Multiple Comparisons

EdPsych 580
C.J. Anderson
Multiple Comparisons

Instead of the Omnibus $\mathcal{F}$–Test or After a Significant $\mathcal{F}$ Test

Outline:

• General comments and definitions.
  • Planned or Post Hoc.
  • Contrasts:
    • Simple and/or complex
    • Orthogonal.
Outline (continued)

• Procedures for simple and/or complex comparisons.
  • Planned orthogonal comparisons (POC).
  • Bonferroni (Dunn).
  • Scheffé.
  • Summary & comparison.
Outline (continued)

• Procedures just for simple comparisons (pairs of means).
  • Least Significant Difference (LSD).
  • Tukey honest significant difference (HSD).
  • Newman-Keuls.
  • Dunnett’s method.
  • Summary & comparison.

• Overall Summary & Comparison (and general recommendations).
General comments and definitions

Planned or Post hoc?

• Usually, we are not simply interested in

\[ H_0 : \mu_1 = \mu_2 = \ldots = \mu_J \]

• Rather, interested in either
  • Start with specific or planned we want to perform \( \rightarrow \) the omnibus \( F \) may not be a necessary step.
  • Want to know where differences lie \( \rightarrow \) the omnibus \( F \) is the first step.
Analytical Comparisons

- An **analytical comparison** is a meaningful comparison between two or more treatment conditions (groups) that are components of a larger experimental design.
- Analytical comparisons are either
  - Planned — before looking at data.
  - Post-hoc — after looking at data.
Analytical Comparisons (continued)

• When comparisons are planned (possibly the reason for doing the study or experiment), this aspect influences the specific method or procedure used to test the comparison.

• If an experimenter does not know beforehand what differences to examine (i.e., an exploratory study), then post hoc tests are performed — designed for “data snooping” or a fishing expedition for significant results.

• The procedures for planned comparisons are generally different than those for post hoc ones.
Contrasts: Simple and/or Complex

- The composite nature of $SS_{between}$ (or $SS_A$),
  $$SS_A = \sum_{j=1}^{J} n_j (\bar{Y}_j - \bar{Y})^2.$$

- For now suppose we have a balanced design,
  $$SS_A = n \sum_{j=1}^{J} (\bar{Y}_j - \bar{Y})^2 = \frac{n \sum_{j<j'} (\bar{Y}_j - \bar{Y}_{j'})^2}{J}$$
  Summation is over all unique pairs of means.

- $SS_A$ equals the average (mean) of the squared differences between the pairs of means.
Example: Fish Data

12 fish were randomly assigned to one of 3 methods of cooking fish. The average flavor rating of three judgements is the dependent variable.

Summary Statistics:

<table>
<thead>
<tr>
<th>Method of cooking fish</th>
<th>n</th>
<th>Mean</th>
<th>Variance</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>5.7083</td>
<td>0.1954</td>
<td>0.4420</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>5.2333</td>
<td>0.3297</td>
<td>0.5742</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>4.8333</td>
<td>0.2115</td>
<td>0.4599</td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>5.258</td>
<td>0.3631</td>
<td>0.6026</td>
</tr>
</tbody>
</table>
Example: Fish Data
Example: Fish Data

ANOVA summary Table:

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>4.6050</td>
<td>2.3025</td>
<td>9.38</td>
<td>.0006</td>
</tr>
<tr>
<td>Error</td>
<td>33</td>
<td>8.1025</td>
<td>0.2455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>35</td>
<td>12.7075</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Demonstration

Fish data,

\[ n = 12 \quad \bar{X}_1 = 5.7083 \quad \bar{X}_2 = 5.2333 \quad \bar{X}_3 = 4.5333 \]

and

\[ SS_{Method} = 4.60 \]

\[
SS_{Method} = \frac{n[(\bar{X}_1 - \bar{X}_2)^2 + (\bar{X}_1 - \bar{X}_3)^2 + (\bar{X}_2 - \bar{X}_3)^2]}{J} \\
= \frac{12[(5.708 - 5.233)^2 + (5.708 - 4.533)^2 + (5.233 - 4.533)^2]}{3} \\
= 4(.2256 + .7656 + .1598) \\
= 4.60
\]
Analytic Comparisons (continued)

