QUALITY AND RELIABILITY ENGINEERING INTERNATIONAL
Qual. Reliab. Engng. Int. 2000; **16**: 423–433

SUMS OF SQUARES AND EXPECTED MEAN SQUARES IN SAS

MICHAEL F. DRISCOLL* AND CONNIE M. BORROR
Arizona State University, Tempe, Arizona, USA

SUMMARY

The four different types of sums of squares available in SAS are considered, and a broad overview is given of how the similarities and dissimilarities between them depend upon the structure of the data being analyzed (for example, on the presence of empty cells). The fixed-effect hypotheses tested by these sums of squares are discussed, as are the expected mean squares computed by SAS procedure GLM.

Primary attention is given to linear models for the analysis of variance. Only two-factor analysis of variance models are explicitly considered, since they are complex enough to illustrate the most important points. Numerical examples are included. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: linear models; tests of hypotheses; SAS procedure GLM

1. INTRODUCTION

Four different types of sums of squares are available in SAS statistical procedures. Which type is appropriate in a given study depends in large part on the structure of the data being analyzed. Practitioners worldwide recognize the analytical and conceptual benefits that accrue to studies with the same number of observations in each cell. But time, money, and other considerations often dictate the use of unequal cell counts. Botched runs or mis-recorded data sometimes result in data whose cell counts are not even proportional or, worse, in data with empty cells. Empty cells may even result by design rather than by happenstance, as when certain treatment combinations are recognized from the start as being inherently infeasible or simply too dangerous to pursue.

As the relationships between cell counts become more complex, the choice of analysis becomes less straightforward. The difficulty of making an informed choice is exacerbated by textbook and computer-manual coverage that is either too sketchy to be helpful or too thorough to be readily understandable. We attempt a middle ground, by providing an overview that shows how the similarities and dissimilarities of the four types of sums of squares available in SAS depend upon the structure of the data being analyzed.

These four types are used, for example, by the SAS linear model procedures REG and GLM, which are described in several sources [1–3]. The type III sums of squares are also used by the linear model procedure MIXED, which is newer [4–7]. Although other authors have done so ([8], for example), we make no comparisons between the sums of squares computed by SAS and those computed by its competitors.

1.1. Terminology

The abbreviation SS will usually be used in place of the words *sum(s) of squares* and MS in place of *mean square(s)*. Also, a SS will be said to *test* a hypothesis. The allusion here is to the null hypothesis tested using the F-ratio obtained by converting the SS into an MS and then dividing it by another MS (typically *MSE*, the mean square for error).

The extra sums of squares notation SS(*more*|*only*) will be used to denote the difference

$$\mathrm{SS}(more|only) = \mathrm{SSE}(only) - \mathrm{SSE}(only, more)$$

in error sums of squares between a full model (having the *only* and the *more* parameters) and a reduced model (having just the *only* parameters). Extra SS are often called reduction sums of squares, because they measure the reduction obtained in SSE by entering the *more* parameters into a model that previously contained just the *only* parameters.

1.2. The model

The specific model considered in Sections 2 through 6 and 8 is the fixed-effects complete factorial

*Correspondence to: Michael F. Driscoll, Department of Mathematics, Arizona State University, PO Box 871804, Tempe, Arizona 85287-1804, USA.

two-factor analysis of variance model with interaction, that is,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

where $i = 1, \ldots, a$, $j = 1, \ldots, b$, and $k = 1, \ldots, n_{ij}$. (Mixed models are considered in Sections 7 and 9.) The symbols μ, α, β, and $\alpha\beta$ will be used to refer respectively to the parameters for the overall effect, the factor A main effects, the factor B main effects, and the interaction effects. The unqualified term *effects* will be used to refer to any or all of these.

Nonetheless, the presentation is designed to address not just this two-factor model, but more complicated models as well. On a first reading, the material might be taken as referring only to this two-factor model (in which case the phrase 'highest-order interactions' would refer only to the $\alpha\beta$'s). On a more general reading, that phrase would refer to second-order interactions $\alpha\beta\gamma$, or third-order interactions $\alpha\beta\gamma\delta$, and so on.

For models like the two-factor model under direct consideration here, it is very common to impose on the parameters the restrictions that the α's sum to zero, that the β's sum to zero, and that the $\alpha\beta$'s sum to zero in every row and column. Indeed, these restrictions are so commonly made that they have come to be called the *usual assumptions*. Here, such restrictions are first used in Section 7.

Numerical examples, including an extended one that illustrates all the main points discussed, are given in Section 8.

2. TYPE I SUMS OF SQUARES

The fundamental decomposition SST = SSR + SSE is the only decomposition that holds in every linear model. It expresses the total SS as the sum of the model (or, regression) SS and the error (or, residual) SS. (A decomposition is often called an orthogonal decomposition, because the SS constituting a decomposition have a geometric interpretation involving perpendicularity. Such SS are also stochastically independent if the errors in the model comprise a random sample from a normal distribution.)

Type I SS, the so-called *sequential* SS, form a decomposition of the model sum of squares SSR. The type I SS are the only type of SS that always have this decomposition property. There are, however, certain rules for the sequencing: for example, interactions in factorial models should be entered only after the corresponding main effects, and factors in nested models should be entered in the order of their nesting.

