1. Multiple Comparisons Procedures

There are three major types of multiple comparisons:

MCA: All-Pairwise Multiple Comparisons

MCB: Multiple Comparisons with the Best

MCC: Multiple Comparisons with a Control

Suppose \( k \) treatments are to be compared in terms of their treatment means (long run average treatment effects), which are denoted by \( \theta_1, \theta_2, \ldots, \theta_k \). In this article, we confine our discussion to oneway (possibly unbalanced) designs only.

All-Pairwise Multiple Comparisons compares all pairs of treatments. For MCA, the parameters of interest are \( \theta_i - \theta_j \) for all \( i \neq j \), the \( k(k-1)/2 \) pairwise differences of treatment means.

Multiple Comparisons with the Best compares each treatment with the (unknown) true best of the other treatments. Thus, for MCB, the parameters of interest are \( \theta_i - \theta_j \) for all \( i \neq j \), the \( \delta \) pairwise differences of treatment means.

Multiple Comparisons with the Best compares each treatment with the (unknown) true best of the other treatments. For MCC, the parameters of interest are \( \theta_i - \min\{\theta_j : j \neq i\} \) for \( i = 1, \ldots, k \), the difference between each treatment and the true best of the other treatments. If \( 0 < \theta_i - \min\{\theta_j : j \neq i\} \), then Treatment \( i \) is the true best treatment, for it is better than every other treatment. In fact, suppose \( \theta_i - \max\{\theta_j : j \neq i\} \), then Treatment \( i \) is not the best, then there is another treatment better than it. On the other hand, if \( 0 < \theta_i - \max\{\theta_j : j \neq i\} \), then Treatment \( i \) is the true best treatment, for it is better than every other treatment. In fact, suppose \( \theta_i - \max\{\theta_j : j \neq i\} > -\delta \) (say), where \( \delta \) is a small positive number; then even though we cannot say Treatment \( i \) is best, we can guarantee that it is at least close to the true best. Conversely, if a smaller treatment effect implies a better treatment, then the parameters of interest are \( \theta_i - \min\{\theta_j : j \neq i\} \) for \( i = 1, \ldots, k \), (again the difference between each treatment and the true best of the other treatments). If \( 0 < \theta_i - \min\{\theta_j : j \neq i\} \), then Treatment \( i \) is not the best. If \( \theta_i - \min\{\theta_j : j \neq i\} < 0 \), then Treatment \( i \) is best. If \( \theta_i - \min\{\theta_j : j \neq i\} < \delta \), where \( \delta \) is a small positive number, then even though we cannot say Treatment \( i \) is the best, we can guarantee that it is at least close to the true best.

Multiple comparisons with a Control compares each (new) treatment with the control only. For MCC, suppose Treatment \( k \) is the control; then the parameters of interest are \( \theta_i - \theta_k \) for \( i = 1, \ldots, k-1 \), the difference between each new treatment and the control.

2. Choice of Multiple Comparisons Procedure

Historically, Multiple Comparisons (MC) has been taken as synonymous with All-Pairwise Multiple Comparisons (MCA). For example, the MEANS option in PROC GLM of the SAS® System contains Scheffé, Tukey, GT2, REGWP, REGWQ, Sidák, Bonferroni, Newman-Keuls, Duncan, Protected and Unprotected LSD, all of which are implemented as MCA procedures (SAS User's Guide: Statistics, Version 5 edition, pp.169-175).

Among these MCA procedures, it is well known that Newman-Keuls, Duncan, and LSD are statistically invalid. The true error rate for Newman-Keuls (i.e. its probability of making at least one incorrect assertion), can be as high as \( 1-(1-\alpha)^{k/2} \). Thus, for 10 treatments say, Newman-Keuls with a nominal error rate of 5% can have a true error rate as high as \( 1-(.95)^{10/2} = 23\% \). The error rates for Duncan and LSD can be even higher.

