Tukey’s procedure
Dunnett’s Procedure
Bonferroni’s Procedure

ST4241 — Design and Analysis of Clinical Trials
Lecture 3: Multiple Comparison (cont.)

Chen Zehua
Department of Statistics & Applied Probability

8:00-10:00 am, Friday, August 19, 2016
Outline

Tukey’s procedure

Dunnett’s Procedure

Bonfferoni’s Procedure
In certain trials of comparing $g$ treatments, the purpose is only to find out the differences in any pair of the treatments.

- Only the particular pairwise contrasts, $\mu_i - \mu_j, 1 \leq i < j \leq g$, are of interest.
- One is not concerned with other contrasts of treatments.
- The overall significance $F$-test is not necessary.
- The appropriate procedure for the pairwise comparison is Tukey’s procedure.
Let $\bar{X}_i, i = 1, \cdots, g$ be the group sample means in a PGD comparing $g$ treatments with equal group sample sizes, i.e., $n_i = n$. Let $\sqrt{\text{WMS}}$ be the within group mean sum of squares.

- The studentized range statistic is defined as:

$$q_{g,n-g} = \sqrt{n}(\max \bar{X}_i - \min \bar{X}_i)/\sqrt{\text{WMS}}.$$  

Under the assumption that $\mu_1 = \cdots = \mu_g$, the distribution of $q_{g,n-g}$ is referred to as the *Studentized range distribution*.

- Let $q_{g,n-g,\alpha}$ denote the upper $\alpha$-quantile of the Studentized range distribution. This quantile is referred to as the *Tukey’s criterion* for pairwise comparison at level $\alpha$.

For values of $q_{g,n-g,\alpha}$, use Table A.5 of Fleiss.
The $Q$-statistic for the contrast $\mu_i - \mu_j$ is given by

$$Q_{ij} = \begin{cases} \frac{\sqrt{n}|\bar{X}_i - \bar{X}_j|}{\sqrt{\text{WMS}}}, & \text{if } n_1 = \cdots = n_g = n; \\ \frac{\sqrt{\hat{n}_{ij}^{(H)}|\bar{X}_i - \bar{X}_j|}}{\sqrt{\text{WMS}}}, & \text{otherwise,} \end{cases}$$

where $\hat{n}_{ij}^{(H)} = \frac{2n_in_j}{n_i+n_j}$.

The contrast $\mu_i - \mu_j$ is significant at level $\alpha$ if

$$Q_{ij} > q_{g,n - g,\alpha}.$$
Tukey’s procedure
Dunnett’s Procedure
Bonfferoni’s Procedure

Difference between $Q$-statistic and $t$-statistic

Recall that the $t$-statistic for the contrast $\mu_i - \mu_j$ is given by

$$T_{ij} = \begin{cases} 
\frac{\sqrt{n} (\bar{X}_i - \bar{X}_j)}{\sqrt{WMS}}, & \text{if } n_1 = \cdots = n_g; \\
\sqrt{\frac{n_i n_j}{n_i + n_j}} \frac{(\bar{X}_i - \bar{X}_j)}{\sqrt{WMS}}, & \text{otherwise}.
\end{cases}$$

Note $|T_{ij}| = Q_{ij}/\sqrt{2}$

- The $t$-statistics can also be used for pairwise comparison using Tukey’s criterion, but $|T_{ij}|$ must be compared with $q_{g,n-g,\alpha}/\sqrt{2}$.

- The contrast $\mu_i - \mu_j$ is significant at level $\alpha$ if

$$Q_{ij} > q_{g,n-g,\alpha}, \quad \text{or} \quad |T_{ij}| > q_{g,n-g,\alpha}/\sqrt{2}.$$
Illustration for the computation of $Q_{ij}$

Sample sizes and means:

<table>
<thead>
<tr>
<th>Group</th>
<th>Ether</th>
<th>Cyclopropane</th>
<th>Thiopental</th>
<th>Spina</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Mean</td>
<td>4.64</td>
<td>4.63</td>
<td>3.53</td>
<td>3.08</td>
</tr>
</tbody>
</table>

\[ WMS = 0.59 \]

For difference $\mu_1 - \mu_2$:

- $C_{ij} = |\bar{X}_1 - \bar{X}_2| = 4.64 - 4.63 = 0.01.$
- $\tilde{n}_{12}^{(H)} = \frac{2 \times 5 \times 7}{5 + 7} = 5.83.$
- \[ Q_{12} = \frac{\sqrt{\tilde{n}_{12}^{(H)}} |\bar{X}_1 - \bar{X}_2|}{\sqrt{WMS}} = \sqrt{5.83} \times 0.01/\sqrt{0.59} = 0.03. \]
The anesthetics trial example (cont.)

