Multiple Testing of General Contrasts: Truncated Closure and the Extended Shaffer-Royen Method

Peter H. Westfall, Texas Tech University
Randall D. Tobias, SAS Institute

Pairwise Comparisons

- ANOVA, $g = 10$ groups, $n = 2$ per group
- Averages: -1.9695, 2.1520, 5.5915, 0.8695, 4.6315, 0.2780, -1.3210, 0.3490, -7.1530, and 3.6220; MSE=1.051976
- $k = 45$ pairwise comparisons
- Bonferroni-Based
  - Bonferroni
  - Holm
  - Shaffer Method 2
- MaxT-Based
  - Tukey
  - Stepdown MaxT-based
  - Royen

SAS Code

```
proc glimmix data=allpairs;
  class g;  model y=g;
  lsmeans g/pdiff adjust=bonferroni;
  lsmeans g/pdiff adjust=bonferroni stepdown(type=free);
  lsmeans g/pdiff adjust=bonferroni stepdown(type=logical);
  lsmeans g/pdiff adjust=tukey;
  lsmeans g/pdiff adjust=simulate stepdown(type=free);
  lsmeans g/pdiff adjust=simulate stepdown(type=logical);
run;
```

Assumptions

- $\widehat{\boldsymbol{\mu}} \sim N_g(\boldsymbol{\mu}, \sigma^2 \boldsymbol{V})$, $\nu\widehat{\sigma}^2/\sigma^2 \sim \chi^2_\nu$, independent of $\widehat{\boldsymbol{\mu}}$
- $\boldsymbol{V}$ is known, positive definite
- $H_i : \boldsymbol{c}_i'\boldsymbol{\mu} = 0$ vs. $H_i' : \boldsymbol{c}_i'\boldsymbol{\mu} \neq 0$, $i = 1, \ldots, k$, for contrast vectors $\boldsymbol{c}_i$
- $T_i = |\boldsymbol{c}_i'\widehat{\boldsymbol{\mu}}|/(\widehat{\sigma}^2 \boldsymbol{c}_i' \boldsymbol{V} \boldsymbol{c}_i)^{1/2}$

Shaffer 2 Method (JASA 81, 826-831)

- Let $H_{(i)}$ correspond with $T_{(i)} = T_{a_i}$.
- Let $S = \{1, \ldots, k\}$. Define
  $\boldsymbol{S}_i = \{K \subset S \mid a_i \in K$ and $H_K \cap H'_{(k)} \cap \ldots \cap H'_{(k-i+2)} \neq \varnothing\}$.
- S2: starting with $i = k$, $H_{(i)}$ is rejected iff $H_{(k)}, \ldots, H_{(i+1)}$ are rejected and $p_{(i)} \leq \alpha/k_i$, where
  $$k_i = \max\{|K|; K \in \boldsymbol{S}_i\}.$$
- Adjusted p-values are $\widetilde{p}_{(i)}^{S2} = \min\{\max_{j \geq i}(k_j p_{(j)}), 1\}$
- S2 is uniformly more powerful tham Holm.

Extended Shaffer 2 Method (JASA 92, 299-306)

- Define $t_{ES}(\alpha, i) = \max\{t(\alpha, K) \mid K \in \boldsymbol{S}_i\}$, where
$$P_0\{\max_{j \in K} T_j \geq t(\alpha, K)\} = \alpha$$
- Starting with $i = k$, $H_{(i)}$ is rejected iff $H_{(k)}, \ldots, H_{(i+1)}$ are rejected and $T_{(i)} \geq t_{ES}(\alpha, i)$.
- Adjusted p-values are
$$\widetilde{p}_{(i)}^{ES2} = \max_{j \geq i}\{\max_{K \in \boldsymbol{S}_j} P_0(\max_{l \in K} T_l \geq t_{(j)})\}$$
- Uniformly more powerful than S2.

Example: All Pairs, *g*=5

i	1	2	3	4	5	6	7	8	9	10
$p_{(i)}$	.001	.001	.006	.010	.015	.020	.039	.060	.266	.392
H	H_{13}	H_{15}	H_{34}	H_{12}	H_{45}	H_{23}	H_{14}	H_{25}	H_{24}	H_{35}

