Much ado about nothing: the mixed models controversy revisited

Viviana B. Lencina
Departamento de Investigación, FM
Universidad Nacional de Tucumán, Argentina

Julio M. Singer
Departamento de Estatística, IME
Universidade de São Paulo, Brazil

Edward J. Stanek III
Department of Biostatistics and Epidemiology, SPH
University of Massachusetts at Amherst, USA

Abstract

We consider a well-known controversy that stems from the use of two mixed models for the analysis of balanced experimental data with a fixed and a random factor. It essentially originates in the different statistics developed from such models for testing that the variance parameter associated to the random factor is null. The corresponding hypotheses are interpreted as that of null random factor main effects in the presence of interaction. The controversy is further complicated by different opinions regarding the appropriateness of such hypothesis. Assuming that this is a sensible option, we show that the standard test statistics obtained under both models are really directed at different hypotheses and conclude that the problem lies in the definition of the main effects and interactions. We use expected values as in the fixed effects case to resolve the controversy showing that under the most commonly used model, the test usually associated to the inexistence of the random factor main effects addresses a different hypothesis. We discuss the choice of models, and some further problems that occur in the presence of unbalanced data.

Key words: Mixed model, random effect, variance components.

1 Introduction

We consider the analysis of experiments with two factors (A and B) where response is recorded for all (a) levels of factor A and for a randomly selected subset of (b) levels of factor B with r replications for each factor level combination. A typical example is presented in Scheffé (1956a) and discussed by McLean *et al.* (1991); although no details about the data collection process are presented, the authors refer to the study as an industrial experiment with machines (a = 2) taken as the fixed factor and operators (b = 3) taken as the random factor. Each machine produces small parts of the same kind from homogeneous metal pieces, which we assume are the randomly assigned experimental units on which some response variable is measured. The data are reproduced in Table 1 for the sake of self-containment.

Table 1: Data from McLean *et al.* (1991)

Machine (A)	Worker (B)		
	1	2	3
1	51.43	50.93	50.47
	51.28	50.75	50.83
2	51.91	52.26	51.58
	52.43	52.33	51.23

We discuss the appropriate way to interpret and test the hypothesis of null factor B main effects in the presence of AB interaction, motivated by a controversy over two different statistics presumably directed at the same hypothesis. The advocates of each model claim to test the null hypothesis of no factor B effects.

We show that what is being tested is not the same, and that the more commonly used model is not testing the hypothesis that many would consider to be that of null factor B effects. Before we attack the central problem under investigation, the relevance of testing for null main effects (of factor B, for example) in the presence of interaction must be placed into perspective. Some authors like Montgomery (1997, p.174) or Lindman (1992, p.103) claim that this approach should be ruled out and that the analysis of factor B main effects should be performed at each level of factor A. Other authors like John (1971, p.70) or Mason *et al.* (1989, p.124) do not specify strategies to deal with the problem. Cox (1984, p.16) or Neter *et al.* (1996, p.808) identify situations where such alternative may be reasonable, as when the effects of factor B at the various levels of factor A have the same direction, but different magnitudes. Cox (1984) refers to such interactions as quantitative while Neter *et al.* (1996, p.808) term them unimportant. It is under such settings that the question over how to appropriately test the null hypothesis of null factor

B main effects in the presence of AB interaction is relevant.

In particular, suppose that one of the machines in the example of McLean *et al.* (1991) is an old one, requiring more ability to operate while the other is a new one, easier to operate. Suppose further, that more qualified workers tend to produce similar results (on the average) with either machine, but that less qualified workers tend to produce different results when working with the old or the new machine (i.e., a machine x worker interaction is expected). We could still be interested in verifying whether (on the average) there are worker main effects, i.e., whether the average results (across machines) are different among workers. We focus our attention on making inferences for a population of workers and on a formal test of the hypothesis of no worker main effects.

The tests that generate the controversy stem from two common models, the constrained parameters model (*CP*), and the unconstrained parameters model (*UP*) used in the analysis of data with the structure described above. This has been recently discussed by Voss (1999) with reader reactions by Hinkelmann (2000), Wolfinger and Stroup (2000) and a reply by Voss (2000). Wolfinger and Stroup (2000) recommend the analysis based on the UP model and Hinkelmann (2000) emphasizes that the general preference for such a model is mainly explained by the existence of computer software. In light of the reader reactions, Voss (2000) appears to reverse his previous recommendation.

