VARIANCE COMPONENTS ESTIMATION IN A K-WAY NESTED RANDOM EFFECT MODELS

Adilson J. M. da Silva, PhD Student & Researcher
University of Cape Verde, Cape Verde

Abstract

This work aims to presented the k - way nested random models (designs) and discuss the estimators for the variance components in this kind of models presented by Henderson (1953), proposing condition concern their existence, as well as its two proposed modifications presented by Khattree (1999). At the first proposal, where he choose to sacrifice the unbiasedness of the Henderson's one to preserve the nonnegativity of the variance, and after noting (through simulations) that his estimator has better performance than the Henderson's one except for the variance of the error term, which has less mean square error than the corresponding error term of the Khattree's estimator, Khattree (1999), on is proposed modification, replaced the error term by the error term of the Henderson's estimators. Using Henderson's estimators, method of determining hypothesis tests based on Satterwaite (1946) procedure are discussed as well.

Keywords: Nested random models, variance components, henderson's estimators, khattree's estimators, tests of hypothesis

Introduction

The completed *nested models* arise in many experiments and surveys. Suppose, for sake of motivation, that some local government is interested in choose laboratories to administrates urinalyses to human subject, who work caring for plants with hazardous materials, and nextly several analyses may be made from each urine samples, so that the underlying model involves laboratories, human subjects within the laboratories, tests within human subjects, and analysis with tests. Each one of the factor (stage) may be chooses fixed or in a random way. For a variety of reasons, more sublevels (data) may be available for some levels than for others, i.e., the data are unbalanced. For example, there may be some occasions on which some subjects does not appears to his test, or more analysis is needed at different levels of the factor tests, so that the sublevels within each level may varying. In this case the coefficient of a particular *variance component* in the mean squares expectation will vary from one *mean square* for another, leading to strong difficulties in computation of the *variance components estimations* or in performance of *tests of hypothesis*, situation which does not holds for the case when the data are balanced (see Gates and Shiue (1962)).

This work aims to analyses the estimators for the *variance components* proposed separately by Henderson (1953) and Khattree (1999) in a *k - way nested random model* (see, for instance, Gates and Shiue (1962) for notions of this kind of model), proposing condition on their existence, and the performance of tests *of hypothesis* for the *variance components*, taking the data to be unbalanced and assuming that the observations of the last factor levels, which are randomly taken, constitute the (k-1)th factor (see Tietjen and Moore Tietjen (1968) for tests of hypothesis in k - way random nested models). The approach to the construction of the estimators for *variance components* presented here is proposed by Henderson (1953).

General K-Way Nested Model - Basic Notions

A statistical model is said be a k-way nested one if it consists of k factors, say $A_1, \ldots, A_k$ (see the Figure 1), having each one of than some levels, where:

- The levels of the factor A_k are nested within the levels of the factor A_{k-1};

- The levels of the factor A_{k-1} are nested within the levels of the factor A_{k-2};

•••

- The levels of the factor A_{k-3} are nested within the levels of the factor A_{k-2};

- The levels of the factor A_{k-2} are nested within the levels of the factor A_{k-1}.

The effect associated with any factor is the effect which its levels have on the interest response variable.

One supposes now that:

There is $a_{(k+1)j_k \ldots j_1}$ observations nested within the $j_k th$ level of the factor A_k;

The factor A_k has $a_{kj_{k-1}\cdots j_1}$ levels nested within the $j_{k-1}th$ level of the factor A_{k-1} $(j_{k-1} = 1, \ldots, a_{(k-1)j_{k-2}\ldots j_1})$;

The factor A_{k-1} has $a_{(k-1)j_{k-2}\ldots j_1}$ levels nested within the $j_{k-2}th$ level of the factor A_{k-2} $(j_{k-2} = 1, \ldots, a_{(k-2)j_{k-3}\ldots j_1})$;

•••

The factor A_3 has $a_{3j_2j_1}$ levels nested within the $j_2 th$ level of the factor A_2 $(j_2 = 1, \ldots, a_{2j_1})$;

The factor A_2 has a_{2j_1} levels nested within the $j_1 th$ level of the factor A_1 $(j_1 = 1, \ldots, a_1)$;

The factor A_1 has a_1 levels.

