

# Math 530/630 CM 5.3

## Contrasts, post-hoc tests & p-value adjustments



# Two key equations

When I introduced the *t*-test as a general linear model (GLM), I said that "all statistical procedures are basically the same thing":

$$outcome_i = (model) + error_i$$

Where:

- Outcome is your dependent variable (DV; also known as *y*) and
- Model is a linear function or linear combination of your independent variable(s) (IVs) (also known as *x*'s)

$$DV_i = (model) + error_i$$

"Essentially, all models are wrong, but some are useful"-George E. P. Box (1987)

$$deviation = \sum (observed - model)^2$$

# The sample mean

$$DV_i = (1b) + error_i$$

$$DV_i = (model) + error_i$$

Coefficient=1 is implied.

# Correlation

$$DV_i = (bIV_i) + error_i$$

$$DV_i = (model) + error_i$$

# Simple linear regression

$$DV_i = (b_0 + b_1 IV_i) + error_i$$

$$DV_i = (model) + error_i$$

Where:

- $b_0$  is the intercept term and
- $b_1$  is the slope.

# *ttest*

$$DV_i = (b_0 + b_1 IV_i) + error_i$$

$$DV_i = (model) + error_i$$

Where:

- $b_0$  is the intercept term ( $\bar{y}_{group1}$ )  
and
- $b_1$  is the slope ( $\bar{y}_{group2} - \bar{y}_{group1}$ )

# Analysis of Variance

$$DV_i = (b_0 + b_1 IV1_i + b_2 IV2_i) + error_i \quad DV_i = (model) + error_i$$

# GLM logic

- The simplest model we can ever conceive of for predicting a DV of interest is the mean of that DV.
- This is called the null (or reduced) model:
  - Simple linear regression:  $y_i = b_0 + \epsilon_i$
  - ANOVA:  $y_{ij} = \mu_{\bullet\bullet} + \epsilon_{ij}$  where  $\mu_{\bullet\bullet}$  is the ***grand mean***

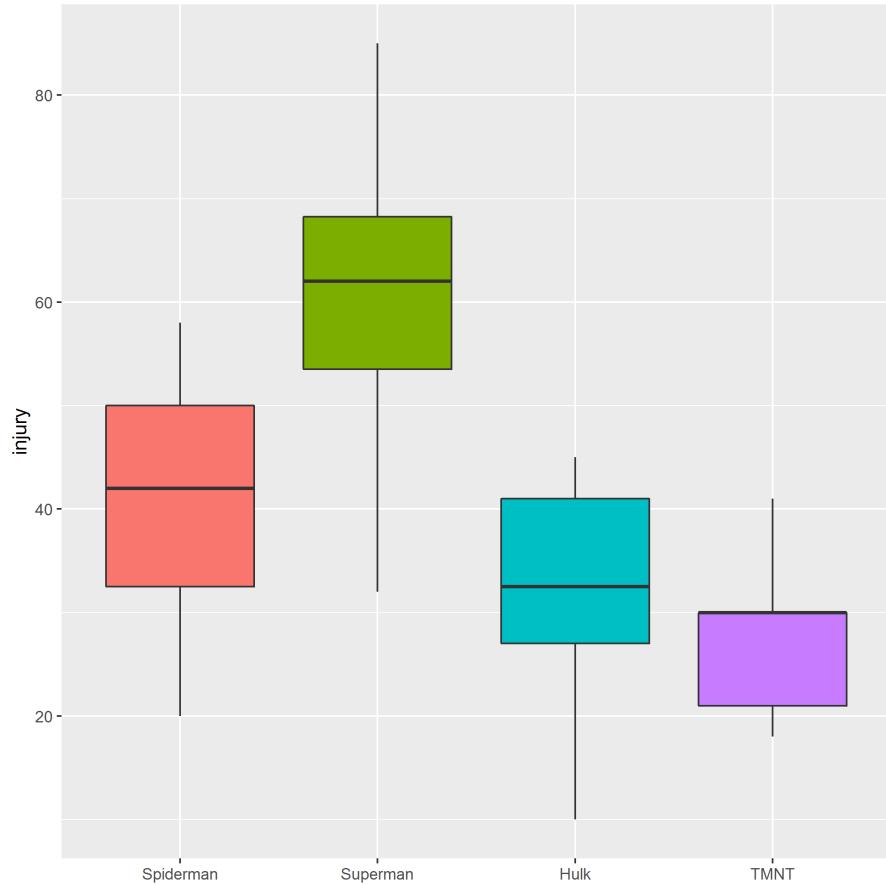
■ N.B. What is *not* present in these equations?

