MULTIPLE COMPARISONS AND CONTRASTS

RECAP

- Oneway ANOVA with model comparison approach
  - $E_r - E_f = SS_{Between}$
  - $SS_{Between}$ = differences in means
  - SS_{Within} = individual differences + random error
- Effect sizes
  - Eta-squared rules of thumb:
    - .01 - .059 = small
    - .06 - .139 = medium
    - >.14 = large
  - PRE rules of thumb:
    - 0 - .1 = small/weak
    - .1 - .4 = medium/moderate
    - >.4 = large/strong

INTRODUCTION TO MULTIPLE COMPARISONS

WHEN COMPARING MORE THAN TWO GROUPS…

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ ⟵ Omnibus
- The alternative hypothesis is NOT…
  - $H_A: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$
  - $H_A: \mu_1 \neq \mu_2 = \mu_3 = \mu_4$

T-TESTS VERSUS ONE-WAY ANOVA

- You could run separate t-tests comparing any two of the four groups

$\mu_1 \neq \mu_2$	$\mu_2 \neq \mu_3$	$\mu_3 \neq \mu_4$
$\mu_1 \neq \mu_3$	$\mu_2 \neq \mu_4$	
$\mu_1 \neq \mu_4$		

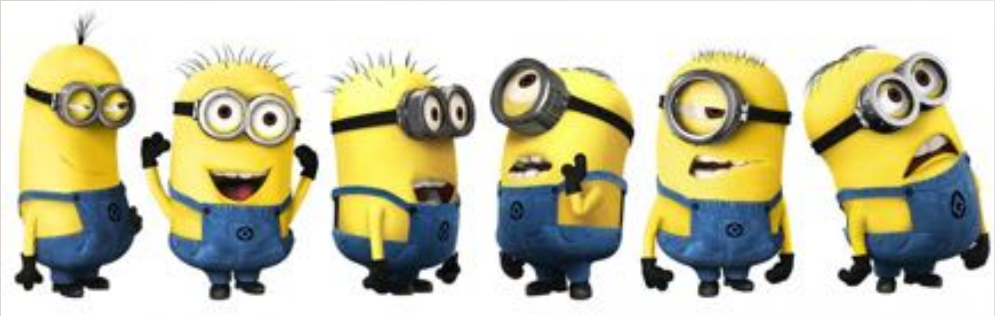
Number of tests would be a(a-1)/2

FAMILY-WISE ERROR

- Estimate the Type I error rate across all tests
- $1 - (1 - \alpha)^{a(a-1)/2}$ = the family wise error rate
- $1 - (.95)^6 = 1 - .735 = .265$

WE ARE FAMILY

- Here we have chosen an “obvious” family of tests



$\mu_1 \neq \mu_2$	$\mu_2 \neq \mu_3$	$\mu_3 \neq \mu_4$
$\mu_1 \neq \mu_3$	$\mu_2 \neq \mu_4$	
$\mu_1 \neq \mu_4$		

WE ARE FAMILY

- Here we have chosen an "obvious" family of tests

$\mu_1 \neq \mu_2$		
$\mu_1 \neq \mu_3$		
$\mu_1 \neq \mu_4$		

WE ARE FAMILY

- Here we have chosen an "obvious" family of tests

$\mu_1 \neq \mu_2$	$\mu_2 \neq \mu_3$	$\mu_3 \neq \mu_4$
$\mu_1 \neq \mu_3$	$\mu_2 \neq \mu_4$	
$\mu_1 \neq \mu_4$		
$\mu_1 \neq Mean(\mu_2, \mu_3, \mu_4)$		

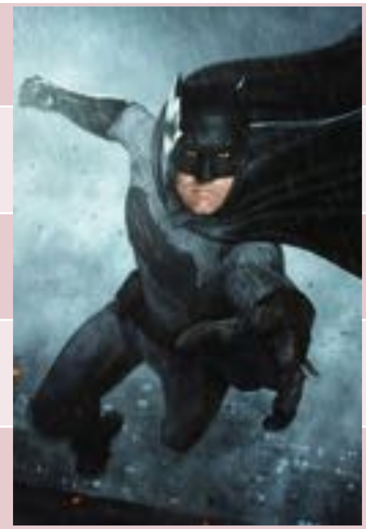
MORE BROADLY…

- Some contend that the "family" of tests could mean…
  - All tests related to a specific research question, within a single study
  - All tests using the same data
  - All tests related to a specific research question, across all studies
  - All tests reported in a paper
  - Any of the above plus the tests that were run for that topic but not reported

TODAY'S EXAMPLE

- Thanksgiving is one of the bigger box office weekends of the year
- Far more important a research question than anything we have considered yet

TODAY'S EXAMPLE

Adam West	Michael Keaton	Christian Bale	Ben Affleck
			

TODAY'S EXAMPLE

Adam West	Michael Keaton	Christian Bale	Ben Affleck
2	3	1	4
3	1	2	4
4	3	1	2
4	3	1	2
2	3	1	4