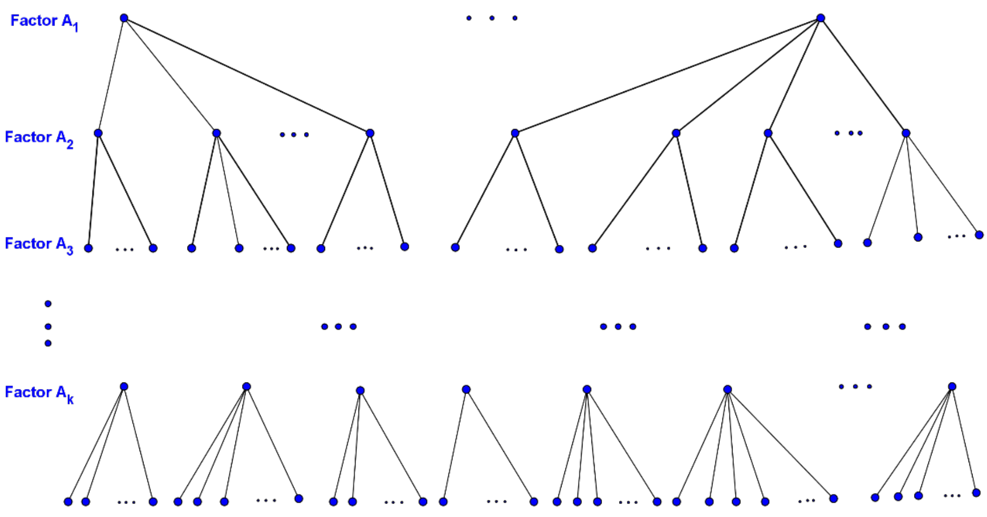

Figure 1: k-way nested model with unbalanced data, with factors $A_1, \ldots, A_k$. The observations (sublevels) nested within the different levels of the factor A_k are assumed to constitute the factor A_{k+1}.

So, a k-way nested random model can be written as:

$$\begin{aligned} y_{j_1 \ldots j_{k+1}} &= \mu + \beta_{j_1} + \beta_{j_2(j_1)} + \beta_{j_3(j_1 j_2)}, \ldots, \beta_{j_k(j_1 \ldots j_{k-1})} + \beta_{j_{k+1}(j_1 \ldots j_k)} \\ &= \mu + \beta_{j_1} + \sum_{i=2}^{k+1} \beta_{j_i(j_1 \ldots j_{i-1})}, \end{aligned} \quad \textbf{(0.1)}$$

with

$$\begin{cases} j_1 = 1, \ldots, a_1, \\ j_i = 1, \ldots, a_{ij_{i-1} \ldots j_1}, \, i = 2, \ldots, k+1, \end{cases}$$

where

$y_{j_1 \ldots j_{k+1}}$ is the $j_{k+1}th$ observation of the $j_k th$ level of the factor A_k nested within the $j_{k-1}th$ level of the factor A_{k-1} nested within the $j_{k-2}th$ level of the factor A_{k-2} nested … nested within the $j_3 th$ level of the factor A_3 nested within the $j_2 th$ level of the factor A_2 nested within the $j_1 th$ level of the factor A_1;

μ represents the general mean;

β_{j_1} is the random effect due to the $j_1 th$ level of the factor A_1;

$\beta_{j_2(j_1)}$ is the random effect due to the $j_2 th$ level of the factor A_2 nested within the $j_1 th$ level of the factor A_1;

$\beta_{j_3(j_1 j_2)}$ is the random effect due to the $j_3 th$ level of the factor A_3 nested within the $j_2 th$ level of the factor A_2 nested within the $j_1 th$ level of the factor A_1;;

●●●

$\beta_{j_k(j_1 \cdots j_{k-1})}$ is the random effect due to the $j_k th$ level of the factor A_k nested within the $j_{k-1}th$ level of the factor A_{k-1} nested … nested within the $j_3 th$ level of the factor A_3 nested within the $j_2 th$ level of the factor A_2 nested within the $j_1 th$ level of the factor A_1;

$\beta_{j_{k+1}(j_1 \cdots j_k)}$ is the random error due to the observation $y_{j_1 j_2 \cdots j_{k+1}}$.

