

The two-way mixed model: A long and winding controversy

Manuel Ato García¹, Guillermo Vallejo Seco² and Alfonso Palmer Pol³

¹ Universidad de Murcia, ² Universidad de Oviedo and ³ Universitat de les Illes Balears

Abstract

Background: With the 2-way mixed model, one a fixed factor and the other random, the procedure followed to test statistical significance of the random factor has been the focus of a heated controversy in theoretical and applied statistics, and the debating continues even now. One of the main consequences of this controversy is that the position defended in the classical ANOVA texts on the hypothesis of the significance of the random effect is not the same as that defended in almost all of the professional statistical software programs. **Method:** In this paper, we deal with a detailed analysis about the controversy of mixed model and the decision about one of two basic options, the non restrictive and the restrictive model. **Results:** Three key questions we consider to go beyond the controversy are: (1) the two classical models are equivalent, (2) the marginality principle do not allow to test main effects in presence of interactive significant effects and (3) the relevance of linear mixed approach to analyze models with fixed and random effects. **Conclusions:** We propose the simple solution of using the mixed linear approach with REML estimation instead of the classical linear approach, which is really unapplicable in this context.

Keywords: mixed approach, linear mixed model, mixed model controversy.

Resumen

El modelo mixto de dos factores: una larga y tortuosa controversia.

Antecedentes: en el modelo mixto de dos factores, con un factor fijo y el otro aleatorio, la forma de probar la significación del factor aleatorio ha sido objeto de una enconada controversia en la estadística teórica y aplicada que todavía hoy sigue siendo objeto de polémica. Una de las consecuencias más sorprendentes de esta contro-versia es que la posición que se defiende en los textos clásicos de ANOVA sobre la prueba de hipótesis del factor aleatorio no es la misma que la defendida en casi todos los programas estadísticos profesionales. **Método:** en este trabajo se aborda un análisis detallado de la controversia sobre el modelo mixto y la decisión de adoptar una de las dos opciones básicas, el modelo no restrictivo o el modelo restrictivo. **Resultados:** las cuestiones clave que se consideran para trascender esta controversia son: (1) las dos opciones básicas son matemáticamente equivalentes, (2) el principio de marginalidad no permite probar efectos principales en presencia de efectos interactivos significativos y (3) la pertinencia del enfoque lineal mixto para analizar modelos con efectos fijos y aleatorios. **Discusión:** en este trabajo se propone como solución a la controversia la utilización del enfoque lineal mixto con estimación REML en menoscabo del enfoque lineal clásico, que resulta inaplicable en este contexto.

Palabras clave: enfoque mixto, modelo lineal mixto y controversia del modelo mixto.

When analyzing data from a factorial design $A \times B$, the applied researcher has many possible alternatives to consider, depending on the nature of the factors (fixed vs. random), the availability of the design structure (replicated vs. non-replicated) and the number of replicates (balanced vs. unbalanced) of the treatment structure.

Assuming that the design is replicated and balanced, the most common situation in psychological research arises when the two factors are fixed (i.e., the levels of both factors are arbitrarily selected by the researcher), and the result is Model I or fixed effects model, as classified by Eisenhart (1947). Using notation of effects (Ato & Vallejo, 2007, 187-190; Palmer, 2011, 19), Model I structural equation

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + e_{ijk} \quad (1)$$

assumes that the systematic components α_j , β_k and $(\alpha\beta)_{jk}$ are fixed constants with σ - restrictions

$$\sum_{j=1}^a \hat{\alpha}_j = 0, \sum_{k=1}^b \hat{\beta}_k = 0, \sum_{j=1}^a (\hat{\alpha}\hat{\beta})_{jk} = 0,$$

for all j and $\sum_{k=1}^b (\hat{\alpha}\hat{\beta})_{jk} = 0$, for all k , and the random components

are residuals $e_{ijk} \sim NID(0, \sigma^2)$. The expected mean squares E(MS) and F-ratios are shown in Table 1.

Rarely used in psychology, except in some sectors of psycholinguistic and meta-analytic research, are studies where both factors are random (i.e., the levels of both factors are selected at random), the structural equation of Model II or random effects model (Eisenhart, 1947) is

$$Y_{ijk} = \mu + A_j + B_k + (AB)_{jk} + e_{ijk} \quad (2)$$

where all components are independent random variables: $A_j \sim NID(0, \sigma_a^2)$, $B_k \sim NID(0, \sigma_b^2)$, $(AB)_{jk} \sim NID(0, \sigma_{ab}^2)$. Table 2 shows E(MS) and F-ratios for this model.