We show that Voss's apparent initial conclusion is correct and recommend use of the *CP* model for testing the hypothesis of interest here. Our results are based on a definition of the random factor main effects that corresponds to the classical definition under fixed effects models and their relation to the associated parameters. We begin by defining the *UP* and *CP* models, and briefly discussing the ample literature surrounding this controversy in Section 2. Keeping in mind that main effects may be easily understood as changes in expected values, in Section 3 we first define main effects and interactions under a general model, then indicate how they are related to the corresponding parameters of the *UP* and *CP* models. We show that the former is not compatible with the hypothesis of interest. In Section 4, we simulate an example to illustrate our conclusions. Section 5 is devoted to computational aspects and extensions to the case where the numbers of replications are not equal. Finally, we present a general discussion in Section 6.

2 Definition of the Models and Background

Adopting the usual ANOVA notation, the *unconstrained parameters* (*UP*) *model* in Voss's (1999) terminology, may be expressed as

$$Y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + E_{ijk}, \tag{1}$$

$i = 1, ..., a$, $j = 1, .., b$ and $k = 1, ..., r$, with $\sum_{i=1}^{a} \alpha_i = 0$ and $B_j \sim N(0, \sigma_B^2)$, $(\alpha B)_{ij} \sim N(0, \sigma_{\alpha B}^2)$ and $E_{ijk} \sim N(0, \sigma^2)$ representing independent random variables. Under model (1), α_i is associated with the fixed level i of factor A, B_j is associated with the j^{th} randomly selected level of factor B and $(\alpha B)_{ij}$ is associated with the interaction of level i of factor A with the j^{th} randomly selected level of factor B. In his formulation, Voss (1999) does not consider the restrictions on the fixed effects, but we prefer to include them for identifiability purposes. Many authors (Searle (1971), Milliken and Johnson (1984) and the SAS (1990) software, for example) take $H_0 : \sigma_B^2 = 0$ as the hypothesis of no factor B main effects, concluding that $MSB/MSAB$ is the appropriate test statistic.

The competing model, termed the *constrained parameters* (*CP*) *model* by Voss (1999) corresponds to

$$Y_{ijk} = \eta + \tau_i + D_j + (\tau D)_{ij} + E_{ijk}, \tag{2}$$

$i = 1, ..., a$, $j = 1, .., b$ and $k = 1, ..., r$, with $\sum_{i=1}^{a} \tau_i = 0$, $D_j \sim N(0, \sigma_D^2)$, $(\tau D)_{ij} \sim N(0, (a-1)\sigma_{\tau D}^2/a)$ and $E_{ijk} \sim N(0, \sigma^2)$ representing independent random variables; additionally, $\sum_{i=1}^{a} (\tau D)_{ij} = 0$, $j = 1, .., b$, which implies $Cov((\tau D)_{ij}, (\tau D)_{i'j}) = -\sigma_{\tau D}^2/a$ for $i \neq i'$ and is zero otherwise. Under model (2), τ_i is associated with the fixed level i of factor A, D_j is associated with the j^{th} randomly selected level of factor B and $(\tau D)_{ij}$ is associated with the interaction of level i of factor A with the j^{th} randomly selected level of factor B. The hypothesis $H_0 : \sigma_D^2 = 0$ is taken as the hypothesis of no factor B main effects; the standard test rejects H_0 if MSB/MSE is sufficiently large. This is the approach considered in Neter *et al.* (1996), Montgomery (1991) and others.

The tests obtained under the *UP* and the *CP* models (see Voss (1999), for example) may generate conflicting results for testing what is thought of as the hypothesis of null factor B main effects and this is the source of the controversy.

Both the *UP* and *CP* models are special cases of the following model presented by Scheffé (1959)

$$Y_{ijk} = \mu + \alpha_i + M_{ij} + E_{ijk}, \tag{3}$$

$i = 1, ..., a$, $j = 1, .., b$ and $k = 1, ..., r$, with $\sum_{i=1}^{a} \alpha_i = 0$, $M_{ij} \sim N(0, p+q)$, $Cov(M_{ij}, M_{i'j}) = q$, $i \neq i'$, $Cov(M_{ij}, M_{i'j'}) = 0$, for all $j \neq j'$ and $E_{ijk} \sim N(0, \sigma^2)$ denoting independent random variables which in turn are independent of the M_{ij}.

