Family-Wise Error Rate (function of V we like to control)

- The **family-wise error rate** is defined as the probability of committing a Type I Error in *any* of the M hypothesis tests.

$$\underbrace{FWER}_{\text{at least one Type I error}} = \Pr(V \geq 1)$$

- If each of the M tests are carried out with a significance level α, the FWER will be much greater than α

- Boole's inequality provides an upper bound:

$P(V \geq 1) = P(\bigcup_{k=1}^{M} \text{Type I error on test } k)$

$\leq \sum_{k=1}^{M} P(\text{Type I error on test } k) = M\alpha \; (\leq 1)$

- If we're willing to assume that the M tests are independent then:

$P(V \geq 1) = 1 - P(\underbrace{V = 0}_{\text{no type I errors}})$

$= 1 - P(\bigcap_{k=1}^{M} \text{no type I error on test } k)$

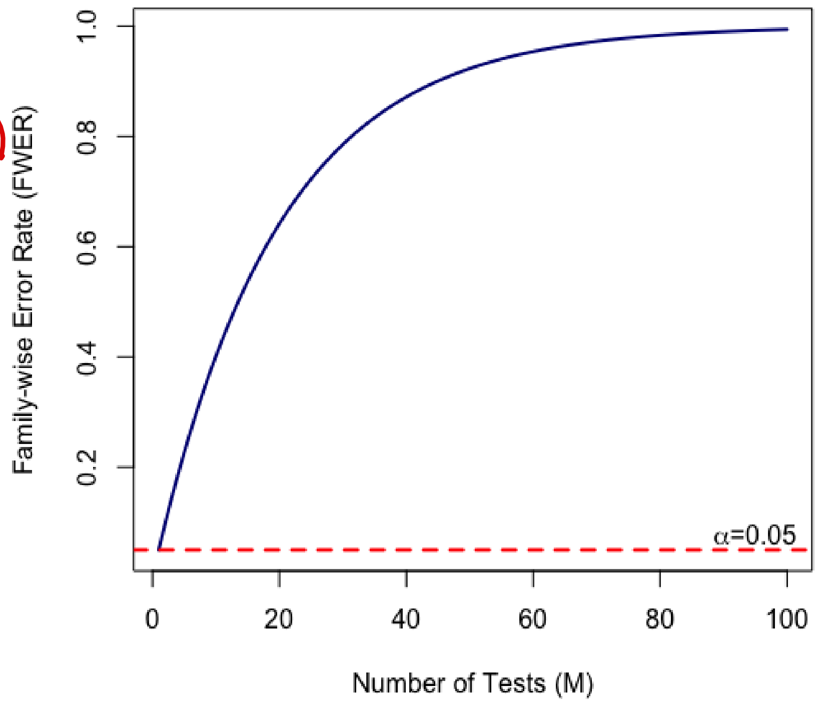
$= 1 - \prod_{k=1}^{M} (1 - \alpha)$

$= 1 - (1 - \alpha)^M$

$\lim_{M \to \infty} FWER = 1$

- This error rate, as a function of M is plotted below:

(assuming independence)



- A common value of M is $\binom{m}{2}$: the number of pairwise comparisons necessary to compare each condition to every other condition.
  - Example: $m = 5$

- Available to us are a variety of different statistical techniques that may be used to ensure the FWER does not exceed some threshold

$$FWER \leq \alpha^\star \in [0, 1]$$

- **Some general notation:**
  - Denote the M null hypotheses as: $H_{0,1}, H_{0,2}, \ldots, H_{0,M}$
  - Denote their corresponding p-values as: $p_1, p_2, \ldots, p_M$

- **A specific example:**
  - Suppose $M = 4$ hypotheses are tested and the resulting p-values are $p_1 = 0.015$, $p_2 = 0.029$, $p_3 = 0.008$, and $p_4 = 0.026$

The Bonferroni Correction

- This is the simplest method

- Reject $H_{0,k}$ if
$$p_k \leq \frac{\alpha^\star}{M}$$
for $k = 1, 2, \ldots, M$

test all hypos with sig level $\frac{\alpha^*}{M}$

- The procedure ensures $FWER \leq \alpha^\star$

(Boole's Ineq.)

