Shayle Searle's Contributions to the Evolution of the SAS System

Robert N. Rodriguez, Russell D. Wolfinger, Randall D. Tobias, Oliver Schabenberger
SAS Institute Inc., SAS Campus Drive, Cary, NC 27513 USA

Prepared for Conference in Honor of Shayle R. Searle on the Occasion of his Retirement, Cornell University, Ithaca, New York, August 9–10, 1996

Updated March 2005, for the 14th International Workshop on Matrices and Statistics, Massey University, Auckland

Abstract

For over twenty years, SAS statistical software has grown and evolved in response to the needs of its audience. One of the most influential members of this audience has been Shayle Searle. Our presentations acknowledge Shayle's many contributions to SAS users and to the developers of the linear models procedures in the SAS System.

Contributions to SAS Users (RNR)

This celebration of Shayle's career coincides with the 20th anniversary of SAS Institute, and so it is particularly appropriate to look back on his many contributions to the evolution of SAS software for linear models. Figure 1 provides a chronology of these activities.

Since the early 1970s, when Shayle's book *Linear Models* was published, there has been substantial growth in the SAS System for statistical analysis. The SAS72 manual, which was the only SAS documentation available at the time, contained 140 pages for statistical procedures, and it was first distributed to around 60 sites. The software itself consisted of around 35,000 lines of code. Today, the documentation for SAS/STAT software spans over 2800 pages, the code is measured in millions of lines, and the number of SAS users is estimated at over 3.5 million across 60 countries.

Likewise, the number of SAS employees has grown during this period. Figure 2 shows the building on Hillsborough Street in Raleigh, North Carolina, which was the Institute's home until 1980, when it moved to the present-day SAS campus in Cary.

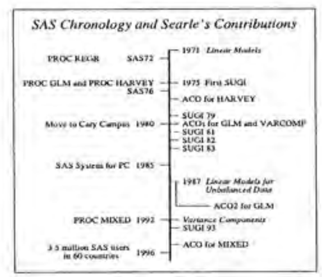

Figure 1: SAS Chronology and Searle's Contributions

Currently, 2100 employees are located at the Cary campus, which is shown in Figure 3, and the Institute has over 4000 employees worldwide.

Shayle was present at the first SAS Users Group International (SUGI) conference, which took place in Kissimmee, Florida, in 1975. Linear models and variance components analysis were topics of primary interest to the attendees. The GLM procedure was being developed by Dr. James Goodnight, the Institute president, and there

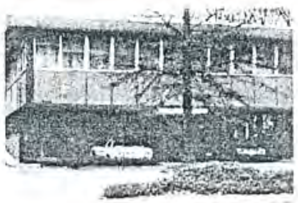

Figure 2: SAS Office Prior to 1980

was considerable discussion concerning the appropriate analysis of variance for unbalanced data. Both Shayle and Jim recall that the participants could not agree about which types of sums of squares the procedure should provide in its default output. In the end, Jim declared that he would include all of them. Although this resolved the discussion, it left open some important questions concerning the interpretation and applicability of the various sums of squares.

Fortunately, Shayle continued to think about these issues, and his expository presentations at subsequent SUGI conferences have provided valuable illumination for many SAS users. (Of course, many other individuals have also made significant suggestions and expository contributions in this area, as indicated in the Acknowledgments section of the *SAS/STAT User's Guide.*)

Figure 3: SAS Campus in 1996

I have not been able to locate any photographs taken during the Kissimmee meeting. However, Figure 4 shows Jim and Shayle resuming their discussion at the 1983 SUGI conference.

Figure 4: Jim Goodnight and Shayle Searle

Shayle was one of the earliest members of the statistics profession to recognize both the disadvantages and the advantages of the computer as a tool for statistical work. A typical starting point for Shayle's SUGI papers was the observation that "understanding something that has been computed is more of a problem than computing what we understand"; see Searle (1983). These papers addressed the need for better understanding of computer output, and they provided highly detailed explanations of linear models computations that made use of numerically simple examples, which are a hallmark of his expository writing.