From (3), it follows that $Var(M_{1j}, \ldots, M_{aj}) = p\mathbf{I}_a + q\mathbf{J}_a$ where $\mathbf{I}_a$ denotes an $(a \times a)$ identity matrix and $\mathbf{J}_a$ denotes an $(a \times a)$ matrix with all elements equal to 1. Under the *UP* model, we have $M_{ij} = B_j + (\alpha B)_{ij}$ with $B_j \sim N(0, \sigma^2_B)$ and $(\alpha B)_{ij} \sim N(0, \sigma^2_{\alpha B})$ so that $p = \sigma^2_{\alpha B}$ and $q = \sigma^2_B$. On the other hand, under the *CP* model, $M_{ij} = D_j + (\tau D)_{ij}$ with $D_j \sim N(0, \sigma^2_D)$, $(\tau D)_{ij} \sim N(0, (a-1)\sigma^2_{\tau D}/a)$, $Cov((\tau D)_{ij}, (\tau D)_{i'j}) = -\sigma^2_{\tau D}/a$ so that $p = \sigma^2_{\tau D}$ and $q = \sigma^2_D - \sigma^2_{\tau D}/a$.

Taking $\mu = \eta$, $\alpha_i = \tau_i$, $\sigma^2_B = \sigma^2_D - \sigma^2_{\tau D}/a$ and $\sigma^2_{\alpha B} = \sigma^2_{\tau D}$, Hocking (1973) shows that the *UP* and *CP* models are equivalent. Therefore, the hypothesis $H_0 : \sigma^2_B = 0$ under the *UP* model is equivalent to $H_0 : \sigma^2_D - \sigma^2_{\tau D}/a = 0$ under the *CP* model; alternatively, the hypothesis $H_0 : \sigma^2_D = 0$ under the *CP* model is equivalent to $H_0 : \sigma^2_B + \sigma^2_{\alpha B}/a = 0$ under the *UP* model. This indicates that $H_0 : \sigma^2_B = 0$ and $H_0 : \sigma^2_D = 0$ are not equivalent and that the controversy is not related to the choice between the two contending models but to the definition of the random factor main effects and its expression in terms of the components of each model.

Hocking (1973) addresses this issue, but he suggests that the choice between the two competing hypotheses, $H_0 : \sigma^2_B = 0$, under the *UP* model or $H_0 : \sigma^2_D = 0$, under the *CP* model commonly interpreted as that of no factor B main effects should be made by the researcher in the light of the subject matter problem under investigation. Voss (1999), on the other hand, attempts to resolve the controversy by embedding both models in a finite population setting (which he terms a super-population) in which all possible b^* levels of the random factor (B) are included and defines the main effects of B as in a fixed effects model. This basic idea dates back to Tukey (1949) and Scheffé (1956b). Voss (1999) relates either the random variable D_j in the *CP* model, or the random variable $\overline{B}_j + \overline{(\alpha B)}_{.j} - \overline{B}. - \overline{(\alpha B)}_{..}$, with $\overline{(\alpha B)}_{.j}$, $\overline{B}.$ and $\overline{(\alpha B)}_{..}$ denoting $a^{-1} \sum_{i=1}^{a} (\alpha B)_{ij}$, $b^{*-1} \sum_{j=1}^{b^*} B_j$

and $(ab^*)^{-1} \sum_{i=1}^{a} \sum_{j=1}^{b^*} (\alpha B)_{ij}$ respectively, in the *UP* model to the B main effects. Then he shows via conditioning on the b selected levels of the random factor that under either model, $E(MSB)$ measures error variability plus "main effects of B", concluding that the appropriate test statistic for null factor B main effects is MSB/MSE in both cases. Voss (1999) neither defines the main effects of the random factor explicitly , nor extends results to the case where the random factor has infinite levels.

Wolfinger and Stroup (2000) comment that placing both the *UP* and the *CP* models under a finite population setup does not resolve the controversy because this is where the *CP* model has its theoretical basis. They recommend use of the *UP* model coupled with a likelihood-based inference approach and discuss the hypotheses in terms of the covariance structure. They do not try to define or interpret the main effects of the random factor nor the variance components in terms of potentially observable quantities. The lack of a definition for the main effects of the random factor B seems to be the major cause of the misunderstanding. The objective of this paper is to specify the relation between the main effects of the random factor B and the parameters in both models.

3 Definitions of main effects and interactions in mixed models

Using model (3), we may define main effects and interactions in terms of expected values, just as in the case of fixed effects models. First, we clarify the underlying probability structure. Let Ω denote the set of all levels of factor B and Θ denote the set of all experimental units. Then, let Y_{ijk}, M_{ij} and E_{ijk} be random variables defined on the space $\Omega \times \Theta$ such that

i) For a given $(\omega, \theta) \in \Omega \times \Theta$, neither the response nor the random error term depend on the order of selection of ω and θ, i.e.,

$$Y_{ijk}(\omega, \theta) = Y_{ij'k'}(\omega, \theta) = Y_{i**}(\omega, \theta),$$

