Cognitive development study

- Cognitive test administered to grades 1-5
  - 15 children per grade
- Average scores increase approx linearly
  - Correlation between grade & \( \bar{Y} \) = 0.97
- Use ANOVA to evaluate group differences

Cognitive development study

Check constant variance assumption

\[
> \text{bartlett.test(score~grade,dfl3)} \\
\text{Bartlett test of homogeneity of variances} \\
data: score by grade \\
\text{Bartlett's K-squared} = 6.6227, \text{df} = 4, \text{p-value} = 0.1572
\]

Do not reject null hypothesis that variances are equal
Cognitive development study
check normality assumption

> shapiro.test(residuals(aov.01))
Shapiro-Wilk normality test
data:  residuals(aov.01)
W = 0.9871, p-value = 0.6482

> qqnorm(residuals(aov.01))
> qqline(residuals(aov.01))

Do not reject null hypothesis that residuals are Normal

Cognitive development study

> load(file=url('http://pnb.mcmaster.ca/bennett/psy710/datasets/contrasts.rda'))
> df3$grade <- factor(df3$grade,ordered=FALSE)
> options(contrasts=c("contr.sum","contr.poly")) # IMPORTANT!!
> aov.01 <- aov(score~grade,df3)
> anova(aov.01)

Analysis of Variance Table
Response: score
Df  Sum Sq Mean Sq F value  Pr(>F)
grade      4  1361.2  340.31  2.2578 0.07152 .
Residuals 70 10550.9  150.73

- the effect of grade was not significant
- do not reject the null hypothesis of no difference among group means

Cognitive development study

> library(effectsize)
> cohens_f(aov.01)

# Effect Size for ANOVA
Parameter | Cohen's f | 95% CI
--- | --- | ---
grade | 0.1 | [0.00, Inf]

> omega_squared(aov.01)

# Effect Size for ANOVA
Parameter | Omega2 | 95% CI
--- | --- | ---
grade | 0.06 | [0.00, 1.00]

Note that 1-sided 95% CIs for effect size & association strength include zero

Cognitive development study
estimate power assuming medium effect size (f = 0.25)

> library(pwr)
> pwr.anova.test(k=5,n=15,
+ f=0.25,sig.level=.05,
+ power=NULL)

Balanced 1-way anova power calculation
k = 5
n = 15
f = 0.36
sig.level = 0.05
power = 0.35

NOTE: n is number in each group
Cognitive development study

estimate power assuming $f = 0.36$

```r
> library(pwr)
> pwr.anova.test(k=5,n=15,
+     f=0.36,sig.level=.05,
+     power=NULL)
Balanced 1-way anova power calculation
k = 5
n = 15
f = 0.36
sig.level = 0.05
power = 0.67
NOTE: n is number in each group
```

alternatives to ANOVA

```r
> oneway.test(score~grade,data=df3)
1-way analysis of means (not assuming equal variances)
data:  score and grade
F = 3, num df = 4, denom df = 35, p-value = 0.04

> kruskal.test(score~grade,data=df3)
Kruskal-Wallis rank sum test
data:  score by grade
Kruskal-Wallis chi-squared = 9, df = 4, p-value = 0.07
```

K-W Null Hypothesis:
- groups were sampled from the same distribution
- if we assume distributions have same shape & scale:
  - then $H_0$ is that group medians are equal

Omnibus vs. Focussed F tests

$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$

$H_1 : \alpha_j \neq 0$

- A significant omnibus F test tests a very general hypothesis
  - $H_0$: all group means are equal; $H_1$: not all means are equal
  - $H_0$: all group effects are zero; $H_1$: not all group effects are zero
- Significant F doesn’t tell us how group means differ
- Generality of omnibus F often comes at cost of reduced power

linear contrasts/comparisons
Omnibus vs. Focussed F tests

Omnibus F test is not significant:

```r
> lm.01 <- lm(score~grade,data=df3)
> anova(lm.01)
```

Analysis of Variance Table
---
Response: score

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>grade</td>
<td>4</td>
<td>1361.2</td>
<td>340.31</td>
<td>2.2578</td>
</tr>
<tr>
<td>Residuals</td>
<td>70</td>
<td>10550.9</td>
<td>150.73</td>
<td></td>
</tr>
</tbody>
</table>

Focussed tests provide more power

Linear contrasts often more appropriate & more powerful

```r
> c1 <- c(-1/2,-1/2,1/3,1/3,1/3)  
# (g1 & g2) vs (g3 & g4 & g5)
> c2 <- c(-1,1,0,0,0)  
# g1 vs g2
> c3 <- c(0,0,-1/2,1/2)  
# g3 vs (g4 & g5)
> c4 <- c(0,0,0,-1,1)  
# (g4 vs g5)
> contrasts(df3$grade) <- cbind(c1,c2,c3,c4)
> fractions( contrasts(df3$grade) )
```

```
c1  c2  c3  c4
-1/2 -1  0  0
1/2  1  0  0
1/3  0 -1  0
1/3  0  1/2 -1
1/3  0  1/2  1
```