Tutorials like these continue to be of benefit to SAS users, and so a brief review of Shayle's SUGI papers is appropriate at this point. His 1979 paper noted that users often compute F-statistics without first formulating a valid hypothesis. This backwards approach then leads to the question, "What hypothesis is it that I am testing?" This paper explains the derivation and use of the four types of estimable functions in the GLM procedure, and it clarifies the limitations and utility of GLM output for various situations involving the two-way crossed classification model with unbalanced data.

Shayle's 1981 paper dealt with quirks in linear model computations, special issues involving the HARVEY, GLM, and VARCOMP procedures, and the notion of "least squares means" (the beginning of a path pursued later in this presentation by Randy Tobias.) Shayle's 1982 paper (with G. F. S. Hudson) returned to the problem of understanding the hypotheses that are tested by the GLM

procedure, and it explored the nature and limitations of the hypotheses chosen for Type IV sums of squares.

Shayle's 1983 paper advocated broader use of analysis of covariance, and it covered situations involving several covariates, intraclass slopes of all forms, and unbalanced data. Shayle's 1993 paper discussed the history of analysis of variance computations, and it distilled his views on the analysis of unbalanced data which are presented at length in his 1987 book *Linear Models for Unbalanced Data.* This paper is particularly helpful for its recommendations concerning the use of the GLM procedure with unbalanced data, and it distinguishes between the all-cells-filled case and the some-cells-empty case.

Having attended all of Shayle's SUGI presentations, I can personally testify that they de-mystified difficult concepts and equipped attendees to make more effective use of the software. I recall returning from the 1983 conference to the General Motors Research Laboratories, where I worked at the time, and being asked on short notice to analyze a set of automobile accident data in response to a question posed by a top-level GM executive about the relative safety of American and Japanese cars. The analysis proved to be an immediate application of what I had learned from Shayle's talk!

SAS users have also benefited greatly from Shayle's annotated computer outputs (ACOs); see Figure 1. This series, which is available through the Biometrics Unit at Cornell, complements Shayle's books and our SAS manuals by providing detailed explanations of the output from various linear models procedures. The most recent addition is a 1995 ACO on the MIXED procedure.

During my years as a SAS user and developer, some of the lasting lessons that I have learned from Shayle are that

- users should not rely on algorithms that are detached from data specifics
- standard output can be hard to understand, and consequently labeling and terminology are critical. In software documentation, simple examples and detailed explanations are essential.
- statistical software needs to provide much more than statistical computing!

Shayle, thank you for all your lessons. I wish you a happy, productive retirement, and speaking on behalf of the many SAS users whom you have helped, I hope that you will continue to find opportunities to teach us.

Shayle Searle's Influence on a Young Statistician (RDW)

"The application of generalized inverse matrices to linear statistical models is of relatively recent occurrence." Thus begins Searle's classic 1971 text *Linear Models*, a text that also marked the beginning of my formal graduate training in statistics. As a quiet but confident new graduate student at NC State in 1985 with a B.S. in mathematics, I felt my previous training in linear algebra would be perfectly adequate to handle my initial coursework in linear statistical modeling. I was mistaken. Having never even heard of a generalized inverse, Shayle Searle had me scrambling with the very first sentence of his book.

Soon thereafter Searle defines a "...generalized inverse of a matrix A as any matrix G that satisfies the equation AGA = A." A little later we read that this definition "...does not define G as 'the' generalized inverse of A but as 'a' generalized inverse..." and we are then launched directly into a formal treatment of the various kinds generalized inverses and their properties. Such is the style of Searle, getting directly to the point at hand and applying it in as practical a way as possible. Extra descriptions and precise examples are included freely in attempt to make the point perfectly clear.

This attempt was successful in my case, but not without many hours of pondering and discussion with instructors and fellow students over such concepts as consistency, definiteness, quadratic forms, estimability, testability, sum-to-zero constraints, and last (but certainly not least!), the mixed model. Mine was a baptism by fire into the wonderful theory of linear models, and it has truly formed the foundation upon which most of my present statistical activity rests.

