Intermediate Applied Statistics
STAT 460

Lecture 10, 10/1/2004

Instructor:
Aleksandra (Seda) Slavkovic
ses@stat.psu.edu

TA:
Wang Yu
wangyu@stat.psu.edu

One-way ANOVA

A small ANOVA table can be made. The table is shown below.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPOUND</td>
<td>4</td>
<td>401.3</td>
<td>100.3</td>
<td>5.02</td>
<td>0.002</td>
</tr>
<tr>
<td>Error</td>
<td>45</td>
<td>899.2</td>
<td>20.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>1300.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Individual 95% CIs For Mean
Based on Pooled StDev

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>CIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>10.693</td>
<td>4.819</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>6.050</td>
<td>2.915</td>
<td>(--- ---*------)</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>8.636</td>
<td>3.291</td>
<td>(-------*- -----)</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>9.798</td>
<td>5.806</td>
<td>(------ *-------)</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>14.706</td>
<td>4.863</td>
<td></td>
</tr>
</tbody>
</table>

Pooled StDev = 4.470

ANOVA

- So far we have talked about testing the null hypothesis
  \( H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k \)
  versus the alternative hypothesis
  \( H_a: \) at least one pair of groups has \( \mu \neq \mu \).

Next, we want to know where the difference comes from?

Approaches to Comparisons

A. Pairwise comparisons
B. Multiple comparison with a control group
C. Multiple comparison with the “best” group
D. Linear contrasts

- Each of these procedures, (e.g. pairwise confidence intervals for the difference) are available in addition to corresponding tests, in most statistical packages.

All Pairwise Comparisons

- In this approach, we don’t just want to know whether there is any significant difference, we want to know which groups are significantly different from which other groups.

- The most obvious approach to testing pairwise comparisons is just to do a t-test on each pair of groups. The problem with this is sometimes called “compound uncertainty” or “capitalizing on chance.”

- \( \text{Note: By significantly we mean statistically and demonstrably (statistical significance), not greatly or importantly (practical significance).} \)
If there are five independent parts and each fail 5% of the time, then the whole machine fails $1 - (0.95)^5 = 23\%$ of the time!

If we want to do all possible pairwise comparisons, then the number of tests will be

$$c = \frac{k(k-1)}{2}$$

where $k$ is the number of groups.

If there are 5 groups then there are 10 tests!

One possible answer is to control individual type I error risk for planned comparisons (a few comparisons which were of special interest as you were planning the study) and control experiment-wide type I error risk for unplanned comparisons (those you are making after the study has been done just for exploratory purposes).

Suppose for now that there aren’t any comparisons of special interest, we just want to look at all possible pairwise comparisons. Then we definitely should try to control experiment-wide error.

A simple approach to trying to control experiment-wide error:
1. First do an overall $F$-test to see if there are any differences at all.
2. If this test does not reject $H_0$ then conclude that no differences can be found.
3. If it does reject $H_0$, then go ahead and do all possible two-group $t$-tests. (You might use a standard deviation pooled over all groups instead of each pair separately, but otherwise act just as in the two-sample case.)
Method 1: Fisher’s Protected LSD (Least Significant Difference)

- The only problem with that approach is that it doesn’t work. At least according to some experts, the LSD procedure does not really control experiment-wide error. You still have the multiple comparisons problem.

Method 2: Bonferroni Correction

- One way to do this is to use a “Bonferroni correction” on \( \alpha \). The idea here is that we first set a family-wise \( \alpha \), say .05, and then figure out how small the individual \( \alpha' \) would need to be in order to keep the family-wise type I error rate at \( \alpha \). For example, if \( \alpha' = .01 \) and \( c = 5 \), then \( \alpha = .05 \).

Bonferroni Inequality:

For independent events \( A_1, \ldots, A_n \) of equal probability,

\[
P(A_1, A_2, A_3, \ldots, A_n) = 1 - (1 - P(A)) \leq c P(A_1)
\]

Method 2: Bonferroni Correction

- So if we just set \( \alpha^* = \frac{\alpha}{c} \), then the experiment-wide error rate will be controlled.

An advantage of this approach is that it is a simple idea (just divide up your acceptable risk equally among all the tests you want to do).

A disadvantage is that it can be inefficient. The \( \frac{\alpha}{c} \) can become quite small, e.g., .05/21 ≈ .002.

Method 3: Tukey’s (HSD) Tests

- Tukey’s “Honest Significant Difference” test takes a different approach that does not use the t or F distributions but the distribution of the “studentized range” or Q statistic.

- In a Tukey test, the Q distribution is used to find a boundary on differences between group averages, such that 95% of the time if the null hypothesis was true, no pair of groups would have a difference larger than this boundary.

\[
(Q_{max} - Q_{min}) \leq M \times SE
\]

- So any pair of groups with a difference larger than this critical value is declared significantly different.