and

$$E_{ijk}(\omega, \theta) = E_{ij'k'}(\omega, \theta) = E_{i**}(\omega, \theta).$$

ii) For a given ω, the random effect term depends neither on θ nor on the order of selection of ω, i.e.,

$$M_{ij}(\omega,\theta) = M_{ij'}(\omega,\theta) = M_{i*}(\omega,\theta).$$

iii) For a given ω, the random effect term has a null expected value, i.e.,

$$E_\Theta(E_{ijk}(\omega,\cdot)) = 0.$$

Then, under the assumptions of model (3), we have $E_{\Omega\times\Theta}(Y_{ijk}(\cdot,\cdot)) = \mu + \alpha_i$, $\mu = a^{-1}\sum_{i=1}^{a} E_{\Omega\times\Theta}(Y_{ijk}(\cdot,\cdot))$, $E_{\Omega\times\Theta}(M_{ij}(\cdot,\cdot)) = 0$ and $E_\Theta(M_{ij}(\omega,\cdot)) = M_{i*}(\omega)$.

The main effect of level i of factor A may be defined as

$$\frac{1}{br}\sum_{j=1}^{b}\sum_{k=1}^{r} E_{\Omega\times\Theta}(Y_{ijk}(\cdot,\cdot)) - \frac{1}{abr}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} E_{\Omega\times\Theta}(Y_{ijk}(\cdot,\cdot)) = \alpha_i,$$

and the hypothesis of null factor A main effects reduces to $\alpha_i = 0$, for all $i = 1,\ldots,a$, or equivalently, to $\sum_{i=1}^{a}\alpha_i^2 = 0$.

Similarly, for every $\omega \in \Omega$, the main effect of level ω of factor B may be defined as

$$\frac{1}{abr}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} E_\Theta(Y_{ijk}(\omega,\cdot)) - \frac{1}{abr}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} E_{\Omega\times\Theta}(Y_{ijk}(\cdot,\cdot)) = \frac{1}{a}\sum_{i=1}^{a} M_{i*}(\omega) = \overline{M}_{\cdot *}(\omega).$$

This definition of factor B main effects is similar to that of fixed factor models. It is the difference between the expected response for a fixed level ω of factor B and the overall expected response. In the example described in the Introduction, this corresponds to the difference between the expected response across both machines for worker ω and the overall expected response (across all workers and both machines). Therefore, testing that the factor B main effects are null corresponds to testing that $\overline{M}_{\cdot *}$ is null for (almost) all $\omega \in \Omega$, which, in turn, is equivalent to testing that $Var_\Omega(\overline{M}_{\cdot *}(\cdot)) = 0$, or equivalently that $q + a^{-1}p = 0$ if we refer to model (3) variance components. Under the *UP* model, it follows that $Var_\Omega(\overline{M}_{\cdot *}(\cdot)) = 0$ corresponds to $\sigma_B^2 + \sigma_{\alpha B}^2/a = 0$; under the CP model, on the other hand, $Var_\Omega(\overline{M}_{\cdot *}(\cdot)) = 0$ is equivalent to $\sigma_D^2 = 0$ so that in both cases, the appropriate test is based on MSB/MSE.

We may also define the interaction between level i of factor A and level ω of factor B as

$$\frac{1}{br}\sum_{j=1}^{b}\sum_{k=1}^{r}[E_\Theta(Y_{ijk}(\omega,\cdot)) - E_{\Omega\times\Theta}(Y_{ijk}(\cdot,\cdot))] - \frac{1}{abr}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}[E_\Theta(Y_{ijk}(\omega,\cdot)) - E_{\Omega\times\Theta}(Y_{ijk}(\cdot,\cdot))] =$$

$$M_{i\cdot}(\omega) - \overline{M}_{\cdot *}(\omega).$$

As in fixed factor models, this interaction measures how the effect of level ω of factor B under level i of factor A differs from the overall effect of level ω of factor B, so that null interaction corresponds to $\sum_{i=1}^{a} Var_\Omega[M_{i\cdot}(\cdot) - \overline{M}_{\cdot *}(\cdot)] = 0$, or explicitly, to $p = 0$ under model (3). Again using (1) and (2), we may show that this corresponds to $\sigma^2_{\alpha B} = 0$ under the *UP* model or to $\sigma^2_{\tau D} = 0$ under the *CP* model.

At this point we observe that the controversy occurs only in the presence of interaction ($\sigma^2_{\alpha B} > 0$ or $\sigma^2_{\tau D} > 0$). If we agree with the definition of factor B main effects, the *UP* model does not allow us to test the hypothesis that they are null in the presence of interaction, since $\sigma^2_B + a^{-1}\sigma^2_{\alpha B} = 0$ and $\sigma^2_{\alpha B} > 0$ are incompatible conditions.