Following Sahai and Ojeda (2005), one assumes the β's to be mutually and completely uncorrelated variables with means zero and variance $Var(\beta_{j_1}) = \sigma_1^2$, \$ $Var(\beta_{j_i(j_1,\ldots,j_{i-1})}) = \sigma_i^2$ $Var(\beta_{j_i(j_1,\ldots,j_{i-1})}) = \sigma_i^2$, $i = 2, \ldots, k$, and $Var(\beta_{j_{k+1}(j_1,\ldots,j_k)}) = \sigma_e^2$ $Var(\beta_{j_{k+1}(j_1,\ldots,j_k)}) = \sigma_e^2$. This last one is the variance of the error term. Here, $\sigma_1^2, \ldots, \sigma_k^2$ $\sigma_1^2, \ldots, \sigma_k^2$ and σ_e^2 σ_e^2 are known as the *variance components* of the response variable.

The Analysis of Variance: The Expected Mean Squares

In order to establish the analysis of variance (ANOVA) sum of squares for each factor, one firstly provides the sums of the number of levels for different factors, as well as the sums of observations at different levels.

Conveniently, one denotes the number of levels at factor A_i as

$$a_{i\bullet} = \sum_{j_{i+1}} \cdots \sum_{j_k} a_{(k+1) j_k \cdots j_1}, \, i = 1, \ldots, k-1,$$

$$a_{k\bullet} = a_{(k+1) j_k \ldots j_1},$$

and the total number of levels (observations) in the sample by $a_{0\bullet} = \sum_{j_1} a_{1\bullet}$.

The general sum of observations, denoted by $y_{0\bullet}$, is

$$y_{0\bullet} = \sum_{j_1} \cdots \sum_{j_{k+1}} y_{j_1 \ldots j_k j_{k+1}}$$

and the sums of observations at different levels is given by:

$$y_{i\bullet} = \sum_{j_{i+1}} \cdots \sum_{j_{k+1}} y_{j_1 \ldots j_k j_{k+1}}, \, i = 1, \ldots, k. \quad (0.2)$$

Thus, the sum of squares for the factor A_1 is given by

$$SS_{A_1} = \sum_{j_1} \frac{y_{1\bullet}^2}{a_{1\bullet}} - \frac{y_{0\bullet}^2}{a_{0\bullet}}; \quad (0.3)$$

The one for the remaining factors A_i, $i = 1, \ldots, k$, by

$$SS_{A_i} = \sum_{j_1} \cdots \sum_{j_i} \frac{y_{i\bullet}^2}{a_{i\bullet}} - \sum_{j_1} \cdots \sum_{j_{i-1}} \frac{y_{i-1\bullet}^2}{a_{i-1\bullet}}, \; i = 2, \ldots, k, \quad (0.4)$$

and the sum of squares for the errors is given by

$$SS_e = \sum_{j_1} \cdots \sum_{j_{k+1}} y_{j_1 \cdots j_k j_{k+1}}^2 - \sum_{j_1} \cdots \sum_{j_k} \frac{y_{k\bullet}^2}{a_{(k+1) j_k \cdots j_1}}.$$

See Sahai and Ojeda (2005) for some additional explanation.

Having in mind the total number of observations at different factors A_i, $i = 1, \ldots, k$, the degrees of freedom, d_i, $i = 1, \ldots, k+1$, at each sources of variation is computed as follows:

$$d_1 = a_1 - 1; \quad d_2 = \sum_{j_1} a_{2(j_1)} - a_1;$$

$$d_i = \sum_{j_1} \cdots \sum_{j_{i-2}} \left(\sum_{j_{i-1}} a_{i(j_{i-1} \cdots j_1)} - a_{i-1(j_{i-2} \cdots j_1)} \right), \quad i = 3, \ldots, k+1; \qquad \mathbf{(0.5)}$$

is the *degrees of freedom* among the levels of the factor A_1, d_i, $i = 1, \ldots, k$, the *degrees of freedom* among levels of the factor A_i nested within the factor A_{i-1}, and d_{k+1} the one among the error factor.

Thus, the *mean square* (which is obtained by dividing the *sum of squares* by its corresponding *degrees of freedom*), denoted here by MS_{A_i}, $i = 1, \ldots, k+1$, can be written as

$$MS_{A_i} = \frac{SS_{A_i}}{d_i}, \; i = 1, \ldots, k,$$

and

$$MS_{A_{k+1}} = \frac{SS_{A_e}}{d_{k+1}}.$$

Now, in what follows, one presents the result concerning *expected mean square*. Such result, which is presented here as a proposition, can be found at Sahai and Ojeda (2005) or, for instance, at Searle at al. (2006).

