Omnibus vs. Focussed F tests

Omnibus F test is not significant:

```r
> lm01 <- lm(y~g);
> anova(lm01);
```

Analysis of Variance Table

<table>
<thead>
<tr>
<th></th>
<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>g</td>
<td>6</td>
<td>3431</td>
<td>571.8</td>
<td>1.726</td>
<td>0.13</td>
</tr>
<tr>
<td>Residuals</td>
<td>63</td>
<td>20873</td>
<td>331.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Yet a linear comparison of g7 to g1-g6 is significant:

```r
> my.weights <- c(-1, -1, -1, -1, -1, 6)
> my.contrast <- linear.comparison(y, g, c.weights = my.weights)
> my.contrast[[1]]
[1] 7.21771
> my.contrast[[1]]$p
[1] 0.00921941
```

Linear Comparisons

- focussed comparison among group means
- can be a more direct test of hypothesis of interest
- contrasts are more powerful, but less general, than omnibus F
**Hypotheses Evaluated by Contrast**

\[
\text{H}_0:\ \sum_{j=1}^{a} c_j \mu_j = 0
\]

\[
\text{H}_1:\ \mu_k \neq \frac{1}{a}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6)
\]

**General Form of Linear Contrast**

\[
F = \frac{(\Psi^2)/ \sum_{j=1}^{a} (c_j^2/n_j)}{MS_W} \quad \text{df} = (1, N-a)
\]

Value of \(\Psi^2\) is divided by weighted sum of squared contrasts weights

This normalization means that multiplying weights by \(k\) has no effect on the magnitude of the contrast (though it may change sign) and doesn’t alter \(F\) or \(p\) values

Same results obtained by contrast weights \(w_1 \leftarrow c(-1, -1, -1, -1, 0, 6)\) and \(w_2 \leftarrow c(1/3, -1/3, -1/3, 1)\)

**One-tailed Linear Comparisons**

Use one-tailed t test

\[
\text{H}_0:\ \mu_k \leq \frac{1}{a}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6)
\]

\[
\text{H}_1:\ \mu_k > \frac{1}{a}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6)
\]

In R, use \texttt{pt()} command to link \(t\) value to \(p\) value:

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

\[
> \text{pt}(2.68, df=63, lower.tail=FALSE)
\]

\[
[1] \ 0.00469164
\]
Unequal Group Variances

- Assume equal variance in different groups
- F test for contrasts is not robust to violation of equal variance assumption
- When groups have unequal variances, use different method to calculate F
- Correcting for unequal var reduces denominator df (↑, hence, power)

\[
F = \frac{(\bar{\Psi}^2)/\sum_{j=1}^{n}(c_j/n_j)}{\sum_{j=1}^{n}[(c_j/n_j)^2]/(n_j - 1)}
\]

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

Effect Size

Cohen’s d (for a contrast)

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

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

Expresses Psi in terms of the number of standard deviations of population error distribution

Association Strength

- Proportion of Between-Groups variation accounted for by contrast
- With equal n, equals squared correlation between contrast weights & group means

\[
R^2_{\text{orth}} = \frac{SS_{\text{contrast}}}{SS_B}
\]

- Proportion of total variation accounted for by contrast

\[
R^2_{\text{effective}} = \frac{SS_{\text{contrast}}}{SS_{\text{Total}}}
\]

- 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)

\[
R^2_{\text{contrast}} = \frac{SS_{\text{contrast}}}{(SS_{\text{contrast}} + SS_W)}
\]

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:

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

the sum of the \((a-1)\) \(SS_{\text{contrast}}\)s will equal \(SS_B\)

Complete set of orthogonal contrasts divides \(SS_B\) into independent pieces of variation, and the average of the contrast \(F\) values will equal the omnibus \(F\).

