ANOVA & Pairwise Comparisons

- ANOVA for multiple condition designs
- Pairwise comparisons and RH Testing
- Alpha inflation
- LSD and HSD procedures

H0: Tested by ANOVA

- Regardless of the number of IV conditions, the H0: tested using ANOVA (F-test) is …
  - “all the IV conditions represent populations that have the same mean on the DV”
- When you have only 2 IV conditions, the F-test of this H0: is sufficient
  - there are only three possible outcomes …
    T=C  T<C  T>C & only one matches the RH
- With multiple IV conditions, the H0: is still that the IV conditions have the same mean DV…
  T₁ = T₂ = C but there are many possible patterns
  - Only one pattern matches the Rh:

Omnibus F vs. Pairwise Comparisons

- Omnibus F
  - overall test of whether there are any mean DV differences among the multiple IV conditions
  - Tests H0: that all the means are equal
- Pairwise Comparisons
  - specific tests of whether or not each pair of IV conditions has a mean difference on the DV
- How many Pairwise comparisons ??
  - Formula, with k = # IV conditions
    # pairwise comparisons = \([k * (k-1)] / 2\)
  - or just remember a few of them that are common
    - 3 groups = 3 pairwise comparisons
    - 4 groups = 6 pairwise comparisons
    - 5 groups = 10 pairwise comparisons
How many Pairwise comparisons – revisited !!

There are two questions, often with different answers…

1. How many pairwise comparisons can be computed for this research design?
   • Answer $\rightarrow \frac{k \times (k-1)}{2}$
   • But remember $\rightarrow$ if the design has only 2 conditions the Omnibus-F is sufficient; no pairwise comparisons needed

2. How many pairwise comparisons are needed to test the RH:?
   • Must look carefully at the RH: to decide how many comparisons are needed
   • E.g., The ShortTx will outperform the control, but not do as well as the LongTx
     • This requires only 2 comparisons
       ShortTx vs. control       ShortTx vs. LongTx

Process of statistical analysis for multiple IV conditions designs

■ Perform the Omnibus-F
   – test of H0: that all IV conds have the same mean
   – if you retain H0: -- quit

■ Compute all pairwise mean differences (next page)

■ Compute the minimum pairwise mean diff

■ Compare each pairwise mean diff with minimum mean diff
   – if mean diff > min mean diff then that pair of IV conditions have significantly different means
   – be sure to check if the “significant mean difference” is in the hypothesized direction !!!

Using the Pairwise Computator to find the mmd for BG designs

Using these values to make pairwise comparisons
Using the Pairwise Computator to find mmd for WG designs

K = # conditions
N = n

Example analysis of a multiple IV conditions design

<table>
<thead>
<tr>
<th></th>
<th>Tx1</th>
<th>Tx2</th>
<th>Cx</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
<td>40</td>
<td>35</td>
</tr>
</tbody>
</table>

For this design, $F(2,27)=6.54$, $p < .05$ was obtained.

We would then compute the pairwise mean differences.

- Tx1 vs. Tx2 10
- Tx1 vs. C 15
- Tx2 vs. C 5

Say for this analysis the minimum mean difference is 7

Determine which pairs have significantly different means

<table>
<thead>
<tr>
<th></th>
<th>Tx1 vs. Tx2</th>
<th>Tx1 vs. C</th>
<th>Tx2 vs. C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sig Diff</td>
<td>Sig Diff</td>
<td>Not Diff</td>
<td></td>
</tr>
</tbody>
</table>

What to do when you have a RH:

The RH: was, “The treatments will be equivalent to each other, and both will lead to higher scores than the control.”

Determine the pairwise comparisons, how the RH applied to each ...

<table>
<thead>
<tr>
<th></th>
<th>Tx1 = Tx2</th>
<th>Tx1 &gt; C</th>
<th>Tx2 &gt; C</th>
</tr>
</thead>
</table>

For this design, $F(2,42)=4.54$, $p < .05$ was obtained.

Compute the pairwise mean differences.