Proposition 1.

Consider all the results established up to now. Then, the expected mean square at each source of variation is given by

$$E(MS_{A_i}) = EMS_{A_i} = \begin{cases} \sigma_e^2 + c_{i,k}\sigma_k^2 + c_{i,k-1}\sigma_{k-1}^2 + \ldots + c_{i,i}\sigma_i^2, & i = 1, \ldots, k, \\ \sigma_e^2 & i = k+1, \end{cases} \quad (0.6)$$

where $c_{i,s}$, $i \le s \le k$, *are given by*

$$c_{i,s} = \sum_{j_1} \cdots \sum_{j_s} a_{s\bullet}^2 \left[\frac{1}{a_{i\bullet}} - \frac{1}{a_{i-1\bullet}} \right] \frac{1}{d_i}, \; i \le s \le k. \quad (0.7)$$

The proof of the above result is very tedious and expensive in what concern the time to perform it, so one will not give it. Instead of that, one recommends Gates and Shiue (1962) or Sahai and Ojeda (2005) for some more details.

Note 1.

It must be noted that the condition $a_{0\bullet} \neq 0 \neq a_{i\bullet}$, $i = 1, \ldots, k$ *, which means every factor has always at least one level, must hold in order to ensure the existence of the* $c_{i,s}$, $i \leq s \leq k$.

The system of equation (1.6) can be rewritten in the matrix notation as follows:

$$[EMS_{A_1} EMS_{A_2} \vdots EMS_{A_k} EMS_{A_{k+1}}] = \begin{bmatrix} 1 & c_{1,k} & c_{1,k-1} & \cdots & c_{1,3} & c_{1,2} & c_{1,1} \\ 1 & c_{2,k} & c_{2,k-1} & \cdots & c_{2,3} & c_{2,2} & 0 \\ 1 & c_{3,k} & c_{3,k-1} & \cdots & c_{3,3} & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & c_{k-1,k} & c_{k-1,k-1} & \cdots & 0 & 0 & 0 \\ 1 & c_{k,k} & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \sigma_e^2 \\ \sigma_k^2 \\ \sigma_{k-1}^2 \\ \vdots \\ \sigma_3^2 \\ \sigma_2^2 \\ \sigma_1^2 \end{bmatrix}$$

(0.8)

Estimation of the Variance Components

One of the most common procedure for the *estimation of variance components* is the one suggested by Henderson (1953) through is three variations known as *method* 1, *method* 2 , and *method* 3, especially because of its simplicity in what concern the computational implementation (even on a hand-held calculator), and umbiaseness. Although the three methods common underlying idea is to form the (observed) *mean squares* for different factors and then equating them to their respective *expected mean squares* (in some case with some readjustment), leading to a system of linear equations, which solved in the *variance components* gives the corresponding estimator, the scenario in this paper is appropriate for the *method* 1, since all terms β_{j_1} and $\beta_{j_i(j_1 \cdots j_{i-1})}$, $i = 2, \ldots, k+1$, are regarded as random variables. Such estimators, as well as their existence and consistence, are discussed at the subsection which comes next (Subsection 3.1).

Despite its good performance, the Henderson's estimators for the *variance components* abdicate the nonnegativity of variance, situation which is approached by Khattree (1999) on its proposed estimators. The approach proposed by Khattree (1999) which preserve the nonnegativity of variance constitutes a modification to the Henderson's estimators. This is discussed at Subsection 3.2.

Estimators Proposed By Henderson

On this subsection one will present the estimator proposed by Henderson (1953) (making use of its method 1) to obtain the estimators for the *variance components* in models with (completely) random designs, which is the case of the one discussed here (see model (1.1)), and proposes some additional condition over that model in order to get the desired estimators.

Recall the result concerning the *expected mean squares* (the system of equations (1.6) or its matricial notation (1.8)).

