What if you want to control the Type I errors in your "family" of m tests. (So far in this class, you controlled the Type I error for m = 1 test at level alpha). Controlling Type II errors is done by maximizing power (through sample size, through better estimators, etc) so we will ignore that for this discussion.

Why is it important to control the Type I error rate for the m tests? Why not just set alpha like we've been doing all along?

Let's say alpha = 5% and m = 30 and $m = m_0$ which means all null hypotheses are true. Thus, all rejections constitute Type I errors. What is the probability you make at least one Type I error?

Decision

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

$$V = R \sim Bin(m, \alpha)$$

$$P(R > 0) = P(V > 0)$$

$$= 1 - P(R = 0) = 1 - \binom{m}{0}\alpha^0(1-\alpha)^m$$

$$= 1 - (1 - 5\%)^{30} = 76\% = FWER$$

If you don't "do something" chances are, you will make at least one mistake almost certainly as m increases.

Maybe you can't afford to have so many mistakes / such a high probability of making mistakes. What can be done?

Define Family-Wise Type I Error Rate (FWER) as the probability of at least one Type I error ("false negative" or "false discovery")

$$FWER := P(V > 0)$$

If you can show that $FWER \le FWER_0 \underset{\text{e.g.}}{=} 5\%$ for every $m_0 \in \{0, ..., m\}$ that is called "strong control of the FWER". This is difficult. We're not going to do it.
If you can show that $FWER \le FWER_0 \underset{\text{e.g.}}{=} 5\%$ for $m = m_0$ (i.e. V = R and no alternatives are true) then this is called "weak control of the FWER". We'll focus on this.

What exactly is a "family of m tests"? A family is any logical "collection of inferences for which it is meaningful to take into account some combined measure of error" or a "collection of tests where we wish to prevent 'data dredging' which is finding spurious relationships" or to "ensure a correct overall decision on the collection".

The following is called the "Bonferroni Correction" for FWER. It does not require independence among the m hypothesis tests.

Let R_1 = 1 if first null is rejected and 0 if retained
Let R_2 = 1 if second null is rejected and 0 if retained
⋮
Let R_m = 1 if mth null is rejected and 0 if retained

$R = \sum_{\ell=1}^{m} R_\ell$, α is Type I error setting for an individual test

$$P(R > 0) = P(R_1 = 1 \cup R_2 = 1 \cup \ldots \cup R_m = 1) \le \sum_{\ell=1}^{m} P(R_\ell = 1) = m\alpha$$

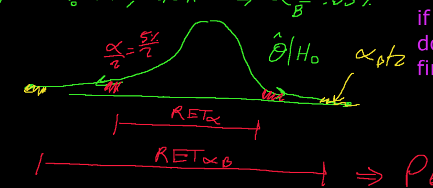
$$P(A_1 \cup \ldots \cup A_n) = \sum P(A_i) - \sum P(A_i, A_j) + \sum P(A_i, A_j, A_k) - + - \ldots$$
$$\le \sum P(A_i)$$

Boole's Inequality

$$\Rightarrow P(R > 0) \le m\alpha = FWER_0 \Rightarrow \alpha = \frac{FWER_0}{m}$$

Bonferroni, 1936 which is super duper conservative... if you employ it, don't expect to find many things!

e.g. $FWER_0 = 5\%,\ m = 100 \Rightarrow \alpha_B = .05\%$



$\Rightarrow Pow \downarrow\downarrow\downarrow$

If we assume the tests are independent we can do slightly better. Recall from above that R ~ Bin(m, α) so...

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

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

Dunn-Sidak Correction, 1967

$m = 100$,
$\alpha_B = .05\%$, $\alpha_{DS} = .0513\%$ which is slightly better but nothing to write home about

We will now talk about the Simes Procedure (1986) which is similar to the Holm step-down procedure (1979). In Bonferroni and Dunn-Sidak we used one "adjusted" alpha level for all tests. Here, we use a different alpha in each test and this gives us more power. We first run all tests and collected $pval_1$, $pval_2$, ..., $pval_m$. Remember, the pval measures the "strength of the rejection" e.g. a pval of 0.0001 is much "more of a rejection than a pval of 0.01 even though they're both "statistically significant" at 5%. Now, sort the pvals $p_{(1)} \le p_{(2)} \le ... \le p_{(m)}$ where $p_{(1)}$ is the smallest pval and $p_{(m)}$ is the largest pval. Then, locate:

$$a_* := \max\left\{a : p_{(a)} \le FWER_0 \frac{a}{m}\right\}$$

or let it be zero if the max doesn't exist i.e. that rhs condition is never fulfilled. This is called "linear step-up". At a = 1, the rhs is the Bonferroni level, at a = 2, it's larger, at a = 3, larger,... and a = m it is $FWER_0$. Equivalently you can also calculate adjusted pvals as follows:

$$a_* := \max\left\{a : p'_{(a)} := \frac{m}{a} p_{(a)} \le \underset{5\%}{\underbrace{FWER_0}}\right\}$$

Then, you reject all tests with p vals $p_{(1)}$, ..., $p_{(a_*)}$ and retain all other tests. In this procedure you will likely get more rejections hence more power with the same FWER control.

Then in the 1990's people started to ask the question "is FWER really what we should be so worried about?" Maybe instead you should care about the proportion of false discoveries i.e. the False Discovery Proportion (FDP) *not* whether you made one or more! Let's control the expected FDP which we will call False Discover Rate (FDR),

ignore the case where R = 0

$$FDR := E[FDP] = E\left[\frac{V}{R} \mathbb{1}_{R>0}\right]$$

We don't control FDP directly since V/R is a rv and we need to pick metrics about the rv to control. How about the expectation? That seems like a good place to start. Hence FDR control.

So let's say we control $FDR \le FDR_0$ (= 5% for example)
and we do m = 1000 tests and get r = 100 rejections. Then we expect ≤ 5 of these rejections to be Type I errors and ≥ 95 to be justified rejections (discoveries). Expect means if you ran m = 1000 test family many times.

If $m = m_0$, then FWER = FDR since V = R then

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

$$FDR = E[FDP] = FWER$$

FDR applies in cases where $m_0 < m$ unlike the weak control of FWER we discussed before.

Thm. Benjamini & Hochberg (1995) proved the Simes procedure controls FDR for any m_0. In fact they proved

$$FDR = \frac{m_0}{m} FDR_0 \le FDR_0$$

so the FDR could be substantially smaller than advertised depending on m_0. Proof beyond scope of course.

Proof that FDR has higher power than FWER i.e. it's always more liberal in allowing test rejections.

$$\mathbb{1}_{V>1} \ge \frac{V}{R} \mathbb{1}_{R>0}$$

$$E[\mathbb{1}_{V>1}] \ge E\left[\frac{V}{R} \mathbb{1}_{R>0}\right] := FDR$$

$$\| $$

$$FWER := P(V > 1)$$

$$\Longrightarrow FDR \le FWER$$

$\Longrightarrow$ when using Simes, you always get more rejections