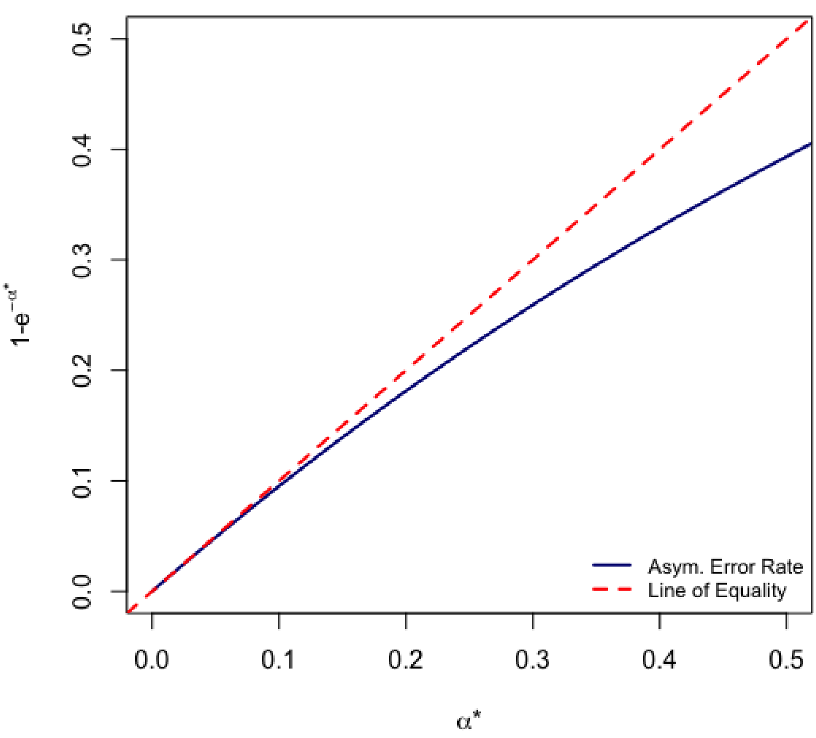
indv. Type I error under B.C.

- When independence is assumed the Bonferroni-corrected FWER becomes $1 - (1 - \frac{\alpha^\star}{M})^M$

$\to 1 - e^{-\alpha^*}$

$\approx \alpha^*$

for small α^*



- **Four-test Example:** $p_1 = 0.015, p_2 = 0.029, p_3 = 0.008, p_4 = 0.026$
  - Suppose that we wish to ensure $FWER \leq \alpha^\star = 0.05$

$FWER = \alpha^* = 1-(1-\alpha)^M$

$(1-\alpha)^M = 1-\alpha^* \implies \alpha = 1-(1-\alpha^*)^{\frac{1}{M}}$

The Šidák Correction

- This approach exploits the FWER formula derived when we assumed the M tests were independent.

- Reject $H_{0,k}$ if

$$p_k \leq 1 - (1-\alpha^\star)^{\frac{1}{M}}$$

for $k = 1, 2, \ldots, M$

- This is actually not much different from the Bonferroni correction since

$$\frac{\alpha^\star}{M} \overset{(\leq)}{\approx} 1 - (1-\alpha^\star)^{\frac{1}{M}}$$

  - For instance, take $\alpha^\star = 0.05$ and $M = 10$

- **Four-test Example:** $p_1 = 0.015, p_2 = 0.029, p_3 = 0.008, p_4 = 0.026$
  - Suppose that we wish to ensure $FWER \leq \alpha^\star = 0.05$

Holm's "Step-Up" Procedure

- The Bonferroni and Šidák corrections methods are very strict for large M
  - In these cases *most* null hypotheses will not be rejected

Type II errors ↑↑

- Ideally we would have an approach that is less strict but still controls the FWER at some $\alpha^\star$

- This is exactly what Holm's Procedure gives us!
  1. Order the M p-values from smallest to largest:

$$p_{(1)}, p_{(2)}, \ldots, p_{(M)}$$

     where $p_{(k)}$ is the k^{th} smallest p-value.