Received: January 14, 2012 • Accepted: May 14, 2012

Corresponding author: Manuel Ato García

Facultad de Psicología

Universidad de Murcia

30080 Murcia (Spain)

e-mail: matogar@um.es

Table 1
E(MS) and F-ratios of fixed-effect model (Model I)

| Source | Component | Mean square | E(MS) | F-ratio |
|----------|-----------|-------------|-------------------------------|------------|
| A | Fixed | MSA | $\sigma_e^2 + nb\theta_a^2$ | MSA / MSE |
| B | Fixed | MSB | $\sigma_e^2 + na\theta_b^2$ | MSB / MSE |
| AB | Fixed | MSAB | $\sigma_e^2 + n\theta_{ab}^2$ | MSAB / MSE |
| Residual | Random | MSE | σ_e^2 | |

Table 2
E(MS) and F-ratios of random-effects model (Model II)

| Source | Component | MS | E(MS) | F-ratio |
|----------|-----------|------|--|------------|
| A | Random | MSA | $\sigma_e^2 + n\sigma_{ab}^2 + nb\sigma_a^2$ | MSA / MSAB |
| B | Random | MSB | $\sigma_e^2 + n\sigma_{ab}^2 + na\sigma_b^2$ | MSB / MSAB |
| AB | Random | MSAB | $\sigma_e^2 + n\sigma_{ab}^2$ | MSAB / MSE |
| Residual | Random | MSE | σ_e^2 | |

An intermediate situation arises when one factor is fixed (e.g., a levels of factor A are selected arbitrarily) and the other factor is random (e.g., b levels of factor B are selected at random). The result is Model III or mixed effects model (Eisenhart, 1947). Some situations in psychology (e.g., agreement studies) use a mixture of random (subjects) and fixed (raters) factors. More complicated situations use hierarchical linear models (Oliver, Rosel, & Jara, 2000; Vallejo, Arnau, & Bono, 2009).

There are three possible ways (the third being a variant of the second) appropriate for defining the two-way mixed model (McLean, Sanders, & Stroup, 1991; Schwarz, 1993). The first way is

$$Y_{ijk} = \mu + \alpha_j + B_k + (\alpha B)_{jk} + e_{ijk} \quad (3)$$

where α_j are fixed constants and $B_k \sim NID(0, \sigma_B^2)$ and $e_{ijk} \sim NID(0, \sigma_e^2)$, for all $i = 1, \dots, n, j = 1, \dots, a$ y $k = 1, \dots, b$, are independent random variables. In this model, the interactive effects are also considered as random: $(\alpha B)_{jk} \sim NID(0, \sigma_{\alpha B}^2)$ and their E(MS) are built on the principle of considering as random any term that contains at least one random component without any restriction, which is also known as the *non-restricted mixed model* (Ato & Vallejo, 2007, p. 232) or *mixed model IIIa*. The E(MS) and F ratios of model IIIa are shown in Table 3.

The second way is

$$Y_{ijk} = \mu + \alpha_j^* + B_k^* + (\alpha B)_{jk}^* + e_{ijk} \quad (4)$$

Table 3
E(MS) and F-ratios of non-restricted mixed model (Model IIIa)

| Source | Component | MS | E(MS) | F-ratio |
|----------|-----------|------|--|------------|
| A | Fixed | MSA | $\sigma_e^2 + n\sigma_{ab}^2 + nb\theta_a^2$ | MSA / MSAB |
| B | Random | MSB | $\sigma_e^2 + n\sigma_{ab}^2 + na\sigma_b^2$ | MSB / MSAB |
| AB | Random | MSAB | $\sigma_e^2 + n\sigma_{ab}^2$ | MSAB / MSE |
| Residual | Random | MSE | σ_e^2 | |

where $B_k^* \sim NID(0, \sigma_{B^*}^2)$, $(\alpha B)_{jk}^* \sim N(0, (a-1/a) \sigma_{\alpha B^*}^2)$ are also independent of $e_{ijk} \sim NID(0, \sigma_e^2)$, but in this case, it is assumed that the interactive effects are subject to restriction $\sum_{j=1}^a (\hat{\alpha B})_{jk}^* = 0$,

for all k , which is also known as the *restricted mixed model* (Ato & Vallejo, 2007, 232-238) or *mixed model IIIb*. The E(MS) for this model are built on the principle of considering as random only those interactive terms resulting from combining a basic factorial effect (whether fixed or random) with another factorial effect which must be strictly random, and are shown in Table 4. A major consequence of this restriction is the presence of correlation, given that the covariance between interactive terms for the same level of B and different levels of fixed factor A is different from zero:

$$\text{cov}((\alpha B)_{jk}^*, (\alpha B)_{j'k}^*) = -\sigma_{(\alpha B)^*}^2 / (a-1), \text{ for all } j \neq j' \quad (5)$$

It is somewhat strange to use restrictions on a random variable that represents interaction effects (although it greatly facilitates the interpretation of the random effects for factor B), so many authors use the model without including any restrictions.

The third way is a variant of Model IIIb, which appears in some reference texts and is based on the same principle (given a factorial effect, fixed or random, the interaction term must be considered as random if it results from being combined with another random factorial effect), but it does not use restriction. The result is very similar to the restricted mixed model of Table 4, but the E(MS) ignore the restriction $\sum_{j=1}^a (\hat{\alpha B})_{jk}^* = 0$, producing a particular

result which is sometimes called *mixed model IIIc* to distinguish it from the mixed model IIIb. Although in this paper we do not often use this variant of restricted mixed model, its E(MS) and F ratios are displayed in Table 5.