Rearranging the matrices involved in (1.8), such result can be, equivalently, rewritten as follows:

$$\begin{bmatrix} EMS_{A_1} \\ EMS_{A_2} \\ EMS_{A_3} \\ \vdots \\ EMS_{A_{k-1}} \\ EMS_{A_k} \\ EMS_{A_{k+1}} \end{bmatrix} = \begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & \cdots & c_{1,k-1} & c_{1,k} & 1 \\ 0 & c_{2,2} & c_{2,3} & \cdots & c_{2,k-1} & c_{2,k} & 1 \\ 0 & 0 & c_{3,3} & \cdots & c_{3,k-1} & c_{3,k} & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_{k-1,k-1} & c_{k-1,k} & 1 \\ 0 & 0 & 0 & \cdots & 0 & c_{k,k} & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \sigma_3^2 \\ \vdots \\ \sigma_{k-1}^2 \\ \sigma_k^2 \\ \sigma_e^2 \end{bmatrix} \qquad (0.9)$$

Now, let

$M = \left[MS_{A_1}\ MS_{A_2}\ MS_{A_3} \ldots MS_{A_{k-1}}\ MS_{A_k}\ MS_{A_{k+1}} \right]^{\bullet}$

be the vector whose the entries are the *mean squares* of the different factors,
$E = \left[EMS_{A_1}\ EMS_{A_2}\ EMS_{A_3} \ldots EMS_{A_{k-1}}\ EMS_{A_k}\ EMS_{A_{k+1}} \right]^{\bullet}$

the vector whose the entries are the *expected mean squares* of the different factors, and

$\sigma = [\sigma_1^2\ \sigma_2^2\ \sigma_3^2 \ldots \sigma_{k-1}^2\ \sigma_k^2\ \sigma_e^2]^{\bullet}$

the vector of the variances of the effects due to different factors (including the one for the error term). In order to find the estimates for the *variance components,* one must solves the following system of linear equation in σ :

$E(M) = E = C\sigma, \qquad (0.10)$

where

$$C = \begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & \cdots & c_{1,k-1} & c_{1,k} & 1 \\ 0 & c_{2,2} & c_{2,3} & \cdots & c_{2,k-1} & c_{2,k} & 1 \\ 0 & 0 & c_{3,3} & \cdots & c_{3,k-1} & c_{3,k} & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_{k-1,k-1} & c_{k-1,k} & 1 \\ 0 & 0 & 0 & \cdots & 0 & c_{k,k} & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}. \qquad (0.11)$$

The system (1.10) yields the (unique) consistent solution

$\hat{\sigma} = C^{-1}M$,

with $\hat{\sigma} = \left[\hat{\sigma}_1^2\ \hat{\sigma}_2^2\ \hat{\sigma}_3^2 \ldots \hat{\sigma}_{k-1}^2\ \hat{\sigma}_k^2\ \hat{\sigma}_e^2 \right]^{\bullet}$, if the matrix C is nonsingular (invertible), having, therefore, the inverse C^{-1}. So one has only to ensure the nonsingularity of the matrix C .

Note 2.

One should note that, in fact, the Henderson's estimator is unbiased. Indeed, $E(\hat{\sigma}) = C^{-1}E(M) = C^{-1}C\sigma = \sigma.$

Proposition 2.

The necessary and sufficient condition for the matrix C in the equation (1.10) to be nonsingular is

$a_{i-1\cdot} \neq a_{i\cdot},\ i = 1, \ldots, k.$

Proof. *Since the matrix C is a triangular one, it is nonsingular if and only if all its diagonal elements are nonzero (this is ensured by the Theorem of Lipschutz (1991)). But this holds if and only if*

$$c_{i,i} = \sum_{j_1} \dots \sum_{j_i} \left[a_{i\bullet} - \frac{a_{i\bullet}^2}{a_{i-1\bullet}} \right] \frac{1}{d_i} \neq 0, i = 1, \dots, k,$$

which, by its turn, holds if and only if $a_{i-1\bullet} \neq a_{i\bullet}, a_{0\bullet} \neq 0 \neq a_{i\bullet}, i = 1, \dots, k$. See Note 1 for the latter inequality explanation

Thus, in order to guarantee that the condition $a_{i-1\bullet} \neq a_{i\bullet}$, $i = 1, \dots, k$, holds, and, therefore, the existence of C^{-1}, it must be assumed that each factor A_i, $i = 1, \dots, k$, has more than one level and each level must have some (sub) levels nested within. One should note that the levels nested within the levels of the factor A_k are, clearly, the observations.

In the next subsection it is discussed the notable modification to the Henderson's estimators proposed by Khattree (1999).