It has now been about 11 years since that intense first semester, and yet Searle's work continues to be influential. A case in point: In December of 1994, Bob, Randy, and I were asked by the journal *Statistics and Computing* to respond to a new article by John Nelder entitled "The statistics of linear models: back to basics." One quote from that article that certainly caught our attention was the following (with Greek symbols referring to a standard two-way ANOVA model):

> ...the Type III SS used in SAS (SAS Institute Inc., 1985) uses a quadratic form whose non-centrality parameter is symmetric in the α_i; this SS (i) loses power if used in any test, and (ii) is obtained by constraining the γ_{ij} margins to zero. It thus corresponds to an uninteresting hypothesis.

How to respond to such a criticism, considering that Type III tests form some of the principal default output from PROC GLM and PROC MIXED? Fortunately we had in hand *Linear Models for Unbalanced Data* (Searle 1987). It not only provides a thorough discussion of the four types of sums of squares (to which we refer in our reply), but Section 4.6 even goes so far as to refer to the Type III test for the two-way model as "..interesting, reasonable, useful..." In addition to this (and unbeknownst to us), Shayle composed his own reply to the article, including a nice discussion of the whole issue (see the June 1995 issue of *Statistics and Computing* for both replies).

Searle also makes the following remark in his reply to Nelder:

> And why...does 'model selection and prediction' have any connection whatever with trying 'to justify Type III sums of squares.'? I don't get it at all.

After a personal discussion with Nelder in June, it appears that Shayle has gotten to the heart of the issue with this query. The resolution is that Nelder is criticizing the use of Type III tests during the *model selection* phase of statistical analysis, whereas they may be appropriate during the *inference* phase once a model has been selected. As the dust is settling, we are grateful to have Searle's insight in these matters.

I had the privilege of meeting Shayle just before Nelder's article appeared as Shayle was working on a new Annotated Computer Output (ACO) about PROC MIXED. My wife Liz and I breakfasted with him one morning during the ASA meetings in San Francisco. We had barely begun discussing the ACO when Shayle surprised both us and our waiter by demonstrating what he referred to as "the proper way to serve tea." An adjoining remark was along the lines of "These #*%@ Americans never get it right." That taken care of, Shayle proceeded to provide me with a fully laundry list of suggestions regarding the output and labels from PROC MIXED. He was particularly critical of the output labeled "Mixed Model Equations", that not only had a mistake in it, but, as Shayle put it, "aren't a set of equations at all" because the final row of the output corresponded to the y-partition of the matrix that PROC MIXED sweeps. If you run PROC MIXED today with the MMEQ option you'll see that this row no longer appears.

A few years ago, we helped Shayle typeset the new ACO, which is now available from the Biometrics Unit at Cornell. During the course of this project, we ordered a complete set of the ACOs for reference. It was then I discovered, with much chagrin, the existence of the solutions manual to *Linear Models*. The chances are pretty good I would have emptied my somewhat meager bank account during that first semester of graduate school to obtain a copy of this book. Anyway, the ACOs have been very interesting, and since interacting with Shayle, I've been able to incorporate several more of his "suggestions" into PROC MIXED and plan to include more in the future. It's been quite a privilege working with him.

With the collective experience assembled in this room, I can't resist posing a question for discussion. It's one that has been recently sent to me by a colleague who has been running some interesting simulation studies. The question is this: Should variance component estimates in PROC MIXED be allowed to go negative by default? (They are currently constrained to be nonnegative by default but can be left unconstrained by using the NOBOUND option.) What piqued my interest in this question is that my colleague's simulation results show that the F-tests for fixed effects which are nested within a particular random effect are much too conservative in cases where this random effect's variance component is constrained to be nonnegative. In other words, PROC MIXED isn't printing true F-tests in this case. On the other hand, traditional precedents and model definitions argue against the unconstrained default.

