

ON INTERACTION VARIANCE COMPONENTS IN MIXED MODELS

H. O. Hartley and S. R. Searle*

Texas A. & M. University

Summary

An inconsistency regarding variance components for interactions between fixed and random effects and their occurrence in expected mean squares is explained and commented upon.

1. Introduction

In the analysis of variance of data from a rows-by-columns (treatments-by-blocks) environment, there appears to be some inconsistency in the literature about the presence or absence of the interaction variance component in the expectation of the mean square for the random factor in the case of a mixed model. Furthermore, the commonly accepted position (absence of the component) represents a discontinuity with the analysis of unbalanced data. This note draws attention to this discontinuity in the hope, if not of clarifying the issue, at least of bringing it to the attention of interested readers.

Since much of the literature about two-way classifications is in terms of treatments and blocks (randomized complete blocks designs with replication), we consider the problem within this framework. In doing so, however, we think of blocks as being any other factor crossed with treatments, such as varieties,

*On leave from Cornell University, 1968-9

breeds or processes. In this context, for the case of t treatments and b blocks we consider two situations: (i) the same number of observations on each of the tb treatment-by-block combinations (balanced data) and (ii) varying numbers of observations on each of the tb combinations (unbalanced data). In either case we think of y_{ijk} being the k 'th observation on the i 'th treatment in the j 'th block and write the equation of the model as

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk} \quad (1)$$

where μ is a general mean, τ_i is the treatment effect due to the i 'th treatment, β_j is the block effect due to the j 'th block, $(\tau\beta)_{ij}$ is the interaction effect and e_{ijk} is the customary random error term. The e 's are assumed to have zero mean and variance σ_e^2 , with the covariance between any pair of them being zero. We now consider the problem of defining the terms in (1).

2. Balanced Data

Even in the case of balanced data, in which precisely n observations are made for each of the tb treatment block combinations (the so-called randomized block experiment with cell repetition), disagreements arise as to definition of the model. They can be traced to the postulation of different assumptions. We confine ourselves to mentioning the two most frequently used alternative premises:-

(a) Finite population randomization

and (b) The additive component mixed model.

For (a) reference is made to Kempthorne (1951, p. 532) and in more detail Kempthorne (1957, p. 247). Briefly Kempthorne envisages a finite population

of T treatments and B blocks with TB numbers μ_{ij} representing the true measurements of responses when treatment i occurs on block j. Using the identity

$$\mu_{ij} = \bar{\mu}_{..} + (\bar{\mu}_{i.} - \bar{\mu}_{..}) + (\bar{\mu}_{.j} - \bar{\mu}_{..}) + (\mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \bar{\mu}_{..}) \quad (2)$$

(where the $\bar{\mu}_{..}$, $\bar{\mu}_{i.}$ and $\bar{\mu}_{.j}$ are the customary arithmetic means), and the definitions

$$\begin{aligned} \mu &= \bar{\mu}_{..}, \\ \tau_i &= \bar{\mu}_{i.} - \bar{\mu}_{..}, \\ \beta_j &= \bar{\mu}_{.j} - \bar{\mu}_{..}, \end{aligned} \quad (3)$$

and $(\tau\beta)_{ij} = \mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \bar{\mu}_{..}$,

the identity (2) can be written

$$\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} \quad (4)$$

with the automatic properties

$$\sum_{i=1}^T (\tau\beta)_{ij} = 0 \text{ for all } j, \text{ and } \sum_{j=1}^B (\tau\beta)_{ij} = 0 \text{ for all } i. \quad (5)$$

The randomization procedure of implementing the 'randomized block' experiment would then select the complete set of $t = T$ treatments but only a random sample of b blocks from the total of B blocks and the classical situation of the so-called 'mixed model' arises as the limiting case when $B \rightarrow \infty$ (Treatments 'fixed', and Blocks 'random'). An automatic consequence of the first set of equations in (5) is that neither the 'Block sum of squares'

$$\sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2 \quad (6)$$

(where Σ' denotes summation over just the sampled blocks) nor its expected value contain any 'interaction' components $(\tau\beta)_{ij}$.