Conducting Contrasts with R `lm()`

```r
> # list 1st column of weights (printed as 1 row):
> contrasts(g)[,1] # they are the same as before
> g1 g2 g3 g4 g5 g6 g7
> -1 -1 -1 -1 -1 -1 6
> > my.lm <- lm(y~g) # construct the linear
> > summary(my.lm) # list the coefficients and t tests

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 96.046 | 2.176 | 1.44e-16  *** |
| g2 | -6.153 | 5.756 | -1.07 | 0.2891 |
| g3 | -3.150 | 5.756 | -0.55 | 0.5861 |
| g4 | 0.163 | 5.756 | 0.03 | 0.9775 |
| g5 | 5.502 | 5.756 | 0.86 | 0.3908 |
| g6 | -5.084 | 5.756 | -0.88 | 0.3804 |

Residual standard error: 18.2 on 63 degrees of freedom
Multiple R-squared: 0.141, Adjusted R-squared: 0.0594
F-statistic: 1.73 on 6 and 63 DF, p-value: 0.13
```

Conducting Contrasts with R `aov()`

```r
> my.weights <- c(-1, -1, -1, -1, -1, 6)
> contrasts(g) <- my.weights
> my.aov <- aov(y~g)
> my.aov <- aov(y~g)
Pearson ANOVA

Df Sum Sq Mean Sq F value Pr(>F)
---
g: myContrast 1 3569.9 3569.9 7.22 0.0092 **
g: others 5 1039.0 207.8
Residuals 63 20873.3
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Trend Analysis

- trend analysis is the application of linear contrasts to groups that differ on a quantitative variable
- trend analysis uses the same methods as linear contrasts
- uses particular weights to compute quantitative trends:
  - linear, quadratic, cubic
Example: Rose Data

- Cognitive performance at 5 ages
- \( r = 0.98 \) (mean score vs age)
- 1-way ANOVA:
  - age: \( F(4,45) = 1.67, p = 0.17 \)
  - omnibus F tests general H0
  - More specific question: Does score increase with age?

Example: Rose Data (continued)

R uses \texttt{ordered factors} to represent groups that differ on quantitative variable

\begin{verbatim}
> rose$orderedAge <- factor(rose$age, labels = "a", ordered = TRUE)
> class(rose$orderedAge)

[1] "ordered" "factor"

> class(rose$ageGroup)

[1] "factor"

> unique(rose$orderedAge)

[1] a1 a2 a3 a4 a5

Levels: a1 < a2 < a3 < a4 < a5

> unique(rose$ageGroup)

[1] g1 g2 g3 g4 g5

Levels: g1 g2 g3 g4 g5
\end{verbatim}

Example: Rose Data (continued)

- R uses \texttt{ordered factors} to represent groups that differ on quantitative variable
- Ordered factors are analyzed using polynomial contrasts (contr.poly)

\begin{verbatim}
> contr.poly(n = 5, scores = c(8,8,10,11,12))

  L Q C G

[1,] 0.6325 0.5345 3.162e-01 0.1195

[2,] 0.3162 0.2673 6.325e-01 0.4781

[3,] 0.0000 -0.5345 4.096e-16 0.7171

[4,] 0.3162 -0.2673 6.325e-01 -0.4781

[5,] 0.6325 0.5345 3.162e-01 0.1195

linear quadratic cubic quartic

weights for 4 orthogonal comparisons
\end{verbatim}

Figure 1: Rose data.

Figure 2: Linear, quadratic, cubic, and quartic trends.
Example: Rose Data (continued)

> summary(rose.aov.02, split = list(orderedAge = list(linear = 1, quadratic = 2, cubic = 3, quartic = 4)))

<table>
<thead>
<tr>
<th>Df Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>orderedAge</td>
<td>4</td>
<td>456</td>
<td>114</td>
</tr>
<tr>
<td>orderedAge: linear</td>
<td>1</td>
<td>440</td>
<td>440</td>
</tr>
<tr>
<td>orderedAge: quadratic</td>
<td>1</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>orderedAge: cubic</td>
<td>1</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>orderedAge: quartic</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Residuals</td>
<td>45</td>
<td>3065</td>
<td>114</td>
</tr>
</tbody>
</table>

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Example: Rose Data (continued)

\[
\begin{align*}
\text{R}(\text{alerting}) &= 440/456 = 0.96 \\
\text{R}(\text{effect size}) &= 440/3065 = 0.14
\end{align*}
\]

Example: Rose Data (continued)

- 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 groups are not constant

Example: Rose Data (continued)

> contr.poly(n=5, scores=cb(8,9,10,11,12))

\[
\begin{array}{cccc}
1 & -0.3624555 & 0.5345225 & -3.162278e-01 \\
2 & -0.3162278 & -0.2672612 & 6.324555e-01 \\
3 & 0.0000000 & -0.3545225 & -4.095972e-16 \\
4 & 0.3162278 & -0.2672612 & -6.324555e-01 \\
5 & 0.6324555 & 0.5345225 & 3.162278e-01
\end{array}
\]

Example: Rose Data (continued)

> ages1 <- c(8,9,10,11,12)
> lin1 <- contr.poly(n=5, scores=ages1)[,1]
> plot(ages1,lin1,'b')

Example: Rose Data (continued)

> ages2 <- c(8,9,10,12,15)
> lin2 <- contr.poly(n=5, scores=ages2)[,1]
> plot(ages2,lin2,'b')
Example: Rose Data (calculating $\psi$)