A careful comparison of the E(MS) in Tables 3 and 4 will show that the essence of the problem lies in the different way they present the random factor B. The E(MS) of the IIIa model include the term $n\sigma_{\alpha B}^2$, which is not included in the E(MS) of model IIIb.

Table 4
E(MS) and F-ratios of restricted mixed model (Modelo IIIb)

| Source | Component | MS | E(MS) | F-ratio |
|----------|-----------|------|--|------------|
| A | Fixed | MSA | $\sigma_e^2 + n \left(\frac{a}{a-1} \right) \sigma_{\alpha B}^2 + nb\theta_a^2$ | MSA / MSAB |
| B | Random | MSB | $\sigma_e^2 + na\sigma_b^2$ | MSB / MSE |
| AB | Random | MSAB | $\sigma_e^2 + n \left(\frac{a}{a-1} \right) \sigma_{(\alpha B)^*}^2$ | MSAB / MSE |
| Residual | Random | MSE | σ_e^2 | |

Table 5
E(MS) and F-ratios of restricted mixed model (Model IIIc)

| Source | Component | MS | E(MS) | F-ratio |
|----------|-----------|------|--|------------|
| A | Fixed | MSA | $\sigma_e^2 + n\sigma_{ab}^2 + nb\theta_a^2$ | MSA / MSAB |
| B | Random | MSB | $\sigma_e^2 + na\sigma_b^2$ | MSB / MSE |
| AB | Random | MSAB | $\sigma_e^2 + n\sigma_{(\alpha B)^*}^2$ | MSAB / MSE |
| Residual | Random | MSE | σ_e^2 | |

As a consequence, in the first we used $F = MSB / MSAB$ to test the effect of B factor, while in the second we used instead $F = MSB / MSE$. The problem is obvious. When a researcher uses a mixed model with two factors and wants to test the effect of B, which model should be used, the non-restricted or the restricted form of the mixed model?