# Superhero ANOVA

```
superhero$hero <- factor(superhero$hero,labels=c("Spiderman", "Superman", "Hulk",  
heroaov <- aov(injury~hero,data=superhero)  
summary(heroaov)
```

```
Df Sum Sq Mean Sq F value Pr(>F)  
hero      3   3805   1268.3   6.716 0.00257 **  
Residuals 20   3777    188.9  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Now what?



# Contrasts!

Contrast	Spiderman	Superman	Hulk	TNMT
Spiderman vs. Superman	1	-1	0	0
Spiderman vs. Hulk	1	0	-1	0
Hulk vs. TNMT	0	0	1	-1

These say:

$$+1\mu_{spiderman} - 1\mu_{superman} + 0\mu_{hulk} + 0\mu_{TNMT}$$

$$+1\mu_{spiderman} + 0\mu_{superman} - 1\mu_{hulk} + 0\mu_{TNMT}$$

$$0\mu_{spiderman} + 0\mu_{superman} + 1\mu_{hulk} - 1\mu_{TNMT}$$

# Contrasts!

$$+1\mu_{spiderman} - 1\mu_{superman} + 0\mu_{hulk} + 0\mu_{TNMT} \rightarrow \mu_{spiderman} - \mu_{superman} = 0$$

$$+1\mu_{spiderman} + 0\mu_{superman} - 1\mu_{hulk} + 0\mu_{TNMT} \rightarrow \mu_{spiderman} - \mu_{hulk} = 0$$

$$0\mu_{spiderman} + 0\mu_{superman} + 1\mu_{hulk} - 1\mu_{TNMT} \rightarrow \mu_{hulk} - \mu_{TNMT} = 0$$

# Doing all pairwise contrasts in R

```
pairwise.t.test(superhero$injury, superhero$hero, p.adjust.method="none")
```

Pairwise comparisons using t tests with pooled SD

```
data: superhero$injury and superhero$hero
```

	Spiderman	Superman	Hulk
Superman	0.02222	-	-
Hulk	0.26156	0.00165	-
TMNT	0.12144	0.00056	0.64897

```
P value adjustment method: none
```

Base R's output=just the  $p$ 's, please!

# Doing all pairwise contrasts in R (better)

```
library(multcomp)
mcp <- glht(heroaov, linfct=mcp(hero="Tukey")) # "Tukey" sets up the pairwise compa
#summary(mcp) # Tukeys with single-step adjusted p-value
#summary(mcp, test=univariate()) #like a set of paired-ts (pooled sd 2 groups), bu
summary(mcp, test=adjusted("none")) #Same as pairwise.t.test() above
```

You get:

- the  $p$ 's plus
- the  $t$ 's
- the  $\psi$ 's
- the  $SE_{\psi}$ 's
- oh my!

# Doing all pairwise contrasts in R (better)

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = injury ~ hero, data = superhero)

Linear Hypotheses:

		Estimate	Std. Error	t value	Pr(> t )
Superman - Spiderman == 0		19.667	7.934	2.479	0.022218 *
Hulk - Spiderman == 0		-9.167	7.934	-1.155	0.261564
TMNT - Spiderman == 0		-12.833	7.934	-1.617	0.121436
Hulk - Superman == 0		-28.833	7.934	-3.634	0.001652 **
TMNT - Superman == 0		-32.500	7.934	-4.096	0.000562 ***
TMNT - Hulk == 0		-3.667	7.934	-0.462	0.648969

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
(Adjusted p values reported -- none method)

# Where are these numbers coming from?

$$\text{estimate} = \psi = \frac{\sum_{j=1}^k c_j \bar{y}_j}{\hat{\sigma}_\psi}$$

where...

$$SE_{est} = \hat{\sigma}_\psi = \sqrt{MS_{error} \times \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

And the  $c$  terms are contrast coefficients. Thus a more general formula is:

$$t_{contrast} = \frac{\text{estimate}}{SE_{est}} = \frac{\psi}{\hat{\sigma}_\psi} = \frac{\sum_{j=1}^k c_j \bar{y}_j}{\sqrt{MS_{error} \times \frac{\sum_{j=1}^k c_j^2}{n_j}}}$$

# This is different from the independent samples $t$ test

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)s_{y_1}^2 + (n_2-1)s_{y_2}^2}{n_1+n_2-2}}}$$

Which is usually just written as:

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{(n_1-1)s_{y_1}^2 + (n_2-1)s_{y_2}^2}{n_1+n_2-2}}}$$

# Defining contrast coefficients

We base the  $c$  terms off of your hypotheses. Let's take a look at the superhero data:

```
# A tibble: 4 x 4
  hero      mean     sd     n
  <fct>    <dbl>   <dbl>   <int>
1 Spiderman 40.7 14.0     6
2 Superman   60.3 17.9     6
3 Hulk       31.5 12.8     6
4 TMNT       27.8  8.73    6
```

# Defining contrast coefficients: contrast 1

Contrast	Spiderman	Superman	Hulk	TNMT
Spiderman vs. Superman	1	-1	0	0
$\bar{y}_j$	40.67	60.33	31.5	27.83

$$\psi = \sum_{j=1}^k c_j \bar{y}_j = (1) \times 40.67 + (-1) \times 60.33 = -19.667$$

$$\hat{\sigma}_\psi = \sqrt{188.85 \times \left(\frac{1}{6} + \frac{1}{6}\right)} = \sqrt{188.85 \times .333} = 7.934$$

$$t_{contrast} = \frac{\psi}{\hat{\sigma}_\psi} = \frac{-19.667}{7.934} = -2.479$$

```
2*(1-pt(2.479, 20)) #unadjusted 2-tailed p-value; note the df!
```

# Contrast degrees of freedom

For the pairwise  $t$  tests ( $t_{contrast}$ ), the degrees of freedom for **each** contrast are the same:  $N-k$  (where  $N$ = total participants and  $k$ =number of groups)

If we were to conduct an independent samples  $t$  test, what would the degrees of freedom be?

- The answer is:  $n_1 + n_2 - 2$
- This is the same as  $N-k$ , where  $k$  must always be 2 for the independent samples  $t$ -test.
- In our example contrast, this is the difference between 20 and 10 degrees of freedom

# Increased degrees of freedom

Degrees of Freedom	$\alpha = .05, 2 - tailed$
10	2.228
11	2.201
12	2.179
13	2.160
14	2.145
15	2.131
16	2.120
17	2.110
18	2.101
19	2.093
20	2.086

# Another post-hoc option

The Tukey Honestly Significant Difference (HSD) is a single-step procedure that analyzes all possible pairwise contrasts between group means.

Instead of a  $t$  statistic ( $t_{contrast}$ ), now we have:

- The HSD value for a family of contrasts
- The Tukey statistic,  $q_{contrast}$ , for each individual contrast within the family
- Both will lead you to the same conclusion

# Tukey Honestly Significant Difference (HSD)

$$HSD = q_{tukey} \sqrt{\frac{MSE}{2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

```
mse <- 188.85
ngroups <- 4 #4 groups, so 4 means
dfmse <- 20 #see ANOVA source table
q_tuk <- qtukey(.95, ngroups, dfmse)
q_tuk
```

```
[1] 3.958293
```

$$HSD = 3.958 \times \sqrt{\frac{188.85}{2} \left( \frac{1}{6} + \frac{1}{6} \right)} = 22.36$$

So, any mean difference larger than 22.36 we will consider significant according to Tukey HSD.

# Which differences are > than our HSD of 22.36?

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = injury ~ hero, data = superhero)

Linear Hypotheses:

		Estimate	Std. Error	t value	Pr(> t )
Superman - Spiderman == 0		19.667	7.934	2.479	0.09417 .
Hulk - Spiderman == 0		-9.167	7.934	-1.155	0.66073
TMNT - Spiderman == 0		-12.833	7.934	-1.617	0.39183
Hulk - Superman == 0		-28.833	7.934	-3.634	0.00817 **
TMNT - Superman == 0		-32.500	7.934	-4.096	0.00290 **
TMNT - Hulk == 0		-3.667	7.934	-0.462	0.96644
---					

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Adjusted p values reported -- single-step method)

# But wait!

Those  $p$ -values are different than before!

Old  $p$ 's...

Superman - Spiderman  
0.0222176879  
Hulk - Superman  
0.0016520896

Hulk - Spiderman  
0.2615640877  
TMNT - Superman  
0.0005617408

TMNT - Spiderman  
0.1214364334  
TMNT - Hulk  
0.6489687464

New  $p$ 's...

Superman - Spiderman  
0.093931100  
Hulk - Superman  
0.008121732

Hulk - Spiderman  
0.660697206  
TMNT - Superman  
0.003014060

TMNT - Spiderman  
0.391843953  
TMNT - Hulk  
0.966428897

# Tukey HSD

That's because the  $p$  value is based on a **new** statistic ( $q_{contrast}$ ), but only the  $p$  values in the table  $glht$  provides changed (which is confusing!). Let's see that new statistic...

For any specific contrast we can calculate:

$$q_{contrast} = \frac{|\psi|}{\sqrt{\frac{MSE}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