- For $J = 2$, there is only one pair of means.
- If we reject $H_0 : \mu_1 = \mu_2$, then we know that the means of the two groups are not equal (i.e., they are different).
- For $J > 2$, it’s ambiguous which means are different.
- Suppose you have detailed research hypotheses that you specified beforehand and your only interest is in these questions — Planned Comparisons.
Planned Orthogonal Comparisons

- Calories of hot dogs and type (beef, meat, and poultry). We specifically want to know
  - Whether the non-poultry and poultry dogs have the same average calories.
  - Whether the beef or meat (“combo-dog”) have the same mean calories.

- Methods of cooking fish. Method I is a traditional method and II and III are alternative methods. Want to test
  - Whether the new and old methods are the same.
  - Whether the two new methods are the same.
Planned Orthogonal Comparisons (cont.)

• Translation of the two questions into statistical hypotheses

\[ H_{o(1)} : \quad \mu_1 = \frac{1}{2}(\mu_2 + \mu_3) \]
\[ H_{o(2)} : \quad \mu_2 = \mu_3 \]

• \( H_{o(1)} \) is an example of a complex comparison
  \( \rightarrow \) it involves more than two means

• \( H_{o(2)} \) is an example of a simple comparison
  \( \rightarrow \) it only involves the comparison of two means.
Contrasts

• The two hypotheses written as contrasts
  • $H_{o(1)} : \mu_1 - \frac{1}{2}(\mu_2 + \mu_3) = 0$
  • or equivalently,
    $H_{o(1)} : \mu_1 - \frac{1}{2}\mu_2 - \frac{1}{2}\mu_3 = 0$

• $H_{o(2)} : \mu_2 - \mu_3 = 0$
• or equivalently,
  $H_{o(2)} : 0\mu_1 + 1\mu_2 - 1\mu_3 = 0$

• Analytical comparisons are tested by forming contrasts of the treatment means.
Contrasts (continued)

• Formally, a **contrast** is defined as

\[ \psi = c_1 \mu_1 + c_2 \mu_2 + \ldots + c_J \mu_J = \sum_{j=1}^{J} c_j \mu_j \]

where

• At least two \( c_j \)'s are non-zero.
• \( \sum_{j=1}^{J} c_j = 0 \), which ensures comparisons are independent of the overall mean \( \mu \).
Contrasts (continued)

• Our example:

\[ H_{o(1)} : \mu_1 - \frac{1}{2} \mu_2 - \frac{1}{2} \mu_3 = 0 \quad H_{o(2)} : 0\mu_1 + 1\mu_2 - 1\mu_3 = 0 \]

• The coefficients for these two comparisons,

For comp. 1 \((1, -1/2, -1/2)\)
For comp. 2 \((0, 1, -1)\)

• Alternative coefficients,

For comp. 1 \((2, -1, -1)\)
For comp. 2 \((0, .5, -1.5)\)

• These two sets of coefficients yield the same results.
Contrasts & Orthogonal Contrasts

• Requirement for a contrast,
\[ \sum_{j=1}^{J} c_j = 0. \]

• Requirements for orthogonal contrasts:
  • For each set of coefficients,
\[ \sum_{j=1}^{J} c_j = 0 \]
  • Two set of coefficients, \( c \) and \( c' \),
\[ \sum_{j=1}^{J} c_j c'_j = 0 \]
Orthogonal Contrasts

• In our example, we have comp. 1 is \((1, -1/2, -1/2)\) and comp. 2 \((0, 1, -1)\). Are these orthogonal contrasts?

• What would be a non-orthogonal set of contrasts for this example?
Orthogonal Contrasts (continued)

- Orthogonal comparisons contain **linearly independent** (non-redundant) information.
- The largest number of orthogonal comparisons \( = (J - 1) \).
- If the number of comparisons \( > (J - 1) \), then the comparisons must be redundant.