Although the problem is defined here for the simplest case, i.e., for a replicated and balanced model with two factors, it has greater importance in more complex cases, such as unbalanced and mixed models with more than two factors. The empirical evidence from an example by Keppel (1991, p. 564) serves to illustrate the problem. Using the R program (version 2.14) the script required to obtain a 2x4 factorial ANOVA output with all factors fixed is

```
> Y = c(3,2,1,1,5,3,5,9,10,10,3,3,6,6,4,3)
> A = as.factor(gl(2,8,labels=c("a1", "a2")))
> B = as.factor(rep(gl(4,2,labels=c("b1", "b2", "b3", "b4")),2))
> summary(mod1 = aov(Y ~ A*B))
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|----------|
| A | 1 | 16.00 | 16.000 | 11.6364 | 0.009207 |
| B | 3 | 40.25 | 13.417 | 9.7576 | 0.004752 |
| A:B | 3 | 60.50 | 20.167 | 14.6667 | 0.001291 |
| Residuals | 8 | 11.00 | 1.375 | | |

Note that the F ratios and P values of the ANOVA output assume that this is model I, since all the MS are tested against the residual error term. Therefore, if we want to analyze the mixed model with A fixed and B random, the output would not be correct. Using E(MS) and the mixed model F-ratios for the non-restricted case (see Table 3) to test the effect of B, we obtained $F(3;3) = 13.417/20.167 = 0.665$, $P = .627$, enabling us to conclude that the effect of B is not statistically significant with the mixed model IIIa. Using E(MS) and appropriate F ratios for the restricted case (see Table 4) to test the effect of factor B we obtained $F(3;8) = 13.417/1.375 = 9.758$, $P = .005$, and the conclusion is now that the effect of B is statistically significant with the mixed model IIIb. This latter result is also true with the mixed model IIIc, which disregards the finite correction, because the correction does not essentially affect the construction of F-ratios. However, these three models can produce different estimates of variance components. Using the method of moments or ANOVA (see Searle, Casella, & McCulloch, 1992; Cox & Solomon, 2003), Table 6 presents the estimation of variance components for the three models with the Keppel data, where the discrepancy can be observed in the estimates of the components (and the corresponding proportion) of the variance and the total variance.

Table 6
Components of variance for models IIIa,b,c (Keppel's data, 1991)

| Source | Component | Model IIIa | Model IIIb | Model IIIc |
|----------------|-----------|-------------|-------------|-------------|
| A | Fixed | – | – | – |
| B | Random | -1.687 (0%) | 3.010 (33%) | 3.010 (22%) |
| AB | Random | 9.396 (87%) | 4.698 (52%) | 9.396 (68%) |
| Residual | Random | 1.375 (13%) | 1.375 (15%) | 1.375 (10%) |
| Total variance | | 10.771 | 9.083 | 13.781 |

The presence of discordant results in the F-ratios and the estimation of variance components requires help from the applied researcher. This is the main purpose of this work, where we develop firstly the main discussion of the mixed-model controversy in the academic tradition and the existing divorce between the academic tradition and the applied software afterwards. Finally we advocate settling this fruitless controversy with a solution based on three fundamental aspects of the problem.

The academic debate

The mixed effects model with two factors has been the object of a long and heated debate which, despite the numerous attempts made to resolve it, remains virtually in its same original state. To a large extent, the problem has become chronic as a result of the divorce that seems to exist between the academic and applied contexts.

At the root of the problem seems to be the classic analysis of variance text (Scheffé, 1959), which was the first to submit a thorough development of the mixed 2-factor model. Hartley and Searle (1969) were the first to report the existence of a “discontinuity” in the data analysis with mixed models of 2 factors establishing that the procedures for determining the E(MS) for unbalanced data included the term $n\sigma^2_{(ab)}$, while the vast majority of the advanced texts at the time did not include such a term for balanced data.

Shortly thereafter, a seminal work by Hocking (1973) discussed the three alternative models for the 2-way factor mixed model and concluded that the covariance matrix was the key issue generating the problem.

In the same vein, McLean, Sander & Stroup (1991) also discussed the main mixed models and proposed the use of the mixed approach to analyze the data because it allows us to assume the existence of a correlation between the random effects. Meanwhile, Schwarz (1993) rigorously defined the three forms of mixed models and conducted a thorough review of the coverage that the professional statistical packages (BMDP, SAS and SPSS) and the main reference texts in the field (Christensen, 1987; Hocking, 1985; Milliken & Johnson, 1984; Neter, Kutner, & Wasserman, 1990; Searle, 1987) gave to mixed models in the early 90's, claiming that most of the reference texts preferred to use the restricted model (or its variant), while the statistical packages prefer the non-restricted model (in the case of SPSS, with incorrect results).

Voss (1999) planned to resolve the dispute between the two basic models appealing to overpopulation finite models to define the random effects model, initially opting for the restricted model. However, two letters to the editor in response to their work (Hinkelmann, 2000; Wolfinger & Stroup, 2000) highlighted the preponderance of the non-restricted model in the statistical packages for its flexibility and pragmatism. In his reply, Voss (2000) changed his initial position concluding that, in the absence of a sound reason to introduce restrictions on the parameters, it was more convenient to use the non-restricted model.