```
psi <- abs(-19.667)
se <- sqrt((mse/2)*(1/6+1/6))
qobs <- psi/se
se
```

```
[1] 5.610258
```

```
qobs
```

```
[1] 3.505543
```

# Finding the $p$ -value for Tukey HSD

So, according to Tukey HSD, we have the following:

$$q_{Spiderman \text{ v. } Superman} = 3.506$$

Because our critical  $q_{tukey}$  was 3.958, we know that our  $q_{contrast}$  of 3.506 is not significant. In fact, the  $p$ -value for our  $q_{contrast}$  is...

```
ptukey(qobs,ngroups,dfmse,lower.tail=F)
```

```
[1] 0.09425035
```

- N.B. #1: This is the same  $p$  value in the  $glht$  output.
- N.B. #2: We reach the same conclusion based on comparing our difference in means to the HSD: for Spiderman vs. Superman,  $19.667 < 22.36$ .

# Note about Tukey HSD in *glht*

We have just observed something important in the *glht* output.

- The  $t$  statistics correspond to our pairwise  $t$  test ( $t_{contrast}$ ). This is a standard  $t$  distributed variable.
- The  $p$ -values correspond to the Tukey Studentized Range Distribution ( $q_{contrast}$ ).

So in our example contrast between Spiderman and Superman:

Distribution of Statistic	Statistic	SE	$p$ -value	tails
Student $t$ Distribution	2.479	7.93	.022	2 (can be 1)
Tukey Studentized Range Distribution	3.506	5.61	.094	always 1

# To sum up...

**Those  $p$ -values are different than before! Yes they are.**

Old  $p$ 's based on Student  $t$  distributed statistic ( $t_{contrast}$ )...

Superman - Spiderman 0.0222176879	Hulk - Spiderman 0.2615640877	TMNT - Spiderman 0.1214364334
Hulk - Superman 0.0016520896	TMNT - Superman 0.0005617408	TMNT - Hulk 0.6489687464

New  $p$ 's based on Tukey Studentized Range statistic ( $q_{contrast}$ )...

Superman - Spiderman 0.094396173	Hulk - Spiderman 0.660736939	TMNT - Spiderman 0.391799285
Hulk - Superman 0.008342815	TMNT - Superman 0.002900001	TMNT - Hulk 0.966431538

# Why not just do a bunch of *t*-tests?

- First, note that pairwise contrasts use the degrees of freedom **based on all groups in your sample**, rather than for just the two groups in the contrast. This increases your statistical power!
- Second, pairwise contrasts (whether  $t_{contrast}$  or  $q_{contrast}$ ) base the standard error estimate (i.e., the denominator) on the weighted mean square error ( $MS_{error}$ ) from the overall ANOVA. Again, this tends to increase power!
- But, each individual pairwise contrast is designed to control the probability of false rejection at  $\alpha$ . Unfortunately, if our data analysis involves many hypothesis tests (and thus many contrasts), the probability of at least one Type I error increases rather sharply with the number of contrasts.

# Probability(at least one error)

For example:

- If there are  $m$  tests
- They are independent
- Each is performed with  $\alpha=.05$
- All alternative hypotheses are **true**
- What is the probability of at least one Type I error?

# Probability refresher

If we have 3 levels of a factor, and want to compare all three to each other:

1. Contrast 1 vs. 2 ( $\alpha = .05$ )
2. Contrast 2 vs. 3 ( $\alpha = .05$ )
3. Contrast 1 vs. 3 ( $\alpha = .05$ )

Probabilities:

- What is the probability that I will **incorrectly reject** all three null hypotheses?

$$P(\alpha) \times P(\alpha) \times P(\alpha) = (.05)(.05)(.05) = .000125$$

- Using the same logic, what is the probability that I will **correctly reject** all three null hypotheses?

$$P(1 - \alpha) \times P(1 - \alpha) \times P(1 - \alpha) = (.95)(.95)(.95) = .8574$$

- Hmm...

# Probability(at least one error)

$$P(\text{Type I error}) = \alpha$$

$$P(\text{no Type I error}) = 1 - \alpha$$

$$P(\text{no Type I errors in } m \text{ tests}) = (1 - \alpha)^m$$

$$P(\text{at least one Type I errors in } m \text{ tests}) = 1 - (1 - \alpha)^m$$

# Family-wise error rate (FWER)

The probability that I make **zero** errors:

$$P(1 - \alpha) \times P(1 - \alpha) \times P(1 - \alpha) = (.95)(.95)(.95) = .8574$$

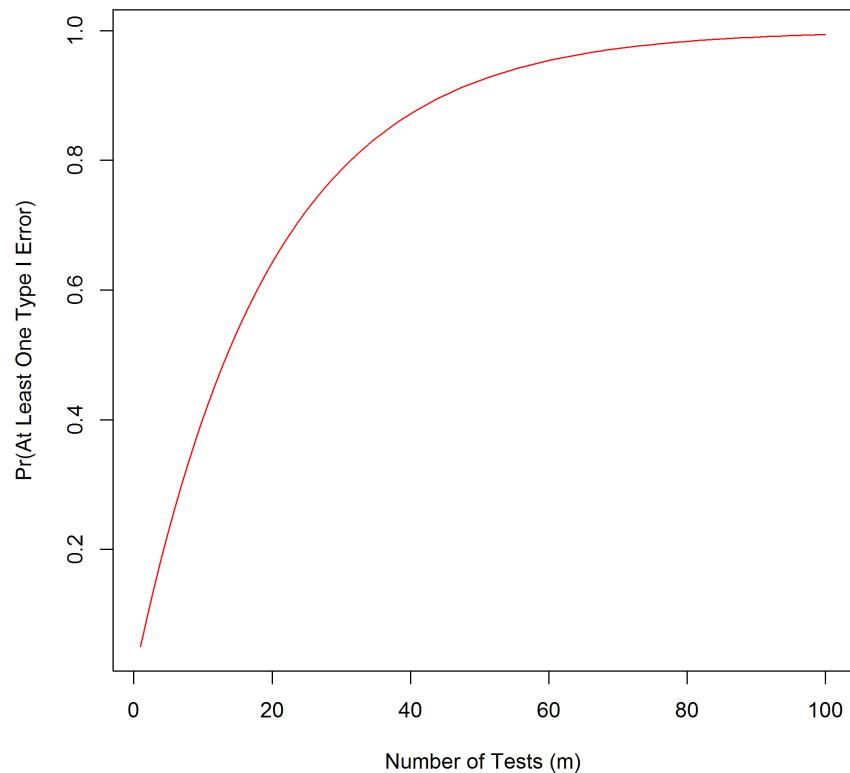
The *family-wise error rate*:

$$P(\text{at least one error}) = 1 - P(\text{no errors}) = 1 - .8574 = .1426 > .05$$

Thus, the probability that I will make at least one Type I error is  $> .05$

$$\alpha_{\text{family}} > \alpha_{\text{per test}}$$

# We can plot this...



**What can we do?**

**Adjust the p-values!**

# Doing all pairwise contrasts in R better\*

