

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2001 - 13th Annual Conference Proceedings

ENUMERATION OF MIXED LINEAR MODELS AND A SAS MACRO FOR COMPUTATION OF CONFIDENCE INTERVALS FOR VARIANCE COMPONENTS

Ann Hess

Hari Iyer

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Hess, Ann and Iyer, Hari (2001). "ENUMERATION OF MIXED LINEAR MODELS AND A SAS MACRO FOR COMPUTATION OF CONFIDENCE INTERVALS FOR VARIANCE COMPONENTS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1215>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

ENUMERATION OF MIXED LINEAR MODELS AND A SAS MACRO FOR COMPUTATION OF CONFIDENCE INTERVALS FOR VARIANCE COMPONENTS

Ann Hess
Hari Iyer
Department of Statistics
Colorado State University
Fort Collins, CO 80523

Abstract

It is well known that there is a one-to-one correspondence between fixed effects ANOVA models involving both crossed and nested factors, and combinatorial objects called “posets”. The enumeration of nonisomorphic posets is an interesting and nontrivial combinatorial problem for which answers are available for posets of order 14 or less, i.e. for fixed effects ANOVA models with 14 or fewer factors. Such an enumeration problem does not appear to have been considered for mixed effects models. In this paper we study this problem and obtain a list of nonisomorphic mixed effects models involving five or fewer factors. For instance, it is shown that there are 576 nonisomorphic mixed effects ANOVA models involving five factors. Confidence intervals for variance components in any of these models, with superior coverage properties than those afforded by the application of the Satterthwaite method, may be computed using the methods discussed in the book “Confidence Intervals for Variance Components” by Burdick and Graybill [2]. To facilitate these calculations, we have written a SAS macro that will compute these confidence intervals for any mixed effects saturated ANOVA model involving five or fewer factors. User input to the SAS MACRO is the actual data set along with a matrix indicating whether each factor is fixed or random and the nesting/crossing configuration among the factors. The operation of the MACRO is illustrated using an example involving real data. Such a MACRO is expected to be extremely useful to practitioners in view of the fact that SAS or any other commonly available statistical software package does not have built in commands for obtaining confidence intervals for variance components discussed in [2]. A copy of this MACRO is available at www.stat.colostate.edu/~hess/MixedModels.htm.

1 Introduction

The mixed model is a special case of the general linear model, where both the mean function and covariance matrix for the data have a linear structure. Specifically, the vector of observations \mathbf{y} is assumed to have moments

$$E[\mathbf{y}] = \mathbf{X}\boldsymbol{\alpha}, \quad (1)$$

$$Var[\mathbf{y}] = \sum_t \phi_t \mathbf{V}_t, \quad (2)$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^t$ and $\boldsymbol{\phi} = (\phi_1, \dots, \phi_q)^t$ are unknown parameters and \mathbf{X} , $\mathbf{V}_1, \dots, \mathbf{V}_q$ are known matrices. Note that any mean vector and covariance matrix can be written in this form, but we are interested in models in which the parameters $\alpha_1, \dots, \alpha_p, \phi_1, \dots, \phi_q$ are functionally independent (except perhaps some linear estimability conditions among the α_i) and relatively few

parameters are needed to obtain this form. Such general covariance structures arise by assuming that some of the effects are random.

In the discussion that follows, we will often refer to “balanced” data and “saturated” models. For balanced data, at each factor level combination, the same number of responses are available. A saturated model includes all possible main effect and interaction terms.

The following example illustrates the general situation. It involves fixed and random factors and also crossing and nesting among factors.

Example 1: An experiment was designed to test the effectiveness of different growing regimes on crop yield. All combinations of two Fertilizers, two Pesticides and two Water volumes were considered. Two growing Areas were randomly assigned to each of the eight treatments. Within each area, 3 samples were taken and the crop yield was recorded. Note that Fertilizer, Pesticide, and Water are fixed factors and Area is a random factor. The response for the r th replicate in the l th area for the i th fertilizer with the j th pesticide and the k th water treatment is:

$$y_{ijklr} = \mu + f_i + p_j + fp_{ij} + w_k + fw_{ik} + pw_{ik} + fpw_{ijk} + a_{l(ijk)} + \epsilon_{ijklr}.$$

It is assumed that

$$a_{l(ijk)} \sim N(0, \phi_l), \epsilon_{ijklr} \sim N(0, \phi_\epsilon).$$

We further assume that the random variables are mutually independent.

2 Fixed Effects Models and Hasse Diagrams

In this section we will establish the one to one correspondence between fixed effects models with n factors and posets with n elements. First we establish some definitions and proofs pertaining to posets.