The last known attempt to resolve this controversy was outlined in two papers published by Lencina, Singer & Stanek III (2005) and Lencina & Singer (2006), who emphasized the formal similarity of the two mixed model forms IIIa and IIIb with model I and in line with Voss's proposal, they showed that the ratio $F = MSB / MSE$ behaved like an F statistic, accurate both within the non-restricted and the restricted model. In response to this proposal, Nelder (2008) noted that the hypothesis to test the effect of B has

no practical interest and demonstrated how to develop a consistent solution to the problem based on two basic arguments: 1) the principle of marginality and 2) the need to use restrictions on the estimates but not on the parameters. In their reply, Lencina, Singer & Stanek III (2008) defended their initial proposal calling for the non-universality of the principle of marginality and the need for restrictions on parameters to facilitate their interpretability.

The academic-applied divorce

The divorce is becoming increasingly apparent between the academic context, where the most representative statistical texts are created, and the applied computing context, which outlines the successive versions of professional statistical packages used by applied researchers. While teaching is the main focus in the first context, based on the consistent demonstration of the arguments, the second is focused on pragmatism and the practical applicability of all analytical techniques.

Many of the reference texts on analysis of variance and most influential experimental designs in psychological research until the end of the last century provided an irregular treatment of the mixed model with 2 factors. While some design texts presented an acceptable level (Montgomery, 1984), in others it was superficial (e.g., Keppel, 1982; Kirk, 1982; Winer, 1971; Winer, Brown, & Michels, 1991) or even null (e.g., Myers, 1979; Myers & Well, 1991). Other general texts provided a much more detailed treatment (for example, Glass & Hopkins, 1996; Hays, 1988; Neter & Wasserman, 1985). Moreover, following the academic tradition, they all defended the use of the restricted model.

The treatment of the 2-factor mixed model has changed significantly in the reference texts at the beginning of this century and it is now much more uniform and detailed (e.g., Hinkleman & Kempthorne, 2004; Howell, 2010; Maxwell & Delaney, 2004; Montgomery, 2009; Oehlert, 2010). All of them continue to use the restricted model, but the existence of the non-restricted model is at least simply mentioned, leaving it to the discretion of the researcher to use one or the other.

The situation is quite different in the computational context. Almost all professional statistical packages (e.g., SAS, SPSS, SPLUS, GENSTAT) use the non-restricted model based on pragmatic criteria, such as the lack of flexibility of the restricted model and its difficulties in dealing with unbalanced data. The E(MS) and variance components are estimated assuming the model is non-restricted, so users are forced to manually calculate F tests if they want to use the restricted model. The most important exception is the MINITAB package, version 16.2, which gives users the option to estimate the parameters with either of the two basic models.

Three key aspects to settle the controversy

The core of the controversy about the mixed model with two factors is, assuming a fixed factor (A), how to test the effect of the random factor (B). The two basic options are the non-restricted model (mixed model IIIa) and the restricted model (mixed model IIIb), which can achieve different results. But the key question may not be which of the two models should be used, but whether there is a solution to the dispute which does not require the use of one of the two alternatives. There are three clarifications worth noting in this direction.

The two basic models are equivalent

Searle (1971) and Searle, Casella and McCullagh (1992), on the one hand, and Hocking (1973, 1985), on the other, have shown the mathematical relationships between the models of equations (3) and (4) and their equivalence. Following Searle, Casella and McCullagh (1992, pp. 126-128), the equation (3) can be rewritten as follows

$$Y_{ijk} = (\mu' + \bar{\alpha}_j) + (\alpha_j - \bar{\alpha}_j) + (B_k + (\alpha B)_k) + ((\alpha B)_{jk} - (\alpha B)_k) + e_{ijk} = \mu' + \alpha'_j + B'_k + (\alpha B)_{jk} + e_{ijk} \quad (6)$$

thus obtaining again the equation (4), which proves the close link between models IIIa and IIIb. In addition, given that estimates of the variance of the random factor B and the AB interaction are different for the two models, it is very simple to convert between the two models estimating their respective variance components.

Using the ANOVA method (see Searle, Casella, & McCulloch, 1992) to estimate the variance components for the random terms, and assuming that $\hat{\sigma}_e^2 = MSE$ is the same in both cases, the estimates of variance components for the non-restrictive (model IIIa) are for data from Keppel (1991):

$$\hat{\sigma}_B^2 = (MSB - MSAB) / an = (13.417 - 20.167) / 4 = -1.687$$

$$\hat{\sigma}_{\alpha B}^2 = (MSAB - MSE) / n = (20.167 - 1.375) / 2 = 9.396 \quad (7)$$

and for the restrictive (model IIIb):

$$\hat{\sigma}_{B^*}^2 = (MSB - MSE) / an = (13.417 - 1.375) / 4 = 3.010$$

$$\hat{\sigma}_{\alpha B^*}^2 = [(a-1)MSAB - MSE] / an = [(1)(20.167 - 1.375)] / 4 = 4.698 \quad (8)$$

Note that with respect to model IIIa, the estimation of variance for model IIIc does not change and so $\hat{\sigma}_{\alpha B}^2 = \hat{\sigma}_{\alpha B^*}^2$.