Among the other MCA methods, Tukey's method is generally the most efficient. (Dunnett 1980) Until recently, there lacked a proof of the statistical validity of Tukey's method for unbalanced oneway designs. However, such a proof has now been provided by Hayter (1984). Thus, if all pairwise comparisons are of interest, then Tukey is the method of choice.

In many situations, it is not necessary to compare all pairs of treatments. The purpose of many comparative experiments is to eliminate the treatments that are not the true best, and to assess whether the sample best treatment can be declared the true best. For these experiments, it is only necessary to compare each treatment with the true best of the other treatments, while comparisons among the inferior treatments themselves are unnecessary. (To give a morbid example, if all the patients under Treatments 3 and 5 died shortly after the treatments were administered, it is probably not of primary interest to see which of Treatments 3 and 5 is less undesirable.) The appropriate method in these situations is Multiple Comparisons with the Best, proposed by Hsu (1981, 1984a,b) and implemented in PROC RSMCB. By making only the comparisons that are actually needed, MCB can significantly out-perform Tukey's method in these situations.

If the experimenter is primarily interested in comparing each new treatment with a control, but not among the new treatments themselves, then the appropriate inference procedure is Dunnett's MCC method.

3. PROC RSMCB

PROC RSMCB is the (simultaneous) computer implementation of Multiple Comparisons with the Best (MCB) inference and Ranking and Selection (RS) inference.

Multiple Comparisons with the Best (MCB) compares each treatment with the true best of the other treatments using a simultaneous confidence intervals approach. Ranking and Selection (RS) decides which of the
treatments that appear inferior can be inferred to not be the true best (Subset Selection), and whether the sample best treatment can be inferred to be the true best treatment (Indifference Zone Selection) using a multiple decisions approach. The connection between RS and MCB is as follows: inferring a treatment to be worse than the true best of the other treatments is equivalent to inferring it not to be the true best. Likewise, inferring a treatment to be better than the best of the other treatments is equivalent to inferring it to be the true best treatment. An interesting fact is that (Hsu 1981, 1984a,b), if one gives MCB inference at confidence level $1 - \alpha$, and Subset Selection inference at confidence level $1 - \alpha$, and Indifference Zone Selection inference at confidence level $1 - \alpha$, then the overall confidence level (i.e. the probability that all three inferences are correct simultaneously) is still $1 - \alpha$. PROC RSMCB accordingly provides all three inferences simultaneously, for balanced as well as unbalanced oneway designs.

For Ranking and Selection inference, PROC RSMCB computes $R$- and $S$-values. The $R$-value of a treatment is the smallest $\alpha$ for which that treatment is rejected (not selected) by Gupta’s (1965) Subset Selection. Treatments with $R$-values less than the user’s $\alpha$ are inferred to not be the true best. The $S$-value is the smallest $\alpha$ for which the sample best treatment is selected by Bechhofer’s (1954) Indifference Zone Selection (modified to be applicable to single-stage experiments with unknown variance; see Hsu 1984a). If the $S$-value of that treatment is smaller than the user’s $\alpha$, then it is inferred to be the true best treatment. R- and S-values were proposed in Hsu (1984a) as Ranking and Selection analogues of the $P$-value in hypothesis testing. The exact expressions for $R$- and $S$-values given in that article are double integrals dependent on the sample statistics as well as the sample size configuration. They are evaluated at execution time by PROC RSMCB.