These sequential SS are basic in another sense: they can be used to provide specific formulas for SS of other types, as will be seen shortly for type II. For types III and IV, the process requires reformulating analysis of variance models as regression models (a task left unpursued here).

For a two-factor model using *balanced* or *proportional* data (that is, data having all cell counts equal or, at least, proportional across rows and columns), the order in which α and β are entered is not important, since in either case the type I SS are

$$\mathrm{SS}(\alpha|\mu) = \mathrm{SS}(\alpha|\mu, \beta)$$
$$\mathrm{SS}(\beta|\mu) = \mathrm{SS}(\beta|\mu, \alpha)$$
$$\mathrm{SS}(\alpha\beta|\mu, \alpha, \beta)$$

But, even in this desirable situation, the hypotheses tested by $\mathrm{SS}(\alpha|\mu)$ and $\mathrm{SS}(\beta|\mu)$ are contaminated with interaction parameters.

The allusion here is to the estimable parameter functions needed to specify the hypothesis tested by a particular SS. The term *estimable parameter function* refers to any linear combination of the parameters whose value can be estimated unbiasedly using some linear combination of the response values Y_{ijk} [9, Section 8.4; 10, Section 5.4]. In a regression (or full-rank) model, all linear combinations of the parameters are estimable. But any model of less than full rank has some non-estimable parameter functions.

For example, the hypothesis tested by $\mathrm{SS}(\alpha|\mu)$ in the present two-factor model can be expressed only in terms of estimable parameter functions that involve the interactions $\alpha\beta$ as well as the main effects α. Briefly said: the hypothesis tested by $\mathrm{SS}(\alpha|\mu)$ is contaminated by interaction effects. (More detail is given in Sections 6 and 7.)

For non-proportional data, the order of effect specification is important. (Of all the types, type I SS depend most heavily on the order of effect specification.) Each effect is adjusted only for those effects previously entered, as in the decomposition

$$\mathrm{SS}(\beta|\mu)$$
$$\mathrm{SS}(\alpha|\mu, \beta)$$
$$\mathrm{SS}(\alpha\beta|\mu, \beta, \alpha)$$

of SSR, produced by the statement 'MODEL Y = B A A*B / SS1 ;' in procedure GLM. (The types of SS used by GLM are controlled via options of the form SS1, SS2, SS3, and SS4, the defaults being SS1 and SS3.)

For unbalanced data—whether proportional or not!—the hypotheses tested by these type I SS involve the cell counts, and they also depend on the order in which the effects are entered into the model [11, Section 4.9.h], as can be verified using the process outlined in Section 6. And, although there is no contamination from effects specified earlier in the sequence, hypotheses tested are often contaminated by effects specified later in the sequence. In particular, the hypothesis tested by the sequential sum of squares $SS(\beta|\mu)$ from the decomposition last displayed is contaminated by α effects.

Since type I SS are appropriate primarily for balanced, proportional, or purely nested models—with effects that are specified in an appropriate sequence—other types of SS are needed.

3. TYPE II SUMS OF SQUARES

Type II SS are sometimes called *partial* SS. Like the type I SS, they are readily described using the extra sums of squares principle. And they agree with the type I SS for balanced or proportional data.

Each type II SS pertains to a given effect (that is, to a certain set of main effects or of interactions) after all other effects have been added to the model, except those that contain the given effect. (The idea of containment is easily illustrated: an interaction effect contains the corresponding main effects and a nested effect contains any effect in which it is nested.) For the specific model being considered here, the type II SS consist of

$$\begin{gathered} SS(\alpha|\mu,\beta) \\ SS(\beta|\mu,\alpha) \\ SS(\alpha\beta|\mu,\alpha,\beta) \end{gathered}$$

which is *not*, in general, a decomposition of the model sum of squares SSR.

The hypothesis tested by a given type II SS is contaminated by containing effects, even for balanced data, but it will not involve parameters from non-containing effects. Therefore, the hypothesis tested by a type II SS for an effect not contained in any other effect will involve only parameters associated with the effect in question. This latter result holds, in particular, for the highest-order interactions in a model.

For unbalanced data, the hypotheses tested by type II SS involve the cell counts. Even for effects not contained in other effects, these hypotheses are not the same as those that would be tested by the type II (or, equivalently, the type I) SS if the data were balanced. The reason is that, for balanced data, the common cell counts factor out of the relevant estimable parameter functions. (A careful demonstration of this fact requires derivations of the kinds discussed in Sections 6 and 7.)

4. TYPE III AND TYPE IV SUMS OF SQUARES

If the data are balanced, the SAS type III and type IV SS both coincide with types I and II. If not, the SAS type III and type IV sums of squares are not conveniently described using extra SS, since doing so requires that the model be explicitly reformulated as a regression (or full-rank) model. Readers interested in pursuing this issue should begin with the several 'Regression Approach to . . . ' sections of [12]. The more theoretical treatment in [13] uses cell-mean models and Moore–Penrose inverses, while supplying some numerical examples.

The primary rationale behind types III and IV is that the hypotheses they test are the same for all models having the same set of estimable parameter functions. Consequently, the hypotheses tested do not depend on the actual values of *positive* cell counts. If *all* the cell counts are positive, then the type III and type IV SS are the same, and the hypotheses they test are the hypotheses that would be tested by the type I SS were the data balanced. Still, the hypothesis tested by a given type III or type IV SS involves parameters from any containing effect (but is not otherwise contaminated).