**Are these comparisons orthogonal?**

<table>
<thead>
<tr>
<th></th>
<th>( \mu_1 )</th>
<th>( \mu_2 )</th>
<th>( \mu_3 )</th>
<th>( \mu_4 )</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(b)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>(c)</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td></td>
</tr>
</tbody>
</table>
Testing Comparisons

Need the sum of squares for each comparison:

\[
SS_{comp} = \frac{n(\sum_{j=1}^{J} c_j \bar{Y}_j)^2}{\sum_{j=1}^{J} (c_j)^2} = \frac{n(\hat{\psi})^2}{\sum_{j=1}^{J} (c_j)^2}
\]

Fish example, for the first comparison,

- Numerator

\[
\hat{\psi}_1 = 1\bar{Y}_1 - .5\bar{Y}_2 - .5\bar{Y}_3 = 5.7083 - .5(5.2333) - .5(4.8333) = .675
\]

- Denominator

\[
\sum_{j=1}^{J} (c_j)^2 = (1)^2 + \left(-\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 = 1.5
\]
Example Testing Comparisons

• $SS_{comp1}$ is
  \[
  SS_{comp1} = \frac{12(0.675)^2}{1.5} = 3.645
  \]

• For the second comparison,
  \[
  \hat{\psi}_2 = (0)\bar{Y}_1 + (1)\bar{Y}_2 + (-1)\bar{Y}_3 = 5.2333 - 4.8333 = .40
  \]
  \[
  \sum_{j=1}^{J}(c_j)^2 = (0)^2 + (1)^2 + (-1)^2 = 2
  \]
  So
  \[
  SS_{comp2} = \frac{12(.40)^2}{2} = .960
  \]
Example Testing Comparisons

Summing all of this in an ANOVA table,

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p–value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>(2)</td>
<td>(4.605)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comp. 1</td>
<td>1</td>
<td>3.645</td>
<td>3.645</td>
<td>14.85</td>
<td>.0005</td>
</tr>
<tr>
<td>Comp. 2</td>
<td>1</td>
<td>.960</td>
<td>.960</td>
<td>3.91</td>
<td>.0564</td>
</tr>
<tr>
<td>Within (error)</td>
<td>33</td>
<td>8.1025</td>
<td>.2455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected total</td>
<td>35</td>
<td>12.7075</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example Testing Comparisons

- Using the fact that $F = t^2$, an equivalent test of the comparisons uses $t$–ratios:

\[
t = \frac{\hat{\psi}}{s_\hat{\psi}} = \frac{\hat{\psi}}{\sqrt{M S_e \left( \sum_{j=1}^{J} \frac{c_j^2}{n_j} \right)}}
\]

This has Student’s $t$–distribution with $\nu = \nu_e$.

- For our example, for comparison 1:

\[
t = \frac{.675}{\sqrt{\frac{.2455}{12} (1.5)}} = \frac{.675}{\sqrt{.0307}} = 3.85
\]

- Note that $(3.85)^2 = 14.85 \ldots$ as it should.
Confidence Interval for Comparisons

• Forming a \((1 - \alpha)100\%\) confidence interval for the contrast,

\[ \hat{\psi} \pm (1 - \alpha/2) t_{\nu} s_{\hat{\psi}} \]

• 95\% confidence interval for comparison 1 is

\[ .675 \pm .975 t_{33} s_{\hat{\psi}} \]

\[ \pm 2.0345(\sqrt{.0307}) \]

\[ \pm .356 \quad \longrightarrow (.319, 1.031) \]

• As a hypothesis test:
  If 0 is in the interval, then retain \(H_{0}\).
  If 0 is not in the interval, then reject \(H_{0}\).
Confidence Interval for Comparisons

- For comparison 2:
  \[ \hat{\psi}_2 = 0 \bar{Y}_1 + 1 \bar{Y}_2 - 1 \bar{Y}_3 = 5.2333 - 4.8333 = 0.40 \]
  \[ \sum_{j=1}^{J} (c_j)^2 = (0)^2 + (1)^2 + (-1)^2 = 2 \]
  \[ SS_{comp2} = \frac{12(0.40)^2}{2} = 0.960 \]

- Since \( C = J - 1 = 2 \) orthogonal comparisons,
  \[ SS_{method} = SS_{comp1} + SS_{comp2} \]
  \[ 4.605 = 3.645 + 0.960 \]

- “Partitioned” \( SS_{method} \) into two (non-redundant, independent) parts, each with 1 degree of freedom.
Planned Orthogonal Comparisons