Some of Shayle's thoughts on the problem are on pages 129-131 of *Variance Components* by Searle, Casella, and McCulloch (1992), included six possible courses of action to take when one encounters a negative ANOVA variance component. Shayle, what should PROC MIXED do by default?

Let me also say that *Variance Components* has been helpful to me in my recent work on the Bayesian analysis of variance component models. The exact expressions for the second derivatives of the likelihood were invaluable in confirming my calculations of Jeffrey's prior, and they even revealed an error in equation (5.2.10a) of Box and Tiao (1973).

Shayle, I hope your retirement is in name only, because your work tends to provide me with answers when I need them most; however, I seem to be lagging behind your ideas by about a decade. Thank you for all of the wonderful contributions you've made to statistics!

Recent Developments for Least Squares Means (RDT)

This section of our presentation describes some recent enhancements of the GLM and MIXED procedures for least squares means. Conceptually, the starting point for this

work is the definition for least-squares means provided by Searle, Speed, and Milliken (1980) and Searle (1981).

SAS Facilities for Comparing Independent Means

Multiple comparisons in the analysis of group means have long been included in the statistical tools of the SAS System, via the MEANS statement in the GLM procedure. The first multiple comparisons procedure (MCP) to be implemented was the Waller-Duncan Bayesian test (Waller and Duncan 1969). In the late 1970's many standard MCPs were added, including methods for all pairwise comparisons (Gabriel 1978, Hochberg 1974, Scheffe 1953, Sidak 1967,Tukey 1953), comparisons with a control (Dunnett 1955), and a variety of multiple stage tests (Duncan 1955, Einot and Gabriel 1975, Newman 1939, Ryan 1960, Welsch 1977). Each of the tests was requested by including an option in the MEANS statement, usually the name of the principal researcher responsible for the test. For example,

```
MEANS TMT / BON SIDAK SCHEFFE TUKEY
            GABRIEL GT2 DUNNETT
            DUNNETTU DUNNETTL SNK
            REGWQ REGWF WALLER
            and a few more ...;
```

In addition to the above facilities in the GLM procedure, the SAS Users Group Supplemental Library included the RSMCB procedure, contributed by Aubuchon, Gupta, and Hsu (1985), for ranking and selection and multiple comparisons with the best.

Limitations

The GLM procedure is a premier statistical tool for analyzing unbalanced designs (Milliken and Johnson 1984). However, MCPs usually do not apply in unbalanced cases because most MCPs require independent means with equal variances. The second part of this assumption implies that many of the procedures mentioned above are only defined for groups of equal size. The GLM procedure usually gets around this constraint by using the harmonic mean of the sample sizes when necessary for calculating critical values (Kramer 1956). The independence assumption is more fundamental. While the arithmetic means are always independent under the usual linear model assumptions, with unbalanced data they may not even estimate any model quantities of interest.

For example, consider the unbalanced two-way design given in Table 1. No matter how you look at it, this data exhibits a strong effect due to the columns (F-test $p < 0.0001$) and no significant interaction between rows and columns (F-test $p > 0.7$). But the lack of balance affects how the row effect is interpreted. In a main-effects-only model, there are no significant differences between the row means themselves (Type I F-test $p > 0.7$), but there are highly significant differences between the row means corrected for the column effects (Type III F-test $p < 0.0001$).

17, 28, 19, 21, 19	43, 30, 39, 44, 44	16
21, 21, 24, 25	39, 45, 42, 47	19, 22, 16
22, 30, 33, 31	46	26, 31, 26, 33, 29, 25

Table 1. Unbalanced Two-Way Design

The *least squares means* (LS-means) for rows are in effect the column-corrected row means. In the same way that the Type I F-test tests for differences between the arithmetic row means, the Type III F-test assesses differences between the LS-means; see Table 2. This table clarifies the difference between the results of the Type I and Type III tests.

