International Statistical Review (2008), 76, 1, 134–139 doi:10.1111/j.1751-5823.2007.00022.x

Letter to the Editors

What is the Mixed-Models Controversy?

From: John A. Nelder

Imperial College, London, UK

Two papers have recently been published in this journal which purport to deal with the mixed-models controversy (Lencina *et al.*, 2005; Lencina & Singer, 2006, to be referred to as L1 and L2). In my view they do not represent the current state of thinking on this subject. My own work begins with Nelder (1977) and continues with papers in 1982, and particularly with Nelder (1994, 1995). Further papers are Nelder (1997), whose title includes the phrase 'The great mixed-model muddle', and Nelder (1998). In Nelder (1994) I describe what I regard as three false steps that have generated confusion and show how a consistent treatment may be developed. The two most important ideas are (1) marginality relations between terms in a factorial model, and (2) why constraints must not be put on parameters because they are required on estimates.

1 Marginality

Consider a simple model with two classifying factors A and B. The terms in the full factorial model are 1, A, B and A.B where 1 represents the grand mean. These terms are ordered by their marginality relations; A and B are subspaces of the two-way table of effects A.B and 1 is a subspace of both A and B. We say that 1 is marginal to A and B and both A and B are marginal to A.B. If a model is going to make statistical sense it must respect the marginality relations between the terms, meaning that any term must be preceded by the terms that are marginal to it . Thus 1+A, 1+B, 1 +A+B, 1+A+B+A.B satisfy these conditions but 1+A+A.B does not. This last model implies that the two-way table of effects A.B, which, whether fixed or random, can be completely arbitrary, nonetheless is to have a null margin showing no variation. The probability of this is zero, and hence I contend that this model, though expressible mathematically, makes no statistical sense. These models correspond to statistically uninteresting hypotheses; the fact that they are mathematically expressible does not imply that they make statistical sense. Note that Type III and IV sums of squares in SAS break these marginality rules, and even Scheffé (1959) comes unstuck here.

2 Constraints on Parameters

Every one agrees that estimates of parameters in factorials obtained by least squares will need to be constrained if we are to obtain unique values for them. Thus it is common in the example above to set the sums of the A and B effects to be zero and also the row and column totals for the interaction A.B. The possible constraints are not unique, and any contrast given a value by this procedure is said to be *estimable*. Thus the following are not estimable whatever set of valid constraints is used

$$\mu, \alpha_1, \beta_1, \gamma_{11}, \mu + \alpha_1, \beta_2 + \beta_3, \gamma_{11} - \gamma_{12}.$$

The false step in many expositions of fitting these models is to assume that because constraints have been put on the estimates of the parameters, similar constraints must be put on expectations of the sums of squares in mixed models. Without such constraints the formulae for the expectation of fixed and random terms are entirely consistent, with sums of squares for the fixed effects and the variance components for the random effects having the same form. To the extent that fixed effects can be thought of as random effects with the variance becoming infinite, this is what we should require. For details of these expectations see Tables 1 and 2 in Nelder (1994). It is noticeable that in L1, equation (1), the αs are constrained to sum to zero, whereas in L2, equation (1), they are not. L2 is correct and L1 is incorrect. It is not clear that the authors are aware of this.

Conclusion

The great mixed-model muddle has now been resolved by the recognition of two simple principles, and I commend this approach to the authors.

References

Lencina, V.B. & Singer, J.M. (2006). Measure for measure: exact F tests and the mixed models controversy. *Int. Stat. Rev.*, **74**, 391–402.

Lencina, V.B., Singer, J.M. & Stanek, E.J. (2005). Much ado about nothing: the mixed models controversy revisited. *Int. Stat. Rev.*, **73**, 9–20.

Nelder, J.A. (1977). A reformulation of linear models. *J. Roy. Stat. Soc. A*, **140**, 48–77.

Nelder, J.A. (1982). Linear models and non-orthogonal data. *Utilitas Math.*, **21B**, 141–152.

Nelder. J.A. (1994). The Statistics of linear models: back to basics. *Stat. Comput.*, **4**, 221–234.

Nelder, J.A. (1995). Rejoinder to comments on 'The statistics of linear models: back to basics'. *Stat. Comput.*, **5**, 109–111.

Nelder, J.A. (1997). The great mixed-model muddle is alive and flourishing—alas! *Food Quality and Preference*, **9**, 157–159.

Nelder, J.A. (1998). The selection of terms in response-surface models—How strong is the weak-heredity principle? *Amer. Stat.*, **52**, 315–318.