- Don’t need to perform the overall $F$–test.
- We only considered non-directional alternatives, i.e.,
  
  \[ H_0 : \psi = 0 \quad \text{versus} \quad H_a : \psi \neq 0 \]

- Can do directional tests by using $t$-ratios, e.g.,
  \[ \text{If } H_a : \psi > 0 \text{ and the observed } t > t_{\nu_e}(\alpha) \]
  Then reject $H_0$

- We’ve been assuming equal variances (homogeneous). Modifications exist for unequal variances. See Kirk.
Planned Orthogonal Comparisons

• In planning comparisons, the meaningfulness of the comparisons is most important factor.

• In an ideal world, meaningful comparisons are orthogonal (as they are in our example), but this won’t always be the case.

• e.g., Instead of comp2 (i.e., $\mu_2 - \mu_3$), suppose that cooking methods 1 and 3 are the least expensive, then it would be desirable to test

\[
\frac{1}{2}(\mu_1 + \mu_3) \text{ vs } \mu_2
\]
Planned Orthogonal Comparisons

• Alternative comparisons:
  Comparison 1:  1  −.5  −.5
  Comparison 2:  −.5  1  −.5

• But...  

\[ \sum_{j} c_{j}c_{j}^* = \frac{-1}{2} + \frac{-1}{2} + \frac{1}{4} = \frac{-3}{4} \neq 0 \implies \text{not orthogonal}. \]

• SAS
SAS/Planned Orthogonal Comparisons

SAS/ASSIST (can only do contrast 1 per run):

- Solutions > ASSIST > Data Analysis > ANOVA > Analysis of Variance >
- In ANOVA window: Table, Dependent & Classification.
- “Additional Options” > Model Hypotheses > Contrasts.
- “Select effect” → the factor.
- “Specify number of contrasts” → 1.
- “Specify contrast label” → name the contrast.
- “Supply contrast values” → these are the $c_j$’s.
- “OK”, ”Goback” to main ANOVA window, then RUN
SAS/Planned Orthogonal Comparisons

Program Commands: in the editor window

```
proc glm;
  class method;
  model flavor = method;
  contrast 'Old vs New' method 1 - .5 - .5;
  contrast 'New vs New' method 0 1 -1;
run;
```

Then click “run” on main SAS toolbar.
Planned Orthogonal Comparisons

- The familywise error rate, $\alpha_{\Sigma}$, equals

$$\alpha_{\Sigma} = \text{Prob(at least one Type I error)}$$

$$= 1 - (1 - \alpha)^C$$

$$= 1 - (.95)^2 = .0975$$

where $C$ = the number of comparisons, and $\alpha$ is the per comparison significance level.

- If the comparisons were not orthogonal, then

$$\alpha_{\Sigma} \leq C\alpha = 2(.05) = .10$$

- With POC, set the per comparison Type I error rate. The familywise type I error rate is larger than $\alpha$. 
Planned Orthogonal Comparisons

• Note: We’ve used the term “familywise” instead of the term “experimentwise,” because “experimentwise” is not applicable to ANOVA designs where there are two or more factors.

• If you are doing planned comparisons that are
  • Simple or complex,
  • Orthogonal or not orthogonal,
  • Want to set the familywise error rate $\alpha \Sigma$,

Use the next method...
Bonferroni (Dunn) Method

- Designed for a relatively small number of planned comparisons (simple and/or complex, orthogonal or not) to set a familywise type I error rate, $\alpha_{\Sigma}$.

- This method is like POC, except that the per comparison $\alpha$–level is set to $\alpha = \alpha_{\Sigma}/C$ where $\alpha_{\Sigma}$ is the familywise type I error rate and $C$ is the number of planned comparisons.
Bonferroni Method

- If you are using $F$-statistics, the critical value would be

$$(1 - \alpha \Sigma / C) F_{1, \nu_e}.$$ 

- If you are using $t$ ratios and/or computing confidence intervals, the critical value is

$$(1 - \alpha \Sigma / 2C) t_{\nu_e}.$$
Example: Bonferroni Method

- Test the $C = 2$ comparisons
  
  $H_{o(1)}: \mu_1 - \frac{1}{2}\mu_2 - \frac{1}{2}\mu_3 = 0$ and $H_{o(2)}: 0\mu_1 + 1\mu_2 - 1\mu_3 = 0$

  (which were planned before looking at the data) with a familywise type I error rate equal to $\alpha_\Sigma = .05$.