Linear Contrast Example

H0: (means of grades 3,4,5) = (means of grades 1,2)
H1: (means of grades 3,4,5) ≠ (means of grades 1,2)

```r
> aov.02 <- aov(score~grade,data=df3)
> summary(aov.02, 
+     split=list(grade=list(c1=1,c2=2,c3=3,c4=4)))
```

```
Df SS         MS      F        Pr(>F)
grade     4 1361 340.31   2.2578 0.0715 .
grade: c1  1  753 752.91   4.9950 0.029  *
grade: c2-c4  3  608 202.73  1.3452 0.269
Residuals 70 10551 150.73
```

Linear Contrast Example

H0: (means of grades 3,4,5) = (means of grades 1,2)
H1: (means of grades 3,4,5) ≠ (means of grades 1,2)

```r
> aov.02 <- aov(score~grade,data=df3)
> summary(aov.02, 
+     split=list(grade=list(c1=1,c2=2,c3=3,c4=4)))
```

```
Df SS         MS      F        Pr(>F)
grade     4 1361 340.31   2.2578 0.0715 .
grade: c1  1  753 752.91   4.9950 0.029  *
grade: c2-c4  3  608 202.73  1.3452 0.269
Residuals 70 10551 150.73
```
Trend Analysis Example

linear contrasts can be more powerful & more appropriate tests of null hypothesis

H0: linear trend of score across grade = 0
H1: linear trend of score across grade ≠ 0

<table>
<thead>
<tr>
<th></th>
<th>Lin</th>
<th>Quad</th>
<th>Cubic</th>
<th>Quartic</th>
</tr>
</thead>
<tbody>
<tr>
<td>[g1]</td>
<td>-0.632</td>
<td>0.535</td>
<td>-3.16e-01</td>
<td>0.120</td>
</tr>
<tr>
<td>[g2]</td>
<td>-0.316</td>
<td>-0.267</td>
<td>6.32e-01</td>
<td>-0.478</td>
</tr>
<tr>
<td>[g3]</td>
<td>0.000</td>
<td>-0.535</td>
<td>-4.10e-16</td>
<td>0.717</td>
</tr>
<tr>
<td>[g4]</td>
<td>0.316</td>
<td>-0.267</td>
<td>-6.32e-01</td>
<td>-0.478</td>
</tr>
<tr>
<td>[g3]</td>
<td>0.632</td>
<td>0.535</td>
<td>3.16e-01</td>
<td>0.120</td>
</tr>
</tbody>
</table>

Bennett, PJ PSYCH 710

Chapter 6

Linear Contrasts (Comparisons)

- Contrasts allow us to evaluate focussed hypotheses
  - evaluate specific pattern of differences among group means
- Each contrast is defined by a set of contrast weights
  - weights (c₁, c₂, ... cₐ) specify a pattern of group means
  - value of contrast, ψ, is a weighted combination of group means
    - ψ = c₁ Y₁ + c₂ Y₂ + c₃ Y₃ + c₄ Y₄ + ... cₐ Yₐ

Hypotheses tested with Linear Contrasts

- Linear contrasts are defined by weights
  - must sum to zero
  - sum(1/2, 1/2,-1/3, -1/3,-1/3) = 0
- Multiplying weights by constant produces an equivalent linear contrast
  - w₁ = (1/2, 1/2,-1/3, -1/3,-1/3)
  - w₂ = 6 x w₁ = (3,3,-2,-2,-2)
  - w₁ is equivalent to w₂
Contrasts are defined by weights

Each contrast sums to zero!

\[ g_1 = -1, g_2 = -1, g_3 = -1, g_4 = -1, g_5 = -1, g_6 = 6 \]

Hypotheses Evaluated by a Contrast

\[ H_0: \quad -1(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6) + 6\mu_7 = 0 \]

\[ \mu_7 = \frac{1}{6}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6) \]

H1:

\[ \mu_7 \neq \frac{1}{6}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6) \]

Hypotheses Evaluated by a Contrast

\[ > w_1 \leftarrow c(-2,-2,0,1,1,1) \]
\[ > (w_2 \leftarrow w_1/4) \]
\[ [1] \quad -1/2 -1/2 0 1/4 1/4 1/4 \]
\[ > \# sum(w1) = sum(w2) = 0 \]