<table>
<thead>
<tr>
<th></th>
<th>Tx1 vs. Tx2</th>
<th>Tx1 vs. C</th>
<th>Tx2 vs. C</th>
</tr>
</thead>
<tbody>
<tr>
<td>85</td>
<td>15</td>
<td>30</td>
<td>15</td>
</tr>
</tbody>
</table>
Cont. Compute the pairwise mean differences.
Tx1 vs. Tx2  15   Tx1 vs. C  30   Tx2 vs. C  15
For this analysis the minimum mean difference is 18
Determine which pairs have significantly different means
Tx1 vs. Tx2   No Diff !  Tx1 vs. C  Sig Diff !!  Tx2 vs. C  No Diff !
Determine what part(s) of the RH were supported by the pairwise comparisons …
RH: Tx1 = Tx2       Tx1 > C       Tx2 > C
results       Tx1 = Tx2       Tx1 > C       Tx2 = C
well ? supported supported not supported
We would conclude that the RH: was partially supported !

Your turn !! The RH: was, “Treatment 1 leads to the best performance, but Treatment 2 doesn’t help at all.”

What predictions does the RH make ?
   Tx1 > Tx2       Tx1 > C       Tx2 = C

For this design, F(2,42)=5.14, p < .05 was obtained. The minimum mean difference is 3

Compute the pairwise mean differences and determine which are significantly different.
Tx1 vs. Tx2  7    Tx1 vs. C  4    Tx2 vs. C  2

Your Conclusions ?
Complete support for the RH: !!
Alpha Inflation

- Increasing chance of making a Type I error the more pairwise comparisons that are conducted

Alpha correction

- Adjusting the set of tests of pairwise differences to “correct for” alpha inflation
- So that the overall chance of committing a Type I error is held at 5%, no matter how many pairwise comparisons are made

LSD vs. HSD - 3 Possible Outcomes for a Specific Pairwise Comparison

1. Both LSD & HSD show a significant difference
   - Having rejected H0: with the more conservative test (HSD) helps ensure that this is not a Type I error

2. Neither LSD nor HSD show a significant difference
   - Having found H0: with the more sensitive test (LSD) helps ensure this isn’t a Type II error

- Both of these are “good” results, in that there is agreement between the statistical conclusions drawn from the two pairwise comparison methods

LSD vs. HSD Pairwise Comparisons

- Least Significant Difference (LSD)
  - Sensitive -- no correction for alpha inflation
    - Smaller minimum mean difference than for HSD
    - More likely to find pairwise mean differences
      - Less likely to make Type II errors (Miss)
      - More likely to make Type I errors (False Alarm)

- Honest Significant Difference (HSD)
  - Conservative -- alpha corrected
    - Larger minimum mean difference than for LSD
    - Less likely to find pairwise mean differences
    - More likely to make Type II errors
    - Less likely to make Type I errors

- Golden Rule: Perform both!!!
  - If they agree, there is less chance of committing either a Type I or Type II error !!!
LSD vs. HSD -- 3 Possible Outcomes for a Specific Pairwise Comparison

3 Significant difference from LSD, but no significant difference from HSD
- This is a problem !!!
- Is HSD right & the sigdif from LSD a Type I error (FA)?
- Is the LSD is right & H0: from HSD a Type II error (miss) ?
- There is a bias toward “statistical conservatism” in Psychology -- using more conservative HSD & avoiding Type I errors (False alarms)
- A larger study may solve the problem -- LSD & HSD may both lead to rejecting H0: with a more powerful study
- Replication is the best way to decide which is “correct”

Here’s an example…
A study was run to compare 2 treatments to each other and to a no-treatment control. The resulting means and mean differences were …

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>Tx1</th>
<th>Tx2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based on LSD mmd = 3.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tx1</td>
<td>12.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tx2</td>
<td>14.6</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Cx</td>
<td>19.9</td>
<td>**7.6</td>
</tr>
</tbody>
</table>

Conclusions:
- confident that Cx > Tx1 -- got w/ both lsd & hsd
- confident that Tx2 = Tx1 -- got w/ both lsd & hsd
- not confident about Cx & Tx2 -- lsd & hsd differed
  - next study should concentrate on these comparisons