  2. Starting from $k = 1$ and continuing incrementally, compare $p_{(k)}$ to $\frac{\alpha^\star}{M-k+1}$. Determine $k^\star$ the smallest value of k such that

$$p_{(k)} > \frac{\alpha^\star}{M-k+1}$$

  3. Reject the null hypotheses $H_{0,(1)}, \ldots, H_{0,(k^\star-1)}$ and do not reject $H_{0,(k^\star)}, \ldots, H_{0,(M)}$.

- What's really happening?

$p_{(1)}$ vs. $\frac{\alpha^*}{M}$

$p_{(2)}$ vs. $\frac{\alpha^*}{M-1}$

$\vdots$

$p_{(k)}$ vs. $\frac{\alpha^*}{M-k+1}$

$\vdots$

$p_{(M)}$ vs. α^*

compare each $p_{(k)}$ to a Bonferroni-corrected sig. level based on the number of remaining comparisons

- **Proof** that Holm's Procedure really does control FWER:

WTS: $FWER = P(V \geq 1) \leq \alpha^* \in [0,1]$ for Holm's procedure

let $K_0 \subseteq \{1, 2, \cdots, M\}$ be the subset of indices for true H_0's

$\boxed{H_{0,(1)}, H_{0,(2)}, \cdots \cdots}, \cdots, H_{0,(M-1)}, H_{0,(M)}$

"Rejection region"

sw. inside, there $\exists$ $h \geq 1$ s.t. $H_{0,(h)}$ is the 1^{st} TRUE H_0 that was REJECTED $\Rightarrow p_{(h)} \leq \frac{\alpha^*}{M-h+1}$

$\Downarrow$

All $H_{0,i}$'s before $H_{0,(h)}$ are FALSE (# $= h-1$)

Also, we must have $\underline{h-1 \leq M-M_0}$

$\Leftrightarrow M_0 \leq M-h+1$

$\Leftrightarrow \frac{1}{M_0} \geq \frac{1}{M-h+1}$

$\Leftrightarrow \frac{\alpha^*}{M_0} \geq \frac{\alpha^*}{M-h+1}$

$\Rightarrow p_{(h)} \leq \frac{\alpha^*}{M-h+1} \leq \frac{\alpha^*}{M_0}$

Thus, $FWER = P(\exists k \in K_0 \text{ s.t. } p_k \le \frac{\alpha^*}{m_0})$

(for Holm's)

$$= P\left(\bigcup_{k \in K_0} p_k \le \frac{\alpha^*}{m_0}\right)$$

$$\le \sum_{k \in K_0} P\left(p_k \le \frac{\alpha^*}{m_0}\right)$$

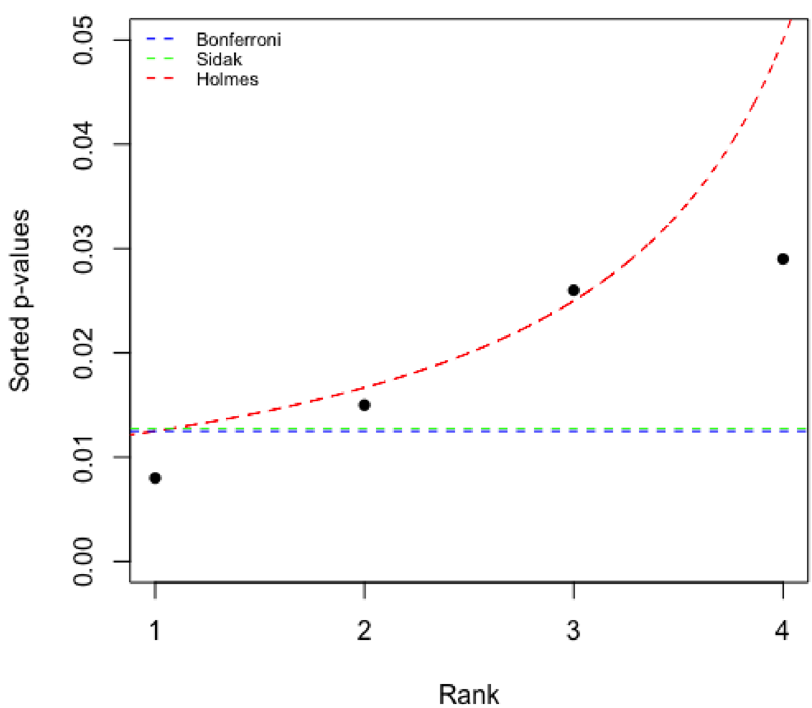
$$= \sum_{k \in K_0} \frac{\alpha^*}{m_0}$$