H0:

\[ -\frac{1}{2}(\mu_1 + \mu_2 + 0 \times \mu_3 + \frac{1}{4}(\mu_4 + \mu_5 + \mu_6)) = 0 \]

\[ \frac{(\mu_3 + \mu_5 + \mu_6) + \mu_7}{4} = \frac{(\mu_1 + \mu_2)}{2} \]

H1:

\[ \frac{(\mu_3 + \mu_5 + \mu_6) + \mu_7}{4} \neq \frac{(\mu_1 + \mu_2)}{2} \]

\[ \frac{(\mu_2 + \mu_4 + \mu_7) + \mu_6}{4} \neq \frac{(\mu_1 + \mu_2)}{2} \]
Hypotheses Evaluated by a Contrast

\[ \text{H}_0 : \mu_1 - 0 \mu_2 - 1 \mu_3 - 1 \mu_4 = 0 \]
\[ 3 \mu_2 - 1 (\mu_1 + \mu_4 + \mu_5) = 0 \]
\[ 3 \mu_1 - 1 (\mu_1 + \mu_4 + \mu_5) = 0 \]
\[ \mu_1 = 1 (\mu_1 + \mu_4 + \mu_5) \]
\[ \mu_1 = \frac{1}{3} (\mu_1 + \mu_4 + \mu_5) \]

\[ \text{H}_1 : \mu_1 \neq \frac{1}{3} (\mu_1 + \mu_4 + \mu_5) \]

\[ \text{With equal } n \text{ per group:} \]
\[ F = \frac{(\Psi^2) / \sum_{j=1}^a (c_j^2/n_j)}{MS_W} \]
\[ F = \frac{\Psi^2 / \sum_{j=1}^a (c_j^2/n_j)}{MS_W} \]

General Form of Linear Contrast

\[ H_0 : c_1 \mu_1 + c_2 \mu_2 + \cdots + c_a \mu_a = 0 \]
\[ \text{weighted sum of population means equals zero} \]
\[ \sum_{j=1}^a c_j = 0 \]
\[ \text{sum of weights must equal zero} \]
\[ \Psi = \sum_{j=1}^a (c_j \bar{Y}_j) \]
\[ \text{value of contrast equals weighted sum of group means} \]
\[ \text{contrast df = 1} \]

Hypotheses tested with Linear Contrasts

2-tailed tests

\[ H_0 : \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 = \mu_3 + \mu_4 + \mu_5 \]
\[ H_1 : \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 \neq \mu_3 + \mu_4 + \mu_5 \]

1-tailed tests

\[ H_0 : \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 \geq \mu_3 + \mu_4 + \mu_5 \]
\[ H_1 : \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 < \mu_3 + \mu_4 + \mu_5 \]

General Form of Linear Contrast (directional tests evaluated with t statistic)

\[ t = \frac{\Psi}{\sqrt{\sum_{j=1}^a (c_j^2/n_j)}} \]
\[ \text{df = N-a} \]

\[ t^2 = F \]

\[ \text{t statistic more useful for 1-tailed tests} \]

\[ H_0 : \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 = \mu_3 + \mu_4 + \mu_5 \geq 0 \]
\[ H_1 : \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 = \mu_3 + \mu_4 + \mu_5 < 0 \]
\[ H_0 : \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 \geq \mu_3 + \mu_4 + \mu_5 \]
\[ H_1 : \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 < \mu_3 + \mu_4 + \mu_5 \]
General Form of Linear Contrast
(the sign of the weights determines the direction of the test)

\[
t_{\text{statistic}}: \text{the sign of weights matters!}
\]

\[
H_0: \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 - \frac{1}{3} \mu_3 - \frac{1}{3} \mu_4 - \frac{1}{3} \mu_5 \geq 0
\]

\[
H_1: \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 - \frac{1}{3} \mu_3 - \frac{1}{3} \mu_4 - \frac{1}{3} \mu_5 < 0
\]

\[
H_0: \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 - \frac{1}{3} \mu_3 - \frac{1}{3} \mu_4 - \frac{1}{3} \mu_5 \leq 0
\]

\[
H_1: \frac{1}{2} \mu_1 + \frac{1}{2} \mu_2 - \frac{1}{3} \mu_3 - \frac{1}{3} \mu_4 - \frac{1}{3} \mu_5 > 0
\]

equivalent!