The LS-means are defined in the SAS/STAT documentation as "the expected value of ... means that you would expect for a balanced design." A much clearer and more explicit definition is given by Searle, Speed, and Milliken (1980): estimates of population marginal means (PMM's) for a balanced population. This definition makes a clear distinction between the parametric functions being estimated and the statistics used to estimate them.

Row	Arithmetic Mean	Least-squares Mean
1	29.1	26.1
2	29.2	28.4
3	30.2	35.5

Table 2. Means for Unbalanced Two-Way Design

It is clear that multiple comparisons on the arithmetic row means in this case would be misleading. The row LS-means themselves are natural candidates for multiple comparisons, but for unbalanced designs such as the one above, they are correlated, and few of the standard MCPs apply to correlated means. Even in balanced designs, LS-means may be correlated if there are covariates. Furthermore, note that all of the above assumes a general *fixed-effects* linear model with independent and identically distributed (IID) noise. In the context of *mixed models*, where the noise is no longer assumed to be IID, the problem of correlated estimates comes up even when only arithmetic means are being considered.

The GLM procedure is often used to analyze highly unbalanced data. The fact that it offered multiple comparisons only of group averages was a severe incongruity. For example, Hsu (1989) discussed "the misleading inference given by the MEANS option of SAS PROC GLM in general linear models ...". What was called for was multiple comparisons for correlated quantities that would cover LS-means in both mixed and unbalanced fixed-effect models.

A Unified Theory of Multiple Comparisons

Eventually I did find a sufficiently general methodology for multiple comparisons, due mainly to Jason Hsu at Ohio State University. The set-up is described in Hsu (1996): Given estimates $\hat{\theta}_i$, $i = 1, \ldots, k$ of parameters θ_i, $i = 1, \ldots, k$, it is desired to make inferences about a *fixed* subset S of contrasts between the parameters. The variance-covariance matrix $\sigma^2 C$ for the $\hat{\theta}_i$ is assumed to be known up to a constant, for which an independent estimate $\hat{\sigma}^2$ is available. Inference for an individual difference $l'\theta$ (expressed as a linear combination of the parameters) is based on the standardized estimated linear combination

$$t_l = \hat{\sigma}^{-1}\frac{l'(\hat{\theta}-\theta)}{\sqrt{l'Cl}},$$

in such a way that the over-all error rate is controlled. The error rate is

$$P(\max_{l \in S} t_l > d_1^S) \quad \text{and} \quad P(\max_{l \in S} |t_l| > d_2^S) \qquad (1)$$

for one-sided and two-sided inference, respectively. Hypothesis testing calls for calculating the error rate for the observed value of $\max_{l \in S} t_l$, while confidence intervals require fixing the probability and solving for d_1^S or d_2^S, as appropriate.

Let L be the matrix whose rows are the linear combinations in S. Assuming Gaussian noise, the t_l's collectively have a multivariate (central) T distribution. The probabilities in (1) usually must be computed numerically. The full integration requires a $|S|$-fold multiple integral in general, which is infeasible for $|S| > 3$ or so. However, when $L'CL$ has a special form called *one-way structure* by Hsu (1996), the multiple integral collapses to a double integral. Examples of situations with one-way structure are

- when the $\hat{\theta}_i$ are arithmetic means of independent groups (that is, a one-way analysis)
- when S consists of differences between one "control" group and the rest (Dunnett's test)
- when S consists of all (ordered) pairwise differences, and all groups are the same size (Tukey's test)

If numerical computation of (1) is not feasible, various bounds (Bonferroni, Sidak 1967, Worsley 1982) or approximations (Hsu 1996) can be employed.

Thus, by distinguishing between what needs to be computed (the probabilities (1) for various sets of contrasts S) and how it is to be computed (exactly or approximately or conservatively), Hsu provides a framework for a general computational tool. Our implementation matched this distinction; one option is used to specify the general class of MCP to perform (all pairs *versus* comparisons with a control), and another specifies the computational approach.