Now, comparing equations (7) and (8), the link between the two models is obtained noting that

$$\hat{\sigma}_{(\alpha B)^*}^2 = (a-1) \hat{\sigma}_{(\alpha B)}^2 / a \quad (9)$$

$$\hat{\sigma}_{B^*}^2 = \hat{\sigma}_B^2 + \hat{\sigma}_{(\alpha B)}^2 / a \quad (10)$$

and the equivalence between them is evident from (10), because the E(MS) for the non-restricted mixed model (left) and for the restricted mixed model (right) are equal:

$$\hat{\sigma}_e^2 + na\hat{\sigma}_B^2 + n\hat{\sigma}_{\alpha B}^2 = \hat{\sigma}_e^2 + na\hat{\sigma}_{B^*}^2 \quad (11)$$

So it is very important to highlight that, although the two main models are different, they are essentially equivalent.

The marginality principle

Most remarkable in the debate on the mixed model was the response of Professor Nelder (one of the most genuine representatives of the Fisherian tradition at Rothamsted) to what he called "the great disruption of the mixed model". Nelder's position in this debate was initially expressed in a controversial work in which he expressed his dissatisfaction with the way linear models were exposed in reference texts and professional statistical software (Nelder, 1977), a position he reiterated over time (Nelder, 1994; Nelder, 1995; Nelder & Lane, 1995; Nelder, 1998; Nelder, 2008).

His argument is that the components of the mixed ANOVA with two factors (constant or global mean, A, B and A*B) should be ordered according to their terms of marginality, where A and B represent subspaces of A*B, the constant represents a subspace from A and B, and therefore is marginal to A and B, and A and B are in turn marginal to A*B. The principle of marginality implies that for a sensible and meaningful interpretation all statistical models must respect the marginality relations between terms. Some of the consequences of this principle for the mixed model with 2 factors are:

- * If the interaction A*B is included in the model, the main effects and the constant should also be included;
- * If the interaction A*B is significant, the main effects should not be tested or interpreted and instead only the interaction effect should be interpreted. This would be the case of the example of Keppel.
- * If the interaction A*B is not significant, the model should be simplified removing the interactive component and testing the main effects with the resulting additive model.

Statistical orthodoxy in general recommends respecting the principle of marginality, but it should be noted that there are some very rare exceptions to the principle. The most flagrant case is that of a model with significant interaction and non significant main effects, which graphically would correspond to an X-shaped non-ordinal interaction. Another extreme situation occurs when the sums of the squares for the main effects are considerably larger than those of the interaction. For these cases, Nelder (1994, 2008) recommends using a suitable solution. Thus, in the first case, a main effect can be tested in the presence of interactions, taking the interaction as an error term and using the reciprocal value of F with its corresponding degrees of freedom. In the second case, a test of main effects also uses the interaction as an error term in order to determine whether additional variation exists with respect to that shown by the interaction component.

The principle of marginality is closely associated with another heated controversy on the use of type I, II, III and IV sums of squares, particularly with unbalanced models (see Herr, 1986; Langsrud, 2003; Nelder & Lane, 1995; Vallejo, Fernández, & Livacic-Rojas, 2010), which was probably also one of the causes of divorce between the academic and applied fields.

The usefulness of the mixed approach

The most sensitive statistical procedure for analyzing the data from a mixed model with 2 factors is currently Mixed Linear Modeling (MLM), which associated with the method of REstricted Maximum Likelihood estimation (REML), represents a comprehensive approach that includes regression, ANOVA and ANCOVA as special cases and can also be generalized to categorical response variables. The REML method produces unbiased results and its purpose is to divide the likelihood function into two components: the first includes all the fixed parameters and the second includes all the parameters of variance and covariance of random effects. The MLM approach with REML estimation is an extremely flexible, which has become the standard for analyzing empirical data with any combination of fixed and random effects (see Verbeke & Molenberghs, 1997). However, the demand for classical ANOVA tables, more familiar and manageable for the

applied researcher has probably slowed the expected popularization of the mixed approach.

A typical output of the MLM approach to data from Keppel (1991) is shown in Table 8. Note that the results are shown separately for fixed and random effects. For fixed factor A, we report the value of the conventional F test, which shows that is not statistically significant. For the random factor B, the interaction A*B and the residual component, the report includes the corresponding variance component without the null hypothesis testing controversy.