calculating contrasts with aov & emmeans

Conducting Contrasts with R aov()

\[
> \text{contrasts(df3$grade) <- cMat}
\]

\[
> \text{fractions(contrasts(df3$grade))}
\]

\[
\text{myC1 myC2 myC3 myC4}
\]

\[
g1 -1/2 -1 0 0
\]

\[
g2 -1/2 1 0 0
\]

\[
g3 1/3 0 -1 0
\]

\[
g4 1/3 0 1/2 -1
\]

\[
g5 1/3 0 1/2 1
\]

\[
> \text{aov.02 <- aov(score~grade,data=df3)}
\]

\[
> \text{summary(aov.02,}
\]

\[
+ \text{split=list(grade=list(c1=1,c2=2,c3=3,c4=4)))}
\]

\[
\begin{array}{lrrrr}
\text{Df} & \text{Sum Sq} & \text{Mean Sq} & \text{F value} & \text{Pr(>F)} \\
\text{grade} & 4 & 1361 & 340.3 & 2.258 & 0.0715 \\
\text{grade: c1} & 1 & 753 & 752.9 & 4.995 & 0.0286 * \\
\text{grade: c2} & 1 & 251 & 250.6 & 1.662 & 0.2015 \\
\text{grade: c3} & 1 & 210 & 210.5 & 1.396 & 0.2413 \\
\text{grade: c4} & 1 & 147 & 147.3 & 0.977 & 0.3262 \\
\text{Residuals} & 70 & 10551 & 150.7 & \\
\end{array}
\]

Conducting Contrasts with R aov()

\[
> \text{aov.02 <- aov(score~grade,data=df3)}
\]

\[
> \text{summary(aov.02,}
\]

\[
+ \text{split=list(grade=list(c1=1,c2=2,c3=3,c4=4)))}
\]

\[
\text{Linear contrasts assessed with F tests}
\]

Perform ANOVA with aov

Write ANOVA table with summary(), but split results for grouping variable into separate lines for different contrasts

split = list(factor.name=list(contrast.name.1=1, contrast.name.2=2,...))
Conducting Contrasts with emmeans

emmeans = estimated marginal means

> # create emmeans object
> library(emmeans)
> aov.01 <- aov(score~grade, data=df3)
> aov.em <- emmeans(aov.01, specs="grade")
>
grade | emmean   | SE  | df | lower.CL | upper.CL
------ | -------- | ---- |---- |---------- |----------
g1    | 91.3     | 3.17| 70  | 85.0      | 97.6      
g2    | 97.1     | 3.17| 70  | 90.8      | 103.4     
g3    | 97.6     | 3.17| 70  | 91.3      | 103.9     
g4    | 100.0    | 3.17| 70  | 93.7      | 106.3     
g5    | 104.3    | 3.17| 70  | 98.1      | 110.7     

Confidence level used: 0.95

Conducting Contrasts with linear.comparison

linear.comparison() & emmeans() yield same results

> source(url("http://pnb.mcmaster.ca/bennett/psy710/Rscripts/linear_contrast_v2.R"))
[1] "loading function linear.comparison"
> y <- df3$score
> g <- df3$grade
> myContrast1 <- linear.comparison(y, g, c.weights = myContrasts, var.equal=T)

[1] "computing linear comparisons assuming equal variances among groups"
[1] "C 1: F=6.495, t=2.235, p=0.029, psi=6.467, CI=(0.367,12.568), adj.CI=(-0.952,13.887)"
[1] "C 2: F=1.462, t=1.289, p=0.202, psi=5.780, CI=(-5.714,17.274)"
[1] "C 3: F=1.396, t=1.182, p=0.241, psi=4.588, CI=(-2.544,11.719), adj.CI=(-5.366,14.542)"
[1] "C 4: F=1.077, t=0.989, p=0.326, psi=4.432, CI=(-2.955,11.819), adj.CI=(-7.662,15.926)"

trend analysis
Trend Analysis

- the analysis of trends uses the same methods as linear contrasts
- weights are designed to evaluate specific differences across groups:
  - linear, quadratic, cubic, etc.
- weights must sum to zero
- weights can be calculated using R’s contr.poly function
  - useful when differences between levels on group variable are not constant

```r
> contrasts(df3$grade) <- contr.poly(n=5,scores=1:5)
> round(cor(polyWeights),digits=2)
```

Trend Analysis Example

- trends are linear contrasts
- weights are designed to evaluate specific differences across groups:
  - linear, quadratic, cubic, etc.
- weights must sum to zero
- weights can be calculated using R’s contr.poly function
  - useful when differences between levels on group variable are not constant

```r
> summary(aov.trends)
> aov.trends <- aov(score~grade,data=df3)
```
Trend Analysis Example

trends are linear contrasts

> # emmeans poly method uses polynomial contrasts
> # and assumes equally-spaced levels on grouping factor
> # ?poly.emmc for details

> aov.01.em <- emmeans(aov.01.specs="grade")
> contrast(aov.01.em, method="poly")