2.1 Posets, Hasse Diagrams, and Adjacency Matrices

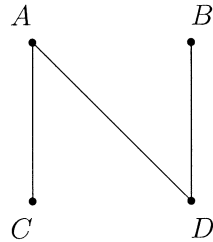
A partial order on a set, X , is a collection, R , of ordered pairs (x, y) with $x \in X$ and $y \in X$, such that

- $(x, x) \in R$ for all $x \in X$ (reflexivity),
- $(x, y), (y, x) \in R$ imply $x = y$ (antisymmetry), and
- $(x, y), (y, z) \in R$ imply $(x, z) \in R$ (transitivity).

Instead of $(x, y) \in R$, we often write $x \leq y$. The pair (X, R) is called a partially ordered set, or poset for short, see [3, p. 187].

A Hasse diagram is a convenient graphical representation of a poset. The Hasse diagram of a poset P is a graph drawn such that each vertex corresponds to a point of the poset, and for each pair (x, y) such that $x \leq y$ the vertices representing x and y are joined by an edge (such that x is ‘below’ y) provided that $x \leq y$ is a nonredundant relation in R . The relation $x \leq z$ in R is called redundant if $\exists y \in X, y \neq x, y \neq z$, such that $x \leq y$ and $y \leq z$. Otherwise the relation $x \leq z$ is said to be nonredundant. By convention, the relation $x \leq x$ is a redundant relation.

Consider the poset N such that $n = 4$ and $C \leq A, D \leq A, D \leq B$. The Hasse Diagram corresponding to this poset is shown below.



The following theorems related to posets and Hasse diagrams will be useful later.

Theorem 1: Giving a Hasse diagram H on n vertices is equivalent to giving a poset P on n elements, [5, p. 61].

Theorem 2: Any subset S of a poset P is itself a poset under the same relation (restricted to S), [1, p. 1].

Two posets are considered isomorphic if one can be transformed into another by simply rearranging the labelling of the elements. The number of unlabelled or nonisomorphic posets with $n \leq 14$ elements has been derived. These results are shown in Table 1, [4]. A formula to count the number of unlabelled posets with n elements is not known to exist, and even estimates of the number of unlabelled posets with n elements as $n \rightarrow \infty$ is a difficult combinatorial problem, [5, p. 59].

A poset can also be represented using an adjacency matrix. If P is a poset with n elements, the adjacency matrix of P is the $n \times n$ matrix whose (i, j) entry is equal to 1 if $j \leq i$ and 0 otherwise. Note that if $j \leq i$ the (i, j) entry of the adjacency matrix is 1, but the (j, i) entry is 0. The adjacency matrix for a poset P is denoted as $A(P)$. Note that by the reflexivity of posets, the diagonal elements should all be 1. However, we will adapt the convention of putting 0's along the diagonal. The adjacency matrix for the poset N such that $n = 4$, and $C \leq A, D \leq A, D \leq B$, is given by

$$A(N) = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

where the row and column indices are A, B, C, D in that order.

2.2 The Relationship between Posets and Fixed Effects Models

In this section we will establish the relationship between posets and fixed effects models. We are interested in enumerating the nonisomorphic fixed effects models with $n \leq 5$ factors, and it will be convenient to use already established information about the number of unlabelled posets with $n \leq 5$ elements.

Proposition 1 There is a one to one correspondence between posets and fixed effects models. Hence, the number of nonisomorphic posets with n elements is equal to the number of nonisomorphic fixed effects models with n factors.

Table 1: Number of Unlabelled Posets with n Elements

n	Unlabelled Posets		
0	1		
1	1		
2	2		
3	5		
4	16		
5	63		
6	318		
7	2,045	1972	J. Wright
8	16,999	1977	S.K. Das
9	183,231	1984	R.H. Mohring
10	2,567,284	1990	J.C. Culberson and G.J.E. Rawlins
11	46,749,427	1990	J.C. Culberson and G.J.E. Rawlins
12	1,104,891,746	1991	C. Chaunier and N. Lygeros
13	33,823,827,452	1992	C. Chaunier and N. Lygeros
14	1,338,193,159,771	2000	N. Lygeros and P. Zimmermann*

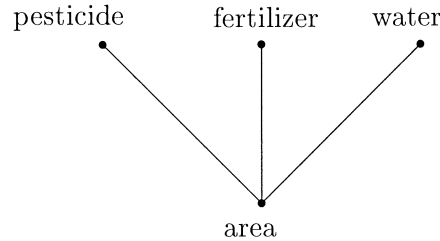
*<http://www.desargues.univ-lyon1.fr/home/lygeros/poset.html>

This can be seen when we consider a poset P with n elements, and a fixed effects model F with n factors. For distinct i and j , if $j \leq i$ in the poset P , this corresponds to factor j nested in factor i in the fixed effects model F . If $j \not\leq i$ and $i \not\leq j$ in P , this corresponds to factor i crossed with factor j in the fixed effects model F .