What distinguish the type III and type IV SS are their secondary rationales in the presence of empty cells. The distinction is somewhat technical. Type III SS are defined so that the SS for a given effect is orthogonal to the SS for any containing effect, thus providing a partial decomposition property. On the other hand, type IV SS are designed to exhibit a certain uniformity, namely that the coefficient weights in the estimable parameter functions relevant to the effect are equally spread over the parameters for the highest-order containing effects (if any) that correspond to non-empty cells.

Even when there are empty cells, the type III and type IV SS differ only for those effects that are contained in other effects. But, in general, a type III or type IV SS and the hypothesis it tests depend heavily on any empty cells (even if the data are otherwise balanced or proportional), since the presence of empty cells changes the set of estimable parameter functions. The situation is so complex that type IV SS and the corresponding hypotheses are sometimes not uniquely determined. (When this occurs, SAS gives an appropriate warning message.) It may then be necessary to specify the desired hypothesis in detail, by using a CONTRAST statement.

The type III SS are widely used, as they should be, although some [14] would disagree. But there is some evidence that they are inappropriate for certain kinds of nesting [15].

5. SUMMARY OF THE FOUR TYPES

The differences and similarities between the four types of SS can be summarized in various ways. It is convenient to begin by separating the discussion of regression models from that of over-parameterized models, that is, of models having full rank from that of models with less than full rank.

5.1. Regression models

For regression models, it is easy to summarize relationships between the four types of SS: types II, III, and IV are all the same. Consequently, procedure REG reports only types I and II SS. The information that REG reports using type I SS depends upon the specific order in which the predictor variables are listed in the MODEL statement. Consequently, type I SS are often said to refer to variables *added-in-order*. The type II SS for a given predictor gives information about that predictor in a model that already contains all the other predictors. So, type II SS in procedure REG are frequently said to pertain to variables *added-last*.

5.2. Over-parameterized models

The major distinctions between the four types of SS arise in the context of the over-parameterized models available to procedures GLM and MIXED. If the data are balanced, the four types of SS reduce to the same single set of sums of squares. So the different types need be distinguished only when the data are unbalanced, that is, when the cell counts are not all equal.

Summary according to data layout. Table 1 [also see 1, Section 4.3.3] provides a first summary. It states the equalities that always occur among the SS obtained from the statement 'MODEL Y = A B A*B / SS1 SS2 SS3 SS4 ;' in procedure GLM, for data layouts categorized according to relationships between cell counts, namely:

- *balanced*: cell counts n_{ij} all equal;
- *proportional*: $n_{ij}/n_{il} = n_{kj}/n_{kl}$ for all i, j, k, l;
- *non-systematic*: n_{ij} not proportional, but all positive;
- *chaotic*: some $n_{ij} = 0$.

Table 1. Equalities among types of SS

Data layout	Source	Equalities
Balanced	A	I = II = III = IV
	B	I = II = III = IV
	AB	I = II = III = IV
Proportional	A	I = II III = IV
	B	I = II III = IV
	AB	I = II = III = IV
Non-systematic	A	III = IV
	B	I = II III = IV
	AB	I = II = III = IV
Chaotic	A	(all unequal)
	B	I = II
	AB	I = II = III = IV

The adjectives 'balanced' and 'proportional' are standard terminology, and non-systematic is commonly used. It is neither standard nor common to refer to data being chaotic (most often the reference is to data with some *empty cells* or *missing observations*). But perhaps chaotic should be the preferred term, since empty cells can present real difficulties in interpretation of estimable parameter functions and testable hypotheses, even for data that are otherwise balanced or proportional.

Table 1 indicates, then, that for balanced data no choice is necessary and that for proportional data the choice is effectively between type I SS and type III SS. Note that, for non-systematic or chaotic data, the type I and type II SS are equal for factor B but not for factor A. This is a consequence of the order of their specification ('A B A*B' as opposed to 'B A A*B') in the MODEL statement. Note, too, that the four types of SS are all equal for the interaction AB. (In general, the four types of SS are equal for the highest-order interaction in the model, provided that the model is complete in the sense of including all possible interactions of all possible orders.)

Table 1 can be loosely but conveniently summarized by saying that (a) the second equality in the string I = II = III = IV breaks when the data become unbalanced, or worse; (b) the first equality breaks when the data become non-systematic, or worse; and (c) the third equality breaks when the data become chaotic.

Summary by count and import. To obtain a second summary, consider the relevance of the cell counts. If these counts in unbalanced data are representative of the population—that is, if the treatments or the

Table 2. Appropriate sums of squares

	Classes of equal import	Classes of unequal import
Data are balanced for classes	Types I–IV	Contrast
Data are unbalanced for classes	Type III or Type IV	If same, type II If not, contrast

factor levels should be given weighted importance in accordance with the respective cell or level counts—then the type I and type II SS are pertinent and the type III and type IV SS are not pertinent. On the other hand, if the cell counts in unbalanced data are not representative of the population, then the type I and type II SS are not pertinent. If, in addition, the treatments or factor levels should be given equal weight, then the type III and type IV SS are pertinent.