Estimators proposed by Khattree

As seen at the preceding sections and subsections, the Henderson's estimator $\hat{\sigma} = C^{-1}M$ which is a consistent solution to the system $M = C\sigma$ do not preserve the nonnegativity of the variance, preserving instead the unbiasedness. On its suggested modification to Henderson's estimator, Khattree (1999) (see Khattree (1998) as well), unlike Henderson (1953), choose to sacrifice the unbiasedness of its components to guarantee their nonnegativity.

Namely, on its approach, he considered the problem:

$$\min_{\sigma \geq 0} \left\| M - E(M) \right\| = \min_{\sigma \geq 0} \left\| M - C\sigma \right\|, \qquad (0.12)$$

with M, C and σ defined above, and $\|\cdot\|$ (although it may be an appropriate norm) taken to be the Euclidean norm $\|x\| = (x^{\bullet} x)^{\frac{1}{2}}$.

In an equivalent way, the problem (1.12) can be stated as

$$\min_{\sigma \geq 0} \frac{1}{2} \sigma^{\bullet} C^{\bullet} C\sigma, \qquad (0.13)$$

which is a quadratic problem with linear constrains $\sigma \geq 0$. Such problem, as did Lemke (1962), using the primal-dual notion relationship of (1.13) with another optimization problem, in our case can be posed in a different way: *find σ and u* such that

$$\sigma - (C^{\bullet} C)^{-1} u = \hat{\sigma}, \sigma, u \geq 0, \sigma^{\bullet} u = 0, \qquad (0.14)$$

with $\hat{\sigma}$ the Henderson's estimator (see subsection 3.1).

Let $\tilde{\sigma} = \left[\tilde{\sigma}_1^2 \; \tilde{\sigma}_2^2 \; \tilde{\sigma}_3^2 \dots \tilde{\sigma}_{k-1}^2 \; \tilde{\sigma}_k^2 \; \tilde{\sigma}_e^2 \right]^{\bullet}$ be such solution on σ.

Through simulation using a 3-way nested random model, Khattree (1998) remarked (Khattree and Gill (1988) and Ahrens at al (1981) made the same remarks, although in others contexts) that the *mean square error* of its suggested *variances components* estimators $\tilde{\sigma}_i^2$ is generally smaller than the correspondent one of the estimator $\hat{\sigma}$ suggested by Henderson, $\tilde{\sigma}_i^2$, except for the case of the error *variance component* estimator $\tilde{\sigma}_e^2$. For that case he remarked that the result

$$E\left[\left(\sigma_e^2-\tilde{\sigma}_e^2\right)^2\right]\leq E\left[\left(\sigma_e^2-\hat{\sigma}_e^2\right)^2\right]$$

holds in generally, that is $\tilde{\sigma}_e^2$ has in generally less *mean square error* than $\hat{\sigma}_e^2$.

In order to get a estimator with higher performance, Khattree (1998) (see Khattree (1999) as well) combines the “goodness” of his suggested one with the “goodness” of the one suggested by Henderson (1953) performing a modification which applied to our case can be stated as follow.

Let $\sigma_{(1)}=\left[\sigma_1^2\ \sigma_2^2\ \sigma_3^2\ldots\sigma_{k-1}^2\ \sigma_k^2\right]^{\bullet}$, that is the k first *variance components* (dismissing the one of the error term), $M_{(1)}=\left[MS_{A_1}\ MS_{A_2}\ MS_{A_3},\ldots,MS_{A_{k-1}}\ MS_{A_k}\right]^{\bullet}$, the *mean square* of the different first k factors, and

$$C=\begin{bmatrix}C_{(11)} & \mathbf{1}\\ 0 & 1\end{bmatrix}, \qquad (0.15)$$

where $\mathbf{1}$ is an unitary vector with dimension equal to the row number of the sub matrix $C_{(11)}$, so that

$$\sigma=\left[\sigma_{(1)}\ \sigma_e^2\right]^{\bullet},\ M=\left[M_{(1)}\ MS_{A_{k+1}}\right],$$

and once the system of equation (1.10) can be rewritten as

$$\begin{aligned}M_{(1)}&=C_{(11)}\sigma_{(1)}+\sigma_e^2\\ MS_{A_{k+1}}&=\sigma_e^2,\end{aligned},$$