- Instead of using
  
  $(1-.05)F_{1,33} = .95 \quad F_{1,33} = 4.1393$

  we would use

  $(1-.05/2)F_{1,33} = .975 \quad F_{1,33} = 5.5147$
Example: Bonferroni Method

• For $t$–ratios and/or computing confidence intervals for the contrasts, instead of using

$$(1-.05/2)t_{33} = .975 \quad t_{33} = \sqrt{4.1393} = 2.0345$$

we would use

$$(1-.05/2(2))t_{33} = .9875 \quad t_{33} = \sqrt{5.5147} = 2.3483$$

• Fish example: Since our test statistics were

\[
\begin{align*}
\text{Comp 1:} & \quad t = 3.85 & & \text{and} & & F = 14.85 \\
\text{Comp 2:} & \quad t = 1.98 & & \text{and} & & F = 3.91
\end{align*}
\]

• In this example, our conclusions are the same as they were with the POC; namely, reject $H_{o(1)}$ but retain $H_{o(2)}$. 
Bonferroni “Critical” Values

- Use the standard tables of percentiles of Student’s $t$ distribution but use the appropriate $\alpha$ value.
- Special tables of Student’s $t$ distribution that have $t$ values for more $\alpha$ levels.
- Use online density calculator at UCLA web-site (get $F$ or $t$ values).
- Use the $p$-value program on course web-site (get $F$ or $t$ values).
Issues Regarding Planned Comparisons

- Are they orthogonal or not?
- The number of comparisons should be limited to a “small” number for the Bonferroni method where small is less than or equal to \((J - 1)\).
- \((J - 1)\) is the maximum number of orthogonal comparisons that you could test.
- With the Bonferroni methods, as the number of comparisons \(C\) increases, it’s harder to reject \(H_o\), and the power decreases.
- So only use Bonferroni when \(C\) is “small” \(\leq (J - 1)\).
Scheffé Method

This method can be used for

- Planned or un-planned (Post Hoc) comparisons.
- Simple or complex ones.
- Any number of comparisons.
- Setting familywise type I error rate to a desired level.

The Scheffé method is like Bonferroni in that it only involves using a different critical value for a test statistic.
Scheffé Method

• For Scheffé’s method, the critical value for an $F$ statistic is

$$F_{crit} = (v_b)(1-\alpha_\Sigma)F_{v_b,v_e} = (J-1)(1-\alpha_\Sigma)F(J-1),v_e$$

• For $t$ statistics and/or confidence intervals, take the square root of $F_{crit}$.

• e.g., For the first comparison in the methods of cooking fish example, since

$$0.95F_{2,33} = 3.2849$$

The critical value we need is

$$F_{crit} = (3-1)(3.2849) = 6.5698 \text{ or } t_{crit} = \sqrt{6.5698} = 2.563$$
Comparison of Methods (simple &/or complex)

95% confidence intervals for example comparison 1:

- Planned Orthogonal Comparisons:
  \[ \hat{\psi}_1 = .675 \pm 2.0345(\sqrt{.0307}) \]
  \[ \pm .356 \quad \Rightarrow (.319, 1.031) \]

- Bonferroni: For \( C = 2 \),
  \[ .675 \pm 2.348(\sqrt{.0307}) \]
  \[ \pm .411 \quad \Rightarrow (.264, 1.086) \]

- Scheffé:
  \[ .675 \pm 2.563(\sqrt{.0307}) \]
  \[ \pm .449 \quad \Rightarrow (.226, 1.124) \]
Comparison of Methods (simple &/or complex)

- Scheffé is the most conservative, has the lowest type I error rate, and has the lowest power.
- Scheffé is the most flexible and can be used with any number of planned or post hoc simple and/or complex comparisons.
- If the number of comparisons is “large,” Bonferroni can be more conservative than Scheffé.
- POC is the most powerful, but also has the highest type I error rate.
- POC requires the comparisons to be orthogonal (so sum of $SS_{comp}$ add up to $SS_{between}$).
Simple Comparisons

