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

$$H_o : \mu_1 = \mu_2 = \ldots = \mu_J$$

- Rather, interested in either
  - Start with specific or planned we want to perform $\longrightarrow$ the omnibus $\mathcal{F}$ may not be a necessary step.
  - Want to know where differences lie $\longrightarrow$ the omnibus $\mathcal{F}$ 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.

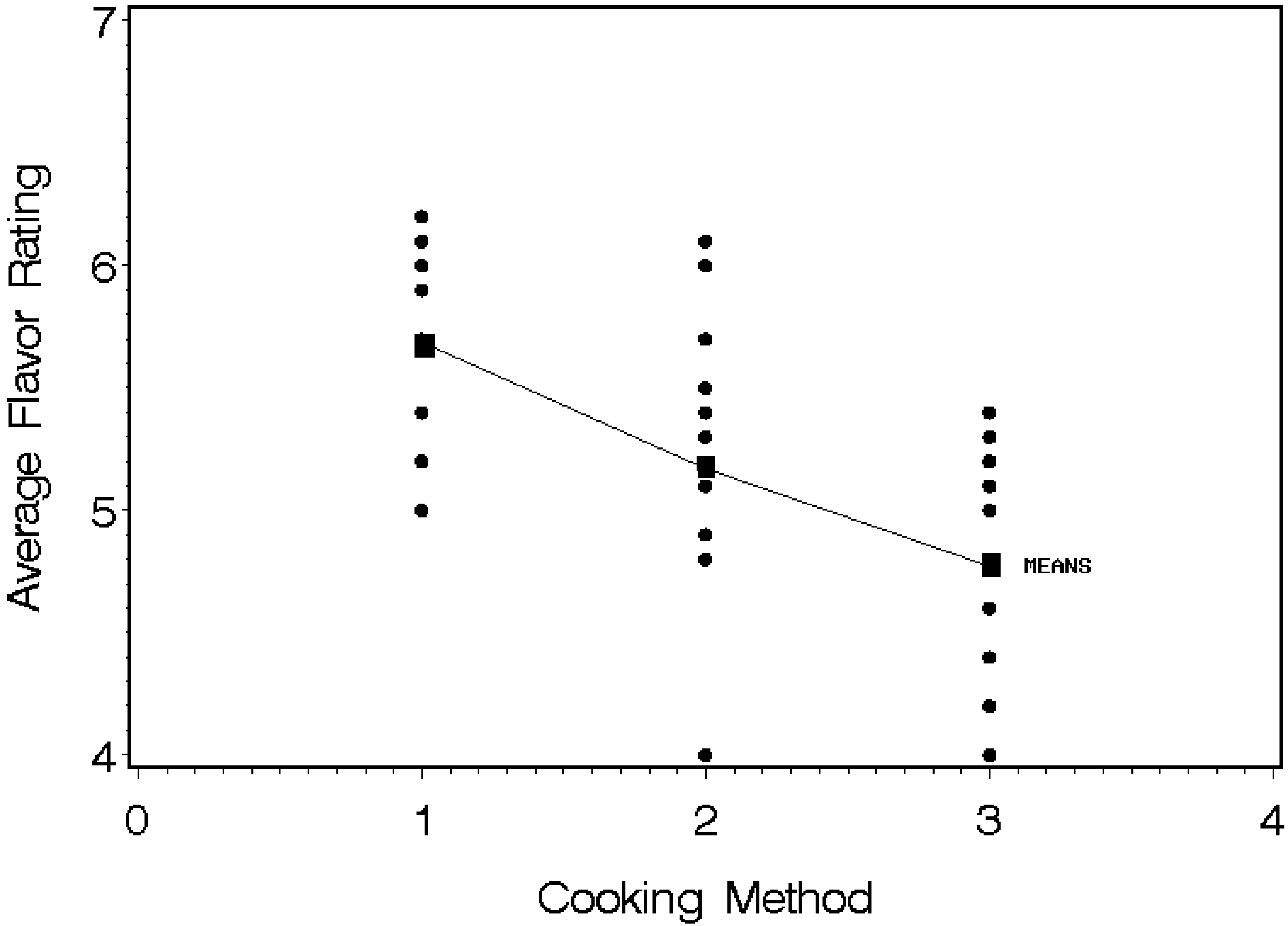
Example: Fish Data

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:

Method of cooking fish	n	Mean	Variance	Std Dev
1	12	5.7083	0.1954	0.4420
2	12	5.2333	0.3297	0.5742
3	12	4.8333	0.2115	0.4599
Total	36	5.258	0.3631	0.6026

Example: Fish Data



Example: Fish Data

ANOVA summary Table:

Dependent Variable: Flavor rating

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	4.6050	2.3025	9.38	.0006
Error	33	8.1025	0.2455		
Corrected Total	35	12.7075			

Demonstration

Fish data,

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

and

$$SS_{Method} = 4.60$$

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

Analytic Comparisons (continued)

- For $J = 2$, there is only one pair of means.
- If we reject $H_o : \mu_1 = \mu_2$, 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 $\longrightarrow$ 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

$$H_{o(1)}: \qquad \mu_1 = \frac{1}{2}(\mu_2 + \mu_3)$$
$$H_{o(2)}: \qquad \mu_2 = \mu_3$$

- $H_{o(1)}$ is an example of a complex comparison $\longrightarrow$ it involves more than two means
- $H_{0(2)}$ is an example of a simple comparison $\longrightarrow$ 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 μ.

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 $\quad (1\,, \quad -1/2, \quad -1/2)$

  For comp. 2 $\quad (0\,, \quad 1, \quad ,-1)$

- Alternative coefficients,

  For comp. 1 $\quad (2, \quad -1, \quad -1)$

  For comp. 2 $\quad (0, \quad .5, \quad ,-.5)$

- These two sets of cofficients 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?

	μ_1	μ_2	μ_3	μ_4	sum
(a)	1	-1	0	0	
(b)	0	0	1	-1	
(c)	1	1	-1	-1	

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

$$\sum_{j=1}^{J}(c_j)^2 = (1)^2 + (-\frac{1}{2})^2 + (-\frac{1}{2})^2 = 1.5$$

Example Testing Comparisons

- SS_{comp1} is
$$SS_{comp1} = \frac{12(.675)^2}{1.5} = 3.645$$

- 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,

Source	df	SS	MS	F	p–value
Method	(2)	(4.605)			
Comp. 1	1	3.645	3.645	14.85	.0005
Comp. 2	1	.960	.960	3.91	.0564
Within (error)	33	8.1025	.2455		
Corrected total	35	12.7075			

Example Testing Comparisons

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

$$t = \frac{\hat{\psi}}{s_{\hat{\psi}}} = \frac{\hat{\psi}}{\sqrt{MS_e \left(\sum_{j=1}^{J} \frac{c_j^2}{n_j} \right)}}$$

  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,

$$\hat{\psi} \pm ({}_{1-\alpha/2})t_{\nu_e} s_{\hat{\psi}}$$

- 95% confidence interval for comparison 1 is

$$\begin{aligned} .675 &\pm {}_{.975}t_{33} s_{\hat{\psi}} \\ &\pm 2.0345(\sqrt{.0307}) \\ &\pm .356 \qquad \longrightarrow (.319, 1.031) \end{aligned}$$

- As a hypothesis test:
  If 0 is in the interval, then retain H_o.
  If 0 is not in the interval, then reject H_o.