[*Received February 2007, accepted June 2007*]

Response

Viviana B. Lencina[1], Julio M. Singer[2] and Edward J. Stanek III[3]

[1]*Instituto de Investigaciones Estadísticas, FCE, Universidad Nacional de Tucumán, Argentina*
[2]*Departamento de Estatística, IME, Universidade de São Paulo, Brazil*
[3]*Department of Public Health, SPHHS, University of Massachusetts at Amherst, USA*

1 Introduction

We begin by apologizing for the lamentable omission of Professor Nelder's interesting and controversial papers both in Lencina *et al.* (2005) and in Lencina & Singer (2006). Although the focus of our papers is much narrower than that of his articles, we feel that any manuscript touching on the topic of parametrization and definition of hypotheses in linear mixed models should necessarily refer to his numerous contributions to this field.

Essentially, we compare two stochastic models (one with constrained parameters, denoted *CP* and the other with unconstrained parameters, denoted *UP* by Voss (1999)) that are extensively employed to analyze data from balanced experiments with a fixed and a random factor. The two models generate different test statistics presumably directed at the same hypothesis, namely that of the non-existence of the random factor main effect in the presence of the interaction. Differing opinions over which statistic should be used is what we call the 'mixed-model controversy' following Voss (1999). Our objective is to provide an answer to this very specific question. To retain this focus, we do not discuss the general issue of model building or add to a debate over the reasonableness of the hypothesis under investigation. We explicitly state this in the abstract of Lencina *et al.* (2005).

The focus of Professor Nelder's comments is on broader issues that we avoid in either paper, and not on the specific problem we address. In particular, he

(1) considers the hypothesis under investigation of no practical interest since it does not satisfy what he calls the 'marginality principle';
(2) claims that constraints should not be put on the model parameters and therefore, that the *CP* model should be discarded.

Since such concerns relate to the narrower problem we address, we appreciate the opportunity to place our results in this context and further contribute to the discussion.

2 Is the "Marginality Principle" Universal?

Professor Nelder asserts that the hypothesis under dispute is not interesting since it assumes null marginal effects for one factor in a two-way table when interaction is present. We agree that in most cases, his assertion is true. However, this is not a universal rule and many authors, including several discussants (Lindley, Cox, Preece, Tukey, Frane and Jennrich) of Nelder (1977) have provided examples where the hypothesis of null random effects in the presence of interaction is reasonable. We consider an additional example in Lencina *et al.* (2005).

Nelder (1995) admits to being dogmatic on this issue. Since a null marginal effect in the presence of interaction violates his dogma, he claims that those who discuss such a problem are not in line with the 'current state of thinking on this subject'. For those who follow Professor Nelder's point of view, the discussion ends here, since the controversy exists only if we admit interest in such hypothesis.

Given the numerous practical examples where the hypothesis makes sense, we adopt a more flexible position and address the mixed-model controversy in our articles. This is in line with Lindsey (1995), who states that 'one should avoid being dogmatic, laying down certain rules which must always be followed'.

3 Should We Place Constraints on Parameters?

Professor Nelder pointed out an inconsistency in the specification of the fixed effects in Lencina *et al.* (2005, equation (1)), where the corresponding parameters are constrained, and Lencina & Singer (2006, equation (1)), where they are not. He asserts that while the latter is correct, the former is not. Although we prefer the constrained version of the model, both alternatives are correct and do not interfere in the theoretical results presented in either paper.

Professor Nelder emphasizes the use of constraints for the estimability purposes, whereas we advocate the use of constraints for the interpretability of parameters. If, for example, we add the constraint $\sum_{i=1}^{a} \alpha_i = 0$ to the fixed-effects parameters in Lencina & Singer (2006, equation (1)), the terms μ and α_i, $i = 1, \dots, a$ have clear interpretations. If left unconstrained, they are meaningless.

This seems to be a recurrent theme. Nelder (1998, equation (1)) uses the terms 'grand mean', 'row effects', 'column effects' and 'interaction between rows and columns' to refer to parameters that are not defined; since the estimating equations obtained from the minimization of $\sum(y_{ijk} - \nu - \alpha_i - \beta_j - \gamma_{ij})^2$ have infinite solutions, the one adopted by Nelder (1998, equation (2)) is based on a specific set of constraints on the estimates. How are these constraints chosen? The answer is clear, if we specify the parameters in his expression (1) by placing the constraints