The computed values for all the pairs

<table>
<thead>
<tr>
<th>Comparison</th>
<th>$C_{ij}$</th>
<th>$\hat{n}_{ij}^{(H)}$</th>
<th>$Q_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 vs. 2</td>
<td>0.01</td>
<td>5.83</td>
<td>0.03</td>
</tr>
<tr>
<td>1 vs. 3</td>
<td>1.11</td>
<td>6.43</td>
<td>3.66</td>
</tr>
<tr>
<td>1 vs. 4</td>
<td>1.56</td>
<td>6.15</td>
<td>5.04$^a$</td>
</tr>
<tr>
<td>2 vs. 3</td>
<td>1.10</td>
<td>7.88</td>
<td>4.02$^a$</td>
</tr>
<tr>
<td>2 vs. 4</td>
<td>1.55</td>
<td>7.47</td>
<td>5.52$^a$</td>
</tr>
<tr>
<td>3 vs. 4</td>
<td>0.45</td>
<td>8.47</td>
<td>1.71</td>
</tr>
</tbody>
</table>

$^a Q_{ij} > q_{4,25,0.05} = 3.89$

Conclusion: The differences between 1 and 4, 2 and 3, 2 and 4 are significant at level 0.05.
Several treatments versus a control

In certain trials, people is interested only in comparing several treatments (new) with a particular (standard) treatment. There are $p$ treatment groups to be compared with a control group. The data is summarized as follows:

<table>
<thead>
<tr>
<th>Group</th>
<th>Size</th>
<th>Mean</th>
<th>Variance</th>
<th>$L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$n_0$</td>
<td>$X_0$</td>
<td>$s_0^2$</td>
<td>—</td>
</tr>
<tr>
<td>1</td>
<td>$n_1$</td>
<td>$X_1$</td>
<td>$s_1^2$</td>
<td>$L_1$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>$p$</td>
<td>$n_p$</td>
<td>$X_p$</td>
<td>$s_p^2$</td>
<td>$L_p$</td>
</tr>
</tbody>
</table>

Let $\mu_j, j = 0, 1, \ldots, p$ denote the mean responses of group $j$. Only the following $p$ contrasts are of one’s interests:

$$C_j = \mu_j - \mu_0, \ j = 1, \ldots, p.$$
A total of 60 cockrels were assigned at random to receive either no treatment or one of three drugs in diets. These 60 experimental birds were sacrificed at the end of the experiment and the fat content of the breast muscle was measured. The experiment is to see whether the treatments can reduce the fat. The data is summarized below:

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample size</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>15</td>
<td>2.580</td>
<td>0.258</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>2.461</td>
<td>0.409</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>2.232</td>
<td>0.381</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>2.573</td>
<td>0.348</td>
</tr>
</tbody>
</table>

In this example, we only need to compare each of the three treatments with the control group of no drugs.
Dunnett’s criterion

- The test statistics for Dunnett’s procedure are \( t \) statistics.
- For comparison of treatment \( j \) with the control, the statistic is given by

\[
L_j = \frac{\bar{X}_j - \bar{X}_0}{s} \sqrt{\frac{n_0 n_j}{n_0 + n_j}}, \quad j = 1, \ldots, p,
\]

where

\[
s^2 = \frac{\sum_{j=0}^{p} (n_j - 1)s_j^2}{n - p - 1} = \text{WMS}, \quad n = \sum_{j=0}^{p} n_j.
\]

- The Dunnett’s criterion is given by \( d_{p, \nu, \alpha} \), where \( p \) is the number of treatment groups (not including the control group), \( \nu = n - p - 1 \), \( \alpha \) is the size of the test.
- The \( d_{p, \nu, \alpha} \) values are given in Table A. 6 of Fleiss.
The comparison procedure

- The criterion \( d_{p, \nu, \alpha} \) for one-sided tests and two-sided tests are different. For one-sided tests, each \( |L_j| \) is compared with the one-sided criterion, and for two-sided test, each \( |L_j| \) is compared with the two-sided criterion.
- If \( |L_j| > d_{p, \nu, \alpha} \), \( C_j \) is claimed significant at level \( \alpha \).
- The tabled criterion are for the case \( n_0 = n_1 = \ldots n_p \). For unequal group sizes, for contrast \( C_j \), the criterion are adjusted by a factor \( m = 1 + 0.07(1 - n_j/n_0) \). E.g., if the tabled criterion is \( d_{p, \nu, \alpha} \), the adjusted criterion is \( md_{p, \nu, \alpha} \).
- The 100(1-\(\alpha\))% simultaneous confidence intervals for the \( p \) contrasts are constructed as

\[
(\bar{X}_i - \bar{X}_0) \pm md_{p, \nu, \alpha} s \sqrt{\frac{n_0 + n_i}{n_0 n_i}}.
\]
Cockrel example (cont.)