This broad but not-quite-parallel distinction—type I or II SS on the one hand and type III or IV SS on the other hand—has a counterpart in comparisons of treatment and factor means. The least-squares estimates of marginal means (computed under the LSMEANS statement) are often appropriate in unbalanced designs, just as class and subclass arithmetic means (computed under the MEANS statement) are usually suitable in balanced designs. Simply put, the least-squares means in an unbalanced design are the values of class or subclass means that would be expected if the design was balanced. In other words, MEANS are to LSMEANS as type I and II SS are to type III and IV SS.

So when testing for factor A main effects, factor B main effects, or interaction effects in the two-factor fixed-effects analysis of variance model, which type of SS is suitable depends on the relevance of the cell counts. The appropriate choices are indicated in Table 2, which should be self-explanatory except for two things. First, the qualification *if same* in the table refers to situations in which the relative counts for the classes agree with the relative importance of the classes. Second, the entry *contrast* indicates situations in which none of the four types of SS are appropriate. In such a situation, a test of the desired hypothesis must be requested via an appropriate CONTRAST statement.

Care must be taken when testing hypotheses using chaotic data, since the estimability of parameter functions is dependent upon the positions of the empty cells and, as a consequence, the hypotheses tested by type IV SS can depend on the otherwise irrelevant order in which the levels of the factors appear in the data. In such a design, it is safer to do desired hypothesis testing via explicit CONTRAST statements. See [11,16,17] for examples and further details.

The discussion of types I, II, III, and IV analyses in [18] includes numerical examples using balanced (Ch. 8), non-systematic (Ch. 10), and chaotic (Ch. 14) data.

5.3. *Covariance models*

The types of SS that GLM uses by default, namely types I and III, play an especially convenient role for analysis of covariance models: (a) the type I SS are those which would have been obtained had the covariates not been used and (b) the type III SS are the sums of squares obtained after adjusting the response for the covariates. Furthermore, the MEANS and LSMEANS statements generate estimates that are, respectively, unadjusted and adjusted for the covariates.

6. DETERMINATION OF HYPOTHESES

Although procedure MIXED currently has only type III SS available for inference about fixed effects, procedure GLM has all four types. By specifying options of the form E1, E2, E3, and E4 in the MODEL statement, one can obtain the actual forms of the estimable parameter functions needed to specify the fixed-effect null hypotheses tested by the corresponding set of SS. The descriptions of these hypotheses are obtained by equating the various members of appropriate sets of estimable functions to zero.

These derivations are technical and tedious to perform, so this presentation can be omitted on a first reading. But the procedure is important, because it is occasionally the only way to determine what hypothesis is tested by a given SS.

To begin, recall that the properties of a SS do not depend on any parameter restrictions that are present in the model. Neither does the hypothesis that the SS tests (as will become clearer shortly). However, the manner in which the hypothesis being tested might be *written down* in terms of parameter functions can depend heavily on the parameter restrictions included in the model.

To outline the procedure for obtaining the specific form of a null hypothesis involving fixed effects, consider any particular SS of any particular type (either I, II, III, or IV). Then the method is as follows.

1. Use the corresponding option (E1, E2, E3, or E4) in the MODEL statement to obtain the *coefficient forms* necessary to guarantee that a linear combination of the parameters is an estimable parameter function. These coefficient forms are listed by SAS procedures GLM and MIXED as linear expressions in variables denoted L1, L2, etc. When these L-variables are set equal to specific constants, each coefficient form yields a specific number, and thus a specific estimable parameter function is obtained. For a given effect, the number of such L- variables will be the same as the numerator degrees of freedom for testing that effect.
   Note. The constants in the coefficient forms produced by SAS are decimal fractions. Using the numbers of levels in different factors as a rough guide and multiplying by appropriate constants, it is usually possible to convert these fractions into integers that may make the coefficient forms, and the estimable parameter functions they produce, easier to understand.
2. One by one, write down the specific estimable functions obtained by setting one of these L-variables equal to 1 and all the others equal to zero. This gives a set of linearly independent estimable parameter functions involving the parameters associated with the given effect (and probably some parameters associated with other effects). The size of this linearly independent set equals the degrees of freedom in the given SS.
3. Set all of these linearly independent parameter functions equal to zero, and then simplify the resulting statements as much as possible (which might not be very much). This set of conditions on the parameters is the null hypothesis being tested by the given SS, expressed without requiring or using any parameter restrictions in the model.
4. Now make use of any parameter restrictions included in the model to further simplify the description of the null hypothesis being tested—that is, to further simplify the form of the estimable parameter functions that specify this hypothesis.

This lengthy procedure is useful if one has any doubt about the specific hypotheses being tested. Some additional feeling for the procedure can be obtained from related discussions in [15,17].

It is an instructive exercise to use this procedure to obtain the hypotheses tested in a situation as simple as the complete factorial two-factor analysis of variance model with fixed effects and balanced data. The issue here is to show that, *under the usual assumptions*, the hypotheses tested by the type I SS can be described as

$\mathrm{SS}(\alpha|\mu)$ tests H_o: the α's all equal zero

$\mathrm{SS}(\beta|\mu)$ tests H_o: the β's all equal zero

$\mathrm{SS}(\alpha\beta|\mu)$ tests H_o: the $\alpha\beta$'s all equal zero,

in which no hypotheses are contaminated with unwanted effects. It is even more instructive to take such a model with balanced data, delete some points to obtain proportional data, delete more points to produce non-systematic data, delete still more points to achieve chaotic data, and then determine for each of these four data sets the actual form of the hypotheses tested by SS of a given type. Some further guidance can be obtained from illustrations in [18, Sections 10.3 and 10.5].