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0.$$

The constraints on the fixed-effects parameters are better motivated if we consider a finite population set-up. In the setting outlined in Lencina *et al.* (2005) suppose that N homogeneous metal pieces are processed in each of a machines to produce screws and that we are interested in their lengths, assumed to be observed without error. Let the length of screw s produced by machine i ($i = 1, \dots, a$) be y_{is}, a non-stochastic value. The mean length of the screws produced by machine i is $\mu_i = N^{-1} \sum_{s=1}^{N} y_{is}$. Letting $\mu = (aN)^{-1} \sum_{i=1}^{a} \sum_{s=1}^{N} y_{is}$ represent the grand mean length, one way to define machine effects is to consider the difference between the mean length of the pieces produced by each machine and the grand mean, namely, $\alpha_i = \mu_i - \mu, i = 1, \dots, a$. Because $\mu = a^{-1} \sum_{i=1}^{a} \mu_i$, there is a natural constraint between the parameters α_i, $i = 1, \dots, a$, and it is $\sum_{i=1}^{a} \alpha_i = 0$. This constraint results from the definition of the grand mean and thus is not related to the estimation. In fact, up to here, there are no random terms involved. It is possible to define machine effects differently; if for example, one wishes to refer the results to those of a specific machine, say 1, then the effect of machine i, $i = 2, \dots, a$ could be defined as $\alpha_i = N^{-1} \sum_{s=1}^{N} y_{is} - \mu$ with $\mu = N^{-1} \sum_{s=1}^{N} y_{1s}$ denoting the mean length of the screws produced by machine 1. On the other hand, if a different number of pieces is manufactured in different machines, the grand mean could be defined as $\mu = \sum_{i=1}^{a} N_i \mu_i / \sum_{i=1}^{a} N_i$; then the constraint between the α_i is $\sum_{i=1}^{a} N_i \alpha_i = 0$. These constraints are neither arbitrary nor inconsistent; they correspond to different definitions of machine effects and do not depend on the sample data. We believe that parameters in statistical models should mimic those in the population; by doing so, the restrictions will emerge naturally. This approach has been considered by Hinkelmann & Kempthorne (1994), for example.

Different sets of constraints, as for example, deviations from means constraints or reference cell constraints do not alter the estimates of estimable linear combinations of parameters. The advantage of placing constraints on parameters is that it clarifies their interpretation. Rather than being motivated by estimability, constraints on parameters define interpretable models which we believe should precede the estimation and hypothesis-testing phases of the analysis. Both in teaching and consulting, we feel that starting off with well-defined parameters will help to specify both the operational objectives of the problem and the analysis strategy. In this process, the meaning of the parameters plays a fundamental role. This position has been also expressed by many authors, including Professor Cox in his reply to Nelder (1977).

Nelder (1994) also mentions cases where constraints placed on random-effects parameters lead to statistical inconsistencies by requiring that the sum of independent normal random variables with non-zero variance be identically equal to zero. We agree with his criticism and note that this is not the case with the *CP* model because the constrained random variables, namely,

$(\tau D)_{ij}$ subject to $\sum_{i=1}^{a}(\tau D)_{ij} = 0$ are not independent. In fact, we assume that $(\tau D)_{ij} \sim N[0, (1 - a^{-1})\sigma^2_{\tau D}]$ with $\text{Cov}[(\tau D)_{ij}, (\tau D)_{i'j}] = -a^{-1}\sigma^2_{\tau D}, i \neq i'$ so that the constraint is naturally satisfied.

In the context of a two-way fixed-effects study, placing constraints on estimates (as advocated by Nelder), or on parameters will produce identical estimates of estimable quantities. In the context of a two-way mixed-model, the two approaches lead to different expected mean squares for the term commonly taken to represent the main random effects in the presence of interaction. How to test for null random main effects in a mixed-model is where the mixed model controversy occurs.

One of Professor Nelder's main argument for using unconstrained parameters in a mixed model is that this approach leads to ANOVA tables with the same patterns for the expected mean squares irrespective of the nature (fixed or random) of the effects included in the model. Obtaining the results is definitely facilitated under such an approach, but we do not see why this feature should be used to justify the choice of the models. In giving such advice, more emphasis seems to be placed on form than on substance. In particular, Nelder (1994) argues that to avoid confusion, *UP* models should be used since they produce simple patterns for the expected mean squares for fixed and random effects. This, however, masks the difference in the interpretation of the *UP* and *CP* model parameters. We advocate that the decision on the appropriate test should rely on a clear identification of the hypothesis of null random effects (in the presence of an interaction). The difference between the variance components corresponding to 'random effects' under the *CP* and the *UP* mixed models has been discussed by Hocking (1973) and Voss (1999), with results summarized by Lencina *et al.* (2005). The parameter used to represent the random-effect variance in the *CP* model, σ^2_D, is different from the parameter representing the 'random-effects variance', σ^2_B, in the *UP* model. In fact, $a\sigma^2_D = a\sigma^2_B + \sigma^2_{\alpha B}$. This difference is not accounted for in Nelder (1994, Table 2). Instead, the same notation is used for the 'random effects variance' in the expected mean squares table for the *CP* and *UP* models. Using the notation from Lencina *et al.* (2005), the equivalent representation of the expected mean squares for a mixed model comparable to Nelder (1994, Table 2) is given in Table 1.