- Computation:
  
  \[ s^2 = \frac{(15 - 1)(0.258^2 + 0.409^2 + 0.381^2 + 0.348^2)}{(60 - 4)} \]
  
  \[ = 0.125, s = 0.3536. \]

  \[ L_1 = \sqrt{\frac{15}{2}}\left(2.461 - 2.58\right)/0.3536 = -0.92. \]

  \[ L_2 = \sqrt{\frac{15}{2}}\left(2.232 - 2.58\right)/0.3536 = -2.70. \]

  \[ L_3 = \sqrt{\frac{15}{2}}\left(2.573 - 2.58\right)/0.3536 = -0.05. \]

- The one-sided criterion \( d_{3,56,0.05} = 2.106 \) which is obtained by interpolation between \( d_{3,40,0.05} \) and \( d_{3,60,0.05} \) as follows:

  \[ d_{3,56,0.05} = 2.13 + \frac{2.10 - 2.13}{60 - 40}(56 - 40) = 2.106. \]

- Conclusion: Drug 2 has a significant effect to reduce the fat content of the animals.
Some design issues

Optimal allocation of group sizes
- The variance of $\bar{X}_i - \bar{X}_0$ is proportional to $\frac{n_0 + n_i}{n_0 n_i}$. A reduction of this variance has the effect of increasing the power of the comparison.
- Let $n_1 = \cdots = n_p$. The optimal group sizes are
  \[ n_0 = n \cdot \frac{1}{(1 + \sqrt{p})}, \quad n_i = n \cdot \frac{1}{(p + \sqrt{p})}. \]

Randomization scheme with optimal group sizes
- Note that $n_0 / n_i = \sqrt{p}$. The randomly permuted blocks scheme is carried out in batches of size $(p + \sqrt{p})r$ ($r \leq 4$). Within each block, the first $r \sqrt{p}$ subjects are assigned to control group, the next $r$ assigned to treatment 1, \cdots
- Since $\sqrt{p}$ is not necessarily an integer, some adjustment is needed to minimize the gap between theoretical optimal size and the actual allocated size.
Cockrel example revisited

A optimal design for the cockrel example is as follows:

- Optimal group size:
  \[
  n_0 = \frac{60}{1 + \sqrt{3}} = 21.96, \quad n_j = \frac{60}{3 + \sqrt{3}} = 12.68, \quad j = 1, 2, 3.
  \]

Take \( n_0 = 21 \) and \( n_j = 13 \).

- Random allocation scheme: Block permutation with block sizes:
  \[
  \begin{array}{c|ccccc}
    \text{Block} & 1 & 2 & 3 & 4 & 5 \\
    \text{Size}   & 15 & 15 & 10 & 12 & 8 \\
  \end{array}
  \]

Block size calculation: \( 15 = 3(3 + 2) \), \( 10 = 2(3 + 2) \), \( 12 = 3(3 + 1) \), \( 8 = 2(3 + 1) \).
Cockrel example revisited (cont.)

- **Subjects in blocks**
  - block 1: 1-15; block 2: 16-30; block 3: 31-40;
  - block 4: 41-52; block 5: 53-60.

- **Subject Allocation**

<table>
<thead>
<tr>
<th>Block</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>10 6 2 5 11 3</td>
</tr>
<tr>
<td>2</td>
<td>30 29 16 18 28 19</td>
</tr>
<tr>
<td>3</td>
<td>34 36 35 31</td>
</tr>
<tr>
<td>4</td>
<td>47 52 45</td>
</tr>
<tr>
<td>5</td>
<td>59 58</td>
</tr>
</tbody>
</table>
Suppose there are $K$ prespecified contrasts which are the only concern of the investigator. The overall type I error rate $\alpha$ for the $K$ contrasts can be controlled by Bonferroni’s method.

- Bonferroni’s procedure to control the overall type I error rate by adjusting the individual type I error rate. If the overall rate is $\alpha$, see, for the contrast $k$, the rate is controlled at $\alpha_k (< \alpha)$, subject to that the individual error rates sum up to $\alpha$, i.e.,

$$\sum_{k=1}^{K} \alpha_k = \alpha.$$

- The allocation of the $\alpha_k$’s can be determined by a consideration of the seriousness of the type I error associated with each of the contrasts. In general, $\alpha_k$ can be taken equal for all $k$, i.e.,

$$\alpha_k = \alpha/K.$$
Bonferoni’s criterion

- The test statistics for pre-determined contrasts are $t$-statistics.
- If the overall type I error rate is to be controlled at $\alpha$ and there are a total of $K$ contrasts, the Bonferroni’s criterion is given by
  $$t_{n, -g, \alpha/2K}.$$

*The rationale of Bonferroni criterion*

Let $A_k$ be the event that contrast $k$ is erroneously claimed significant. The probability that at least one such event occurs, which is the overall error rate to be controlled, is given by

$$Pr(\bigcup_{k=1}^{K} A_k) \leq \sum_{k=1}^{K} Pr(A_k) = \sum_{k=1}^{K} \alpha_k.$$