equivalently,

$$M_{(1)}-MS_{A_{k+1}}=C_{(11)}\sigma_{(1)}+\sigma_{(e)}^2,$$

the problem (1.12) can now be stated as

$$\min_{\sigma\geq 0}=\left\|M^*-C_{(11)}\sigma_{(1)}\right\|, \qquad (0.16)$$

with $M^*=M_{(1)}-MS_{A_{k+1}}$, which amounts to find (following Lemke (1962)) $\sigma_{(1)}$ and s such that

$$\sigma_{(1)}-(C_{(11)}^{\bullet}C_{(11)})^{-1}s=\hat{\sigma}_{(1)},\ \sigma_{(1)},s\geq 0,\ \sigma_{(1)}^{\bullet}s=0. \qquad (0.17)$$

Let $\tilde{\sigma}_{(1)}$ be the solution in $\sigma_{(1)}$. Thereby the Khattree suggested estimator for the *variance components* is then

$$\tilde{\tilde{\sigma}}=\left[\tilde{\sigma}_{(1)}\ \hat{\sigma}_e^2\right]^{\bullet},$$

with $\hat{\sigma}_e^2$ the Henderson's estimator for the error term.

Hypothesis Tests for the Variance Components

Tietjen and Moore (1968) proposed a method to constructing an approximate (pseudo) F-tests to test the hypothesis

$$H_0^r:\sigma_r^2=0\ vs\ H_1^r:\sigma_r^2>0,\ r\in\{1,\ldots,k+1\} \qquad (0.18)$$

in nested models with unbalanced data, based on Satterwaite (1946) procedure for testing a linear combination of the mean squares. To perform such tests one construct the ratio $\frac{MS_{A_r}}{D_r}$, where MS_{A_r} is the mean squares error at the factor r, having expectation $EMS_{A_r} = c_{r,r}\sigma_r^2 + c_{r,r+1}\sigma_{r+1}^2 + \ldots + c_{r,k-1}\sigma_{k-1}^2 + c_{r,k}\sigma_k^2 + \sigma_e^2$,

and $D_r = \sum_{i=r+1}^{k+1} l_i MS_{A_i}$ an appropriate linear combination (presented below: equation (1.20)) of the mean squares

$MS_{A_{r+1}}, \ldots, MS_{A_{k+1}}$,

having expected value, i.e., $E(D_r)$, given by

$E(D_r) = c_{r,r+1}\sigma_{r+1}^2 + c_{r,r+2}\sigma_{r+2}^2 + \ldots + c_{r,k-1}\sigma_{k-1}^2 + c_{r,k}\sigma_k^2 + \sigma_e^2$

which has an approximate F distribution with appropriate *degrees of freedom*. The *degree of freedom* of the numerator is d_r (see result (1.5)). To calculate those for the denominator, considering C_r the rth row of the matrix C (see the matrix C in (1.11)), clearly $EMS_{A_r} = C_r\sigma$, so that

$D_r = C_r\hat{\sigma} - c_{r,r}\hat{\sigma}_r^2 = MS_{A_r} - c_{r,r}\hat{\sigma}_r^2$

is the desired denominator of the ratio above.

The *degrees of freedom* for the denominator are usually given by

$$d_r^* = \frac{\left(\sum_{i=r+1}^{k+1} l_i MS_{A_i}\right)^2}{\sum_{i=r+1}^{k+1}\left(\frac{\left(l_i MS_{A_i}\right)^2}{di}\right)}. \qquad (0.19)$$

Let assume now that $\sigma_{k+1}^2 = \sigma_e^2$ and $\hat{\sigma}_{k+1}^2 = \hat{\sigma}_e^2$.

Taking $c_{i,j}^*$ to be the ith row and jth column element of the matrix C^{-1}, i.e., the inverse matrix of the matrix C, noting that

$$D_r = c_{r,r+1}\hat{\sigma}_{r+1}^2 + c_{r,r+2}\hat{\sigma}_{r+2}^2 + \ldots + c_{r,k-1}\hat{\sigma}_{k-1}^2 + c_{r,k}\hat{\sigma}_k^2 + \hat{\sigma}_{k+1}^2$$
$$= c_{r,r+1}\sum_{j=1}^{k+1} c_{r+1,j}^* MS_{A_j} + c_{r,r+2}\sum_{j=1}^{k+1} c_{r+2,j}^* MS_{A_j} + \ldots + c_{r,k}\sum_{j=1}^{k+1} c_{k,j}^* MS_{A_j} + \sum_{j=1}^{k+1} c_{k+1,j}^* MS_{A_j},$$