- Only test pairs of means (simple contrasts).
- If look at all possible means, you have \( J(J - 1)/2 \) tests. We would not use
  - Planned orthogonal comparisons, because the constraints/comparisons would not be orthogonal.
  - Bonferroni, because the number of comparisons would be “large” (so Bonferroni would have lower power than Scheffé).
  - Scheffé’s Method, because there are better/more powerful methods for examining all possible means or a sub-set of all possible means.
Procedures for Simple Comparisons

We will cover

- Protected Lease Significant Difference (LSD)
- Tukey’s Honest Significant Difference (HSD)
- Dunnett’s test
- Newman-Keuls
Protected Lease Significant Difference

- Sometimes called “Fisher’s method”

- If you **reject** the overall $F$ test from the ANOVA, then to test the equality of pairs of means, $H_0 : \mu_j = \mu_{j'}$ versus $H_a : \mu_j \neq \mu_{j'}$

Compute $t$–ratio,

$$t = \frac{\bar{Y}_j - \bar{Y}_{j'}}{s(\bar{Y}_j - \bar{Y}_{j'})}$$

where $s(\bar{Y}_j - \bar{Y}_{j'}) = \sqrt{MS_e \left( \frac{1}{n_j} + \frac{1}{n_{j'}} \right)}$

- Compare the obtained $t$ statistic to Student’s $t$ with $\nu_e$ degrees of freedom.
Example: LSD

- For the methods of cooking fish data:

  \[ H_o : \mu_1 = \mu_2 \quad \text{versus} \quad H_a : \mu_1 \neq \mu_2 \]

  \[
  t = \frac{(5.7083 - 5.2333)}{\sqrt{0.24553 \left( \frac{1}{12} + \frac{1}{12} \right)}} = \frac{0.4753}{0.20229} = 2.350
  \]

- Since \( 0.975t_{33} = 2.035 \) is less than the obtained test statistic, reject \( H_o \). It appears that methods 1 and 2 differ.
Example: LSD

- Alternatively, compute the confidence interval for the difference between means for methods 1 and 2:

  \[
  (\bar{Y}_1 - \bar{Y}_2) \pm .975 t_{\nu} s(\bar{Y}_1 - \bar{Y}_2)
  \]

  \[
  .4753 \pm 2.035(.20229)
  \]

  \[
  .4753 \pm .4116 \rightarrow (.064, .887)
  \]

- Since this design is balanced,

  \[
  s(\bar{Y}_j - \bar{Y}_{j'}) = \sqrt{MS_e \left( \frac{1}{n_j} + \frac{1}{n_{j'}} \right)} = \sqrt{MS_e \left( \frac{2}{n} \right)}
  \]

- The half width of the interval is \( .975 t_{\nu} s(\bar{Y}_j - \bar{Y}_{j'}) \)
Minimum Significant Difference

• The half width of the interval is $0.975 t_{\nu e} s (\bar{Y}_j - \bar{Y}_{j'})$ is the same for all pairs of means.
• For our example the half width is $0.4116$
• Any method for comparing pairs of means, such a width (i.e., difference between means) is know as the "Minimum Significant Difference"
• For any pair of means whose difference is greater than this difference will lead to rejecting the null hypothesis that the means are equal.
Minimum Significant Difference

- In the LSD procedure, the minimum significant difference equals

\[ 0.975 t_{\nu e} S ( \bar{Y}_j - \bar{Y}_{j'} ) \]

- This difference is referred specifically to as the “Least Significant Difference”.
Tukey’s Honest Significant Difference

or “wholly significant difference” or HSD.

• This is a post hoc method designed for testing all pairs of means while controlling the familywise Type I error rate.

• The test statistic is the same as for LSD,

\[ q_T = \frac{\bar{Y}_j - \bar{Y}_j'}{s(\bar{Y} - \bar{Y})} \]

except that now

\[ s(\bar{Y}_j - \bar{Y}_j') = \sqrt{\frac{M S_e}{2} \left( \frac{1}{n_j} + \frac{1}{n_{j'}} \right)} \]
Tukey’s Honest Significant Difference

- Test the hypotheses, compare the $q_T$ statistic to a Studentized range statistic $q$, which depend on $\nu_e$, $\alpha_\Sigma$, and $J$.