7. EXPECTED MEAN SQUARES

This section discusses rules for finding expected mean squares (EMS), that is, expected values of ratios of the form SS/df, in which the sum of squares SS has df degrees of freedom associated with it. Context here is restricted to balanced analysis of variance models involving crossed and nested factors.

For some of these models, the EMS and the rules for finding them as given by many textbooks [12, Appendix D, for example] are different from the EMS reported by procedure GLM of SAS. The purpose here is to explain why there are two sets of EMS for some analysis of variance models, and how it is that each set can be considered correct.

There are two sets of EMS for some models because there are two versions of the rules for finding EMS. The difference between these two versions relates to the way the model is parameterized, specifically to the presence or absence in the model of assumptions (like $\sum_i \alpha_i = 0$) that certain summations of parameters equal zero. The two versions of the rules differ only slightly, namely, in how they treat the variance components from interactions between fixed and random factors.

The two versions of the rules for finding EMS will herein be called the USL-rules and the SAS-rules. (These terminologies are emphatically *not* standard, but they are convenient.) The correspondence is that the USL-rules pertain to over-parameterized models incorporating the usual assumptions as restrictions and the SAS-rules pertain to over-parameterized models that include no restrictions on the parameters.

It is sufficient for illustration to consider a balanced two-factor analysis of variance model with interaction, in which factor A has fixed effects and factor B has random effects. (This is the same kind of model as considered in [8].) The model equation for this mixed model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

where $i = 1, \ldots, a$, $j = 1, \ldots, b$, and $k = 1, \ldots, n$.

Under the usual assumptions, which are used by many textbooks, the over-parameterized model states that:

(1) μ is a constant;
(2) the α_i's are constants subject to the restriction that they sum to zero;
(3) the β_j's are i.i.d. normal $(0, \sigma^2_\beta)$;
(4) the $(\alpha\beta)_{ij}$'s are i.i.d. normal $(0, (a-1)\sigma^2_{\alpha\beta}/a)$, subject to the restrictions that for each j their sum over i equals zero;
(5) the ε_{ijk}'s are i.i.d. normal $(0, \sigma^2)$;
(6) the β_j's, $(\alpha\beta)_{ij}$'s, and ε_{ijk}'s are all independent.

The over-parameterized model with no restrictions is the same as this except that:

(a) the summation restrictions in (2) and (4) are dropped;
(b) the variance of the $(\alpha\beta)$'s is taken to be $\sigma^2_{\alpha\beta}$ instead of $(a-1)\sigma^2_{\alpha\beta}/a$.

This unrestricted form of the over-parameterized model is the one used by SAS procedure GLM.

Application of the USL-rules to this mixed two-factor model produces the EMS given by many textbooks, specifically

$$\begin{aligned}
E(\text{MSA}) &= \frac{nb}{a-1}\sum_i \alpha_i^2 && + n\sigma^2_{\alpha\beta} + \sigma^2\\
E(\text{MSB}) &= \quad + na\sigma^2_\beta && + \sigma^2\\
E(\text{MSAB}) &= && + n\sigma^2_{\alpha\beta} + \sigma^2\\
E(\text{MSE}) &= && + \sigma^2
\end{aligned}$$

whereas the SAS-rules produce the EMS

$$\begin{aligned}
E(\text{MSA}) &= \frac{nb}{a-1}\sum_i (\alpha_i - \alpha_\bullet)^2 && + n\sigma^2_{\alpha\beta} + \sigma^2\\
E(\text{MSB}) &= \quad + na\sigma^2_\beta && + n\sigma^2_{\alpha\beta} + \sigma^2\\
E(\text{MSAB}) &= && + n\sigma^2_{\alpha\beta} + \sigma^2\\
E(\text{MSE}) &= && + \sigma^2
\end{aligned}$$

in which $\alpha_\bullet$ denotes $(1/\alpha)\sum_i \alpha_i$.

The difference between the USL-rules and the SAS-rules is easily indicated by describing how to get from the EMS produced by the former to those produced by the latter. First, apply the USL-rules as if the model had all effects random, then change the variance components for fixed main effects and interactions into quadratic forms in the corresponding parameters. (As is suggested by the first term in the last formula for $E(\text{MSA})$, each such quadratic form is defined to be the sum of the squares of the deviations of the corresponding parameters about their arithmetic mean, divided by the degrees of freedom for that effect.)

In the model being considered here, only the A main effects are fixed. After the USL-rules have been applied to obtain the EMS of the two-factor model with both A and B random, the variance component σ^2_α is replaced by the quadratic form

$$\frac{1}{a-1}\sum_i (\alpha_i - \alpha_\bullet)^2.$$

These are the EMS obtained from SAS procedure GLM for the mixed two-factor model (by using the statement 'RANDOM B A*B / Q ;').

The presence or absence of the usual assumptions affects the EMS in the model and, consequently, affects the form of the F-statistics used to test certain hypotheses. In the model of this section, for example, the statistic for testing

$$H_o : \sigma^2_\beta = 0 \quad \text{against} \quad H_a : \sigma^2_\beta > 0$$

is $F = \text{MSB}/\text{MSE}$ in the model incorporating the usual assumptions, whereas it is $F = \text{MSB}/\text{MSAB}$ in the model without restrictions.