4 Simulated illustration

We simulate some situations with different patterns of factor B main effects and AB interactions under model (3) and use them to illustrate their relations to the parameters in the *CP* and *UP* models. For simplicity and without loss of generality, we consider only two levels for the fixed factor A. Reasoning as in fixed effects models, we interpret factor B main effects as differences between the expected responses at factor B levels and the overall expected response; if such differences are all null, we say that the factor B main effects are null. Similarly, AB interactions may be interpreted as differences between the expected responses at the two levels of factor A for each level of factor B . If all such differences are equal, we say that the AB interaction is null. The patterns we consider are described in Table 2.

The expected responses at the different levels of factor B may be generated from (3) disregarding the variability of the unit effects E_{ijk}. We arbitrarily chose different values for μ, α_i, p and q and generate 100 expected responses (corresponding to 100 levels of factor B) for each case in Table 2. In the context of the industrial experiment described in Section 1, the expected response for worker ω in machine i is $\mu + \alpha_i + M_{i.}(\omega)$ so that the required sample of expected responses may be obtained by generating a random sample of size 100 from a normal distribution with mean $\mu + \alpha_i$ and variance $p + q$. Theoretically, an infinite number of expected responses could be generated. The results (ordered by the

Table 2: Patterns of factor B main effects and AB interactions

Case	Main effects and interaction pattern	Restrictions on parameters		
		Scheffé´s Model	*UP* Model	*CP* Model
1	B main effects + AB interaction	$p > 0$, $q > 0$	$\sigma_B^2 > 0$, $\sigma_{\alpha B}^2 > 0$	$\sigma_D^2 > \sigma_{\tau D}^2/2 > 0$
2	B main effects + AB interaction	$p > 0$, $q = 0$	$\sigma_B^2 = 0$, $\sigma_{\alpha B}^2 > 0$	$\sigma_D^2 > 0$, $\sigma_{\tau D}^2 > 0$
3	Null B main effects + AB interaction	$p > 0$, $q = -a^{-1}p$	Incompatible	$\sigma_D^2 = 0$, $\sigma_{\tau D}^2 > 0$
4	B main effects + null AB interaction	$p = 0$, $q > 0$	$\sigma_B^2 > 0$, $\sigma_{\alpha B}^2 = 0$	$\sigma_D^2 > 0$, $\sigma_{\tau D}^2 = 0$

magnitude of factor B expected responses) are presented in Figures 1, 2, 3, and 4.

Figures 1 and 2 indicate non-null factor B main effects (different expected responses across machines for different workers in reference to the example) in the presence of AB interaction (varying differences in expected response under machines 1 and 2 for different workers). Since the values in Figure 2 were generated with $\sigma_B^2 = 0$, it seems clear that the corresponding hypothesis is not related to null factor B main effects. Figure 3, on the other hand, depicts a situation with null factor B main effects (equal expected responses across machines 1 and 2 for different workers) in the presence of AB interaction. This corresponds to a situation for which the *UP* model is not appropriate, but is clearly compatible with the hypothesis $\sigma_D^2 = 0$. The situation presented in Figure 4 indicates that both models are equivalent for cases where there are factor B main effects but null AB interactions.

Figure 1: Factor B means at levels A_1 and A_2 of factor A $(p = 32; q = 48)$

Insert Figure 1 here

Figure 2: Factor B means at levels A_1 and A_2 of factor A $(p = 16; q = 0)$

Insert Figure 2 here

Figure 3: Factor B means at levels A_1 and A_2 of factor A ($p = 32; q = 16$)

Insert Figure 3 here

Figure 4: Factor B means at levels A_1 and A_2 of factor A ($p = 0; q = 16$)

Insert Figure 4 here

For each case described in Table 2, we selected a simple random sample of 10 levels of factor B and simulated $r = 6$ replicates from the distribution of the response error with $\sigma_D^2 = 2$, i.e., we simulated the response variable for 6 units under each combination of the levels of factor A and selected levels of factor B. The samples were selected with replacement out of the 100 available levels of factor B to mimic a random selection from an infinite population. In Table 3, we present the mixed model ANOVA results obtained under both the CP and UP models.