```
summary(mcp, test = adjusted("BH")) #Tukey HSD + BH p-adjustment!
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = injury ~ hero, data = superhero)

Linear Hypotheses:

		Estimate	Std. Error	t value	Pr(> t )	
Superman - Spiderman == 0		19.667	7.934	2.479	0.04444	*
Hulk - Spiderman == 0		-9.167	7.934	-1.155	0.31388	
TMNT - Spiderman == 0		-12.833	7.934	-1.617	0.18215	
Hulk - Superman == 0		-28.833	7.934	-3.634	0.00496	**
TMNT - Superman == 0		-32.500	7.934	-4.096	0.00337	**
TMNT - Hulk == 0		-3.667	7.934	-0.462	0.64897	
---						
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						
(Adjusted p values reported -- BH method)						

# But wait!

You changed the  $p$ -values again! Yes I did.

Unadjusted  $p$ 's based on Tukey's HSD...

Superman - Spiderman 0.094396173	Hulk - Spiderman 0.660736939	TMNT - Spiderman 0.391799285
Hulk - Superman 0.008342815	TMNT - Superman 0.002900001	TMNT - Hulk 0.966431538

Tukey HSD  $p$ 's with Benjamini-Hochberg adjustment...

Superman - Spiderman 0.044435376	Hulk - Spiderman 0.313876905	TMNT - Spiderman 0.182154650
Hulk - Superman 0.004956269	TMNT - Superman 0.003370445	TMNT - Hulk 0.648968746

# The Benjamini-Hochberg False Discovery Rate (FDR)

The basic idea of the FDR is to try to achieve the smallest possible fraction of false signals among all those that appear to be true (i.e., significant).