TODAY'S EXAMPLE

Adam West	Michael Keaton	Christian Bale	Ben Affleck
2	3	1	4
3	1	2	4
4	3	1	2
4	3	1	2
2	3	1	4
Ybar = 3	***2.6***	***1.2***	***3.2***

E_R

Value	Prediction	Error	Squared Error	
2	2.5	-.5	0.25	
3	2.5	.5	0.25	
4	2.5	1.5	2.25	
4	2.5	1.5	2.25	
2	2.5	-.5	0.25	
3	2.5	.5	0.25	
1	2.5	-1.5	2.25	
3	2.5	.5	0.25	E_R = 25
3	2.5	.5	0.25	
3	2.5	.5	0.25	
1	2.5	-1.5	2.25	
2	2.5	-.5	0.25	
1	2.5	-1.5	2.25	
1	2.5	-1.5	2.25	
1	2.5	-1.5	2.25	
4	2.5	1.5	2.25	
4	2.5	1.5	2.25	
2	2.5	-.5	0.25	
2	2.5	-.5	0.25	
4	2.5	1.5	2.25	

E_F

Value	Prediction	Error	Squared Error	
2	3	-1	1	
3	3	0	0	
4	3	1	1	
4	3	1	1	
2	3	-1	1	
3	2.6	0.4	0.16	
1	2.6	-1.6	2.56	
3	2.6	0.4	0.16	**E_F = 12.8**
3	2.6	0.4	0.16	
3	2.6	0.4	0.16	
1	1.2	-0.2	0.04	
2	1.2	0.8	0.64	
1	1.2	-0.2	0.04	
1	1.2	-0.2	0.04	
1	1.2	-0.2	0.04	
4	3.2	0.8	0.64	
4	3.2	0.8	0.64	
2	3.2	-1.2	1.44	
2	3.2	-1.2	1.44	
4	3.2	0.8	0.64	

CALCULATING THE F

$$F = \frac{(E_R - E_F) / (df_R - df_F)}{E_F / df_F}$$

$$F = \frac{(25 - 12.8) / (19 - 16)}{12.8 / 16}$$

$F = 5.083$ Critical $F(3, 16) = 2.462$

WHAT DOES THIS TELL US?

- It doesn’t answer the question the Internet was build for
- A priori versus post hoc comparisons

A PRIORI VS. POST HOC

- If a priori (planned ahead):
  - You could use a whole bunch of t-tests. But that doesn't limit the # of tests that you're running! So you get an inflated family wise error rate, and increase your chances of making a Type I error.
    - Rather than running a lot of t-test, you could use a planned contrast (see end of lecture)
  - Another option is to reduce your alpha for each comparison that you make (to compensate)
    - T-test + Bonferonni correction
  - Contrats (and maybe some bonferonni correction, but mostly contrasts alone) is the ideal
- What if you really don't know the outcome? Post-hoc tests!
  - Tukey
  - Scheffe

MULTIPLE COMPARISON CORRECTIONS

PAIRWISE VERSUS OTHER CONTRASTS

PAIRWISE

- Is Christian Bale ranked higher than Adam West?
- Is Ben Affleck ranked lowest of all Batmen?

CONTRAST

- Do Batmen from over 20 years ago rank lower than recent Batmen?
- Do people rank Michael Keaton as better than the average Batman?

IF YOU HAVE A SET OF TESTS…

- Bonferroni Correction
- If you have 6 planned tests, divide .05 evenly amongst the tests

T-TESTS

Batman 1	Batman 2	T-test	P-value
Adam West	Michael Keaton	0.667	.524
	Christian Bale	3.674	.006
	Ben Affleck	-0.302	.771
Michael Keaton	Christian Bale	3.130	.014
	Ben Affleck	-0.949	.371
Christian Bale	Ben Affleck	-3.780	.005

Alpha level = .05 / 6 = .0083

IF YOU DIDN'T CORRECT FOR FAMILYWISE ERROR

Batman 1	Batman 2	T-test	P-value
Adam West	Michael Keaton	0.667	.524
	Christian Bale	3.674	.006
	Ben Affleck	-0.302	.771
Michael Keaton	Christian Bale	3.130	.014
	Ben Affleck	-0.949	.371
Christian Bale	Ben Affleck	-3.780	.005

WHAT ISSUES MIGHT PRESENT WITH THE BONFERRONI ADJUSTMENT?

Perneger (1998; *BMJ*)

WHAT ISSUES MIGHT PRESENT WITH THE BONFERRONI ADJUSTMENT?