For models having only fixed effects, the SAS-rules and the USL-rules agree except for differences in quadratic forms, such as

$$\frac{nb}{a-1}\sum_i (\alpha_i - \alpha_\bullet)^2 \quad \text{versus} \quad \frac{nb}{a}\sum_i \alpha_i^2.$$

For models having only random effects, the SAS-rules and the USL-rules give identical results.

8. NUMERICAL EXAMPLES

We begin with a re-analysis of data on the effects that direction of travel and four differently-striped roadway curves had on travel speed. As in the original analysis [16] and in order to justify use of a two-factor interaction model with 2 and 4 levels per factor and 16 observations per cell, we have ignored the randomized block structure inherent in the experimental design and pretended that each curve–direction combination was negotiated by its own set of 16 drivers.

Table 3. Differences in SS for driver speed example

Source	Reported	Re-computed
Balanced data (I = II = III = IV)		
Model	1475.18	1461.38
Curve	1043.84	1034.56
Direction	92.82	91.13
Curve*direction	338.52	335.69
Error	3542.19	3542.13
Non-systematic data (III = IV)		
Model	1372.35	1359.40
Curve	1049.25	1040.83
Direction	209.69	207.16
Curve*direction	105.25	103.34
Error	2115.65	2115.59
Chaotic data (III)		
Model	1303.55	598.02
Curve	870.84	505.82
Direction	86.65	86.65
Curve*direction	65.16	65.16
Error	1946.71	2037.65

Table 4. Data for battery life example

	Temperature					
Material	15		70		125	
1	130	155	34 *p*	40	20 *p*	70
	74 *p*	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58 *n*	45
3	138	110	174 *n*	120	96 *c*	104 *c*
	168	160	150	139	82 *c*	60 *c*

p—deleted from original data set to produce proportional data set
n—deleted from proportional data set to produce non-systematic data set
c—deleted from non-systematic data set to produce chaotic data set

Table 5. Fundamental ANOVA values

	Model		Error	
Setting	df	SS	df	SS
Balanced	8	59 416.22	27	18 230.75
Proportional	8	52 429.81	24	10 714.25
Non-systematic	8	46 814.04	22	9553.83
Chaotic	7	43 842.79	19	8438.83

The illustrations in [16] of differences among the four types of SS in SAS are to the point but, as the (selected) entries of Table 3 indicate, many of the SS values reported are in error. The differences are minor for the initial balanced data set, as they are for the non-systematic data set they obtained from it by deleting two questionable observations from a particular curve driven in a particular direction. But for the chaotic data set they obtained by additionally deleting all observations taken at another curve–direction combination, there are major discrepancies in the model and curve SS.

8.1. *An extended example*

As a second and more thorough example, consider the data [19, p. 235] in Table 4, on the life times of batteries (measured in hours). The data fit a two-factor interaction model with one factor (type of plate material) at three levels, the other factor (test temperature in degrees Fahrenheit) also at three levels, and four observations per cell. The analyses presented here treat both factors as having fixed effects, although temperature could appropriately be considered a random factor.

As suggested by the footnotes of Table 4, the illustrations that follow are based upon a subset-chain of four data sets: (1) the original *balanced* data set; (2) the *proportional* data set obtained by deleting the smallest observation from each cell in the first row; (3) the *non-systematic* data set obtained by additionally deleting one (randomly selected) observation from each of two other (randomly selected) cells; and (4) the *chaotic* data set obtained by additionally deleting all four observations from yet another (randomly selected) cell.

For the sake of brevity, presentation of analytic results from these four settings concentrates on SS values, suppressing such things as mean squares and F-statistics. Temperature has 2 degrees of freedom, whatever the setting, as does material. Their interaction has 4 degrees of freedom except in the chaotic setting, when it has 3 degrees of freedom. With these quantities and the fundamental ANOVA values given in Table 5, it is trivial to produce the complete ANOVA table for any given setting.

Table 6 gives each of the four types of SS in each of these four settings for each of temperature, material, and the temperature–material interaction. These SS values are from fitting material after temperature, a consideration that potentially affects only the Type I SS. The footnote in the table reports how the Type I SS change when, instead, temperature is fitted after material.

The equal signs in Table 6 emphasize the equalities promised by Table 1. The common values of 4312.48

Table 6. SS from fitting material after temperature

Source		Type I		Type II		Type III		Type IV
Balanced data								
Temperature		39 118.72	=	39 118.72	=	39 118.72	=	39 118.72
Material		10 683.72	=	10 683.72	=	10 683.72	=	10 683.72
Interaction		9613.78	=	9613.78	=	9613.78	=	9613.78
Proportional data								
Temperature		38 124.06	=	38 124.06		37 425.35	=	37 425.35
Material		4312.48	=	4312.48	≠	4312.48	=	4312.48
Interaction		9993.27	=	9993.27	=	9993.27	=	9993.27
Non-systematic data								
Temperature	*b*	35 385.99		35 302.10		36 588.67	=	36 588.67
Material	*a*	2826.53	=	2826.53		3202.42	=	3202.42
Interaction		8601.52	=	8601.52	=	8601.52	=	8601.52
Chaotic data								
Temperature	*e*	34 186.01		27 872.03		28 139.51		31 537.92
Material	*d*	1385.31	=	1385.31		1676.08		3582.68
Interaction		8271.48	=	8271.48	=	8271.48	=	8271.48