Said another way: we estimate the expected proportion of false rejections among all rejected null hypotheses and attempt to keep it under a threshold level.

Using the FDR method, we are trying to control the number of "false discoveries" rather than the number of "false positives."

What is the difference?

# False Positives & False Discoveries

Let's revisit the good ol' decision table. Let  $m$  be the total number of hypotheses tested.

	Decision: Do not reject $H_0$	Decision: Reject $H_0$	Total
$H_0$ is true	$U$ <i>True Negative</i>	$V$ <i>Type I error / False Positive</i>	$m_0$
$H_0$ is false	$T$ <i>Type II error / False Negative</i>	$S$ <i>True Positive</i>	$m - m_0$
Total	$m - R$	$R$	$m$

The false positive rate is:

$$FPR = \frac{V}{m_0}$$

The false discovery rate is:

$$FDR = \frac{V}{R}$$

# False Positives & False Discoveries

The **false positive rate** is:

$$FPR = \frac{\text{number of falsely rejected } H_0\text{s}}{\text{total number of true } H_0\text{s}}$$

The **false discovery rate** is:

$$FDR = \frac{\text{number of falsely rejected } H_0\text{s}}{\text{total number of rejected } H_0\text{s}}$$

Benjamini-Hochberg proposed to keep the latter less than  $\alpha$  such that the maximum FDR is capped at  $q < \alpha$ .

# The FDR Method

Sort your obtained  $p$ -values from lowest to highest, and add the following information:

- $p$  = sorted unadjusted  $p$ -values (these can be based on any statistic, here Tukey HSD)
- $j$  = variable indexing the order of that contrast in the sort ( $j = 1, \dots, m$ )
- $p_{BH}^*$  = critical values for  $p$ , based on Benjamini-Hochberg per contrast adjustments holding  $q < .05$  (see next slides)

	TMNT - Superman	Hulk - Superman	Superman - Spiderman	TMNT - Spiderman
p	0.003	0.008	0.094	0.392
j	1.000	2.000	3.000	4.000
p*BH	0.008	0.017	0.025	0.033
	Hulk - Spiderman	TMNT - Hulk		
p	0.661	0.966		
j	5.000	6.000		
p*BH	0.042	0.050		

# The Benjamini-Hochberg False Discovery Rate

The previous table can be summarized more generally as:

	smallest	→	→	→	largest
$p - values$	$p_1$	$p_2$	$p_3$	$\dots$	$p_m$
$j$	1	2	3	$\dots$	$m$
$p_{BH}^*$	$\frac{1}{m} \times \alpha$	$\frac{2}{m} \times \alpha$	$\frac{3}{m} \times \alpha$	$\dots$	$\frac{m}{m} \times \alpha = \alpha$

Where the  $p_{BH}^*$  threshold for each individual contrast is:

$$\frac{j}{m} \times \alpha$$

# Decision Time

	TMNT - Superman	Hulk - Superman	Superman - Spiderman
p	0.003	0.008	0.094
j	1	2	3
p*BH	0.008	0.017	0.025
Decision	REJECT H0	REJECT H0	DO NOT REJECT H0
	TMNT - Spiderman	Hulk - Spiderman	TMNT - Hulk
p	0.392	0.661	0.966
j	4	5	6
p*BH	0.033	0.042	0.05
Decision	DO NOT REJECT H0	DO NOT REJECT H0	DO NOT REJECT H0

# Bonferroni Method

A different approach to adjusting  $p$ -values per contrast is to control the **false positive rate** by controlling the *family-wise error rate*=  $\alpha$ .

**The idea:** if you want to control your Type I error rate at  $\alpha = .05$  across all  $m$  contrasts, then simply compare your obtained  $p$ -value with a new  $p_{Bonferroni}^*$ :

$$p_{Bonferroni}^* = \frac{\alpha}{m}$$

Where  $m$  is the maximum total number of contrasts you'll need to perform.