<table>
<thead>
<tr>
<th>contrast estimate</th>
<th>SE</th>
<th>df</th>
<th>t.ratio</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td>29.096</td>
<td>10.0</td>
<td>2.903</td>
<td>0.0049</td>
</tr>
<tr>
<td>quadratic</td>
<td>-0.843</td>
<td>11.9</td>
<td>-0.071</td>
<td>0.9435</td>
</tr>
<tr>
<td>cubic</td>
<td>7.321</td>
<td>10.0</td>
<td>0.730</td>
<td>0.4676</td>
</tr>
<tr>
<td>quartic</td>
<td>-6.907</td>
<td>26.5</td>
<td>-0.260</td>
<td>0.7953</td>
</tr>
</tbody>
</table>

Trend Analysis Example

trends are linear contrasts

Can evaluate all higher-order, nonlinear trends with a single F test

> summary(aov.trends,
+     split=list(grade=list(Lin=1,NonLin=2:4)))

<table>
<thead>
<tr>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>grade</td>
<td>Lin</td>
<td>3</td>
<td>1270</td>
<td>1269.9</td>
</tr>
<tr>
<td>grade</td>
<td>NonLin</td>
<td>3</td>
<td>91</td>
<td>30.5</td>
</tr>
<tr>
<td>Residuals</td>
<td>70</td>
<td>10551</td>
<td>150.7</td>
<td></td>
</tr>
</tbody>
</table>

Effect Size for a Linear Comparison

linear contrasts are used to compare two weighted means, so Cohen’s d is appropriate

Cohen’s d (for a contrast)

\[
d = 2\Psi / \left( \sigma \left[ \sum_{j=1}^{a} |c_j| \right] \right)
\]

\[
d = 2\Psi / \left( \sqrt{MSW} \left[ \sum_{j=1}^{a} |c_j| \right] \right)
\]

Expresses \( \Psi \) in terms of the number of standard deviations of population error distribution
In this section, we introduce the notion of orthogonal contrasts. Suppose you are conducting a linear contrast that compares groups and ignores group strength. Your textbook describes one advantage that can vary between 0 and 1. As your book notes, another interpretation of association strength is that it is the correlation between the comparison coefficients and the group means when the group is ignored in our contrast. The value of R^2 has over the other two measures of association.

### Effect Size

Cohen’s d calculation with emmeans & linear.comparison

```r
> library(emmeans)
> aov.01 <- aov(score~grade,data=df3)
> sigma <- sigma(aov.01) # sqrt(MS.resid)
> edf <- df.residual(aov.01) # residual df
> aov.em <- emmeans(aov.01,specs="grade")
> myContrasts <- list(c1=myC1,c2=myC2,c3=myC3,c4=myC4)
> # calculate Cohen's d for each contrast:
> eff.size(aov.em,sigma,edf,method=myContrasts)
```

<table>
<thead>
<tr>
<th>contrast</th>
<th>effect size</th>
<th>SE</th>
<th>df</th>
<th>lower.CL</th>
<th>upper.CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>0.53</td>
<td>0.24</td>
<td>70</td>
<td>0.05</td>
<td>1.01</td>
</tr>
<tr>
<td>c2</td>
<td>0.47</td>
<td>0.37</td>
<td>70</td>
<td>-0.26</td>
<td>1.02</td>
</tr>
<tr>
<td>c3</td>
<td>0.37</td>
<td>0.32</td>
<td>70</td>
<td>-0.26</td>
<td>1.01</td>
</tr>
<tr>
<td>c4</td>
<td>0.36</td>
<td>0.37</td>
<td>70</td>
<td>-0.37</td>
<td>1.09</td>
</tr>
</tbody>
</table>

### Association Strength for a Linear Comparison

- \( R^2_{alerting} = \frac{SS_{contrast}}{SS_B} \) - Proportion of Between-Groups variation accounted for by contrast. With equal n, equals squared correlation between contrast weights & group means.
- \( R^2_{effectsize} = \frac{SS_{contrast}}{SS_{Total}} \) - Proportion of total variation accounted for by contrast.
- \( R^2_{contrast} = \frac{SS_{contrast}}{(SS_{contrast} + SS_{W})} \) - Variation accounted for by contrast relative to the sum of contrast-variation and within-group (error) variation. Not affected by groups that are weighted zero. More resistant to changes in experimental design (e.g., adding or removing groups).