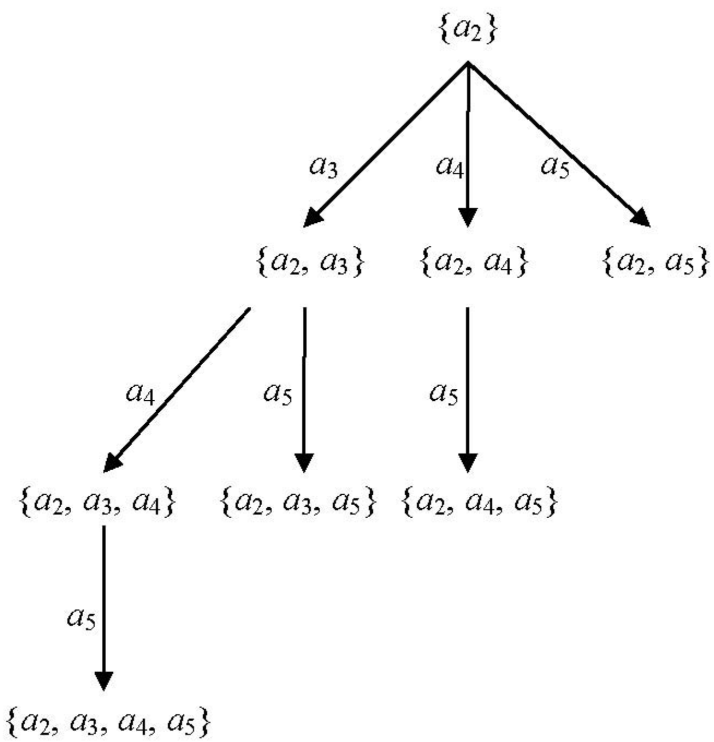
- $\boldsymbol{S}_1 = \{\{1,2,3,4,5,6,7,8,9,10\}\}$ $(k_1 = 10)$
- $\boldsymbol{S}_2 = \{\{2,4,5,7,8,9\},\{2,3,4,8\},\{2,5,6,7\},\{2,3,6,9\}\}$ $(k_2 = 6)$
- $\boldsymbol{S}_3 = \{\{3,4,5,10\},\{3,5,6,8,9,10\}\}$ $(k_3 = 6)$
- $\boldsymbol{S}_4 = \{\{4,7,9,10\},\{4,5\}\}$ $(k_4 = 4)$
- $\boldsymbol{S}_5 = \{\{5,6\},\{5,8,9\}\}$ $(k_5 = 3)$
- $\boldsymbol{S}_6 = \{\{6,7,8,10\}\}$ $(k_6 = 4)$
- $\boldsymbol{S}_7 = \{\{7,8\},\{7,10\}\}$ $(k_7 = 2)$
- $\boldsymbol{S}_8 = \{\{8\}\}$, $\boldsymbol{S}_9 = \{\{9,10\}\}$, $\boldsymbol{S}_{10} = \{\{10\}\}$

S2 and Closure Connection

- Westfall proved FWE control for ES2 directly, not using closure
- Rom and Holland, Hommel and Bernhard showed that S2 is more conservative than full closure with Bonferroni tests for all pairwise comparisons
- Royen noted closed methods are sometimes “nonmonotonic in p-values,” and developed a “truncated closed” procedure: closed testing follows the ordered p-values, stopping as soon as there is insignificance, with application to all pairwise comparisons in ANOVA.

New Results

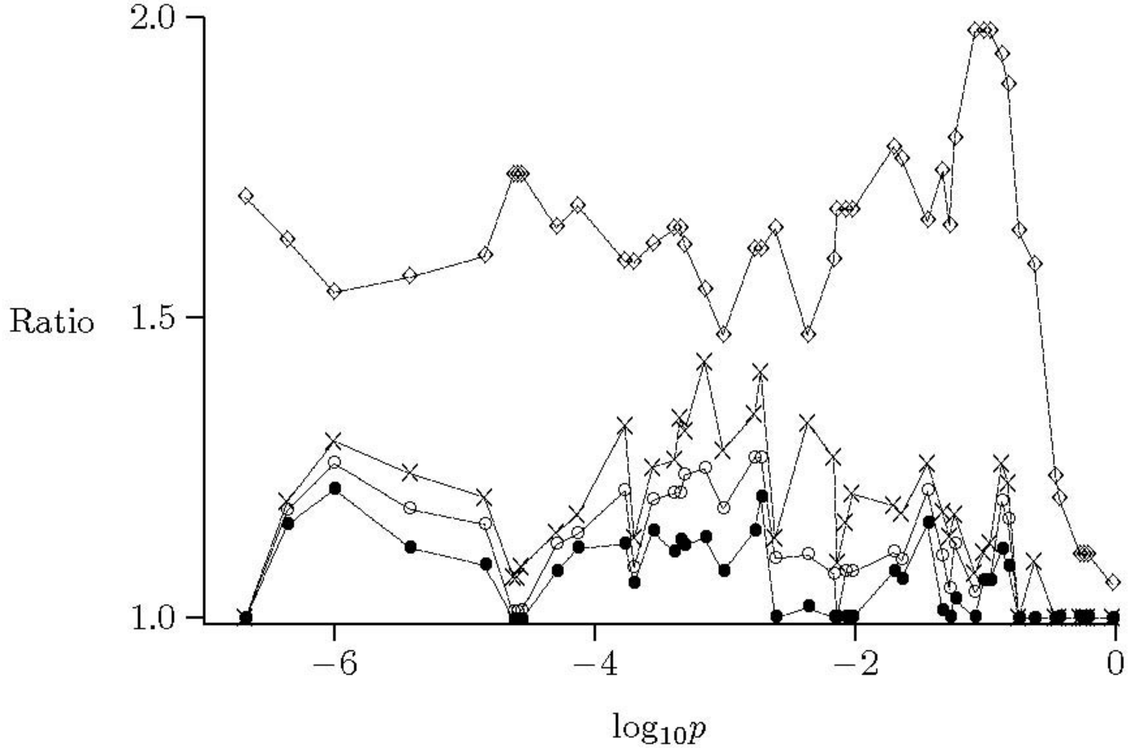
- Westfall and Tobias (a) extend Royen's truncated closure to general contrasts, (b) prove equivalence with ES2, and (c) develop a tree-based branch-and-bound algorithm to
  - prune the all-subsets tree, reducing $O(2^k)$ computational complexity
  - stop the search at early stages if the problem is computationally infeasible, with increasing power for increased search depth.
  - call the method ESR (extended Shaffer-Royen)