**The downside:** comes at a cost of decreasing statistical power (increasing Type II errors) and thus being overly conservative

In our current example, we conduct  $m=6$  contrasts total, so our nominal per contrast  $p$  is:

$$p_{Bonferroni}^* = \frac{.05}{6} = .008$$

# False Positives

Recall that the **false positive rate** is:

$$FPR = \frac{\text{number of falsely rejected } H_0\text{s}}{\text{total number of true } H_0\text{s}}$$

# Bonferroni vs. Benjamini-Hochberg

Let's compare to the BH FDR:

	smallest			→	largest
$p - values$	$p_1$	$p_2$	$p_3$	...	$p_m$
$j$	1	2	3	...	$m$
$p_{BH}^*$	$1 \times \frac{\alpha}{m}$	$2 \times \frac{\alpha}{m}$	$3 \times \frac{\alpha}{m}$	...	$m \times \frac{\alpha}{m} = \alpha$
$p_{Bonferroni}^*$	$\frac{\alpha}{m}$	$\frac{\alpha}{m}$	$\frac{\alpha}{m}$	...	$\frac{\alpha}{m}$

So for the smallest  $p$ -value,  $p_{BH}^*$  will always equal  $p_{Bonferroni}^*$

# Bonferroni vs. Benjamini-Hochberg

Would our decisions have been any different if, instead of controlling the *false discovery rate* via Benjamini-Hochberg, we controlled the *family-wise error rate* via Bonferroni?

	TMNT - Superman	Hulk - Superman	Superman - Spiderman
p	0.003	0.008	0.094
j	1	2	3
p*BH	0.008	0.017	0.025
Decision: BH	REJECT H0	REJECT H0	DO NOT REJECT H0
p*Bonferroni	0.008	0.008	0.008
Decision: Bonferroni	REJECT H0	DO NOT REJECT H0	DO NOT REJECT H0
	TMNT - Spiderman	Hulk - Spiderman	TMNT - Hulk
p	0.392	0.661	0.966
j	4	5	6
p*BH	0.033	0.042	0.05
Decision: BH	DO NOT REJECT H0	DO NOT REJECT H0	DO NOT REJECT H0
p*Bonferroni	0.008	0.008	0.008
Decision: Bonferroni	DO NOT REJECT H0	DO NOT REJECT H0	DO NOT REJECT H0

# Adjusted p-values options

We just covered two  $p$  value adjustment options:

- Benjamini-Hochberg (in R, "BH" or "fdr")
- Bonferroni
- But there are many many more...

From *multcomp*:

"Shaffer" implements Bonferroni-adjustments taking logical constraints into account Shaffer [1986] and "Westfall" takes both logical constraints and correlations among the z statistics into account Westfall [1997]. In addition, all adjustment methods implemented in `p.adjust` can be specified as well.

From *help(p.adjust)*:

```
c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none")
```

# An example:

- As discussed in Benjamini and Yekutieli (2001), Needleman et al (New England Journal of Medicine, 300, 689–695) studied the neuropsychologic effects of unidentified childhood exposure to lead by comparing various psychological and classroom performances between two groups of children differing in the lead level observed in their shed teeth.
- While there is no doubt that high levels of lead are harmful, Needleman's findings regarding exposure to low lead levels, especially because of their contribution to the Environmental Protection Agency's review of lead exposure standards, are controversial.
- The study was attacked on the grounds of methodological flaws, because they analyzed three independent "outcome" variables (or DVs)
  1. Teacher Behavioral Ratings
  2. WISC Scores
  3. Verbal Processing and Reaction Time Scores

# An example

Inputting the  $p$ -values...

```
Teacher <- sort(c(0.003, 0.05, 0.05, 0.14, 0.08, 0.01, 0.04, 0.01, .050, 0.003, 0.003))
WISC <- sort(c(0.04, 0.05, 0.02, 0.49, 0.08, 0.36, 0.03, 0.38, 0.15, 0.90, 0.37, 0.54))
RT <- sort(c(0.002, 0.03, 0.07, 0.37, 0.90, 0.42, 0.05, 0.04, 0.32, 0.001, 0.001, 0.01))
```

Now what happens if we treat analyses for each of the 3 DVs as one "family", setting the FWER=.05 for each family?

# Bonferroni

```
bonf.teacher <- p.adjust(Teacher, method="bonferroni")
bonf.wisc <- p.adjust(WISC, method="bonferroni")
bonf.rt <- p.adjust(RT, method="bonferroni")
```

# Bonferroni

Here, how many would we reject for each family?

- Teacher? WISC? RT?

```
round(bonf.teacher, 2)
```

```
[1] 0.03 0.03 0.03 0.11 0.11 0.44 0.55 0.55 0.55 0.88 1.00
```

```
round(bonf.wisc, 2)
```

```
[1] 0.24 0.36 0.48 0.60 0.96 1.00 1.00 1.00 1.00 1.00 1.00
```

```
round(bonf.rt, 2)
```

```
[1] 0.01 0.01 0.02 0.12 0.36 0.48 0.60 0.84 1.00 1.00 1.00 1.00
```

# Benjamini Hochberg

```
bh.teacher <- p.adjust(Teacher, method="BH")
bh.wisc <- p.adjust(WISC, method="BH")
bh.rt <- p.adjust(RT, method="BH")
```

# Benjamini Hochberg

Here, how many would we reject for each family?