- A Table of the these will be distributed in class.
Tukey’s Honest Significant Difference

- Example: $H_o : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$

\[
q_T = \frac{.4753}{\sqrt{\frac{.2455}{12}}} = \frac{.4753}{.1430} = 3.324
\]

- For $J = r = 3$ and $\alpha_{\Sigma} = .05$, from Table, for $\nu_e = 30$, $q_{crit} = 3.49$, and for $\nu_e = 40$, $q_{crit} = 3.44$.

- From SAS, the critical value for $\nu_e = 33$ is $q_{crit} = 3.470$.

- In any case, do not reject $H_o$. 
Tukey’s Honest Significant Difference

• Alternatively, we can compute the “honest significant difference,”

\[ HSD = q_{crit} s(\bar{Y} - \bar{Y}) = 3.470(0.1430) = 0.4962 \]

and compare it to the difference between pairs of means.

• FYI — Newman-Keuls test
  • Is like Tukey’s HSD except that the per comparison type I error rate is set rather than a familywise type I error rate.
  • This method requires the tests between means to be conducted
Dunnett Test

- Designed for the situation where you want/need to compare each mean with the mean of a control group.
- The test statistic is again a $t$-ratio, specifically

$$t_{\psi} = \frac{\bar{Y}_j - \bar{Y}_{ctrl}}{s(\bar{Y} - \bar{Y})}$$

where

- $s(\bar{Y}_j - \bar{Y}_j')$ is the same as LSD method.
- The critical values are on Table that will be handed out in class (or use SAS).
Example: Dunnett Test

• Since cooking method 1 is the traditional method, we’ll treat it as the “control” and test whether the means for the other methods differ from it.

\[ H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_a : \mu_1 \neq \mu_2 \]

\[
t = \frac{0.4753}{\sqrt{\frac{2(0.2455)}{12}}} = \frac{0.4573}{0.2023} = 2.348
\]

• For \( \alpha = 0.05 \) and \( J = 3 \), the critical value for Dunnett’s test from the table for \( \nu_e = 30 \) is 2.32 and the critical value for \( \nu_e = 40 \) is 2.29.

• From SAS, the critical value for \( \nu_e = 33 \) is 2.31.
Dunnett Test

Since the design is balanced, we can also compute the minimum significant difference for Dunnett’s test:

\[ t_{crit} S(\bar{Y}_j - \bar{Y}_{crtl}) = 2.33(.2023) = .4714 \]
SAS and Mean Comparisons

Can use any of these:

- SAS/ANALYST
- SAS/ASISST (I won’t cover this one in lecture).
- Program commands
Analyst and Mean Comparisons

• Start SAS/Analyst and open sas data set.
• Statistics > ANOVA > Analysis of variance
• Specify dependent and classification (independent) variables.
• Click the “MEANS” button. For each method that you want SAS to perform
  • Select comparison method
  • Select “Main effects”
  • Click on ”Add”
• OK
SAS and Mean Comparisons

Program Commands: in the editor window

```sas
proc glm;
  class method;
  model flavor = method;
  means method / bon scheffe tukey lsd dunnett('1');
run;
```

Then click “run” on main SAS toolbar.

Be sure to explain the output!
Summary of Testing Pairs of Means

- Since the fish example is balanced, we can just look at the differences between the means and compare them to the minimum significant differences from each multiple comparison procedure.

- The differences between the mean flavor ratings of fish cooked by the three different methods are in the table below (i.e., the number in the table \( = \bar{Y}_{col} - \bar{Y}_{row} \)):

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.475</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.875</td>
<td>.400</td>
<td>0</td>
</tr>
</tbody>
</table>
## Summary of Testing Pairs of Means