Confidence Interval for Comparisons

- For comparison 2:

$$\hat{\psi}_2 = 0\bar{Y}_1 + 1\bar{Y}_2 - 1\bar{Y}_3 = 5.2333 - 4.8333 = .40$$

$$\sum_{j=1}^{J}(c_j)^2 = (0)^2 + (1)^2 + (-1)^2 = 2$$

- So $$SS_{comp2} = \frac{12(.40)^2}{2} = .960$$

- Since $C = J - 1 = 2$ orthogonal comparisons,

$$\begin{aligned} SS_{method} &= SS_{comp1} + SS_{comp2} \\ 4.605 &= 3.645 + 0.960 \end{aligned}$$

- "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.,

$$H_o : \psi = 0 \qquad \text{versus} \qquad H_a : \psi \neq 0$$

- Can do directional tests by using t-ratios, e.g.,

If $H_a : \psi \;\; > \;\; 0$ and the observed $t > t_{\nu_e}(\alpha)$
Then reject H_o

- 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

$$\frac{1}{2}(\mu_1 + \mu_3) \qquad \text{vs} \qquad \mu_2$$

Planned Orthogonal Comparisons

- Alternative comparisons:

  Comparison 1: $1 \quad -.5 \quad -.5$

  Comparison 2: $-.5 \quad 1 \quad -.5$

- But...

$$\sum_j c_j c_j^* = \frac{-1}{2} + \frac{-1}{2} + \frac{1}{4} = \frac{-3}{4} \neq 0 \Longrightarrow \quad \text{not orthogonal.}$$

- 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” $\rightarrow$ the factor.
- “Specfy number of contrasts” $\rightarrow$ 1.
- “Specfy contrast label” $\rightarrow$ name the contrast.
- “Supply contrast values” $\rightarrow$ 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, α_Σ, equals

$$\begin{aligned} \alpha_\Sigma &= \text{Prob(at least one Type I error)} \\ &= 1 - (1-\alpha)^C \\ &= 1 - (.95)^2 = .0975 \end{aligned}$$

  where $C =$ the number of comparisons, and α 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 α.

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 α_Σ,

  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, α_Σ.
- This method is like POC, except that the per comparison α–level is set to $\alpha = \alpha_\Sigma / C$ where α_Σ 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

$$_{(1-\alpha_\Sigma/C)}\mathcal{F}_{1,\nu_e}.$$

- 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

$${}_{(1-.05)}\mathcal{F}_{1,33} = {}_{.95}\,\mathcal{F}_{1,33} = 4.1393$$

  we would use

$${}_{(1-.05/2)}\mathcal{F}_{1,33} = {}_{.975}\,\mathcal{F}_{1,33} = 5.5147$$

Example: Bonferroni Method

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

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

we would use

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

- Fish example: Since our test statistics were

Comp 1:	$t = 3.85$	and	$F = 14.85$
Comp 2:	$t = 1.98$	and	$F = 3.91$

- 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 α value.
- Special tables of Student’s t distribution that have t values for more α levels.
- Use online density calculator at UCLA web-site (get $\mathcal{F}$ or t values).
- Use the p-value program on course web-site (get $\mathcal{F}$ 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

$$\mathcal{F}_{crit} = (\nu_b)_{(1-\alpha_\Sigma)}\mathcal{F}_{\nu_b,\nu_e} = (J-1)_{(1-\alpha_\Sigma)}\mathcal{F}_{(J-1),\nu_e}$$

- For t statistics and/or confidence intervals, take the square root of $\mathcal{F}_{crit}$.
- e.g., For the first comparison in the methods of cooking fish example, since

$$_{.95}\mathcal{F}_{2,33} = 3.2849$$

The critical value we need is
$\mathcal{F}_{crit} = (3{-}1)(3.2849) = 6.5698$ or $t_{crit} = \sqrt{6.5698} = 2.563$

Comparison of Methods (simple &/or complex)

95% confidence intervals for example comparison 1:

- Planned Orthogonal Comparisons:

$$\begin{aligned} \hat{\psi}_1 = .675 \quad &\pm \quad 2.0345(\sqrt{.0307}) \\ &\pm \quad .356 \qquad\qquad \Longrightarrow (.319, 1.031) \end{aligned}$$

- Bonferroni: For $C = 2$,

$$\begin{aligned} .675 \quad &\pm \quad 2.348(\sqrt{.0307}) \\ &\pm \quad .411 \qquad\qquad \Longrightarrow (.264, 1.086) \end{aligned}$$

- Scheffé:

$$\begin{aligned} .675 \quad &\pm \quad 2.563(\sqrt{.0307}) \\ &\pm \quad .449 \qquad\qquad \Longrightarrow (.226, 1.124) \end{aligned}$$

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 constrasts/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, $H_o : \mu_j = \mu_{j'}$ 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 ν_e degrees of freedom.