Table 3: Mixed model ANOVA for the simulated data

Case	Tests for null variance components		
	$\sigma_D^2 = 0$	$\sigma_B^2 = 0$	$\sigma_{\alpha B}^2 = 0$ or $\sigma_{\tau D}^2 = 0$
1	F=198.02 (df= 9, 100) $p < 0.001$	F=4.24 (df= 9, 9) $p = 0.021$	F=46.65 (df= 9, 100) $p < 0.001$
2	F=42.72 (df= 9, 100) $p < 0.001$	F=0.94 (df= 9, 9) $p = 0.537$	F=45.51 (df= 9, 100) $p < 0.001$
3	F=0.83 (df= 9, 100) $p = 0.591$	F=0.021 (df= 9, 9) $p = 1$	F=36.57 (df= 9, 100) $p < 0.001$
4	F=38.91 (df= 9, 100) $p < 0.001$	F=72.79 (df= 9, 9) $p < 0.001$	F = 0.535 (df = 9, 100) $p = 0.846$

The conclusions are all consistent with models underlying the generated data. In particular we draw attention to the non-significant result of the test for $\sigma_B^2 = 0$ in Case 2, where the hypothesis of a null factor B main effect is clearly false.

5 Extensions to unbalanced data

Many authors like Wolfinger and Stroup (2000) discourage the use of the *CP* model on account of the lack of flexibility of the available software for the *UP* model as well as of the difficulty in dealing with unbalanced data (i.e., with different numbers of replicates). The developments for unbalanced data are more complex, although the definitions of main effects given in Section 3 still hold. When unbalance is due to unequal numbers of replicates, it is well known that the usual ANOVA test statistic for $\sigma_B^2 = 0$ does not follow an F distribution. Öfversten (1993) proposes exact tests for null variance components in unbalanced mixed linear models that coincide, in balanced cases, to the usual exact F-tests. These tests may be implemented under the *UP* model and with little additional effort, under the *CP* model, but they are not yet available in commercial statistical software packages. Otherwise, we may rely on likelihood ratio tests, but we must recall that under the null hypotheses, the parameters lie on the boundary of the parametric space and the test statistics follow approximate chi-bar distributions, as indicated in Self and Liang (1987) and Stram and Lee (1994).

Although software like Proc GLM and Proc MIXED in SAS are useful tools for the analysis of unbalanced mixed models, they do not provide exact tests for null variance components. In fact, the available (Wald) tests in Proc MIXED are only rough approximations. On the other hand, according to the definitions in Section 3, a test for $\sigma_B^2 = 0$, under the assumption that $\sigma_{\alpha B}^2 > 0$, does not correspond to a test of null random factor main effects.

Using the data in Table 1, we compare the different tests available in SAS with that proposed by Öfversten (1993). The response for worker 2 and machine 2 (52.33) was deleted to induce unbalance in the data. The results are displayed in Table 4 for the balanced data and in Table 5 for the unbalanced data.

Since both the ANOVA and the GLM procedures in SAS rely on the *UP* model, the balanced data tests for $\sigma_B^2 = 0$ and $\sigma_{\alpha B}^2 = 0$ are obtained directly from the output. However, if we really want to test null factor B main effects we should test for $\sigma_B^2 + \sigma_{\alpha B}^2/a = 0$, but such a test is not given directly by SAS. Under the UP model, SSB$/(\sigma^2 + ar\sigma_B^2 + r\sigma_{\alpha B}^2)$ and SSE$/\sigma^2$ are chi-squared independent random variables, so that, under

Table 4: Tests for null variance components for the balanced data in McLean *et al.* (1991)

	Tests for null variance components		
	σ_D^2	σ_B^2	$\sigma_{\alpha B}^2$ or $\sigma_{\tau D}^2$
ANOVA			
proc ANOVA	F=11.95 (df=2,6)	F=3.85 (df=2,2)	F=3.10 (df=2,6)
proc GLM	P=0.0081	P=0.2061	P=0.1188
Exact Test			
Öfversten (1993)	F=11.95 (df=2,6)	F=3.85 (df=2,2)	F=3.10 (df=2,6)
	P=0.0081	P=0.2061	P=0.1188
Wald Test			
proc MIXED	Z=0.92	Z=0.72	Z=0.67
	P=0.3601	P=0.2368	P=0.2526

$H_0 : \sigma_B^2 + \sigma_{\alpha B}^2/a = 0$, MSB/MSE is an exact F statistic, with $b-1$ and $ab(r-1)$ degrees of freedom, and we can use MSB and MSE from the SAS output to compute this test. If we consider the *CP* model, the corresponding test for $\sigma_D^2 = 0$ can be obtained similarly.

For unbalanced data, the GLM procedure produces an approximate F-test for $\sigma_B^2 = 0$ obtained from type III sums of squares by reproducing the balanced data computations, i.e., by using $MSB/MSAB$. Computing the test for $\sigma_D^2 = 0$ under the *CP* model in the same fashion, i.e., using MSB/MSE, we obtain an exact F-test that coincides with that obtained using Öfversten's techniques.