Table 1
Expected mean squares for the CP *and* UP *models.*

Term	*CP* Model *A* Fixed *B* Random	*UP* Model *A* Fixed *B* Random
A	$\sigma^2_{\tau D} + b\sum_{\tau}$	$\sigma^2_{\alpha B} + b\sum_{\alpha}$
B	$a\,\sigma^2_D$	$\sigma^2_{\alpha B} + a\,\sigma^2_B$
AB	$\sigma^2_{\tau D}$	$\sigma^2_{\alpha B}$

This clearly illustrates that the hypothesis $H : \sigma^2_D = 0$ is not equivalent to the hypothesis $H : \sigma^2_B = 0$. Which of them corresponds to a hypothesis of null random effects? We motivated an answer by defining the main effects for the random factor in terms of the potentially observable random variables in the study, similar to what is done for a fixed effect. Our conclusion is that the hypothesis of null random effects (in the presence of the interaction) is $H : \sigma^2_D = 0$ under the *CP* model but is $H : \sigma^2_{\alpha B} + a\,\sigma^2_B = 0$ under the *UP* model. Therefore, the hypothesis $\sigma^2_B = 0$ under the *UP* model does not have the same interpretation as $H : \sigma^2_D = 0$ under the *CP* model.

Furthermore, working in a finite population set-up, Wilk & Kempthorne (1955) obtain covariance structures that are similar to those in the *CP* model. This was our comment in the last paragraph of Lencina *et al.* (2005).

The importance that Professor Nelder places on obtaining the same simple patterns of expected mean squares for fixed and random effects in a mixed model, and on placing constraints on estimates, not parameters, does not appear to clarify what he calls the basic mixed-model muddle. In order to distinguish what test to use for null random effects in the *CP* and *UP* models, the crux of the problem is defining interpretable parameters, which in some cases requires the use of constraints. Thus we disagree with Professor Nelder in his advice to avoid parameter constraints when specifying mixed models. Although different opinions are at stake, we believe that discussions regarding the interpretation of statistical models and their use in practical problems are always welcome. We end by thanking Professor Nelder for calling our attention to such important issues.

Acknowledgments

We are grateful to Consejo de Investigaciones de la Universidad Nacional de Tucumán, Argentina, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil, and the National Institutes of Health (NIH-PHS-R01-HD36848, R01-HL071828-02), USA, for financial support.

References

Hinkelmann, K. & Kempthorne, O. (1994). *Design and Analysis of Experiments: Introduction to Experimental Design*, Vol 1. New York: Wiley.
Hocking, R.R. (1973). A discussion of the two-way mixed model. *Amer. Stat.*, **27**, 148–152.
Lencina, V.B. & Singer, J.M. (2006). Measure for measure: exact F tests and the mixed models controversy. *Int. Stat. Rev.*, **74**, 391–402.
Lencina, V.B., Singer, J.M. & Stanek, E.J. (2005). Much ado about nothing: the mixed models controversy revisited. *Int. Stat. Rev.*, **73**, 9–20.
Lindsey, J.K. (1995). Comments on J.A. Nelder: 'The statistics of linear models: back to basics'. *Stat. Comput.*, **5**, 87–89.
Nelder, J.A. (1977). A reformulation of linear models. With discussion. *J. Roy. Stat. Soc., A*, **140**, 48–77.
Nelder. J.A. (1994). The Statistics of linear models: back to basics. *Stat. Comput.*, **4**, 221–234.
Nelder. J.A. (1995). Rejoinder to cmments on 'The Statistics of linear models: back to basics'. *Stat. Comput.*, **5**, 109–111.
Nelder, J.A. (1998). The great mixed-model muddle is alive and flourishing—alas! *Food Quality and Preference*, **9**, 157–159.
Voss, D.T. (1999). Resolving the mixed models controversy. *Amer. Stat.*, **53**, 352–356.
Wilk, M.B. & Kempthorne, O. (1955). Fixed, mixed and random models. *J. Amer. Stat. Assoc.*, **50**, 1144–1167.