So, the number of nonisomorphic fixed effects models with n factors is equivalent to the number of unlabelled posets with n elements. Hence, the number of fixed effects models with $n \leq 14$ factors is known.

Using this result, we will be able to conveniently represent a given fixed effects model by either a Hasse diagram or an adjacency matrix. If we consider the Hasse diagram corresponding to a given fixed effects model, if factor j is nested in factor i , the vertices corresponding to these two factors will be connected by an edge or a monotone sequence of edges, and the vertex for factor j will be below the vertex for factor i . Conversely, only if two factors are connected by an edge or a monotone sequence of edges in a given Hasse diagram, will there be a nesting relationship between the two factors. A sequence of edges connecting A to B is monotone if it corresponds to a chain of relations of the form $A = y_1 \leq y_2, y_2 \leq y_3, \dots, y_{k-1} \leq y_k = B$ for some k . If we consider the adjacency matrix corresponding to a fixed effects model, if factor j is nested in factor i the (i, j) entry of the adjacency matrix will be 1, otherwise the (i, j) entry will be 0. Hence, if factor i has no factors nested within it, the i th row will be a row of zeros. Note that we can always represent any fixed or mixed effects model using a Hasse Diagram. We return to our original example to illustrate.

Example 1 (continued): The mixed model presented in Example 1 is illustrated by the following Hasse Diagram.



Note that even though Pesticide and Fertilizer are connected by a sequence of edges, this sequence is not monotone, so Pesticide and Fertilizer are not nested but crossed.

Since we are interested in actually creating all possible nonisomorphic fixed effects models with a given number of factors, an algorithm will be required. For this purpose, it is most convenient to represent a given fixed effects model as an adjacency matrix. For the purposes of constructing an algorithm, we will want to construct an $n \times n$ adjacency matrix (corresponding to a fixed effects model with n factors) based on an existing $(n - 1) \times (n - 1)$ adjacency matrix (corresponding to a fixed effects model with $n - 1$ factors). The following proposition will allow us to do this.

Proposition 2 The $n \times n$ adjacency matrix for a given fixed effects model with n factors can be constructed by adding a row and column to an $(n - 1) \times (n - 1)$ adjacency graph corresponding to a fixed effects model with $n - 1$ factors. Specifically, we can create a fixed effects model with n elements by adding a single factor which has no other factors nested within it. This corresponds to adding a final row and column to an existing $(n - 1) \times (n - 1)$ adjacency matrix, where the n th row is a row of zeros.

It is easiest to visualize this proposition in the context of Hasse diagrams. In this context, the proposition states that any Hasse diagram with n vertices can be constructed by adding a single vertex to a Hasse diagram with $n - 1$ vertices. Specifically, we will create the new Hasse diagram by adding a “bottom” vertex to an existing Hasse diagram. By “bottom” vertex we mean that this vertex will not be connected to any vertices “below” itself. In the context of fixed effects models, this will represent a factor with no factors nested within it.

First of all, any Hasse diagram, H^* , must have a bottom vertex. If not, then every vertex of H^* has at least one vertex below it. This would imply that the number of vertices in H^* is infinite, a contradiction. Therefore, let x be a “bottom” vertex of H^* . By removing this vertex and all edges connected to it we get an $n - 1$ vertex Hasse diagram, H , and H^* can be obtained from H by adding back the bottom vertex and an appropriate set of edges.

Hence, we may construct new fixed effects models with n factors by adding a factor (with no other factors nested within it) to a fixed effects model with $n - 1$ factors and defining appropriate nesting relationships of this new factor to the other $n - 1$ factors.

2.3 An Algorithm for Generating Fixed Effects Models with $n \leq 5$ Factors

For the purposes of this project, it is desired to actually construct all fixed effects models with $n \leq 5$ factors. The following algorithm was used to enumerate all fixed effects models with $n \leq 5$ factors.

1. Construct the set of adjacency matrices which represent fixed effects models with n factors. In the actual implementation of this algorithm, we started with the set of 2×2 adjacency matrices which represent all possible fixed effects models with 2 factors. This set, which we will denote by $F2$, is shown below. Note that it would also have been possible to start with the single adjacency matrix that represents the only possible fixed effect model with $n = 1$ factor.

$$F2 = \left\{ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right\}$$

2. Construct a set of candidate adjacency matrices of size $(n + 1) \times (n + 1)$. This set of candidate adjacency matrices is constructed using the following steps.