- Perneger (1998, doi: 10.1136/bmj.316.7139.1236)

Summary points

- Adjusting statistical significance for the number of tests that have been performed on study data —the Bonferroni method—creates more problems than it solves
- The Bonferroni method is concerned with the general null hypothesis (that all null hypotheses are true simultaneously), which is rarely of interest or use to researchers
- The main weakness is that the interpretation of a finding depends on the number of other tests performed
- The likelihood of type II errors is also increased, so that truly important differences are deemed non-significant
- Simply describing what tests of significance have been performed, and why, is generally the best way of dealing with multiple comparisons

IF YOU ONLY CARE ABOUT PAIRWISE…

- Tukey’s HSD (Honestly significant difference)
- What is the minimal difference between two means needed to declare something significantly different?

CALCULATING THE TUKEY

- What is the minimal difference between two means needed to declare something significantly different?
- Usually, we're looking for the q statistic, but we can rearrange the equation!
  - $q = \frac{Mean\ 1 - Mean\ 2}{\sqrt{\frac{MS\ within}{N\ per\ group}}}$
  - $Mean\ 1\ - Mean\ 2 = q\sqrt{\frac{MS\ within}{N\ per\ group}}$

CALCULATING THE TUKEY

- $Mean\ 1\ - Mean\ 2 = q\sqrt{\frac{MS\ within}{N\ per\ group}}$
- MS within = .8
- N per group = 5
- Q = 4.05…get this from a table based on your df and number of groups (or me)
- Mean 1 – Mean 2 = 1.62
- *"the minimal difference between 2 means to declare something significantly different needs to be 1.62"*

BACK TO THE DATA

Adam West	**Michael Keaton**	**Christian Bale**	**Ben Affleck**
Ybar = 3	***2.6***	***1.2***	***3.2***

BACK TO THE DATA

Adam West	Michael Keaton	Christian Bale	Ben Affleck
Ybar = 3	*2.6*	*1.2*	*3.2*

WHEN ALL ELSE FAILS…

- Scheffe test
- Benefits: Can be used for pairwise and/or other contrasts; often used with unequal sample sizes
- Downside: Ridiculously conservative
- Howell (2007)
  - "I can't imagine when I would ever use it, but I include it here because it is such a standard test"

HOW DO YOU DO A CONTRAST?

GENERAL FORMULA FOR A CONTRAST

$$F = \frac{\psi^2}{MS_W \Sigma(c_j^2 / n_j)}$$

ψ is our contrast

c reflects the coefficients for that contrast

IS THERE A RECENCY EFFECT IN OUR DATA?

- Are current Batmen rated as better than past Batmen?
- H_0: Average(Adam West and Michael Keaton) = Average(Christian Bale and Ben Affleck)
- H_A: Average(Adam West and Michael Keaton) $\neq$ Average(Christian Bale and Ben Affleck)

DEFINING THE CONTRASTS

- *First rule of contrasts: Coefficients must equal 0*
- -.5(Adam West and Michael Keaton) vs. .5(Christian Bale and Ben Affleck)
- c's for the contrast: -.5 -.5 .5 .5

CALCULATING ψ

Adam West	Michael Keaton	Christian Bale	Ben Affleck
Ybar = 3	***2.6***	***1.2***	***3.2***

- -.5(3) + -.5(2.6) + .5(1.2) + .5(3.2)
- ψ = -0.6

CALCULATING ψ

Adam West	Michael Keaton	Christian Bale	Ben Affleck
Ybar = 3	***2.6***	***1.2***	***3.2***
-0.5	**-0.5**	**0.5**	**0.5**

- ψ = -0.6
- 2.2 – 2.8

CALCULATING THE F

$$F = \frac{\psi^2}{MS_W \Sigma(c_j^2 / n_j)}$$

CALCULATING THE F

$$F = \frac{\psi^2}{MS_W \Sigma(c_j^2 / n_j)}$$

$$F = \frac{-0.6^2}{0.8\ \Sigma[(-0.5^2 / 5) + (-0.5^2 / 5) + (0.5^2 / 5) + (0.5^2 / 5)]}$$

CALCULATING THE F

$$F = \frac{\psi^2}{MS_W \Sigma(c_j^2 / n_j)}$$

$$F = \frac{-0.6^2}{0.8\ \Sigma[(-0.5^2 / 5) + (-0.5^2 / 5) + (0.5^2 / 5) + (0.5^2 / 5)]}$$

$F =$ 2.25 Critical $F(1, 16) = 4.49$

Cannot reject H_0 that the average of (μ_3 and μ_4) is different from average of (μ_1 and μ_2)

NEXT TIME

- Twoway (factorial) ANOVA

- If your work is with neuroimaging, this is required reading for why multiple comparisons are a problem:
  - Bennett, C. M., Baird, A. A., Miller, M. B., & Wolford, G. L. (2010). Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Arugment For Proper Multiple Comparisons Correction. *Journal of Serendipitous and Unexpected Results*, *1*(1), 1–5.
  - https://teenspecies.github.io/pdfs/NeuralCorrelates.pdf