There were two primary candidates for these facilities: the GLM procedure and the newer MIXED procedure for general linear modeling with non-trivial error structures. The latter case is complicated by the fact that more than one component of the covariance matrix must be estimated. We made the assumption that the covariance was known up to a χ^2-distributed constant (possibly with non-integer degrees of freedom), primarily because there seemed no other way to proceed in general. Simulations indicate that MCP inference made under this assumption has characteristics similar to that for mixed model F-tests based on asymptotics, although a more complete study is in order.

Extensions

While we were working on LS-means we made a number of other enhancements. Our starting point was the Searle, Speed, and Millikin definition of LS-means as the marginal means over a balanced population. As such, LS-means may not be useful if "a balanced population" is unreasonable in a particular application. A new option allows users to specify the population over which the expected marginal means are to be computed. Similarly, covariates had been handled by setting them to their mean values in the analysis dataset, and a new option allows users to set their values directly. A third new option allows users to test for "simple effects" of one or more factors in an interaction. None of these enhancements are directly related to MCP, except that all of the new features taken together go a long way toward making it possible to approach the entire analysis of unbalanced designs from the point of view of LS-means.

Multiple-comparisons-with-the-best or MCB is a type of inference connected with bioequivalence testing, subset selection, and indifference zone selection (Hsu 1996). While not directly covered by the MCP enhancements for LS-means mentioned above, it is related to Dunnett's and Tukey's tests. Therefore, one can write a SAS program which uses the results of these tests to construct MCB confidence intervals and p-values. A macro program to do this in general has been added to the SAS sample library, making this important type of inference generally available.

Conclusion

The addition of multiple comparisons for correlated quantities is a significant enhancement to SAS statistical software. It not only provides multiple comparisons for the full range of linear models that can be fitted with the GLM procedure, but it also leads to an approach to multiple comparisons for mixed models with non-trivial error structures. Directions for further development include

- adding more fundamental types of comparisons, in addition to all pairs and comparisons with a control, and perhaps even a facility for specifying arbitrary comparisons of the model parameters
- adding facilities for multiple-stage tests of correlated means
- providing more techniques that build on the basic computations already available, as do the macros for MCB

This experience demonstrates that a fast way of getting a new statistical technique into commercial software is to make it general enough to convince the developer that he won't have to do the job all over again very soon. The work of Searle, Speed, and Millikin (1980) defining LS-means and the work of Hsu (1996) characterizing multiple comparisons in general were crucial for this development.

2005 Addendum: Further Development for Least Squares Means

The nine years since the 1996 Searle Conference have seen continued development for analysis of least-squares means in SAS. We have found LS-means to be a fundamental tool for asking and answering statistical questions not only about the linear models for which Searle, Speed, and Milliken (1980) and Searle (1981) originally proposed them, but also for mixed, generalized linear, and generalized linear mixed models. We take our cue for this update from the "directions for further development" we discussed in 1996.

Note that all of the enhancements discussed here are implemented in the new GLIMMIX procedure for generalized linear mixed models, available for download to users of SAS 9.1 at `support.sas.com`. In generalized and generalized linear mixed models, LS-means play a pivotal role in estimation and inference. Because of the nonlinearity of the link function, estimating and testing linear combinations of parameter estimates is only meaningful on the linked scale, for which the model effects are assumed to be additive. Because expectations and variances are related in such models, and because data are marginally correlated in mixed models, arithmetic means on the linear scale have no use. In this context, LS-means are crucial to formulating appropriate hypotheses about group comparisons. Simple transformations of these contrasts produce other important inferential quantities, for example, the exponentiation of LS-mean differences in models with logit link produces odds ratios.