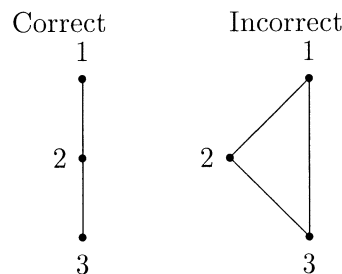
- (a) Construct the 2^n possible vectors of length n comprised only of zeroes and ones. For $n = 2$, this set of vectors is:

$$\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

Adjoin each of these possible vectors (as the $(n + 1)$ st column) to each of the $n \times n$ adjacency matrices described in step 1.

- (b) Adjoin a row vector of zeros of length $n + 1$ (as the $(n + 1)$ st row) to each of the $n \times (n + 1)$ matrices from (a). This creates a set of $(n + 1) \times (n + 1)$ candidate matrices.

3. From the set of $(n + 1) \times (n + 1)$ candidate matrices, identify and remove matrices which represent Hasse diagrams where there is a “redundant” edge. For example, if we consider the poset P with vertices labeled 1, 2, and 3, such that $3 \leq 2, 2 \leq 1$, the Hasse diagram that represents this poset would have an edge connecting vertex 1 to vertex 2 and an edge connecting vertex 2 to vertex 3. An edge that connects vertex 1 directly to vertex 3 would be redundant. See the diagrams below.



In order to remove matrices that have such “redundant” edges, we can use the following process.

- (a) For a given candidate adjacency matrix A of size $(n + 1) \times (n + 1)$, examine the matrix given by $\sum_{k=2}^{k=n} A^k$. Note that (i, j) th entry of the matrix A^k gives the number of monotone paths of length k between vertices i and j . So, the (i, j) th entry of $\sum_{k=2}^{k=n-1} A^k$ indicates the number of monotone paths between vertices i and j with length between 2 and n .

- (b) If the (i, j) th entry of A is 1, hence indicating a path of length 1 between vertices i and j , and the (i, j) th entry of $\sum_{k=2}^{k=n-1} A^k$ is greater than or equal to one, remove matrix A from the list of possible candidate matrices.
4. From the remaining candidate matrices, identify pairs of isomorphic matrices, and remove redundancies. Two fixed effects models are considered isomorphic if one can be transformed into another by simply rearranging the labelling of the factors. Hence, for us, two adjacency matrices will be considered isomorphic if one can be changed into another by permuting rows and columns (using the same permutation for both rows and columns). The matrices B and C would be considered isomorphic if we could obtain B by swapping rows i and j of C and then swapping columns i and j in C . More generally two matrices B and C would be considered isomorphic if we could obtain B by a given permutation of the rows of C and then applying the same permutation to the columns of C . The following steps were used.
- (a) The first matrix from the set of candidates is automatically placed into the acceptance set.
- (b) For each of the remaining candidate matrices, all possible permutations were created from the original matrix. If the intersection of the acceptance set and the set of permuted versions of the candidate matrix is null, then this matrix is added to the acceptance set. Otherwise, it is discarded.
- (c) These steps are repeated until each matrix in the set of candidates has either been discarded or added to the acceptance set.

Using this algorithm, we were able to successfully generate all possible fixed effects models with $n \leq 5$ factors. The number of models with n factors ($n \leq 5$) matched the results for the number of posets with n elements; refer to Table 1 for the results.

3 Mixed Models and Two Colored Hasse Diagrams

Using the results from the previous section, specifically the adjacency matrices that were constructed, we will be able to enumerate mixed models with $n \leq 5$ factors. This problem can be likened to counting the number of non-isomorphic posets with each vertex colored using one of two specified colors—say, red and green. By convention, the green vertices represent fixed effects and the red vertices represent random effects. To the best of our knowledge, this problem has not been studied.

In the following algorithm, we utilize the adjacency matrices corresponding to the fixed effects models that were generated using the algorithm presented in the preceding section. Recall, that the diagonal elements of these adjacency matrices were all zero. In graph theory, a one in the (i, i) position of an adjacency matrix usually indicates a loop at vertex i . However, for our purposes, if a 1 appears in the (i, i) position of the adjacency matrix, we will define the i th factor of the model to be a random factor, while a zero indicates a fixed factor. The following algorithm was used to generate all mixed effects models with n factors for $n \leq 5$.

1. Start with the set of adjacency matrices representing the fixed effects models with n factors.

2. We now construct a set of candidate matrices. For each $n \times n$ fixed effect matrix, create candidate mixed effect matrices by generating all of the possible 2^n combinations of zeroes and ones along the diagonal. Note that the off-diagonal elements remain unchanged.
3. We now remove matrices that violate the constraint that a fixed factor cannot be nested within a random factor. This constraint is discussed below. Recall if the i th factor is random this is indicated by a 1 in the (i,i) position of the adjacency matrix.
 - (a) For a given candidate adjacency matrix A of size $n \times n$, examine the matrix given by $S = \sum_{k=1}^{n-1} A^k$. Note that the (i,j) th entry of the matrix A^k gives the number of paths of length k between vertices i and j . So, the (i,j) th entry of S indicates the number of monotone paths between vertices i and j of any given length.
 - (b) Then for each candidate adjacency matrix A , if $A(i,i) = 1$ and $A(j,j) = 0$ and $S(i,j) \geq 1$ then we remove A from the set of candidate matrices.
 - (c) Continue this process for each candidate matrix.
4. From the remaining candidate matrices, we remove isomorphic copies. In order to accomplish this, we use the process identical to that used to create the set of all nonisomorphic fixed effects models with n factors.