### Association Strength

- \( R^2_{alerting} = \frac{SS_{contrast}}{SS_B} \)
- \( R^2_{effectsize} = \frac{SS_{contrast}}{SS_{Total}} \)
- \( R^2_{contrast} = \frac{SS_{contrast}}{(SS_{contrast} + SS_{W})} \)

Note double brackets \([x]\)!!
Unequal Group Variances

- So far our tests assume equal variance in different groups
- F/t tests for contrasts are not robust to violation of equal variance assumption
- When groups have unequal variances, use a different method to calculate F/t denominator, which is an estimate of population error variance
- Correcting for unequal var reduces denominator df (and, hence, power)

\[
F = \frac{\sum_{j=1}^{a} (c_j / n_j)}{\sum_{j=1}^{a} \left(\frac{c_j^2 s_j^2}{n_j}\right) / \sum_{j=1}^{a} (c_j^2 / n_j)}
\]

\[
df = \frac{\left[\sum_{j=1}^{a} \left(\frac{c_j^2}{s_j^2 n_j}\right)^2 / (n_j - 1)\right]^2}{\left[\sum_{j=1}^{a} \left(\frac{c_j^2}{n_j}\right)^2 / \left(n_j - 1\right)\right]}
\]

Contrasts with unequal variances

The numerator degrees of freedom remains 1, but the denominator degrees of freedom no longer is.

Using this new estimate of violations of the assumption have the e

groups, not just the ones being compared. When the variance is homogeneous across groups, MS = 0 was true. Suppose, instead, we had assumed that

Note that the degrees of freedom (di

erent method for estimating

orthogonal contrasts

\[c o n t r a s t . u n e q u a l . v a r< - l i n e a r . c o m p a r i s o n ( y , g , c . w e i g h t s = l i s t ( c (- 1 , - 1 , - 1 , - 1 , - 1 , - 1 , 6 ) ) , v a r . e q u a l = F A L S E )
\]
In this section we introduce the notion of $\pi$ on the value of the ignored group and ignores group strength. Suppose you are conducting a linear contrast that compares groups.

Comparing Equations 19 and 20 shows that they have the same numerators but different denominators. So, Equation 19 expresses $SS_{\pi}$ as a proportion of the "total variation" among groups.

Another measure of association strength is $R^2$. Recall that $SS_{\pi}$ describes the between-group variability associated with the contrast relative to the total variability among groups.

Your textbook describes several measures of association strength that are based on $SS$. These ideas about orthogonality, and the decomposition of $SS$, are important when we discuss multiple comparisons.

Multiple contrasts that are orthogonal, on the other hand, provide overlapping, correlated, and uninformative pieces of information about group means.

### Orthogonal Contrasts

**Equal $n$:**

\[
\sum_{j=1}^{a} (c_{1j}c_{2j}) = 0
\]

**Unequal $n$:**

\[
\sum_{j=1}^{a} (c_{1j}c_{2j}/n_j) = 0
\]

A set of contrasts is mutually orthogonal if all pairs of contrasts are orthogonal.

Orthogonal contrasts evaluate independent questions about group means.

### Complete Set of Mutually Orthogonal Contrasts

If there are $a$ groups, then the largest set of mutually orthogonal contrasts will have $(a-1)$ contrasts, and:

\[
\sum_{j=1}^{a-1} SS_{contrast,j} = SS_{\pi}
\]

- A complete set of orthogonal contrasts divides $SS_{\pi}$ into independent pieces of variation,
- the sum of the $(a-1)$ $SS_{contrast,j}$ will equal $SS_{\pi}$,
- and the average of the contrast $F$ values will equal the omnibus $F$.

---

**Complete set of orthogonal contrasts**

breaks $SS_{\pi}$ into separate pieces

```r
> cMat <- contrasts(df3$grade)
> # these contrasts/columns are mutually orthogonal:
> summary(aov.10)
> split=list(grade=list(myC1=1,myC2=2,myC3=3,myC4=4))
> summary(aov(score=grade,df3))
```

<table>
<thead>
<tr>
<th>Element</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>myC1</td>
<td>0.83</td>
</tr>
<tr>
<td>myC2</td>
<td>0.00</td>
</tr>
<tr>
<td>myC3</td>
<td>0.00</td>
</tr>
<tr>
<td>myC4</td>
<td>0.00</td>
</tr>
</tbody>
</table>

N.B. Each element in this matrix is the sum of cross-products.
**Controlling False Discovery Rate**

- Instead of controlling $\alpha_{FW}$, control **False Discovery Rate** (FDR):
  - $Q = (# \text{ of false } H0 \text{ rejections}) / (\text{total } H0 \text{ rejections})$
  - FDR = Expected Value[$Q$]
- When all $H0$ are true, controlling $\alpha_{FW}$ and FDR are equivalent
- When some $H0$ are false, FDR-based methods are more powerful

**Multiple Comparisons of Group Means**

- Multiple comparisons inflate Type I error rate
- Generally want to control family-wise Type I error rate by adjusting the per-comparison Type I error rate
- for $C = 100$ comparisons
  - if $\alpha_{PC} = .00051$, then $\alpha_{FW} \leq .05$
- there are several methods for adjusting $\alpha_{PC}$