Multiple Comparisons with the Best (Hsu 1981, 1984a,b) compares each treatment with the true best of the other treatments. For MCB inference, PROC RSMCB computes $100(1 - \alpha)$% simultaneous confidence intervals for the difference between each treatment and the true best of the other treatments. Recall that $\theta_1, \ldots, \theta_k$ denote the treatment effects. To describe the MCB confidence intervals, first consider the balanced oneway model

$$Y_{ia} = \theta_i + E_{ia}, \quad i = 1, \ldots, k, \quad a = 1, \ldots, n,$$

where $E_{11}, \ldots, E_{kn}$ are iid normal with mean 0 and variance $\sigma^2$ unknown. We use the usual notations

$$\bar{Y}_i = \frac{1}{n} \sum_{a=1}^{n} Y_{ia}, \quad i = 1, \ldots, k,$$

$$s^2 = \frac{\sum_{i=1}^{k} \sum_{a=1}^{n} (Y_{ia} - \bar{Y}_i)^2}{(n-1)}$$

for the sample means and the pooled sample variance. If a larger treatment effect is better, then PROC RSMCB computes $100(1 - \alpha)$% simultaneous confidence intervals for $\theta_i - \max_{j \neq i} \theta_j, i = 1, \ldots, k$, of the form

$$\left[ \bar{Y}_i - \max_{j \neq i} \bar{Y}_j - D, \bar{Y}_i - \max_{j \neq i} \bar{Y}_j + D \right]$$

where $-x^+ = x$ if $x$ is negative, 0 otherwise; and $x^+ = x$ if $x$ is positive, 0 otherwise. In (1), $D = d(k, \alpha, \nu)(2/n)^{1/2}$ where $d(k, \alpha, \nu)$ is a critical value that depends on $k$, $\alpha$, and $\nu$, the degrees of freedom for $s^2$ (MSE). These confidence intervals were first derived in Hsu (1984b). Their interpretation is as follows. Suppose a larger treatment effect is better; then any treatment with an upper MCB confidence bound equalling 0 can be inferred to not be the true best, for there is another treatment better than it. A treatment with its MCB lower confidence bound equalling 0 can be inferred to be the true best, for it is better than every other treatment. Even if the MCB lower bound of a treatment is not zero, it is greater than $-\delta$ where $\delta$ is a small positive number, then that treatment may still be considered acceptable if a treatment within $\delta$ of the true best treatment can be considered acceptable. Conversely, if a smaller treatment effect is considered better, then PROC RSMCB computes confidence intervals for $\theta_i - \min_{j \neq i} \bar{Y}_j, i = 1, \ldots, k$, and their interpretation is reversed with respect to the upper and lower bounds. A more in-depth discussion of MCB confidence intervals with examples of their application can be found in User’s Guide to RS-MCB. For unbalanced designs, the expressions for MCB simultaneous confidence intervals, first given in Hsu (1985) and later reproduced in User’s Guide to RS-MCB, depend on the sample statistics, the sample size configuration, and a vector of critical values, which are evaluated as needed by PROC RSMCB at execution time.

Now to indicate more precisely the relationship between RS and MCB, take the case where a larger treatment effect is better. Then it has been shown that (Hsu 1984b) (i) treatments with zero MCB upper bounds are exactly those rejected by Subset Selection; (ii) the sample best treatment has a zero MCB lower bound if and only if it is selected as the true best by Indifference Zone Selection (modified to be applicable to single-stage experiments with unknown variance). Since the MCB confidence intervals are guaranteed to be correct with a probability of at least $1 - \alpha$, therefore MCB, Subset Selection, and Indifference Zone Selection, all three inferences are guaranteed to be correct simultaneously with a probability of at least $1 - \alpha$.

Remark. Based on Hsu (1984a), PROC RSMCB takes a somewhat novel approach to analyzing unbalanced designs. As shown in that article, for unbalanced designs, the ideal MCB inference and the associated RS decisions should be based on a vector of critical values which depends on the sample size configuration. The dimension of this vector is equal to $k$, the number of treatments, and thus is difficult to table in the conventional sense. But clearly this difficulty does not arise if, in the computer implementation, the vector of critical values is computed at execution time, which is the approach taken by PROC RSMCB.