The results differ from those obtained with models IIIa, IIIb and IIIc, but they allow us to conclude firstly that the model typically used is the non-restrictive one, since it is not necessary to introduce any restriction in the estimation of the parameters, and secondly that the effect of factor B is actually null. Furthermore, the estimated variance component for interaction A*B is different for models IIIa, IIIb and IIIc due to the use of the REML estimation criterion. Specifically, assuming a balanced design, if the variance component of B is positive, the method of moments produces the same result as the REML procedure for estimating the variance of the interaction A*B, and if negative, the estimate of the variance of the interaction is also the same if the variance of B is previously subtracted from the variance of A*B (by comparing Tables 7 and 8 it will be seen that this is precisely what happens with Keppel's data). However, this result is different when the design is not balanced.

The differences between the classical approach and the MLM approach are extremely dramatic when the design contains any combination of mixed effects with more than 2 factors and/or is not balanced. With the classical approach, some sources of variation of the mixed model with three or more factors require quasi-F ratios and it is not possible to estimate the degrees of freedom for the denominator by the usual Satterthwaite approximation

Tabla 7
F-ratios and variance components for all effects of models proposed for 1991 Keppel's data, (I: A & B fixed; II: A & B random; III: A fixed & B random)

| Models | A | B | AB |
|---|-----------------------|--|--|
| I | F(1,8)= 11.6; P= .009 | F(3,8)= 9.76; P= .005 | F(3,8)= 14.7; P= .001 |
| II | F(1,3)= .79; P= .439 | F(3,3)= .665; P= .627 | F(3,8)= 14.7; P= .001 |
| IIIa (non restricted) | F(1,3)= .79; P= .439 | F(3,3)= .665; P= .627 $\sigma_b^2 = -1.687 = 0$ | F(3,8)= 14.7; P= .001 $\sigma_{(ab)}^2 = 9.396$ |
| IIIb (restricted) | F(1,3)= .79; P= .439 | F(3,8)= 9.76; P= .005 $\sigma_b^2 = 3.010$ | F(3,8)= 14.7; P= .001 $\sigma_{(ab)}^2 = 4.698$ |
| IIIc (restricted, but ignoring restriction) | F(1,3)= .79; P= .439 | F(3,8)= 9.76; P= .005 $\sigma_b^2 = 3.010$ | F(3,8)= 14.7; P= .001 $\sigma_{(ab)}^2 = 9.396$ |

Table 8
Results from mixed approach with REML estimation

| Source | Component | Random effects | Fixed effects |
|----------|-----------|---|-----------------------|
| A | Fixed | - | F(1,6)= .953; P= .367 |
| B | Random | $\sigma_b^2 = 0$ | - |
| AB | Random | $\sigma_{(ab)}^2 = 7.708$ Z= 1.59; P= .056 | - |
| Residual | Random | $\sigma_e^2 = 1.375$ Z= 2.00; P= .023 | - |

(Ato & Vallejo, 2007, 239), so we cannot test them or estimate their variance components, as revealed by the ANOVA outputs of professional statistical packages. With the MLM approach associated with REML estimation, by separating the fixed and random effects, all sources of variation can be estimated.

Another major drawback of the classical approach to estimating mixed models is that it assumes that all variance components are independent and identically distributed. The mixed approach does not require this assumption because it incorporates the structure and relationships between errors and the variance components of the model.

Conclusions

Taking into account these clarifications, we suggest that the long and winding controversy over the 2-factor mixed model, regarding the different ways to test the effect of factor B with models IIIa, IIIb and IIIc, despite the intense debate that seems have led to a divorce between the academic and the applied environment, should be definitely settled for three reasons. Firstly, it has been shown that the two basic models (IIIa and IIIb) are statistically equivalent, and therefore it is not necessary to raise a dispute. Secondly, using the principle of marginality, the null hypothesis of factor B is meaningless if the interaction A*B is significant and instead, the nature of the interaction must be analyzed in depth (Ato & Vallejo,

2007; Pardo, 2006), or the problem disappears when the interaction A*B is not significant because a more parsimonious additive model is required. Thirdly, the analysis of mixed models must be carried out in its natural setting (i.e., the MLM approach with REML estimation) which, fortunately, all professional statistical packages now incorporate, contrary to the classical approach, which, as we have shown here, is not the best framework for analyzing data from a mixed design. However, this suggestion requires the applied researcher to become familiar with the display of results of the MLM approach with REML estimation, separating the fixed and random effects and interpreting them independently, and to stop demanding output formats based on classical ANOVA tables that produce contradictory and sometimes difficult to interpret results (see Ato & Vallejo, 2007; Palmer & Ato, 2012). The output of the MLM approach displayed by SAS Proc Mixed, SPSS Mixed, Stata xtmixed, GENSTAT Linear Mixed Models and *lme* and *lmer* R packages, in spite of some minor differences, represents at the present time the most satisfactory solution for analyzing data from mixed designs.

Acknowledgments

This work was supported by grants SEJ2006-09025 (M. Ato) and PSI2011-23395 (G. Vallejo) from the Spanish Ministry of Education and Science.

References

- Ato, M., & Vallejo, G. (2007). *Diseños experimentales en psicología*. Madrid: Pirámide.
- Cox, D.R., & Solomon, P.J. (2003). *Components of variance*. Boca Ratón, CA: Chapman and Hall.
- Christensen, R. (2002). *Plane answers to complex questions: The theory of linear models*, 3rd edition. New York, NY: Springer.
- Eisenhart, C. (1947). The assumptions underlying analysis of variance. *Biometrics*, 3, 1-21.
- Glass, G.V., & Hopkins, K.D. (1996). *Statistical Methods in Education and Psychology*. 3rd edition. Boston, MA: Allyn and Bacon.
- Hartley, H.O., & Searle, S.R. (1969). A discontinuity in mixed model analysis. *Biometrics*, 25, 573-576.
- Hays, W.L. (1994). *Statistics*, 5th edition. Florence, KY: Wadsworth Publishing.
