FDR for any m_0 subset of m tests.

$FDR = m_0/m$ $FDR_0 \leq FDR_0$ thus for a small m_0 (which you don't observe), the FDR control is much better than FDR_0