The results for counting the number of mixed models with $n \leq 5$ factors (with the constraint imposed) are shown in the third column of Table 2.

The constraint that a fixed factor may not be nested within a random factor is sensible when we consider examples of “natural” nesting (for example leaves nested within a tree). However, this constraint may be relaxed to accommodate cases where the nesting is “forced”. For example, suppose there are a total of four varieties of wheat with two varieties planted in Field 1 and two varieties planted in Field 2. If variety is fixed and field is random then this example violates the constraint and yet this situation could occur in practice. Note however, that this situation could also be modelled as an example where variety is crossed with field, but we have missing cells.

We can enumerate the mixed models without the constraint imposed. The algorithm is identical to the one above with Step 3 deleted. The results are shown in fourth column of Table 2.

Table 2: Number of Fixed Effect and Mixed Models with n Factors (Counting With and Without the Constraint)

n	Fixed Effects Models	Mixed Models (Constraint)	Mixed Models (No Constraint)
1	1	2	2
2	2	6	7
3	5	22	32
4	16	101	192
5	64	576	1490

The question remains as to whether models that violate the constraint should be permitted. Hocking [6, p. 542] indicates that models that violate the constraint should not be permitted. In

the next section, the list of mixed models created will be used in a SAS macro. SAS documentation does not forbid models that violate the constraint. However, PROC MIXED does not handle these models correctly as the following example illustrates.

Example 2: Variety (fixed) is nested within field (random). Two varieties are applied to two plots each in Field 1. Two other varieties are applied to two plots each in Field 2. Thus, there are a total of eight observations. The data are shown below as part of the SAS code.

The ANOVA is given in Table 3. There is a column for Type 3 expected mean squares provided by PROC MIXED as well as PROC GLM. Note that the expected mean square for Field differs in the two columns. Direct analytical calculation shows that the EMS given by PROC MIXED is incorrect. The correct EMS(Field) is $\text{Var}(\text{Error}) + 4 \text{Var}(\text{Field}) + \text{Q}(\text{Variety}(\text{Field}))$, as given by PROC GLM, where $\text{Q}(\text{Variety}(\text{Field})) = \theta^T A \theta$, with

$$\theta^T = [V_{1(1)}, V_{2(1)}, V_{1(2)}, V_{2(2)}] \quad (3)$$

and

$$A = \begin{bmatrix} 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ -1/2 & -1/2 & 1/2 & 1/2 \\ -1/2 & -1/2 & 1/2 & 1/2 \end{bmatrix}. \quad (4)$$

Note that $V_{i(j)}$, $1 \leq i \leq 2$, $1 \leq j \leq 2$ represent the variety effects within each field.

Table 3: ANOVA Table for Example 2

Source	df	MS	PROC MIXED EMS	PROC GLM EMS
Field	1	0.076050	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{Field})$	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{Field}) + \text{Q}(\text{Variety}(\text{Field}))$
Variety(Field)	2	0.506900	$\text{Var}(\text{Error}) + \text{Q}(\text{Variety}(\text{Field}))$	$\text{Var}(\text{Error}) + \text{Q}(\text{Variety}(\text{Field}))$
Error	4	0.263000	$\text{Var}(\text{Error})$	$\text{Var}(\text{Error})$

The following code was used to obtain the results in Table 3:

```
data Example2;
input Field Variety y;
cards;
1 1 4.98
1 1 5.07
1 2 4.96
1 2 4.39
2 1 5.40
2 1 5.31
2 2 4.95
2 2 4.52 ;
```

```
proc mixed data=Example2 method=type3;
class Field Variety;
model y=Variety(Field);
random Field;

proc glm data=Example2;
class Field Variety;
model y= Field Variety(Field);
random Field;
run;
```

4 Confidence Intervals for Functions of Variance Components

In many cases, we are interested in estimating and obtaining confidence intervals for functions of variance components. Only the mean square terms for random effects are used to estimate variance components. We will use s_t to indicate the sum of squares for factorial effect t and r_t to indicate the corresponding degrees of freedom. Equating the mean squares to their expected mean squares gives a set of equations, which can be solved to estimate variance components. This is described as the AOV method or Method of Moments (MOM) for estimating variance components. For balanced data situations they are also the unconstrained REML estimates.