Differing Type I SS (from analyses fitting temperature after material):
a 2910.42 *b* 35 302.10 *d* 7699.29 *e* 27 872.03

Table 7. p-values for tests on material (after temperature)

Setting		Type I		Type II		Type III		Type IV
Balanced		0.0020	=	0.0020	=	0.0020	=	0.0020
Proportional		0.0173	=	0.0173	≠	0.0173	=	0.0173
Non-systematic	*c*	0.0578	=	0.0578		0.0416	=	0.0416
Chaotic	*f*	0.2360	=	0.2360		0.1789		0.0347

p-values for differing Type I SS (from analyses fitting temperature after material): *c* 0.0537 *f* 0.0021

for the Type II and Type III SS for material in the proportional setting are explicitly marked as unequal (by the presence of an unequal sign rather than the absence of an equal sign), in order to underscore the fact that this equality is not guaranteed. Indeed, since the same sets of SS values arise from fitting temperature after material in the proportional setting, doing so produces equality for the effect fitted first and inequality for the effect fitted second, rather than the other way around. The reason for this behavior is that the proportional data set is balanced with respect to temperature but not with respect to material.

There are four distinct interaction SS values in Table 6, associated with the distinct error SS given in Table 5. The F-values easily computed therefrom have associated p-values that SAS procedure GLM reports—reading down the table and hence down the subset-chain—as 0.0186, 0.0025, 0.0053, and 0.0040. So interaction would likely be considered important in any of these four settings.

But for illustrative purposes, we nonetheless proceed to test for the presence of main effects (without re-analyzing the data sets under no-interaction models). The p-values for the various SS for temperature are all reported as 0.0001, so that temperature is clearly important to battery life. This fact (true whether temperature is fitted before or after material) is a convincing reason to adjust a decision about material by fitting temperature first.

The various SS for material in Table 6 have the p-values shown in Table 7 (which uses the same annotative conventions as Table 6). It is apparent from these p-values that the significance or non-significance of plate material depends upon what type of SS is used and at what level of significance.

8.2. *The example revisited*

The broad view taken thus far is necessary when illustrating the similarities and differences between the various SS in the various settings. Motivated readers can check their understanding by determining—before reading the next paragraph—what differences arise, and for what reasons, if Tables 6 and 7 are changed to reflect the fitting of temperature after material, that is, if substitutions a, b, c are made for the non-systematic setting and substitutions d, e, f are made for the chaotic setting.

Well, the changes in the non-systematic setting are that the Type I SS and Type II SS become unequal for material but equal for temperature, with parallel changes in the associated p-values. These changes occur precisely because the Type I SS pertain to effects *added-in-order* and the Type II SS to effects *added-last*. The same things happen in the chaotic setting, and for the same reason.

In practice, of course, one has only a single setting to analyze and (usually) only one of the four types of SS is pertinent. In the balanced setting, the four types of SS agree and no decision about SS type is required (although recognition of the situation is). That's just one of the many advantages of having a balanced data set.

In the proportional setting, the decision between Types I $=$ II and III $=$ IV can depend upon the meaning (if any) of the proportionalities in the data. The (constructed) pattern of proportionality in these battery data is such that only Types II and III differ, and then only for temperature, which is not a factor of primary interest, so that there are again no important distinctions between the different SS types. If the pattern of proportionality were such that Types II and III differed for material, the present authors would use the Type III SS in a study of what batteries should be produced in the future but would use Type II SS if drawing inferences about an existing stock of batteries for which the sample proportions across material were representative.

In the non-systematic setting, the present authors would always use Types III $=$ IV SS, since there appears to be no reasonable situation in which Type I SS or (remembering that non-systematic implies non-proportional) even Type II SS would be suitable. And it makes a difference, as can be seen in Table 7: the p-value for material after temperature is 0.0578 for Types I $=$ II, it is 0.0416 for Types III $=$ IV, and 0.05 is a commonly-used level of significance.

In the chaotic setting, Types I and II are similarly inappropriate and Types III and IV usually differ. Moreover, the decision between Type III and Type IV may be difficult, since the distinctions between them are, as already noted, rather technical. Yet the choice can be a critical one, as entries in Table 7 demonstrate: material after temperature is non-significant at any level below 0.1789 if the Type III SS of 1676.08 is used, but it is significant at any level above 0.0347 when the Type IV SS of 3582.68 is used. The chaotic setting really can be chaotic.

8.3. *A three-factor example*

A thorough analysis of a three-factor model which has multiple empty cells is given in [18, Ch. 17]. The presentation includes the instructions appropriate for SAS procedure GLM, as well as several ANOVA tables.

9. MORE ABOUT MIXED MODELS

The analysis of variance for mixed models (as opposed to fully fixed or fully random ones) can be subtle and difficult. Some textbooks give incomplete or misleading coverage and some computer programs report incorrect results. The analyses in [20], which include a re-analysis of experimental data first analyzed in [21], explore implications in two specific mixed-model data sets.

See [8,22] for a detailed discussion of these matters and a comparison of the mixed-model features of SAS with those of BMDP and SPSS. See [23] for useful diagrammatic representations of partitions, SS, and EMS in mixed models.