Example: LSD

- For the methods of cooking fish data:

$$H_o : \mu_1 = \mu_2 \qquad \text{versus} \qquad H_a : \mu_1 \neq \mu_2$$

$$t = \frac{(5.7083 - 5.2333)}{\sqrt{.24553\left(\frac{1}{12} + \frac{1}{12}\right)}} = \frac{.4753}{.20229} = 2.350$$

- Since ${}_{.975}t_{33} = 2.035$ 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:

$$\begin{aligned} (\bar{Y}_1 - \bar{Y}_2) &\pm {}_{.975}t_{\nu_e} s_{(\bar{Y}_1 - \bar{Y}_2)} \\ .4753 &\pm 2.035(.20229) \\ .4753 &\pm .4116 \qquad \longrightarrow (.064, .887) \end{aligned}$$

- 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 ${}_{.975}t_{\nu_e} s_{(\bar{Y}_j - \bar{Y}_{j'})}$

Minimum Significant Difference

- The half width of the interval is ${}_{.975}t_{\nu_e} s_{(\bar{Y}_j - \bar{Y}_{j'})}$ is the same for all pairs of means.
- For our example the half width is $.4116$
- Any method for comparing pairs of means, such a width (i.e., difference between means) is know as the

“<u>Minimum Significant Difference</u>”

- 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

$$ {}_{.975}t_{\nu_e} s_{(\bar{Y}_j - \bar{Y}_{j'})} $$

- 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

$$s_{(\bar{Y}_j-\bar{Y}_{j'})} = \sqrt{\frac{MS_e}{2}\left(\frac{1}{n_j}+\frac{1}{n_{j'}}\right)}$$

Tukey's Honest Significant Difference

- Test the hypotheses, compare the q_T statistic to a Studentized range statistic q, which depend on ν_e, α_Σ, 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,"

$$HSD = q_{crit} s_{(\bar{Y}-\bar{Y})} = 3.470(.1430) = .4962$$

  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

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

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.

