Lecture 24
Multiple Testing

STAT 8010 Statistical Methods I
October 18, 2019

Whitney Huang
Clemson University

Why Multiple Testing?

- Recall the overall F-test in (one-way) ANOVA

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_J$$
$$H_a : \text{at least a pair } \mu\text{'s differ}$$

- If we reject H_0 in the overall F-test, then we need to perform multiple ($\binom{J}{2} = \frac{J\times(J-1)}{2}$) pairwise t-tests:

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_a : \mu_1 \neq \mu_2$$
$$H_0 : \mu_1 = \mu_3 \text{ vs. } H_a : \mu_1 \neq \mu_3$$
$$\vdots$$
$$H_0 : \mu_{J-1} = \mu_J \text{ vs. } H_a : \mu_{J-1} \neq \mu_J$$

- In this lecture we will learn how to conduct multiple testing while controlling the family-wise type I error

Review: Type I and II Errors

True State	Decision	
	Reject H_0	Fail to reject H_0
H_0 is true	Type I error	Correct
H_0 is false	Correct	Type II error

Errors in a single hypothesis test:

- The probability of a type I error is denoted by α
- The probability of a type II error is denoted by β

In multiple testing, we have an α chance of making a type I error **on each test**. However we would like to control the family-wise type I error rate

Family-Wise Error Rate (FWER)

> Family-Wise Error Rate (FWER) $\bar{\alpha}$: the probability of making 1 or more type I errors in a set of hypothesis tests

For m independent tests, each with individual type I error rate α, then we have

$$\bar{\alpha} = 1 - (1 - \alpha)^m$$

Suppose we are comparing 3 treatment groups (i.e. $J = 3$) and if we reject $H_0 : \mu_1 = \mu_2 = \mu_3$, then we need to perform $\binom{3}{2} = 3$ hypotheses tests ($m = 3$). Then we will have

$$\bar{\alpha} = 1 - (1 - 0.05)^3 = 0.1426$$

if we use $\alpha = 0.05$ for each test

FWERs for m Independent Tests Each at α Level

	α		
m	0.1	0.05	0.01
1	0.100	0.050	0.010
3	0.271	0.143	0.030
6	0.469	0.265	0.059
10	0.651	0.401	0.096
15	0.794	0.537	0.140
21	0.891	0.659	0.190

$\bar{\alpha}$ increases fairly quickly with J, the number of treatment groups

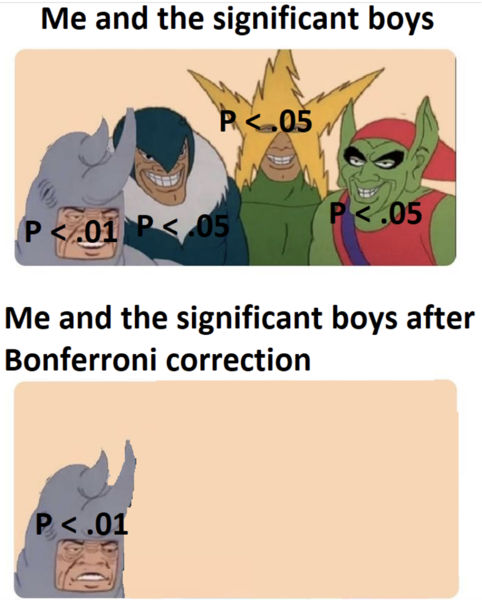
The Bonferroni Correction

If we would like to control the FWER to be α, then we adjust the significant level for each of the m tests to be $\frac{\alpha}{m}$

$$FWER = \mathbb{P}(\cup_{i=1}^{m} p_i \leq \frac{\alpha}{m}) \leq \sum_{i=1}^{m} \mathbb{P}(p_i \leq \frac{\alpha}{m}) = m\frac{\alpha}{m} = \alpha$$

For example, if we have 4 treatment groups ($m = 6$), then we will need to set the significant level for each individual pairwise t-test, to be $\frac{0.05}{6} = 0.0083$ to ensure that FWER is less than 0.05

The Bonferroni Correction Cont'd



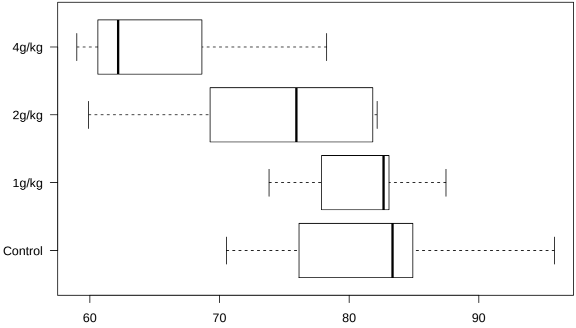
Example

A researcher who studies sleep is interested in the effects of ethanol on sleep time. She gets a sample of 20 rats and gives each an injection having a particular concentration of ethanol per body weight. There are 4 treatment groups, with 5 rats per treatment. She records Rapid eye movement (REM) sleep time for each rat over a 24-period.

Treatment	Control	1g/kg	2g/kg	4g/kg
Mean	82.2	81.0	73.8	65.7
Std	9.6	5.3	9.4	7.9

Recall in last lecture we reject $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

Example: Multiple Testing



P-value	μ_1, μ_2	μ_1, μ_3	μ_1, μ_4	μ_2, μ_3	μ_2, μ_4	μ_3, μ_4
Pooled	0.816	0.202	0.018	0.175	0.007	0.179
Non-pooled	0.818	0.202	0.019	0.185	0.009	0.180