```r
> summary(rose.aov.02, split = list(orderedAge = list(linear = 1, quadratic = 2, cubic = 3, quartic = 4)))

<table>
<thead>
<tr>
<th></th>
<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>orderedAge</td>
<td>4</td>
<td>446.8</td>
<td>111.7</td>
<td>1.87</td>
<td>0.175</td>
</tr>
<tr>
<td>orderedAge linear</td>
<td>1</td>
<td>440.0</td>
<td>440.0</td>
<td>6.60</td>
<td>0.015</td>
</tr>
<tr>
<td>orderedAge quadratic</td>
<td>1</td>
<td>9.9</td>
<td>9.9</td>
<td>0.13</td>
<td>0.720</td>
</tr>
<tr>
<td>orderedAge cubic</td>
<td>1</td>
<td>6.0</td>
<td>6.0</td>
<td>0.08</td>
<td>0.770</td>
</tr>
<tr>
<td>orderedAge quartic</td>
<td>1</td>
<td>1.3</td>
<td>1.3</td>
<td>0.02</td>
<td>0.901</td>
</tr>
<tr>
<td>Residuals</td>
<td>45</td>
<td>3065.6</td>
<td>68.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Finally, it is a simple matter to list the value of $\psi$ and ignoring any remaining nonlinear trends. This analysis can be done by listing the linear and quadratic trends in the split command and grouping the remaining nonlinear trends:

```r
> with(rose, apply(score ~ orderedAge, mean))
```

Better way for trends in balanced designs...

```r
> with(rose, apply(score ~ orderedAge, mean))
```

Example: Rose Data (interpreting $\psi$)

```r
> coef(rose.aov.02) # balanced designs only!

(Intercept) orderedAge.L orderedAge.Q orderedAge.C orderedAge^4
 11.173  440.8  440.0  7463.0  9408.0

> confint(rose.aov.02)
```

<table>
<thead>
<tr>
<th></th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.63</td>
<td>0.53</td>
</tr>
<tr>
<td>orderedAge.L</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
<tr>
<td>orderedAge.Q</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
<tr>
<td>orderedAge.C</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
<tr>
<td>orderedAge^4</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Example: Rose Data (interpreting $\psi$)

```r
> wL <- contr.poly(n=5)[,1]
wQ <- contr.poly(n=5)[,2]
wC <- contr.poly(n=5)[,3]
w4 <- contr.poly(n=5)[,4]
> tmp <- linear.comparisons(y=rose$score,g=rose$orderedAge,c.weights=list(C1=wl,C2=wQ,C3=wC,C4=w4))
```

Example: Rose Data (interpreting $\psi$)

```r
> confint(rose.aov.02)
```

<table>
<thead>
<tr>
<th></th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.63</td>
<td>0.53</td>
</tr>
<tr>
<td>orderedAge.L</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
<tr>
<td>orderedAge.Q</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
<tr>
<td>orderedAge.C</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
<tr>
<td>orderedAge^4</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Example: Rose Data (interpreting $\psi$)

```r
> confint(rose.aov.02)
```

<table>
<thead>
<tr>
<th></th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.63</td>
<td>0.53</td>
</tr>
<tr>
<td>orderedAge.L</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
<tr>
<td>orderedAge.Q</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
<tr>
<td>orderedAge.C</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
<tr>
<td>orderedAge^4</td>
<td>-0.31</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Example: Rose Data (interpreting $\psi$)

```r
> coef(rose.aov.02) # balanced designs only!

(Intercept) orderedAge.L orderedAge.Q orderedAge.C orderedAge^4
 11.173  440.8  440.0  7463.0  9408.0

> wL <- contr.poly(n=5)[,1]
wQ <- contr.poly(n=5)[,2]
wC <- contr.poly(n=5)[,3]
w4 <- contr.poly(n=5)[,4]
> tmp <- linear.comparisons(y=rose$score,g=rose$orderedAge,c.weights=list(C1=wl,C2=wQ,C3=wC,C4=w4))
```