$$H_o : \mu_1 = \mu_2 \qquad \text{versus} \qquad H_a : \mu_1 \neq \mu_2$$

$$t = \frac{.4753}{\sqrt{\frac{2(.2455)}{12}}} = \frac{.4573}{.2023} = 2.348$$

- For $\alpha = .05$ 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

```
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 $= \bar{Y}_{col} - \bar{Y}_{row}$):

	1	2	3
1	0		
2	.475	0	
3	.875	.400	0

Summary of Testing Pairs of Means

Procedure	Minimum Significant Difference	Conclusion
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: $H_o : \mu_1 = \mu_3$
(for $C = 3$)		Retain: $H_o : \mu_1 = \mu_2$, and $H_o : \mu_2 = \mu_3$
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 $H_o : \mu_1 = \mu_3$

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 $H_o : \mu_I = (\mu_2 + \mu_3)/2$. 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)

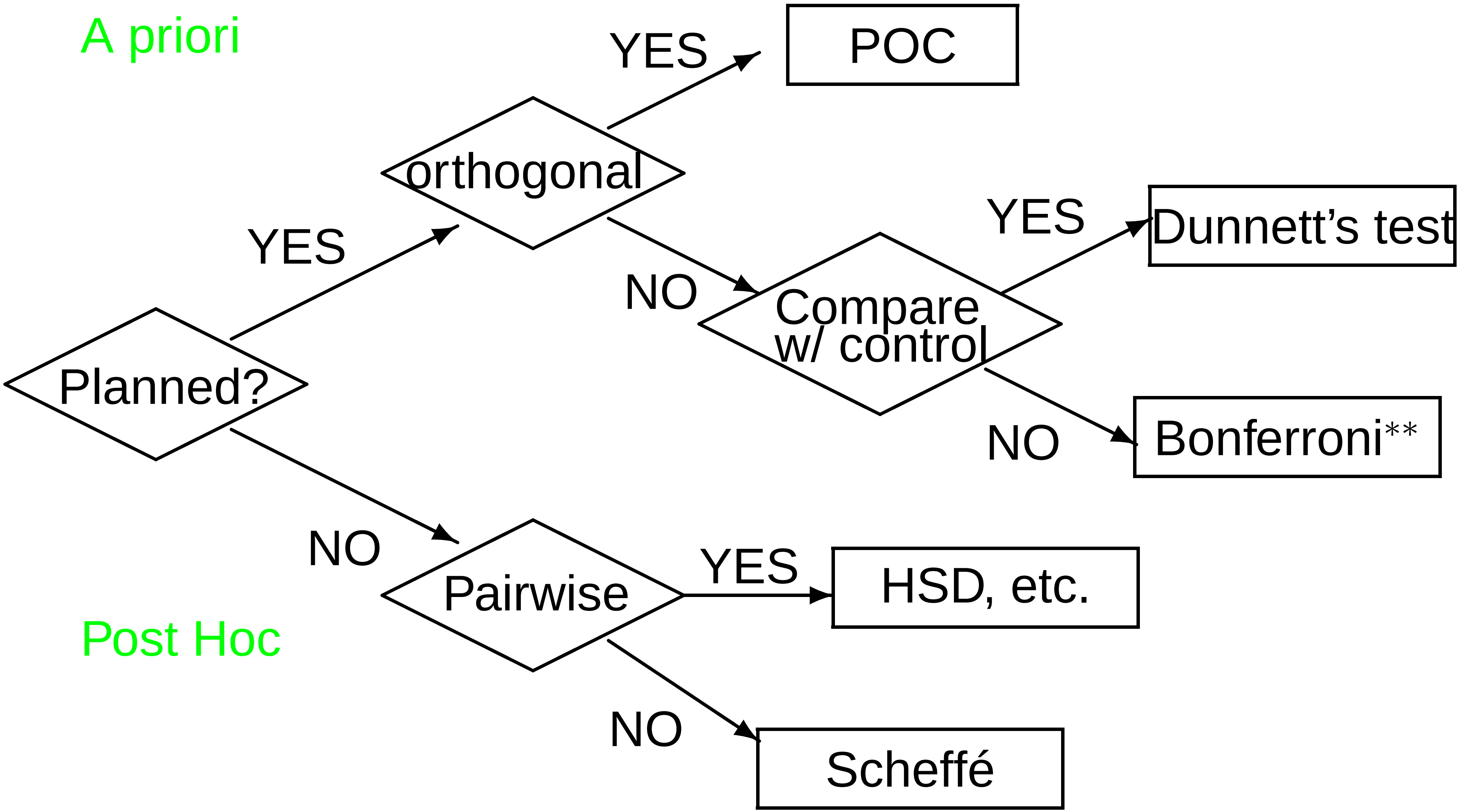
Method	Set Type I Error	Planned or Post Hoc	Comments
Least Significiant Difference (LSD)	Per comparison	Post Hoc	All pairs of means
Tukey honest significant difference (HSD)	Familywise	Post Hoc	All pairs of means.
Newman-Keuls	Per comparison	Post Hoc	All pairs of means
Dunnett Method	Familywise	Planned	One mean with each of the other means

Summary of Simple &/or Complex

Simple and/or Complex Comparisons

Method	Set Type I Error	Planned or Post Hoc	Orthogonal?
Planned Orthogonal Contrasts (POC)	Per comparison	Planned	Yes
Bonferroni (Dunn)	Familywise	Planned (but relatively few)	Yes or No
Scheffé	Familywise	Post Hoc	Yes or No

Schematic Guide to Choosing Test



$** \longrightarrow$ *Unless you have more than J of them.*

Final Comments

- POC and LSD require equal n (balanced design), and there is and 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?