**Corrections for Multiple Comparisons**

- Controlling $\alpha_{FW}$ by adjusting $\alpha_{PC}$:
  - Bonferroni Adjustment (aka Dunn’s Procedure)
  - Holm’s Sequential Bonferroni Test
- Controlling False Discovery Rate (FDR):
  - Benjamini & Hochberg’s (1995) Linear Step-Up Procedure (FDR)
- Relative Power: FDR > Holm’s > Bonferroni
**Multiple Comparisons in R**

Adjust p values with `p.adjust()`

```r
> my.p.values <- c(.127,.08,.03,.032,.02,.001,.01,.005,.025)
> sort(my.p.values)
[1] 0.001 0.005 0.010 0.020 0.025 0.030 0.032 0.080 0.127
> p.adjust(sort(my.p.values),method='bonferroni')
[1] 0.009 0.045 0.090 0.180 0.225 0.270 0.288 0.720 1.000
> p.adjust(sort(my.p.values),method='holm')
[1] 0.009 0.040 0.070 0.120 0.125 0.125 0.125 0.160 0.160
> p.adjust(sort(my.p.values),method='fdr')
[1] 0.009 0.0225 0.0300 0.04114 0.04114 0.04114 0.04114 0.090 0.127
```

Significant tests (alpha/FDR = .05) are highlighted in orange font.

N.B. Sorting p-values is not required.

**Controlling Type I error rate**

```r
> summary(aov.vp,split=list(complexity=list(L=1,Q=2,C=3,q4=4)))

Df SS MS   F    Pr(>F)
complexity          4  1.2709 0.3177 6.214 0.000691 ***
complexity: L    1  0.7441 0.7441 14.552 0.000532 ***
complexity: Q   1  0.4357 0.4357  8.521 0.006100 **
complexity: C   1  0.0477 0.0477  0.933 0.340714
complexity: q4  1  0.0434 0.0434  0.848 0.363286
Residuals         35  1.7897 0.0511
```

**Setting family-wise alpha and FDR**

- Generally, \( \alpha_{FW} \) and FDR are set to 0.01 or 0.05
- larger \( \alpha_{FW} \) may be justified for small number of orthogonal comparisons
  - Bonferroni & Holm tests may reduce power too much
  - perhaps set \( \alpha_{PC} \) to 0.05 or 0.01
    - family-wise Type I error will increase but Type II error will decrease
    - Note: we do this with factorial ANOVA already
All pairwise tests (Tukey HSD)

• Tukey HSD evaluates all pairwise differences between groups
• Is more powerful than Bonferroni method (for between-subj designs)
• Tukey HSD:
  - NOT necessary to evaluate omnibus F prior to Tukey test
  - assumes equal n per group & equal variances
  - Tukey-Kramer is valid with sample sizes are unequal
  - Dunnett’s T3 test is better with unequal n & unequal variances
    [see Kirk (1995, pp. 146-50) for more details]

Tukey HSD (all pairwise differences)

optimal method for evaluating all pairwise differences

assumes equal variances
> TukeyHSD(aov.vp,which=“complexity"")
Tukey multiple comparisons of means
95% family-wise confidence level
Fits: aov(formula = visPref ~ complexity, data = df4)

<table>
<thead>
<tr>
<th>complexity</th>
<th>diff</th>
<th>lwr</th>
<th>upr</th>
<th>p adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>p2-p1</td>
<td>0.1663</td>
<td>-0.159</td>
<td>0.49</td>
<td>0.59</td>
</tr>
<tr>
<td>p3-p1</td>
<td>0.4620</td>
<td>0.137</td>
<td>0.79</td>
<td>0.00</td>
</tr>
<tr>
<td>p4-p1</td>
<td>0.4569</td>
<td>0.132</td>
<td>0.78</td>
<td>0.00</td>
</tr>
<tr>
<td>p5-p1</td>
<td>0.3369</td>
<td>0.012</td>
<td>0.66</td>
<td>0.04</td>
</tr>
<tr>
<td>p3-p2</td>
<td>0.2957</td>
<td>-0.029</td>
<td>0.62</td>
<td>0.09</td>
</tr>
<tr>
<td>p4-p2</td>
<td>0.2906</td>
<td>-0.035</td>
<td>0.62</td>
<td>0.10</td>
</tr>
<tr>
<td>p5-p2</td>
<td>0.1706</td>
<td>-0.154</td>
<td>0.50</td>
<td>0.56</td>
</tr>
<tr>
<td>p4-p3</td>
<td>-0.0051</td>
<td>-0.330</td>
<td>0.32</td>
<td>1.00</td>
</tr>
<tr>
<td>p5-p3</td>
<td>-0.1250</td>
<td>-0.450</td>
<td>0.20</td>
<td>0.80</td>
</tr>
<tr>
<td>p5-p4</td>
<td>-0.1199</td>
<td>-0.445</td>
<td>0.21</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Tukey HSD (all pairwise differences)

emmeans (assumes equal variances)