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Minimum Significant Difference</th>
<th>Conclusion</th>
</tr>
</thead>
</table>
| LSD             | .4116                          | Reject: $H_o: \mu_1 = \mu_2$, and $H_o: \mu_1 = \mu_3$  
Retain: $H_o: \mu_2 = \mu_3$ |
| Tukey (HSD)     | .4962                          | Reject: $H_o: \mu_1 = \mu_3$  
Retain: $H_o: \mu_1 = \mu_2$, and $H_o: \mu_2 = \mu_3$ |
| Bonferroni      | .5102                          | Reject: $H_o: \mu_1 = \mu_3$  
Retain: $H_o: \mu_1 = \mu_2$, and $H_o: \mu_2 = \mu_3$ |
| (for $C = 3$)   |                                |                                                 |
| Scheffé         | .5185                          | Reject: $H_o: \mu_1 = \mu_3$  
Retain: $H_o: \mu_1 = \mu_2$, and $H_o: \mu_2 = \mu_3$ |
| Dunnett*        | .4714                          | Reject: $H_o: \mu_1 = \mu_2$ and $H_o: \mu_1 = \mu_3$  

Multiple Comparisons – p. 63/69
Summary of Testing Pairs of Means

• LSD is the most powerful but has the highest familywise type I error rate \( \leq 3(.05) = .15 \). With LSD, you tend to make fewer type II errors, but more type I errors.

• Tukey’s HSD is the most powerful method for comparing all means that also controls the familywise type I error rate.

• For the situation where all of the means are just compared to the mean of “control” group, Dunnett is more powerful than the rest and also controls the familywise type I error rate.
Summary of Testing Pairs of Means

• While Bonferroni and Scheffé can be used to compare all pairs of means, they are not as powerful as Tukey; therefore, we expect more type II errors with Bonferroni and Scheffé relative to Tukey’s HSD.

• Suppose that we had also planned to test the contrast $H_0 : \mu_I = (\mu_2 + \mu_3)/2$. If we had used Bonferroni, the minimum significant difference for the simple comparison (pairs of means) would have equaled:

\[
(1-.05/2(4))t_{33}s = 2.6421(.2023) = .5345,
\]

which is even larger than the minimum significant difference for Scheffé.
# Summary of Testing Pairs of Means

Simple Comparisons Only (pairs of means)

<table>
<thead>
<tr>
<th>Method</th>
<th>Set Type I Error</th>
<th>Planned or Post Hoc</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Significant</td>
<td>Per comparison</td>
<td>Post Hoc</td>
<td>All pairs of means</td>
</tr>
<tr>
<td>Difference (LSD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tukey honest significant difference (HSD)</td>
<td>Familywise</td>
<td>Post Hoc</td>
<td>All pairs of means.</td>
</tr>
<tr>
<td>Newman-Keuls</td>
<td>Per comparison</td>
<td>Post Hoc</td>
<td>All pairs of means</td>
</tr>
<tr>
<td>Dunnett Method</td>
<td>Familywise</td>
<td>Planned</td>
<td>One mean with each of the other means</td>
</tr>
</tbody>
</table>
# Summary of Simple &/or Complex

Simple and/or Complex Comparisons

<table>
<thead>
<tr>
<th>Method</th>
<th>Set Type I Error</th>
<th>Planned or Post Hoc</th>
<th>Orthogonal?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planned Orthogonal Contrasts (POC)</td>
<td>Per comparison</td>
<td>Planned</td>
<td>Yes</td>
</tr>
<tr>
<td>Bonferroni (Dunn)</td>
<td>Familywise</td>
<td>Planned (but relatively few)</td>
<td>Yes or No</td>
</tr>
<tr>
<td>Scheffé</td>
<td>Familywise</td>
<td>Post Hoc</td>
<td>Yes or No</td>
</tr>
</tbody>
</table>
Schematic Guide to Choosing Test

A priori

Planned?

orthogonal

YES

POC

NO

Dunnett’s test

Compare w/ control

YES

Bonferroni**

NO

HSD, etc.

Pairwise

YES

Scheffé

NO

Post Hoc

** → *Unless you have more than J of them.*
Final Comments

• POC and LSD require equal $n$ (balanced design), and there is an adjustment of Tukey’s HSD for unequal $n$.

• “Familywise” versus “experimentwise”.

• A post hoc procedure can be used with a planned comparison, but may not be the “best” method in terms of power.

• There are lots of other procedures (see Kirk for more).

• Dealing with trade-off between Type I & Type II errors: Suspend judgment?