Öfversten (1993) mentions that his procedure produces unbalanced data exact F-tests for any situation where an exact F-test may be obtained under the corresponding balanced data case. However, it is not clear how to perform the computations required to test $\sigma_B^2 + \sigma_{\alpha B}^2/a = 0$ under the *UP* model.

6 Discussion

The apparent controversy for balanced 2 factor mixed models ANOVA stems from the similarity between the two competing models (*UP* and *CP*) and the usual fixed effects model. The authors advocating one or the other seem to base their conclusion on the variance parameter (σ_B^2 or σ_D^2) associated with the random factor (B_j in the *UP* model and D_j in the *CP* model), and assume that testing whether they are null is equivalent

Table 5: Tests for null variance components for the unbalanced data in McLean *et al.* (1991)

	Tests for null variance components		
	σ^2_D	σ^2_B	$\sigma^2_{\alpha B}$ or $\sigma^2_{\tau D}$
ANOVA			
proc GLM	F=9.77 (df=2,5)	F=5.55 (df=2,2)	F=1.76 (df=2,5)
	P=0.0187	P=0.1527	P=0.2636
Exact Test			
Öfversten (1993)	F=9.77 (df=2,5)	F=9.11 (df=2,2)	F=1.76 (df=2,5)
	P=0.0187	P=0.0989	P=0.2636
Wald Test			
proc MIXED	Z=0.89	Z=0.81	Z=0.29
	P=0.3745	P=0.2100	P=0.3860

to testing for null factor B main effects. Voss (1999) attempts to resolve the controversy by embedding the problem in a setting where the number of levels of the random factor is finite but seems not to address the crux of the problem which, we believe, lies in an appropriate definition of the random factor main effects and interaction.

Characterizing null factor B main effects by asserting that (almost) all factor B expected responses are equal, resolves the controversy. We may recall the work of Scheffé (1959), who defines the random factor main effects under a more general model where $Var(M_{1j}, \ldots, M_{aj}) = \mathbf{\Sigma}$, a symmetric positive semi-definite matrix. The results also hold in this case.

An alternative approach to the problem is to base the analysis on the choice of some special structure for the covariance matrix for the data and to specify hypotheses directly in terms of the dispersion parameters. Under such an approach, no attention is placed on the interpretation of the dispersion parameter, or the main effects of the random factor. We may test if some variance component is zero, but unless the variance component has a meaningful interpretation, it is not clear what is accomplished. According to Wolfinger and Stroup (2000), testing that $\sigma^2_B = 0$ under the *UP* model corresponds to testing whether observations sharing the same level of B are equicorrelated. This is not true, since the covariance between the responses obtained under the same level of factor B is $\sigma^2_B + \sigma^2_{\alpha B}$ if they share the same level of factor A and is σ^2_B otherwise. We have shown that $\sigma_B^2 = 0$ is related to the main effects of factor B but it does not mean that they are null. For such, under this model, we need the additional requirement that $\sigma_{\alpha B}^2 = 0$, but then, there is no controversy at all.

We also note that the choice of the *UP* model advocated by many authors is recommended in view of its computational advantages as well as that of being more easily employed in the analysis of unbalanced cases. McLean *et al.* (1991) suggest that when the purpose is to evaluate estimable functions and the associated standard errors, mixed model procedures (MMP) may be used (for both the *UP* and *CP* models) without any distinction between balanced and unbalanced cases to construct prediction intervals for the random effects. Even in such cases, formal tests of hypotheses might be of interest when an overall picture needs to be drawn.

We believe that the controversy under discussion here is a consequence of the formal similarity between the *CP* and *UP* models and we agree with Wilk and Kempthorne (1955), who state in a more general context, that "Because the models employed are often not explicitly related to the experimental situation, there has been some difficulty in deciding just what the analysis of variance measures". In fact, these authors develop a framework for analysis of variance based on finite population models and on the sampling/randomization characteristics of the underlying experiments in such a way that factor level effects are defined prior to introducing random variables that represent the sampling or randomization. This framework is consistent with the ideas of potentially observable responses that underlies inference in experimental design. Their proposal encompasses the two-factor mixed model as a special case and the main effects for the random factor are tested in the same manner as in the *CP* model. Zyskind (1962) formalized and expanded this setting to a broad variety of experimental and sampling designs so that the results considered here are potentially applicable in more general setups.

Acknowledgements

We are grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil and National Institutes of Health (NIH-R01-HD36848), USA, for financial support. We also appreciate the enlightening comments of an anonymous referee.