(0.20)

where $\hat{\sigma}_i^2$, $i,\ldots,k+1$, are the Henderson *variance components* estimators, which are unbiased as showed the Note 2, reorganizing the term on (1.20) and noting that $\sum_{i=r+1}^{k+1} c_{r,i}c_{i,j}^*$ is the coefficient of MS_{A_j}, and also that except for the absence of the nonzero term $c_{r,r}c_{r,j}^*$ (and the $r-1$ terms each equal to zero), the expression $\sum_{i=r+1}^{k+1} c_{r,i}c_{i,j}^*$ is recognized to be the element at the rth row and jth column of $CC^{-1} = I$, where I is the identify matrix of order $k+1$.

Thus, by adding and subtracting the term $c_{r,r}c_{r,j}^*$ to the expression of D_r, one obtain

$$D_k = -\sum_{i=r+1}^{k+1} c_{r,r}c_{r,i}^* MS_{A_i}. \qquad (0.21)$$

The coefficient of MS_{A_r} is zero since the diagonal elements of C^{-1} are the reciprocals of the diagonal of C. Therefore, the *degrees of freedom* for the denominator are given by

$$d_r^* = \frac{D_r^2}{\sum_{i=r+1}^{k+1}\left(\frac{\left(c_{r,r}c_{r,i}^* MS_{A_i}\right)^2}{di}\right)}. \qquad (0.22)$$

This last expression of d_r^* is easily to compute provided the matrix C and its inverse C^{-1}.

So, the test procedure for testing the hypothesis H_o^r vs H_1^r is based on ratio statistic $\frac{MS_{A_r}}{D_r}$, which follow an approximate F - distribution with d_r and d_r^* *degrees of freedom*.

Conclusion

The both Henderson's estimator and Khattree's estimators for the variance components discussed here should not be seen mutually exclusive, seen instead to complement each other, in the sense that when the data are note highly unbalanced or when certain condition on variance components are not violated (the nonnegativity of the variance, for example), the method to find estimators suggested by Henderson (1953) are adequate, as showed Swallow and Monahan (1984) (among others) through Monte Carlo Simulation. Indeed, the underlying approach to his method still play a central role in the variances approaches since it is never totally dismissed, been, instead of, appropriately modified by many researchers (Blackwell at al.~\cite{Blackwell(1991)} proposed to assign zero to the values of the estimators when the correspondent estimative are negative, but this, clearly, compromise their weak optimality of the unbiasedness). On the other hand, the Khattree's estimator and the Khattree's modificate estimator, which as seen constitute a modification to the Henderson's one, inspite of its good performance, sacrifice the unbiasedness (this is preserved for the Henderson's one) of the variance components to guarantee their nonnegativity, and are not explicitly calculated.

Tests of hypothesis presented here are based on the Henderson's estimators. A numerical example for such a tests can be found at Sahai and Ojeda Sahai(2005).

Acknowledgments

This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through PEst-OE/MAT/UI0297/2011(CMA). It was also partially supported by the Núcleo de Investigação em Matemática e Aplicações (NUMAT) of Universidade de Cabo Verde.

References:

G. Casella, R.L. Berger. Statistical Inference, 2nd ed. Duxbury, 2002.

S.M. Dowdy, D. Chilko. Statistics for Research, 3rd ed. Wiley, 2004.

M. Fonseca, T. Mathew, J.T. Mexia and R. Zmy´slony. Tolerence intervals in a two-way nested model with mixed or random effects. *Statistics: A Journal of Theoretical and Applied Statistics*, 41 (4), 289-300, 2007.

K. Krishnamoorthy, T. Mathew. One-sided tolerance limits in balanced and unbalanced one-way random models based, on generalized confidence intervals. *Technometrics*, 46 (1), 44 - 52, 2004.

D.C. Montgomery. Design and Analysis of Experiments, 7th ed. Wiley, 2008.

G. Sharma, T. Mathew. One-sided and two-sided tolerance intervals in general mixed and random effects models using small-sample asymptotics. *Journal of the American Statistical Association*, 107 (497), 258-267, 2012.