There has been considerable disagreement among statisticians about which kind of mixed model is more appropriate: a restricted model (incorporating, for example, the usual assumptions) or an over-parameterized model with no restrictions. Consideration of EMS for unbalanced designs has suggested that over-parameterized models without restrictions are generally more appropriate than those with restrictions. But, paradoxically, restricted models can be broader than unrestricted ones, in the sense of providing for negative correlations between selected pairs of observations, as opposed to requiring that all such correlations be zero or positive [8, Section 2.1]. A recent review [24] of these disagreements about restrictions versus over-parameterization, done using so-called superpopulation models, resolved the controversy in favor of restricted models.

The output that SAS procedure GLM gives for mixed models is easily misinterpreted, since the

RANDOM statement affects only the way that EMS are computed. By default, GLM uses MSE as the denominator MS for any F-statistic. For example, in an ANOVA table for data under the model of Section 7, procedure GLM would use $F = \mathrm{MSB}/\mathrm{MSE}$, even though that is not the correct statistic for the unrestricted over-parameterized model that GLM uses when it computes EMS. To get the correct test results with procedure GLM, it is necessary to use the TEST option on the RANDOM statement or to specify the appropriate hypothesis (that is, numerator) and error (that is, denominator) SS via a TEST statement.

Another consequence of GLM's default use of MSE as the error MS is that the standard errors produced by the LSMEANS statement underestimate the true standard errors. For these reasons, procedure MIXED is preferred over procedure GLM for most analyses involving mixed models [4].

10. FURTHER READING

Further information about SS and EMS in the general linear model can be found in numerous places. The books [12,18] give readable coverages from an applied point of view; the treatment in [10] is theoretical in nature; and the coverage in [9,11] falls in between.

REFERENCES

1. Freund RJ, Littell RC. *SAS for Linear Models: A Guide to the ANOVA and GLM Procedures*. SAS Institute, Inc.: Cary, NC, 1981.
2. Freund RJ, Littell RC, Spector PC. *The SAS System for Linear Models*. SAS Institute, Inc.: Cary, NC, 1988.
3. *SAS/STAT User's Guide* (Version 6, 4th edn), vols. 1 and 2. SAS Institute, Inc.: Cary, NC, 1990.
4. Latour D, Latour K, Wolfinger RD. *Getting Started With Proc MIXED*. SAS Institute, Inc.: Cary, NC, 1994.
5. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS System for Mixed Models*. SAS Institute, Inc.: Cary, NC, 1996.
6. *SAS/STAT Software: Changes and Enhancements* (through Release 6.11). SAS Institute, Inc.: Cary, NC, 1996.
7. Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics* 1998; **23**:323–355.
8. Schwarz CJ. The mixed-model ANOVA: The truth, the computer packages, the books. Part I: Balanced data. *The American Statistician* 1993; **47**:48–59.
9. Myers RH, Milton JS. *A First Course in the Theory of Linear Statistical Models*. PWS-Kent: Boston, 1991.
10. Searle SR. *Linear Models*. Wiley: New York, 1971.
11. Searle SR. *Linear Models for Unbalanced Data*. Wiley: New York, 1987.
12. Neter J, Kutner MH, Nachtscheim CJ, Wasserman W. *Applied Linear Statistical Models* (4th edn). Irwin: Chicago, 1996.
13. Turner DL. An easy way to tell what you are testing in analysis of variance. *Communications in Statistics—Theory and Methods* 1990; **19**:4807–4832.
14. Nelder JA, Lane PW. The computer analysis of factorial experiments: In memoriam—Frank Yates. *The American Statistician* 1995; **49**:382–385.
15. Searle SR. Analysis of variance computing package output for unbalanced data from fixed-effects models with nested factors. *The American Statistician* 1994; **48**:148–153.
16. Pendleton OJ, von Tress M, Bremer R. Interpretation of the four types of analysis of variance tables in SAS. *Communications in Statistics—Theory and Methods* 1986; **15**:2785–2808.
17. Freund RJ. The case of the missing cell. *The American Statistician* 1980; **34**:94–98.
18. Milliken GA, Johnson DE. *Analysis of Messy Data*, vol. I: *Designed Experiments*. Chapman & Hall: New York, 1992.
19. Montgomery DC. *Design and Analysis of Experiments* (4th edn). Wiley: New York, 1997.
20. Ayres MP, Thomas DL. Alternative formulations of the mixed-model ANOVA applied to quantitative genetics. *Evolution* 1990; **44**:221–226.
21. Pashley DP. Quantitative genetics, development, and physiological adaptation in host strains of fall armyworm. *Evolution* 1988; **42**:93–102.
22. Stanek III EJ, Buonaccorsi JP. Letter to the editor (commenting on [8], with reply). *The American Statistician* 1995; **49**:323–324.
23. Lohr SL. Hasse diagrams in statistical consulting and teaching. *The American Statistician* 1995; **49**:376–381.
24. Voss DT. Resolving the mixed models controversy. *The American Statistician* 1999; **53**:352–356.

Authors' biographies:

Michael F. Driscoll is Associate Professor, Department of Mathematics, PO Box 871804, Arizona State University, Tempe, Arizona 85287-1804. His areas of interest and research are varied, but pertain mainly to the theory and application of the general linear model.

Connie M. Borror is Visiting Assistant Professor, Department of Industrial and Management Systems Engineering, PO Box 875906, Arizona State University, Tempe, Arizona 85287-5906. Her areas of interest and research are experimental design and process control.