> vp.em <- emmeans(aov.vp,specs="complexity")
> contrast(vp.em,method="pairwise",adjust="tukey")

contrast estimate SE df t.ratio p.value
p1 - p2 -0.17 0.113 35 -1.500 0.5900
p1 - p3 -0.46 0.113 35 -4.100 <.0001
p1 - p4 -0.46 0.113 35 -4.000 <.0001
p1 - p5 -0.34 0.113 35 -3.000 0.0400
p2 - p3 -0.30 0.113 35 -2.600 0.0900
p2 - p4 -0.29 0.113 35 -2.600 0.1000
p2 - p5 -0.17 0.113 35 -1.500 0.5600
p3 - p4 0.01 0.113 35 0.00 1.0000
p3 - p5 0.13 0.113 35 1.100 0.8000
p4 - p5 0.12 0.113 35 1.100 0.8300

P value adjustment: tukey method for comparing a family of 5 estimates

Post-hoc comparisons

Scheffe method
Performing a Single Comparison

After plotting data, I decide to compare means of groups 4 & 7 using a t-test:

```
Two Sample t-test
data:  y.4 and y.7
t = 4.165, df = 18, p-value = 0.0005813
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
13.84 42.01
sample estimates:
mean of x mean of y
111.33  83.41
```

Finally, for completeness, I will do the comparison using the linear.comparison command:

```
> my.contrast<-list(c(0,0,0,1,0,0,-1,0));
> c.4vs7 <- linear.comparison(y,g,weights=my.contrast)

[1] "computing linear comparisons assuming equal variances among groups"
[1] "C 1: F=9.915, t=3.149, p=0.002, psi=27.924, CI=(14.560,41.287), adj.CI= (10.245,45.602)"
```

Planned vs. Post-hoc Comparisons

- Previous comparisons were planned
- Last 2 comparisons, made after looking at data, were post-hoc
- Scheffe method is preferred for post-hoc linear contrasts
  - compute contrast with normal procedures
  - evaluate observed F with new critical value:
    - \( F_{\text{Scheffe}} = \frac{(a-1) \times F_{\text{alpha}}}{df2 = N-a} \)
    - \( a = \) number of groups
    - \( F_{\text{alpha}} \) is the F value required for desired alpha
  - \( F_{\text{Scheffe}} \) is "normal" omnibus F x (a-1)
  - alternatively, keep standard F & adjust p values using Scheffe adjustment
- Scheffe method and omnibus F test are mutually consistent

What was wrong with the preceding analyses?

**Answer:** I performed the analyses after inspecting the data and choosing to compare groups 4 & 7 because they looked different which, obviously, inflates Type I error
### Scheffe test
for post-hoc comparisons

<table>
<thead>
<tr>
<th>contrast</th>
<th>estimate</th>
<th>SE</th>
<th>df</th>
<th>t.ratio</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td>0.964</td>
<td>0.253</td>
<td>35</td>
<td>3.815</td>
<td>0.0005</td>
</tr>
<tr>
<td>quadratic</td>
<td>-0.873</td>
<td>0.299</td>
<td>35</td>
<td>-2.919</td>
<td>0.0061</td>
</tr>
<tr>
<td>cubic</td>
<td>-0.244</td>
<td>0.253</td>
<td>35</td>
<td>-0.966</td>
<td>0.3407</td>
</tr>
<tr>
<td>quartic</td>
<td>0.616</td>
<td>0.669</td>
<td>35</td>
<td>0.921</td>
<td>0.3633</td>
</tr>
</tbody>
</table>

P. value adjustment: scheffe method with rank 4

The methods compute normal F and adjust the p value to be consistent with Scheffe method.

Scheffe.rank should be set to degrees of freedom for grouping factor (i.e., a-1)

### Scheffe test
for post-hoc comparisons

<table>
<thead>
<tr>
<th>contrast</th>
<th>estimate</th>
<th>SE</th>
<th>df</th>
<th>t.ratio</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1 &lt;- c(-3, -3, 2, 2, 2)</td>
<td>2.013</td>
<td>0.438</td>
<td>35</td>
<td>4.596</td>
<td>0.0001</td>
</tr>
<tr>
<td>c2 &lt;- c(-1, 1, 0, 0, 0)</td>
<td>0.166</td>
<td>0.113</td>
<td>35</td>
<td>1.471</td>
<td>0.1503</td>
</tr>
</tbody>
</table>

P. value adjustment: scheffe method with rank 4

These methods compute normal F and adjust the p value to be consistent with Scheffe method.

Scheffe.rank should be set to degrees of freedom for grouping factor (i.e., a-1)