4. Example

PROC RSMCB will be included in the Version 5 SUGI Supplemental Library. Exact specifications will be given in the forthcoming manual for that library, and a
write-up with more details and examples is available as Interfacing PROC RSMCB with the SAS® System. PROC RSMCB obtains the sample size, sample mean, and corrected sample sum of squares of each treatment from the current data set. Thus, prior to invoking PROC RSMCB, one can use a PROC SUMMARY followed by a subsetting DATA step to create a data set containing these quantities. Below is an example of such a PROC SUMMARY, DATA, PROC RSMCB sequence. In this example, seven brands of membrane filters are compared in terms of their ability to reveal fecal coliforms in river water. DATA FILTERS below contains the outcome of two experiments, in each of which three filters of each brand were randomly chosen, and 100ml of sample river water was poured through each filter. COUNT1 and COUNT2 were the number of colonies counted on each filter after 24 hours of incubation in a fecal coliform-selective medium, representing the number of bacteria revealed by each filter. The second experiment had some missing values due to colony blurring and spreading. As these filters are used by microbiologists to monitor the quality of drinking water, a brand having a larger mean count is better, as it is better in revealing harmful bacteria. We analyze the outcome of the first experiment only. In the analysis below, for simplicity, we have ignored the possibility that the variables should perhaps be modelled as Poisson rather than normal. The arcsine square-root transformation can be applied to stabilize the variance if one so desires. The SAS program for our example is as follows.

DATA FILTERS;
  INPUT BRAND COUNT1 COUNT2;
  CARDS;
  1 37 69
  1 36 122
  1 35 95
  2 26 118
  2 19 154
  2 9 102
  3 27 171
  3 25 132
  3 31 182
  4 33 122
  4 25 119
  4 26
  5 34 204
  5 23 225
  5 26 190
  6 40 140
  6 36 130
  6 37 127
  7 26 170
  7 33 165
  7 32
;

PROC SUMMARY DATA=FILTERS;
  CLASS BRAND;
  VAR COUNT1 COUNT2;
  OUTPUT OUT=ALL
  N=N1 N2
  MEAN=M1 M2
  CSS=CSS1 CSS2;
DATA STATS;
  SET ALL;
  IF TYPE =1;
PROC RSMCB-LARGEST ALPHA=.05;
  N N1;
  MEAN M1;
  CSS CSS1;

Printed output from PROC RSMCB includes:

1. definition of true best treatment, \( \alpha \), confidence level \((1-\alpha)\), \( S = (\text{MSE})^{1/2} \) and its degrees of freedom, sample means and sample sizes in the same order as they appeared in the data set

2. R- and S-values. Each treatment with an R-value less than the user's \( \alpha \), which is thus inferred to not be the true best, is indicated by an "*" in the REJECT column. If the sample best treatment has an S-value smaller than the user's \( \alpha \) and can thus be inferred to be the true best, then it will have an "*" in the SELECT column corresponding to it.

3. simultaneous 100(1-\( \alpha \))% MCB confidence intervals for the difference between each treatment and the true best of the other treatments. The quantities labelled LOWER and UPPER BOUND under TREATMENT MINUS BEST OF OTHER TREATMENTS are lower and upper confidence bounds for \( \theta_j - \max_{\ell \neq j} \theta_{\ell} \), \( j=1,...,k \), if the LARGEST option was specified or defaulted to. They are lower and upper confidence bounds for \( \theta_j - \min_{\ell \neq j} \theta_{\ell} \), \( j=1,...,k \), if the SMALLEST option was specified.

4. plots of simultaneous MCB confidence intervals.

Omitting the standard SAS Log, the SAS Listing for our example produced by PROC RSMCB is as follows.
RANKING, SELECTION, AND MULTIPLE COMPARISONS WITH THE BEST

TREATMENT WITH LARGEST EFFECT IS TRUE BEST TREATMENT

CONFIDENCE LEVEL = 0.9500  ALPHA = 0.0500

RS-MCB ANALYSIS FOR VARIABLE M1

\[ S = \sqrt{\text{MSE}} = 4.6496 \text{ WITH } 14.0000 \text{ DEGREES OF FREEDOM} \]