$$= m_0 \times \frac{\alpha^*}{m_0} = \alpha^*$$

p-values of true H_0's follow UNIF(0, 1) because how p-values are defined (upper tail cdf of a null dist)

- **Four-test Example:** $p_1 = 0.015, p_2 = 0.029, p_3 = 0.008, p_4 = 0.026$
  - Suppose that we wish to ensure $FWER \leq \alpha^\star = 0.05$

- The decision process for all three of these methods can be visualized by plotting the ordered p-values $p_{(k)}$ vs. their ranks $k = 1, 2, \ldots, M$ and overlay the significance thresholds



- The Bonferroni correction is most strict, followed by the Šidák correction, then by Holm's procedure.

Adjusted p-values (inflate the p-values or shrink the threshold)

- So far we have framed each of the correction procedures above as an adjustment to the significance threshold against which each p-value is compared.

- Alternatively (and equivalently) we could invert this process and frame the decision in terms of a comparison of *adjusted p-values* to $\alpha^\star$

- This is more familiar
  - We just need to adjust our p-values first

- The decisions made with the following adjusted p-values are identical to that achieved by comparing unadjusted p-values to the methods' adjusted significance thresholds
  - Bonferroni: Reject $H_{0,k}$ if $p_k^\star \leq \alpha^\star$ where

$$p_k^\star = Mp_k$$

  - Šidák: Reject $H_{0,k}$ if $p_k^\star \leq \alpha^\star$ where

$$p_k^\star = 1 - (1 - p_k)^M$$

$(1-p_k)^M = 1 - p_k^*$

p_k is way smaller than p_k^*

  - Holm: Reject $H_{0,(k)}$ if $p_{(k)}^\star \leq \alpha^\star$ where

$$p_{(k)}^\star = \max_{j \leq k}\{p_{(j)}(M - j + 1)\}$$

$\alpha^* = 0.05$

k	$p_{(k)}$	$M-k+1$	$p_{(k)} \times (M-k+1)$	$p^*_{(k)}$	$\frac{\alpha^*}{M-k-1}$	Reject?
1	0.008	4	0.032	0.032	0.013	Y
2	0.015	3	0.045	0.045	0.017	Y
3	0.026	2	0.052	0.052	0.025	N
4	0.029	1	0.029	0.052	0.050	N

Sample Size Determination

- So what does all of this mean for power analyses and sample size calculations?

- There is a duality between significance level and power. tradeoff
  - All else equal, reducing a test's significance level will increase the Type II Error rate and hence decrease power
  - Play around with this interactive app to gain comfort with this notion.

- Since all of our correction methods decrease the effective significance levels, the power of such tests is negatively impacted

- In order to maintain power at some pre-specified level, we must compensate by increasing the sample size

- Therefore, the more complicated your experiment (i.e., the more conditions it has), the larger your sample sizes will need to be
  - Such modifications can be accounted for when selecting a sample size
  - The significance level you use in your sample size calculations should be the adjusted one based on whichever correction method you use.
  - This is easier to do with *some* correction methods than others.