More Types of Comparisons

Analysis of Means

In addition to all-pairwise and one- and two-tailed control comparisons, we added the DIFF=ANOM option to the LSMEANS statement of PROC GLIMMIX, for comparing each LS-mean with the average of the LS-means. In the context of one-way analysis of variance, these comparisons, suitably adjusted for multiplicity, are the basis of the *analysis of means* (Ott, 1967). An analysis of means is usually condensed into a graphical display that represents each mean by how much it differs from the overall mean, with decision limits that say whether this difference is significant. Such a plot is also available in PROC GLIMMIX; Figure 5 gives an example of an Analysis of Means plot for the comparison of 16 plant varieties (entries) in a randomized block design where the outcome variable has a binomial distribution.

```
proc glimmix;
   class block entry;
   model y/n = entry / dist=binomial;
   random block;
   lsmeans entry / plot=anomplot
                   adjust=nelson;
run;
```

The decision limits are computed using the distribution described by Nelson (1982), using a factor-analytic covariance approximation described in Hsu (1992). With

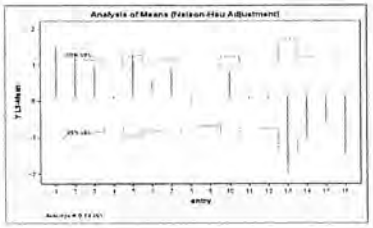

Figure 5: Analysis of Means for Binomial Data

Gaussian data the upper and lower decision limits would be horizontal lines, since the design is balanced. Because the variability of the binomial distribution is tied to its mean, for this data the decision limits vary with the magnitude of the LS-means, here shown on the logistic scale.

Arbitrary LS-mean Contrasts

Besides the "canned" comparison types available via the LSMEANS statement, we added an LSMESTIMATE statement to PROC GLIMMIX, to provide for custom hypothesis tests among the least-squares means. In contrast to the hypotheses tested with the ESTIMATE or CONTRAST statements, the LSMESTIMATE statement allows you to estimate arbitrary linear combinations of the least-squares means. Multiple-row sets of coefficients are permitted, along with multiplicity adjustments for the associated *p*-values and confidence intervals.

Multiple-Stage Tests

The multiplicity adjustments for LS-means comparisons which we reported in 1996 were all *single-stage* procedures. They have the advantage that they yield simultaneous confidence intervals for all comparisons. However, by sacrificing the ability to get simultaneous confidence intervals, you can use multiple-stage tests to improve power while still protecting the familywise error rate. We have added the STEPDOWN option for all multiplicity adjustments to do just that, implementing the ideas of Shaffer (1986) and Westfall (1997).

Building on Basic Computations

In 1999, we published a book on multiple comparisons and multiple tests in SAS (Westfall et al. 1999). As planned, this book was going to cover the traditional MEANS comparisons, plus some home-grown macros for special tests. However, it soon became apparent that the versatility of LS-means for handling unbalanced and/or correlated data made them the right basic tool to use. Using LS-means in conjunction with SAS/IML matrix programming and the (then) new ODS output delivery system, we published very general macros for computing single-stage and multistage adjusted LS-mean comparisons (`%SimIntervals`, `%SimTests`) and for assessing the power of multiple comparisons procedures (`%SimPower`).

References

Aubuchon, J.C., Gupta, S.S., and Hsu, J.C. (1985), "Interfacing PROC RSMCB with the SAS System," Technical Report 313, Department of Statistics, Ohio State University, Columbus, OH.

Box, G. E. P. and Tiao, G. (1973), *Bayesian Inference in Statistical Analysis*, John Wiley & Sons, New York.

Duncan, D.B. (1955), "Multiple Range and Multiple F Tests," *Biometrics*, 11, 1-42.

Dunnett, C.W. (1955), "A Multiple Comparisons Procedure for Comparing Several Treatments with a Control," *Journal of the American Statistical Association*, 50, 1096-1121.

Einot, I. and Gabriel, K.R. (1975), "A Study of the Powers of Several Methods of Multiple Comparisons," *Journal of the American Statistical Association*, 70, 351.

Gabriel, K.R. (1978), "A Simple Method of Multiple Comparisons of Means," *Journal of the American Statistical Association*, 73, 364.

Hochberg, Y. (1974), "Some Conservative Generalizations of the T-method in Simultaneous Inference," *Journal of Multivariate Analysis*, 4, 224-234.