Tree representation of the search space

$S_{i+}^{(0)}$ Covering Sets

- $\boldsymbol{S}_{1+}^{(0)} = \{\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}\}$ $(k_{1+}^{(0)} = 10)$
- $\boldsymbol{S}_{2+}^{(0)} = \{\{2, 3, 4, 5, 6, 7, 8, 9\}\}$ $(k_{2+}^{(0)} = 8)$ $(9$ for free stepdown)
- $\boldsymbol{S}_{3+}^{(0)} = \{\{3, 4, 5, 6, 8, 9, 10\}\}$ $(k_{3+}^{(0)} = 7)$ (8 " " ")
- $\boldsymbol{S}_{4+}^{(0)} = \{\{4, 5, 7, 9, 10\}\}$ $(k_{4+}^{(0)} = 5)$ (7 " " ")
- $\boldsymbol{S}_{5+}^{(0)} = \{\{5, 6, 8, 9\}\}$ $(k_{5+}^{(0)} = 4)$
- $\boldsymbol{S}_{6+}^{(0)} = \{\{6, 7, 8, 10\}\}$ $(k_{6+}^{(0)} = 4)$
- $\boldsymbol{S}_{7+}^{(0)} = \{\{7, 8, 10\}\}$ $(k_{7+}^{(0)} = 3)$
- $\boldsymbol{S}_{8+}^{(0)} = \{\{8\}\}$, $\boldsymbol{S}_{1+}^{(0)} = \{\{9, 10\}\}$, $\boldsymbol{S}_{1+}^{(0)} = \{\{10\}\}$



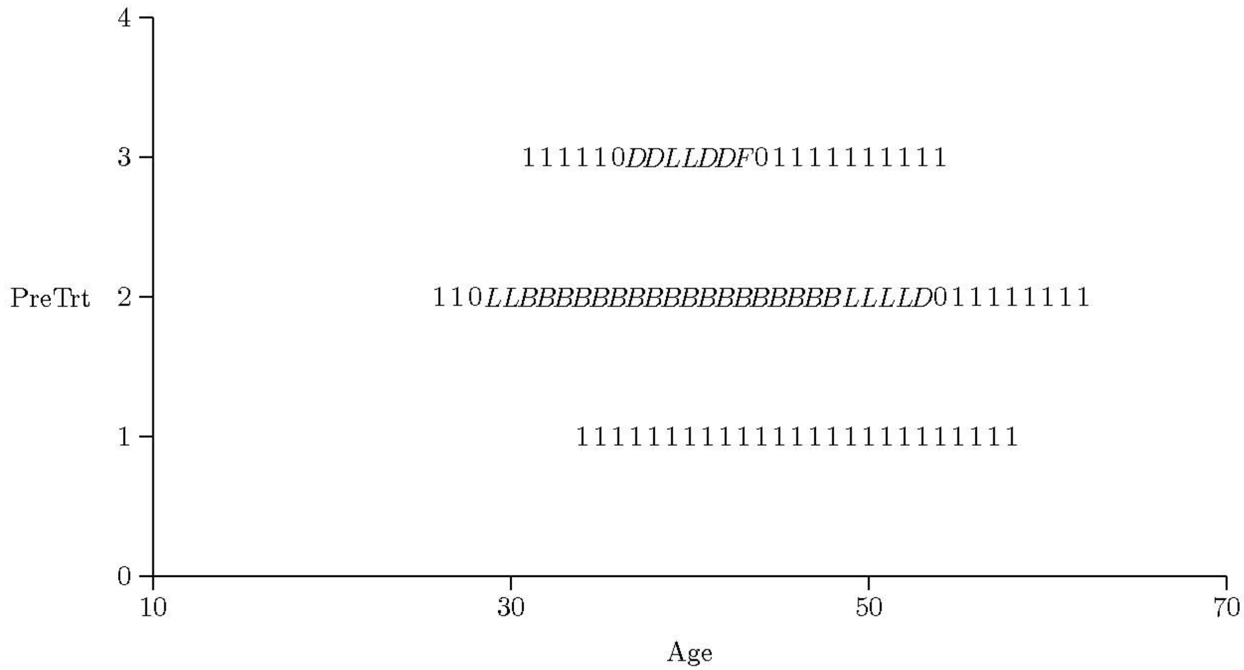
Ratios of adjusted p-values of S2 ($\Diamond$), $S_{i+}^{(0)}$ ($\times$), $S_{i+}^{(1)}$($\circ$), and $S_{i+}^{(2)}$ ($\bullet$) to truncated closure for $g = 10$, $k = 45$ example.