- Teacher? WISC? RT?

```
round(bh.teacher, 2)
```

```
[1] 0.01 0.01 0.01 0.02 0.02 0.06 0.06 0.06 0.06 0.09 0.14
```

```
round(bh.wisc, 2)
```

```
[1] 0.15 0.15 0.15 0.15 0.19 0.30 0.51 0.51 0.51 0.59 0.59 0.90
```

```
round(bh.rt, 2)
```

```
[1] 0.01 0.01 0.01 0.03 0.07 0.08 0.09 0.11 0.43 0.44 0.46 0.90
```

# (Briefly) Non-parametric omnibus

If we wanted to analyze this data using the extension of the Wilcoxon Mann Whitney Rank Sum Test, we would do a Kruskal-Wallis Rank Sum Test.

```
kruskal.test(injury~hero,data=superhero)
```

Kruskal-Wallis rank sum test

```
data: injury by hero  
Kruskal-Wallis chi-squared = 11.731, df = 3, p-value = 0.008363
```

# (Briefly) Non-parametric contrasts

```
pairwise.wilcox.test(superhero$injury,superhero$hero,p.adj=c('BH'))
```

Pairwise comparisons using Wilcoxon rank sum test

data: superhero\$injury and superhero\$hero

	Spiderman	Superman	Hulk
Spiderman	0.095	-	-
Superman	0.288	0.045	-
TMNT	0.095	0.045	0.514

P value adjustment method: BH

# (Briefly) Non-parametric omnibus + contrasts (way better)

```
library(agricolae)
k1 <- kruskal(superhero$injury,superhero$hero,alpha=.05,group=F,p.adj=c('BH'))
```

# (Briefly) Non-parametric omnibus + contrasts (way better)

k1

```
$statistics
  Chisq      p.chisq
 11.73121 0.008363013

$parameters
    test p.adjusted      name.t ntr alpha
Kruskal-Wallis      BH superhero$hero   4  0.05

$means
superhero.injury      rank      std  r Min Max  Q25  Q50  Q75
Hulk      31.50000 9.416667 12.817956 6 10  45 27.0 32.5 41.00
Spiderman 40.66667 13.750000 14.009521 6 20  58 32.5 42.0 50.00
Superman   60.33333 19.916667 17.851237 6 32  85 53.5 62.0 68.25
TMNT      27.83333 6.916667  8.727352 6 18  41 21.0 30.0 30.00

$comparison
                  Difference pvalue Signif.
Hulk - Spiderman      -4.333333 0.2060
Hulk - Superman        -10.500000 0.0079      **
Hulk - TMNT            2.500000 0.4230
Spiderman - Superman   -6.166667 0.0859      .
Spiderman - TMNT       6.833333 0.0739      .
Superman - TMNT        13.000000 0.0023      **
```

# (Briefly) Effect size

```
library(orddom)
hulk <- subset(superhero,select="injury"),hero=="Hulk")
tnmt <- subset(superhero,select="injury"),hero=="TMNT")
orrdom(hulk, tnmt,alpha=.05,paired=FALSE)
```

	ordinal	metric
var1_X	"group 1 (x)"	"group 1 (x)"
var2_Y	"group 2 (y)"	"group 2 (y)"
type_title	"indep"	"indep"
n in X	"6"	"6"
n in Y	"6"	"6"
N #Y>X	"12"	"12"
N #Y=X	"3"	"3"
N #Y<X	"21"	"21"
PS X>Y	"0.5833333333333333"	"0.593459235722213"
PS Y>X	"0.3333333333333333"	"0.406540764277787"
A X>Y	"0.625"	"0.625"
A Y>X	"0.375"	"0.375"
delta	"-0.25"	"-3.66666666666667"
1-alpha	"95"	"95"
CI low	"-0.76689711774936"	"-18.0334552608998"
CI high	"0.463629363759192"	"10.7001219275665"
s delta	"0.35"	"10.9650961388094"
var delta	"0.1225"	"120.233333333333"
se delta	NA	"6.33070120743175"
z/t score	"-0.714285714285714"	"-0.579188078306758"
H1 tails p/CI	"2"	"2"
p	"0.49138699486592"	"0.576959741821073"