Confidence Intervals

- Much like the usual t-intervals, the Bonferroni and Tukey confidence intervals for \( y_{ij} - y_{jk} \) is

\[
y_{ij} - y_{jk} \pm m \times SE
\]

where \( m \) is some multiplier. But \( m \) is chosen differently for each procedure.
\[ \bar{y}_j - \bar{y}_j \pm m \times SE_{\bar{y}_j - \bar{y}_j} \]

For a t-interval or LSD t-interval, \( m = \pm \frac{t}{\sqrt{2}} \)
For a Bonferroni t-interval, \( m = \pm t_{\alpha/2} \) with \( \alpha = 0.05 \)
For a Tukey confidence interval,\( m = \sqrt{\frac{dfs}{dfs - 2}} \times t_{\alpha/2} \)

Method 4: Scheffe’s test
- Also available are Scheffe tests and intervals, which use
  \[ m = \sqrt{(dfs)(dfs - dfe)(1 - \alpha)} \]
- but they lead to a test that may be too conservative.

How to compare specific groups in an ANOVA
- If you have a few planned comparisons of theoretical importance in mind, then you can just do them with a slightly adjusted t-test
- a different value for \( s \), and a different degrees of freedom for the t statistic based on the fact that you pool variance over all groups, not just two.
- If you want to do many comparisons then you may wish to try some means of controlling compound uncertainty (i.e., correcting for multiple comparisons).
- In the case where you want to test for all pairwise differences, choices include Fisher’s LSD (too lenient), Bonferroni, and Tukey (the latter also known as Tukey-Kramer when the sample sizes are different). There are also various other procedures I didn’t (and won’t) mention.

Tukey’s pairwise comparisons
- Family error rate = 0.0500
- Individual error rate = 0.00670
- Critical value = 4.02

Intervals for (column level mean) - (row level mean)

<table>
<thead>
<tr>
<th>Level</th>
<th>COMPOUND = 1 subtracted from:</th>
<th>Level</th>
<th>COMPOUND = 2 subtracted from:</th>
<th>Level</th>
<th>COMPOUND = 3 subtracted from:</th>
<th>Level</th>
<th>COMPOUND = 4 subtracted from:</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Difference: -0.895 SE: 1.999</td>
<td>5</td>
<td>Difference: 1.670 SE: 1.999</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Difference: 4.013 SE: 1.999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This is an abbreviation for...

We are 95% confident according to the Tukey procedure that the signed difference \( \mu_1 - \mu_2 \) is between -1.040 and +10.326 hours. So we can’t conclude that \( \mu_1 \neq \mu_2 \).
Multiple Comparisons with a Control

- A different situation arises when you have a control group and are only interested in the comparison of other groups to the control group, not to one another.
- Then there is something like Tukey’s test, but modified, called Dunnett’s test.

Multiple Comparisons with a Control

- Tukey confidence intervals are for \( \mu_i - \mu_j \) and there is one of them for every pair of distinct groups, which comes out to be \( k^2(k-1)/2 \) of them.
- Dunnett confidence intervals are for \( \mu_i - \mu_{\text{control}} \) and there are only \( k-1 \) of them, one for each noncontrol group.

Multiple Comparisons with a Control

- Suppose compound 5 is thought of as a control group because it is the compound that has been in use previously. We want to know if any of the other compounds have different population mean survival times than compound 5.

Multiple Comparisons with a Control

- Dunnett’s comparisons with a control
- Family error rate = 0.0500
- Individual error rate = 0.0149
- Critical value = 2.53

<table>
<thead>
<tr>
<th>Control</th>
<th>Level COMPOUND</th>
<th>Lower</th>
<th>Center</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>1</td>
<td>1.048</td>
<td>2.013</td>
<td>2.948</td>
</tr>
<tr>
<td>Level 4</td>
<td>3</td>
<td>1.048</td>
<td>2.013</td>
<td>2.948</td>
</tr>
</tbody>
</table>

For compounds 2 and 3, we are confident that their difference from the control is nonzero.

Again we conclude that compounds 2 and 3 are different from the control group (actually, poorer).
Multiple comparison with the best

- There is also a kind of test called Hsu’s multiple comparison with the best (MCB).
- This test tries to decide which groups could plausibly have come from the population with the highest (or lowest, if you prefer) mean.
- The group with the highest (lowest) sample average is automatically the leading candidate, but any group that does not significantly differ from this group is still included as a possible best.
- So Hsu’s confidence intervals, one for each group, would be for μi - “Pwin.”
- This approach is harder to understand than the others, perhaps because it seems to be doing more than one thing at the same time.

Next Lecture

- Linear Contrasts