---

**REJECTION AND SELECTION**

<table>
<thead>
<tr>
<th>TREATMENT</th>
<th>SAMPLE MEAN</th>
<th>SAMPLE SIZE</th>
<th>R-VALUE</th>
<th>REJECT</th>
<th>S-VALUE</th>
<th>SELECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36.0000</td>
<td>3</td>
<td>0.7071</td>
<td></td>
<td>0.7071</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>18.0000</td>
<td>3</td>
<td>0.0004</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>27.6667</td>
<td>3</td>
<td>0.0417</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>28.0000</td>
<td>3</td>
<td>0.0488</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28.3333</td>
<td>3</td>
<td>0.0570</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>37.6667</td>
<td>3</td>
<td></td>
<td></td>
<td>0.7071</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>30.3333</td>
<td>3</td>
<td>0.1370</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**MULTIPLE COMPARISONS WITH THE BEST**

<table>
<thead>
<tr>
<th>TREATMENT</th>
<th>SAMPLE MEAN</th>
<th>SAMPLE SIZE</th>
<th>TREATMENT MINUS BEST OF OTHER TREATMENTS</th>
<th>LOWER BOUND</th>
<th>UPPER BOUND</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36.0000</td>
<td>3</td>
<td>-11.2793</td>
<td>7.9460</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>18.0000</td>
<td>3</td>
<td>-29.2793</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>27.6667</td>
<td>3</td>
<td>-19.6126</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>28.0000</td>
<td>3</td>
<td>-19.2793</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28.3333</td>
<td>3</td>
<td>-18.9460</td>
<td>0.2793</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>37.6667</td>
<td>3</td>
<td>-7.9460</td>
<td>11.2793</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>30.3333</td>
<td>3</td>
<td>-16.9460</td>
<td>2.2793</td>
<td></td>
</tr>
</tbody>
</table>

(*1.00 0)

**TREATMENT MINUS BEST OF OTHER TREATMENTS**

<table>
<thead>
<tr>
<th>TREATMENT</th>
<th>-30.0</th>
<th>-20.0</th>
<th>-10.0</th>
<th>0.0</th>
<th>10.0</th>
<th>20.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(+-----*-------)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(+-----*-------)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(+-----*-------)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(+-----*-------)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(+-----*-------)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>(+-----*-------)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>(+-----*-------)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

764
The LARGEST option in PROC RSMCB specifies that a larger treatment effect implies a better treatment. Treatments 2, 3, and 4 have R-values less than \( \alpha = 0.05 \). Thus, as indicated by the "*" in the REJECT column, they are rejected and inferred not to be the best. Note that the upper bounds on (treatment - best of other treatments) for these three treatments are zero, indicating that for each of these three treatments there exists some treatment better than it, agreeing with the conclusion given by the R-values.

Treatment 6, with a sample mean of 37.6667, is the sample best treatment. However, with an S-value of 0.7071, one cannot assert Treatment 6 to be the true best treatment at \( \alpha = 0.05 \), as indicated by the absence of an "*" in the SELECT column. Hypothetically speaking, had it been decided that a treatment within 10 of the true best treatment is acceptable, then Treatment 6 would have been found acceptable, for its lower bound on (treatment - best of other treatments) equals -7.9460, which is greater than -10.

A more detailed reference on RS-MCB analysis, particularly as it relates to other multiple comparisons methods, is \textit{User's Guide to RS-MCB}. Additional examples of using PROC RSMCB in SAS programs are given in \textit{Interfacing PROC RSMCB with the SAS System}. Both can be obtained by writing to the third author, at Department of Statistics, The Ohio State University, 1958 Neil Ave., Columbus, OH 43210.

Acknowledgement

SSG's research is supported by the Office of Naval Research Contract N00014-84-C-0167 at Purdue University. JCH's research was supported in part by Grant No. 31287, awarded by the National Institute of General Medical Sciences, DHHS, and by a one-year professional leave granted by the Ohio State University.

References