- Herr, D.G. (1986). On the history of ANOVA in unbalanced, factorial designs: The first 30 years. *The American Statistician*, 40, 265-270.
- Hinkelmann, K. (2000). Response to Voss, D.T. (1999), "Resolving the mixed models controversy". *The American Statistician*, 54, 228.
- Hinkelmann, K., & Kempthorne, O. (2008). *Design and analysis of experiments. Vol. I: Introduction to experimental design*, 2nd edition. New York, NY: Wiley.
- Hocking, R.R. (1973). A discussion of the two-way mixed model. *The American Statistician*, 27, 148-152.
- Hocking, R.R. (1985). *The analysis of linear models*. Monterey, CA: Brooks/Cole.
- Howell, D.C. (2010). *Statistical Methods for Psychology*. 7th edition. Canada: Wadsworth
- Keppel, G. (1982). *Design and analysis: A researchers handbook*, 2nd edition. Englewood Cliffs, N.J.: Prentice-Hall.
- Keppel, G. (1991). *Design and analysis: A researchers handbook*, 3rd edition. Englewood Cliffs, N.J.: Prentice-Hall.
- Kirk, R.E. (1982). *Experimental design: Procedures for the behavioral sciences*, 2nd edition. Monterey, CA: Brooks-Cole.
- Kirk, R.E. (1995). *Experimental design: Procedures for the behavioral sciences*, 3rd edition. Monterey, CA: Brooks-Cole.
- Langsrud, O. (2003). ANOVA for unbalanced data: Use type II instead type III sum of squares. *Statistics and Computing*, 13, 163-167.
- Lencina, V.B., Singer, J.M., & Stanek III, E.J. (2005). Much ado about nothing: The mixed models controversy revisited. *International Statistical Review*, 73, 9-20.
- Lencina, V.B., & Singer, J.M. (2006). Measure for measure: Exact F tests and the mixed models controversy. *International Statistical Review*, 74, 391-402.
- Lencina, V.B., Siner, J.M., & Stanek III, E.J. (2008). Response to Nelder's "What is the mixed-model controversy?". *International Statistical Review*, 76, 135-139.
- Maxwell, S.E., & Delaney, H.D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.
- Maxwell, S.E., & Delaney, H.D. (2004). *Designing experiments and analyzing data: A model comparison perspective*. 2nd edition. Mahwah, NJ: LEA.
- McLean, R.A., Sanders, W.L., & Stroup, W.W. (1991). A unified approach to mixed linear models. *The American Statistician*, 45, 54-64.
- Milliken, G.A., & Johnson, D.E. (1984). *Analysis of messy data. Vol. I: Designed experiments*. New York, NY: Van Nostrand.
- Milliken, G.A., & Johnson, D.E. (2009). *Analysis of messy data. Vol. I: Designed experiments*. 2nd edition. New York, NY: Van Nostrand.
- Montgomery, D.C. (1984). *Design and analysis of experiments*, 2nd edition. New York, NY: Wiley.
- Montgomery, D.C. (2009). *Design and analysis of experiments*, 7th edition. New York, NY: Wiley.
- Myers, J.L. (1979). *Fundamentals of experimental design*. 3rd edition. Boston, MA: Allyn and Bacon.
- Myers, J.L., & Well, A.D. (1991). *Research design and statistical analysis*. New York, NY: Harper-Collins Publishers.
- Nelder, J.A. (1977). A reformulation of linear models. *Journal of the Royal Statistical Society, A*, 140, 48-77.
- Nelder, J.A. (1994). The statistics of linear models: Back to basics. *Statistics and Computing*, 4, 221-234 (Discussion, vol. 5, 1995, 85-111).

- Nelder, J.A. (1998). The great mixed-model muddle is alive and flourishing, alas! *Food Quality and Preference*, 9, 157-159.
- Nelder, J.A. (2008). What is the mixed-models controversy? *International Statistical Review*, 76, 134-139.
- Nelder, J.A., & Lane, P.W. (1995). The computer analysis of factorial experiments: In memoriam-Frank Yates. *The American Statistician*, 49, 382-385.
- Neter, J., & Wasserman, W. (1984). *Applied Linear Statistical Models*, 3rd edition. Chicago: Irwin.
- Oehlert, G.W. (2010). *A first course in design and analysis of experiments*. Creative Commons.
- Oliver, J.C., Rosel, J., & Jara, P. (2000). Modelos de regresión multinivel: aplicación en psicología escolar. *Psicothema*, 12, 487-494.
- Palmer, A. (2011). *Análisis de datos en diseños experimentales*. Palma de Mallorca: Edicions UIB.
- Palmer, A., & Ato, M. (2012). *Análisis de datos en los diseños con efectos aleatorios*. Palma de Mallorca: Edicions UIB.
- Pardo, A. (2006). La interacción entre factores en el análisis de varianzas: errores de interpretación. *Psicothema*, 19, 343-349.
- Scheffé, H. (1959). *The analysis of variance*. New York, NY: Wiley.
- Schwarz, C.J. (1993). The mixed-model ANOVA: The truth, the computer packages, the books. Part I: balanced data. *The American Statistician*, 47, 48-59.
- Searle, S.R. (1971). *Linear models*. New York, NY: Wiley.
- Searle, S.R., Casella, G., & McCulloch, C.E. (1992). *Components of variance*. New York, NY: Wiley.
- Vallejo, G., Arnau, J., & Bono, R. (2009). Construcción de modelos jerárquicos en contextos aplicados. *Psicothema*, 20, 830-838.
- Vallejo, G., Fernández, P., & Livacic-Rojas, P. (2010). Analysis of unbalanced factorial designs with heteroscedastic data. *Journal of Statistical Computation & Simulation*, 80, 75-88.
- Verbeke, G., & Molenberghs, G. (Eds.) (1997). *Linear mixed models in practice. A SAS-oriented approach*. New York, NY: Springer.
- Voss, D.T. (1999). Resolving the mixed model controversy. *The American Statistician*, 53, 352-356.
- Voss, D.T. (2000). Reply to Voss, D.T. (1999), "Resolving the mixed models controversy". *The American Statistician*, 54, 229-230.
- Winer, B.J. (1971). *Statistical principles in experimental design*, 2nd edition. New York, NY: McGraw Hill.
- Winer, B.J., Brown, D.R., & Michels, K.M. (1991). *Statistical principles in experimental design*, 3rd edition. New York, NY: McGraw Hill.
- Wolfinger, R.D., & Stroup, W.W. K. (2000). Response to Voss, D.T. (1999), "Resolving the mixed models controversy". *The American Statistician*, 54, 228-229.