If we let λ and ϕ represent the vectors of expected mean squares and variance components, then the relation between these vectors is

$$\lambda = L\phi \quad (5)$$

for a suitable matrix L . By ordering the elements of ϕ appropriately, we note that L can be written as an upper triangular matrix. Thus we see that there is a one-to-one mapping between variance components and the expected mean squares corresponding to random effects.

The estimate of the vector of variance components is then given by

$$\hat{\phi} = L^{-1}\hat{\lambda} \quad (6)$$

where $\hat{\lambda}$ is the vector of mean squares. Note that the estimates of the expected mean squares are nonnegative, but the estimates of the variance components may sometimes be negative.

In addition to estimating single variance components, one may be interested in estimating a linear function of variance components. Either way, the quantity of interest may be expressed as a linear function of expected mean square terms

$$\gamma = \sum_q c_q \lambda_q. \quad (7)$$

An unbiased estimate of γ is given by

$$\hat{\gamma} = \sum_q c_q \frac{s_q}{r_q}. \quad (8)$$

Viewed in this way, we can classify the estimation of linear functions of variance components as follows:

1. $\gamma = c_q \lambda_q$ for some q ,
2. $\gamma = \sum_q c_q \lambda_q$, $c_q > 0, q > 1$,
3. $\gamma = \sum_{q=1}^P c_q \lambda_q - \sum_{r=1}^N d_r \lambda_r$, where $c_q, d_q > 0, P \geq 1, N \geq 1$.

Note that in the following discussion, it is assumed that we have a balanced design. Although the following methods can be used when the design is not balanced, the performance of these methods is untested.

CASE 1. ($\gamma = c_q \lambda_q$) Since $\frac{s_q}{\lambda_q}$ has a chi-squared distribution with r_q degrees of freedom, an exact $1 - 2\alpha$ confidence interval on γ is

$$\left[c_q \frac{s_q}{\chi_{1-\alpha, r_q}^2}, c_q \frac{s_q}{\chi_{\alpha, r_q}^2} \right] \tag{9}$$

where χ_{α, r_q}^2 is the $(1 - \alpha)$ th percentile of a χ^2 random variable with r_q degrees of freedom (i.e., right-tail area = α), [2].

CASE 2. ($\gamma = \sum_q c_q \lambda_q$, $c_q > 0, q > 1$) In this case, there are multiple approximate methods for computing confidence intervals for γ . These include Satterthwaite's Procedure, Welch's Procedure, and the Modified Large-Sample (MLS) Procedure. Among these, the MLS Procedure, proposed by Graybill and Wang, has been found to have superior coverage probabilities. See [2] for details.

The MLS method is a modification of Welch's Procedure which is exact for special cases. This method is exact when all but one of the λ_q are zero or when all but one of the r_q approach infinity. The Graybill-Wang two-sided $1 - 2\alpha$ interval on γ is

$$\left[\hat{\gamma} - \sqrt{\sum_q G_q^2 c_q^2 \left(\frac{s_q}{r_q}\right)^2}, \hat{\gamma} + \sqrt{\sum_q H_q^2 c_q^2 \left(\frac{s_q}{r_q}\right)^2} \right] \tag{10}$$

where

$$G_q = 1 - \frac{r_q}{\chi_{1-\alpha, r_q}} \tag{11}$$

and

$$H_q = \frac{r_q}{\chi_{\alpha, r_q}} - 1. \tag{12}$$

CASE 3. ($\gamma = \sum_{q=1}^P c_q \lambda_q - \sum_{r=1}^N d_r \lambda_r$, where $c_q, d_q > 0$) In this case, the method proposed by Ting, Burdick, Graybill, Jeyaratnam, and Lu is recommended. We will refer to this as the TBGJL method. The $1 - \alpha$ lower confidence bound on γ is given by

$$L = \hat{\gamma} - \sqrt{V_L} \tag{13}$$

where

$$V_L = \sum_{q=1}^P G_q^2 c_q^2 \left(\frac{s_q}{r_q}\right)^2 + \sum_{r=1}^N H_r^2 d_r^2 \left(\frac{s_r}{r_r}\right)^2 + \sum_{q=1}^P \sum_{r=1}^N G_{qr} c_q d_r \frac{s_q s_r}{r_q r_r} + \sum_{q=1}^{P-1} \sum_{t>q}^P G_{qt}^* c_q c_t \frac{s_q s_t}{r_q r_t} \tag{14}$$

$$G_q = 1 - \frac{r_q}{\chi_{1-\alpha}^2} \tag{15}$$

$$H_q = \frac{r_q}{\chi_{\alpha, r_q}} - 1 \tag{16}$$

$$G_{qr} = \frac{(F_{1-\alpha, r_q, r_r})^2 - G_q^2 F_{1-\alpha, r_q, r_r}^2 - H_r^2}{F_{1-\alpha, r_q, r_r}} \tag{17}$$

$$G_{qt}^* = \left[\left(1 - \frac{r_q + r_t}{\chi_{1-\alpha, r_q+r_t}^2}\right)^2 \frac{(r_q + r_t)^2}{r_q r_t} - \frac{G_q^2 r_q}{r_t} - \frac{G_t^2 r_t}{r_q} \right] / (P - 1). \tag{18}$$