Hsu, J.C. (1989), "Multiple Comparison Procedures," American Statistical Association Tutorial Notes.

Hsu, J. C. (1992), "The Factor Analytic Approach to Simultaneous Inference in the General Linear Model," *Journal of Computational and Graphical Statistics*, 1, 151-168.

Hsu, J.C. (1996), *Multiple Comparisons: Theory and Methods*, Chapman and Hall.

Hudson, G. F. S. and Searle, S. R. (1982), "Hypothesis Testing with Type IV Sums of Squares of the Computer Routine SAS GLM", *Proceedings of the Seventh Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC, 521-527.

Kramer, C.Y. (1956), "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications," *Biometrics*, 12, 307-310.

Milliken, G.A. and Johnson, D.E. (1984). *Analysis of Messy Data*, Chapman and Hall.

Nelson, P. R. (1982), "Exact Critical Points for the Analysis of Means," *Communications in Statistics*, 11, 699-709.

Newman, D. (1939), "The Distribution of the Range in Samples from a Normal Population," *Biometrika*, 31, 20-30.

Ott, E. R. (1967), "Analysis of Means—A Graphical Procedure," *Industrial Quality Control*, 101-109. Reprinted in *Journal of Quality Technology*, 15 (1983), 10-18.

Ryan, T.A. (1960), "Multiple Comparisons in Psychological Research," *Psychological Bulletin*, 57, 318-328.

Scheffe, H. (1953), "A Method for Judging All Contrasts in the Analysis of Variance," *Biometrika*, 40, 87-104.

Searle, S. R. (1971), *Linear Models*, John Wiley & Sons, New York.

Searle, S. R. (1979), "Relationships between the estimable functions of SAS GLM output for unbalanced data and the hypotheses tested by traditional-style F-statistics," *Proceedings of the Fourth Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC. 196-207.

Searle, S. R. (1981), "Quirks in Linear Model Calculations with Unbalanced Data", *Proceedings of the Sixth Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC. 38-45.

Searle, S. R. (1983), "Analysis of Covariance Using SAS GLM", *Proceedings of the Eighth Annual SAS Users Group International Conference*, SAS Institute Inc., Cary. NC. 747-759.

Searle, S. R. (1987), *Linear Models for Unbalanced Data*, John Wiley & Sons, New York.

Searle, S. R. (1993), "Analysis-of-Variance Computing: Then and Now, With Reference to Unbalanced Data", *Proceedings of the Eighteenth Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC. 1077-1088.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, John Wiley & Sons, New York.

Searle, S.R., Speed, F.M., and Millikin, G.A. (1980), "Population Marginal Means in the Linear Model: An Alternative to Least Squares Means," *The American Statistician*, 34, 216-221.

Shaffer, J. P. (1986), "Modified Sequentially Rejective Multiple Test Procedures," *Journal of the American Statistical Association*, 81:329-335.

Sidak, Z. (1967), "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions," *Journal of the American Statistical Association*, 62, 626-633.

Tukey, J.W. (1953), "The Problem of Multiple Comparisons," Unpublished manuscript.

Waller, R.A. and Duncan, D.B. (1972), "Computations of Bayesian *t*-Values for Multiple Comparisons," *Journal of Statistical Computation and Simulation*, 75, 169-172.

Welsch, R. E. (1977), "Stepwise Multiple Comparison Procedures," *Journal of the American Statistical Association*, 72, 359.

Westfall, P. H. (1997), "Multiple Testing of General Contrasts Using Logical Constraints and Correlations," *Journal of the American Statistical Association*, 92:299-306.

Westfall, P. H., Tobias, R. D., Rom, D., Wolfinger, R. D., and Hochberg, Y. (1999), *Multiple Comparisons and Multiple Tests Using the SAS System*, SAS Institute Inc., Cary, NC.

Worsley, K. J. (1982), "An Improved Bonferroni Inequality," *Biometrika*, 69, 297-302.