Optional Exercises:

- Proofs: 11, 12, 19
- R Analysis: 9, 14, 15, 16(f,g), 17(f,g,h), 23(g,h,k,l,m)
- Communication: 1(a), 1(b)

Primer on Logistic Regression

- Linear regression is an effective method of modeling the relationship between a single response variable (Y), and one or more explanatory variables ($x_1, x_2, \ldots, x_p$)
  - However, ordinary linear regression assumes that the response variable follows a normal distribution (i.e., $Y \sim N(\mu, \sigma^2)$)
  - When the response variable is binary, this assumption is no longer valid

- When we have a binary response, the Bernoulli distribution (i.e., $Y \sim BIN(1, \pi)$) is a much more appropriate distributional assumption
  - But ordinary linear regression is no longer appropriate
  - Instead we use **logistic regression**

- In the context of a linear regression model, the model is formulated so that the expected response (given the values of the explanataory variables) is equated to the **linear predictor** $\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$:

$$\mathrm{E}[Y|x_1, x_2, \ldots, x_p] = \mu = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

- In the context of logistic regression we also want to relate the expected response to the linear predictor
  - But now $\mathrm{E}[Y] = \pi \in [0, 1]$
  - And equating π and $\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$ does not make sense

- Instead we relate the linear predictor to $\mathrm{E}[Y] = \pi$ through a monotonic differentiable **link function** that maps $[0, 1] \to \mathbb{R}$
  - Logistic regression arises when this link function is chosen to be the **logit** function:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

  - Inverting this yields the expected response (given the values of the explanataory variables):

$$\mathrm{E}[Y|x_1, x_2, \ldots, x_p] = \pi = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}$$

- To interpret β_0 we set each explanatory variable to zero (i.e., $x_1 = x_2 = \cdots = x_p = 0$)
  - We see that β_0 is the **log-odds** that $Y = 1$ when $x_1 = x_2 = \cdots = x_p = 0$

  - Equivalently, e^{β_0} is the **odds** that the response would equal 1 when $x_1 = x_2 = \cdots = x_p = 0$

- The interpretation of β_j, for $j = 1, 2, \ldots, p$, is uncovered by considering the logistic regression equation for different values of x_j
  - Let π_x be the value of π when $x_j = x$ and let π_{x+1} be the value of π when $x_j = x + 1$

$$\begin{aligned} \log\left(\frac{\pi_{x+1}}{1-\pi_{x+1}}\right) - \log\left(\frac{\pi_x}{1-\pi_x}\right) &= (\beta_0 + \beta_1 x_1 + \cdots + \beta_j(x+1) + \cdots + \beta_p x_p) \\ &\quad -(\beta_0 + \beta_1 x_1 + \cdots + \beta_j x + \cdots + \beta_p x_p) \\ &= \beta_j \end{aligned}$$

  - Thus:

$$\log\left(\frac{\pi_{x+1}}{1-\pi_{x+1}} \Big/ \frac{\pi_x}{1-\pi_x}\right) = \beta_j$$

    and so β_j is interpreted as a **log-odds ratio** comparing the odds that $Y = 1$ when $x_j = x + 1$ vs. when $x_j = x$ (all else being equal)

  - Equivalently, e^{β_j} is interpreted as the **odds ratio**, comparing the odds that $Y = 1$ when $x_j = x+1$ vs. when $x_j = x$ (all else being equal)

- Parameter estimation in a logistic regression model is typically carried out with **maximum likelihood estimation**
  - This means that the $\widehat{\beta}$'s are maximum likelihood estimates, whose corresponding estimators have nice properties, such as

$$\widetilde{\beta} \,\dot{\sim}\, \mathrm{N}\left(\beta, \frac{1}{J(\beta)}\right)$$

  - A consequence of this is that hypotheses of the form

$$H_0 : \beta_j = 0 \text{ vs. } H_A : \beta_j \neq 0$$

    are done with *Z-tests* with test statistics given by

$$t = \frac{\widehat{\beta}_j}{\mathrm{SE}\left[\widehat{\beta}_j\right]}$$

  - In order to test hypotheses about several β's being simultaneously equal to zero, we use *likelihood ratio tests.*