If $P = 1$ then G_{qt}^* is defined to be zero. The upper $1 - \alpha$ confidence bound for γ is given by

$$U = \hat{\gamma} + \sqrt{V_U} \tag{19}$$

where

$$V_U = \sum_{q=1}^P H_q^2 c_q^2 \left(\frac{s_q}{r_q}\right)^2 + \sum_{r=1}^N G_r^2 d_r^2 \left(\frac{s_r}{r_r}\right)^2 + \sum_{q=1}^P \sum_{r=1}^N H_{qr} c_q d_r \frac{s_q}{r_q} \frac{s_r}{r_r} + \sum_{r=1}^{N-1} \sum_{u>r}^N H_{ru}^* c_r c_u \frac{s_r}{r_r} \frac{s_u}{r_u} \tag{20}$$

$$H_{qr} = \frac{(1 - F_{\alpha, r_q, r_r})^2 - H_q^2 F_{\alpha, r_q, r_r}^2 - G_r^2}{F_{\alpha, r_q, r_r}} \tag{21}$$

$$H_{ru}^* = \left[\left(1 - \frac{r_r + r_u}{\chi_{1-\alpha, r_r+r_u}^2}\right)^2 \frac{(r_r + r_u)^2}{r_r r_u} - \frac{G_r^2 r_r}{r_u} - \frac{G_u^2 r_u}{r_r} \right] / (N - 1). \tag{22}$$

If $N = 1$ then H_{ru}^* is defined to be zero.

Ting, Burdick, Graybill, Jeyaratnam, and Lu used computer simulation to demonstrate that this interval provides confidence coefficients close to the stated levels over a wide range of conditions; see [10].

There are alternatives to the TBGJL method. Let

$$\gamma = \sum_{q=1}^P c_q \lambda_q - \sum_{r=1}^N d_r \lambda_r = \theta_P - \theta_N. \tag{23}$$

One alternative to the TBGJL method is to apply chi-squared approximations to $\hat{\theta}_P$ and to $\hat{\theta}_N$, then use Howe's method to obtain a confidence interval for $\theta_P - \theta_N$. Refer to [7] for a discussion of Howe's method.

5 A SAS Macro for Computing Confidence Intervals on Linear Functions of Variance Components

At the time of this work, SAS PROC MIXED output generally includes only asymptotic confidence intervals on variance components. Specifically, for parameters that have a lower bound constrained by zero, a Satterthwaite approximation is used to construct a confidence interval. The $1 - 2\alpha$ confidence interval for σ^2 is

$$\left[\frac{\nu \hat{\sigma}^2}{\chi_{1-\alpha, \nu}^2}, \frac{\nu \hat{\sigma}^2}{\chi_{\alpha, \nu}^2} \right] \tag{24}$$

where $\nu = 2Z^2$, and Z is the Wald statistic $\hat{\sigma}^2/se(\hat{\sigma}^2)$. For all other parameters, Wald Z-scores and normal quantiles are used to construct the limits. See [8] for details.

Better methods exist for computing confidence intervals for variance components. Hence, it is believed that a macro that would provide superior intervals would be useful.

Specifically, the goal of the following macro is that the user need only input a data set, the factor names, and an adjacency matrix describing the model. The macro would fit a saturated model to the data and provide an ANOVA table as well as estimates and confidence intervals for the variance components. Additional estimates and confidence intervals for linear functions of variance components can be provided. The following algorithm was used to accomplish these goals.

1. Start with the list of adjacency matrices describing all mixed models with up to five factors (not including the error term). We chose to include only those mixed models which obey the constraint that a fixed factor may not be nested within a random factor.
2. The user input matrix (or a permutation of this matrix) is matched to one of the matrices in the list from Step 1.
3. A (saturated) model statement corresponding to the user input matrix is run through PROC MIXED, generating an ANOVA table with Expected Mean Squares.
4. A matrix (L) defining the linear relationships between EMS and variance components is generated.
5. Using the matrix, L , from above, the appropriate method (Exact, MLS, TBGJL) is used to obtain a confidence interval for each of the variance components.