Response Surface Comparisons

- Compare health for Active and Placebo at various initial health (PreTrt) and Age values. Estimates:

$$\widehat{y}_A = 1.96204 + 0.42977 \cdot \textbf{PreTrt} - 0.05410 \cdot \textbf{Age}/100,$$
$$\widehat{y}_P = 1.54004 + 0.53557 \cdot \textbf{PreTrt} - 1.73130 \cdot \textbf{Age}/100.$$

- Comparisons made at (PreTrt, Age) $\in \{\{1,2,3,4,5\} \times \{10,11,\ldots,70\}\}$ ($k = 5 \times 61 = 305$).
- B & B search finds every element of $\boldsymbol{S}_i$ despite $k = 305$.
- SAS GLIMMIX with CONTRAST statement used for analysis



Significant (0.05 level simultaneous) differences using Benjamini-Hochberg FDR-controlling method (B), Liu et al.'s method (L), the discrete grid approximation to Liu et al. (D), free combination covering sets (F), $S_{i+}^{(0)}$, and $S_{i+}^{(1)} (= S_i$ here).

Timing Issues

- 26 minutes to find all relevant subsets in the response surface example (not $O(2^{305})$)
- Full method feasible for all pairwise comparisons with $g = 11$ ($k = 55$), $\boldsymbol{S}_{i+}^{(2)}$ is feasible with as many as 19 groups ($k = 171$), $\boldsymbol{S}_{i+}^{(1)}$ is feasible with as many as 32 groups ($k = 496$), and $\boldsymbol{S}_{i+}^{(0)}$ is feasible with as many as 80 groups ($k = 3160$).
- Donoghue developed an algorithm for S2 that is faster, but limited.

Conclusions

- ESR is a truncated closed testing procedure
- Branch-and-bound provides conservative, but computable approximations, that are usually much more powerful than standard alternatives.
- The method is fully computable in response surface comparisons, even with large numbers of comparisons.
- Available in PROC GLIMMIX for arbitrary contrasts and general model structures (approximate analyses used for more general models)
- General strategy: choose depth as large as compute resources allow and use the resulting multiplicity adjustment.

References

Donoghue, J. R. (2004), Implementing Shaffer's Multiple Comparison Procedure for a Large Number of Groups, in *Recent Developments in Multiple Comparison Procedures*, Institute of Mathematical Statistics Lecture Notes-Monograph Series, Vol. 47, Y. Benjamini, F. Bretz, and S. Sarkar, eds., 1–23.

Holm, S. (1979), A Simple Sequentially Rejective Multiple Test Procedure, *Scandinavian Journal of Statistics*, 6, 65–70.

Hommel, G., and Bernhard, G. (1999), Bonferroni Procedures for Logically Related Hypotheses, *Journal of Statistical Planning and Inference*, 82, 119–128. Royen, T. (1989), Generalized Maximum Range Tests for Pairwise Comparisons of Several Populations, *Biometrical Journal*, 31, 905–929.

Shaffer, J. P. (1986), Modified Sequentially Rejective Multiple Test Procedures, *Journal of the American Statistical Association*, 81.

826–831.

Rom, D. M., and Holland, B. (1995), A New Closed Multiple Testing Procedure for Hierarchical Families of Hypotheses, *Journal of Statistical Planning and Inference*, 46, 265–275.

Royen, T. (1989), Generalized Maximum Range Tests for Pairwise Comparisons of Several Populations, *Biometrical Journal*, 31, 905–929.

Shaffer, J. P. (1986), Modified Sequentially Rejective Multiple Test Procedures, *Journal of the American Statistical Association*, 81, 826–831.

Westfall, P.H. and Tobias, R.D. (2007). Multiple Testing of General Contrasts: Truncated Closure and the Extended Shaffer-Royen Method, *Journal of the American Statistical Association* 102: 487–494.