References

Cox, D.R. (1984). Interaction. *International Statistical Review*, **52**, 1-31.

Hinkelmann, K. (2000). Resolving the Mixed Models Controversy, Comments. *The American Statistician*, **54**, 228.

Hocking, R.R. (1973). A Discussion of the Two-Way Mixed Model. *The American Statistician*, **27**, 148-152.

John, P.W.M. (1971). *Statistical Design and Analysis of Experiments*. New York: Macmillan.

Kempthorne, O. (1957). *An Introduction to Genetic Statistics*. New York: Wiley.

Lindman, H.R. (1992). *Analysis of Variance in Experimental Design*. New York: Springer-Verlag.

Mason, R.L, Gunst, R.F. and Hess, J.L. (1989). *Statistical Design and Analysis of Experiments with Applications to Engineering and Science*. New York: Wiley.

McLean, R.A., Sanders W.L. and Stroup, W.W. (1991). A Unified Approach to Mixed Linear Models. *The American Statistician*, **45**, 54-63.

Milliken, G.A. and Johnson, D.E. (1984). *Analysis of Messy Data: Vol. 1. Designed Experiments*. Belmont, CA: Lifetime Learning Publications.

Montgomery, D.C. (1997). *Design and Analysis of Experiments, 5th edition*. New York: Wiley.

Neter, J., Wasserman, W., Kutner, M.H. and Nachtsheim, C.J. (1996). *Applied Linear Statistical Models, 4th ed.*. Homewood, Ill: Irwin.

Öfversten, J. (1993). Exact Test for Variance Components in Unbalanced Mixed Linear Model. *Biometrics*, **49**, 45-57.

SAS Institute. Inc. (1990). *SAS Users Guide: Statistics (Version 6).* Cary, NC.

Scheffé, H. (1956a). A 'Mixed Model' for the Analysis of Variance. *The Annals of Mathematical Statistics*, **27**, 23-36.

Scheffé, H. (1956b). Alternative Models for the Analysis of Variance. *The Annals of Mathematical Statistics*, **27**, 251-271.

Scheffé, H. (1959) *The Analysis of Variance*. New York: Wiley.

Searle, S.R. (1971). *Linear Models.* New York: Wiley.

Self, S.G. and Liang, K-Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Test Under Nonstandard Conditions. *Journal of the American Statistical Association*, **82**, 605-610.

Stram, D.O. and Lee, J.W. (1994). Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics*, **50**, 1171-1177.

Tukey, J.W. (1949). Interaction in a Row-by-Column Design. Memorandum Report 18, Statistical Research Group, Princeton University.

Voss, D.T. (1999). Resolving the Mixed Models Controversy. *The American Statistician*, **53**, 352-356.

Voss, D.T. (2000). Resolving the Mixed Models Controversy, Reply. *The American Statistician*, **54**, 229-230.

Wilk, M.B. and Kempthorne, O. (1955). Fixed, Mixed and Random Models. *Journal of the American Statistical Association*, **50**, 1144-1167.

Wolfinger, R. and Stroup, W.W. (2000). Resolving the Mixed Models Controversy, Comments. *The American Statistician*, **54**, 228.

Zyskind, G. (1962). On structure, relation, $\boldsymbol{\Sigma}$, and expectation of mean squares. *Sankhya: The Indian Journal of Statistics, Series A*, **24**, 115-148.

Résumé

Nous considérons une controverse bien-connue provenant de l'emploi de deux modèles mixtes avec un facteur fixé et un facteur aléatoire pour l'analyse de données expérimentales. Le problème provient essentiellement des statistiques differentes developpées à partir de ces modèles pour tester que le paramètre de variance associé au facteur aléatoire est nul. Les hypothèses correspondantes sont interprétées comme celles de l'inexistence de l'effet principal du facteur aléatoire en présence de l'intéraction entre les deux facteurs. La controverse est rendue plus complexe encore par les différentes opinions sur la propriété de cette hypothèse. En admettant que le choix est sensé, nous montrons que les statistiques usuelles obtenues à partir des deux modèles s'adressent réellement à des hypothèses différentes et nous arrivons à la conclusion que le problème est causé par les définitions des effets principaux et de l'interaction. Nous utilisons des valeurs moyennes comme dans le cas de modèles à effets fixés pour résoudre la controverse et nous montrons que selon le modèle plus utilisé, le test généralement associé a l'inexistence de l'effet principal du facteur aléatoire en présence de l'intéraction s'adresse, en realité, à une hypothèse différente. Nous discutons du choix des modèles et d'autres problèmes qui interviennent en présence de données non-équilibrées.