Example 3: Suppose an experiment is conducted to examine the causes of variation in milk production. At each of two randomly selected farms, three milking machines were selected. Five cows were assigned to each machine, and the amount of milk produced was recorded for each of three days. All the factors are random, and cow is nested within machine which is nested within farm.

SAS User Input:

```
/*Step 1: Entering Variable Names and Number of Observations */
%let V1=Farm;
%let V2=Machine;
%let V3=Cow;
%let Response=Y;
%let n=90;

/*Step 2: Entering the Balanced Data*/
data Cows;
input Farm Machine Cow y;
cards;
1 1 1 2.37
...
2 3 5 3.07 ;
```

Applied Statistics in Agriculture

```

/*Step 3: Entering the Adjacency Matrix*/
data matrix;
input Farm Machine Cow;
cards;
1 1 0
0 1 1
0 0 1 ;

/*Step 4: Requesting a CI for the sum of Variance Components.*/
data Lfuncs;
input Farm MachineFarm CowMachineFarm Residual;
cards;
1 1 1 1 ;
    
```

SAS Macro Output:

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
Farm	1	0.645160	0.645160	Var(Residual) + 3 Var(Cow(Farm*Machine)) + 15 Var(Machine(Farm)) + 45 Var(Farm)
Machine(Farm)	4	1.669182	0.417296	Var(Residual) + 3 Var(Cow(Farm*Machine)) + 15 Var(Machine(Farm))
Cow(Farm*Machine)	24	2.014187	0.083924	Var(Residual) + 3 Var(Cow(Farm*Machine))
Residual	60	5.031600	0.083860	Var(Residual)

The following are approximate 95 % Confidence Intervals
 (except when noted as exact)

PARAMETER	METHOD	LB	ESTIMATE	UB
Farm	TBGJL	-0.061383	0.0050637	14.586546
Machine(Farm)	TBGJL	0.0035236	0.0222247	0.2238027
Cow(Farm*Machine)	TBGJL	-0.01739	0.0000215	0.0270036
Residual	Exact	0.060405	0.08386	0.1242931

LFCNS	PARAMETER	METHOD	LB	ESTIMATE	UB
1 1 1 1	Farm	GWMLS	0.0867974	0.1111699	14.69615
	Machine(Farm)				
	Cow(Farm*Machine)				
	Residual				

Note that this macro can be used even with unbalanced data, but methods used for computing confidence intervals for variance components are further approximations (because the sums of squares

are no longer independent chi-squared random variables) of already approximate methods. The methods considered here remain largely untested for unbalanced data.

Summary

The number of nonisomorphic posets up to order 14 has been derived. Hence the number of nonisomorphic fixed effects models has also been derived because there is a one-to-one correspondence between posets and fixed effects models. Using the list of fixed effects models we have been able to enumerate mixed models with up to five factors. Finally, we created a SAS MACRO to compute confidence intervals for variance components from any of these (saturated) models using the methods discussed in the book “Confidence Intervals for Variance Components” by Burdick and Graybill [2]. A copy of this MACRO is available at www.stat.colostate.edu/~hess/MixedModels.htm.

References

- [1] Garrett Birkhoff. *Lattice Theory*. American Mathematical Society, 3rd edition, 1967.
- [2] Richard K. Burdick and Franklin A. Graybill. *Confidence Intervals on Variance Components*. Marcel Dekker, Inc., New York, NY, 1992.
- [3] Peter J. Cameron. *Combinatorics: Topics, Techniques, Algorithms*. Cambridge University Press, Cambridge, United Kingdom, 1994.
- [4] Claude Chaunier and Nik Lygeros. The number of orders with thirteen elements. *Order*, 9:203–204, 1992.
- [5] Louis Comtet. *Advanced Combinatorics: The Art of Finite and Infinite Expansions*. Reidel Publishing Company, Dordrecht, Holland, 1974.
- [6] Ronald R. Hocking. *Methods and Applications of Linear Models: Regression and the Analysis of Variance*. John Wiley and Sons, Inc., New York, NY, 1996.
- [7] W.G. Howe. Approximate confidence limits on the mean of $x+y$ where x and y are two tabled independent random variables. *Journal of the American Statistical Association*, 69:789–794, 1974.
- [8] SAS Institute. *SAS OnlineDoc Version 8*. SAS Institute, Inc., Cary, NC, 1999.
- [9] George A. Milliken and Dallas E. Johnson. *Analysis of Messy Data, Volume I: Designed Experiments*. Van Nostrand Reinhold, New York, NY, 1984.
- [10] Naitee Ting, Richard K. Burdick, Franklin A. Graybill, S. Jeyaratnam, and Tai-Fang C. Lu. Confidence intervals on linear combinations of variance components that are unrestricted in sign. *Journal of Statistical Computing and Simulation*, 35:135–143, 1990.