By contrast the additive component mixed model (b) postulates a model of the form (4) in which μ and the t numbers τ_i , $i = 1, 2, \dots, t$ are constants, the β_j represent a random sample of b values from an infinite population with mean 0 and variance σ_β^2 and the $(\tau\beta)_{ij}$ are an independent random sample of tb values from a population with mean 0 and variance $\sigma_{\tau\beta}^2$. As a consequence the expectation of the block sum of squares now involves $\sigma_{\tau\beta}^2$ and is given by

$$E \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2 = (b - 1)(t n \sigma_\beta^2 + n \sigma_{\tau\beta}^2 + \sigma_e^2) \quad (7)$$

Since the approach (a) does not postulate an additive model whilst the model (b) does, the former is usually preferred in most basic textbooks. Indeed, the only textbooks using (b) which have come to our notice are Mood (1950) and Steel and Torrie (1960). However, since there are two major difficulties in following approach (a) when analyzing unbalanced data and because these difficulties are not just mathematical but conceptual we must here (in the balanced case) raise certain questions about approach (a).

(i) Since obviously the identity (2) and definitions (3) are not unique the question arises why they and no other identity and definitions are regarded appropriate for the drawing of inferences. Indeed Kempthorne (1957, p. 257) mentions that the same approach can be applied to the so-called two-way classification with proportional cell frequencies in which $n_{ij} = n_i m_j$ observations are made in the (i, j) treatment-block combination. However in this case a

different identity is used in that the $\bar{\mu}_{i.}$, $\bar{\mu}_{.j}$ and $\bar{\mu}_{..}$ of (2) are then weighted means of the μ_{ij} with weights m_j , n_i and $n_i m_j$ respectively, resulting in definitions of the μ , τ_i and β_j different from (3). Under these conditions, the definition of 'treatment effects' τ_i (presumably a basic property of the finite population of μ_{ij}) would depend on the n_i and m_j i.e. on the whim of the experimenter.

(ii) Why should we compute the customary ANOVA sum of squares for inferential purposes? Let us first give an answer to this question when we adopt the model (b) and the additional assumption that $\beta_j \sim N(0, \sigma_\beta^2)$ for all j and $(\tau\beta)_{ij} \sim N(0, \sigma^2)$ for all i and j . It is well known that the ANOVA statistics then arise as optima from various principles of estimation and hypothesis testing (maximum likelihood, sufficiency and likelihood ratio principles). In contrast, for the finite population (a), it is well known that we have the customary criteria of unbiasedness. For example

$$s_b^2 = \frac{b}{\sum_{j=1}^b} (\bar{\mu}_{.j} - \bar{\mu}_{..})^2 / (b - 1) \quad (8)$$

is an unbiased estimator of the corresponding population variance component if defined by

$$S_b^2 = \frac{b}{\sum_{j=1}^b} (\bar{\mu}_{.j} - \bar{\mu}_{..})^2 / (B - 1) \quad (9)$$

where the $\bar{\mu}_{..}$ in (8) and (9) are respectively means of the b sampled and B population $\bar{\mu}_{.j}$. Moreover it has recently been shown [Hartley and Rao (1968)] that s_b^2 is an U.M.V. estimator of S_b^2 . It is quite conceivable, therefore, that at least in the case of balanced data analysis a rationale of reasonable-

ness can be provided for the ANOVA sum of squares. We summarize the comparison of the two premises (a) and (b) for balanced data in Table 1.

Table 1. Comparison of two alternative premises for analyzing balanced data in a mixed model.

Mode of comparison	Premise	
	(a)	(b)
	Finite population model	Additive component model
Assumption in model	None	Highly parametric
Definitions of elements in model	Arbitrary but 'reasonable'	Implied in model
Justification of ANOVA statistics	Fairly well established	Well established

Whilst the above comparison would appear to favor (a) we must postpone our evaluation until after the discussion of unbalanced data.

3. Unbalanced Data

Kempthorne (1957, p. 257) describes the analysis of unbalanced data by the finite population randomization approach as an 'important unsolved problem in statistics'. Whilst the mathematical difficulties in this problem may indeed be considerable we must consider the even more important conceptual questions (i) and (ii) raised in Section 2. We do this for the two premises (a) and (b).

(a) Finite population model

(i) How can we define both an identity corresponding to (2) and associated elements corresponding to (3) and (4)? It is of course possible to define weighted averages of the μ_{ij} with the cell frequencies n_{ij} as weights. However, there then immediately arises an inconsistency in that for a particular i a μ_{ij} with $n_{ij} = 0$ would be given no weight in (2) whilst in the definition (3) of τ_i it would be included with weight 1. It may be that we should define parameter functions τ_i , β_j and $(\tau\beta)_{ij}$ which in some sense are 'relevant' for the description of the finite population of μ_{ij} . If this is done we would not normally expect to obtain relations for either unweighted sums as in (5) [since certain (i, j) combinations may be known to be 'impossible'] or indeed weighted sums with weights depending on the sampling design.

(ii) Assuming that we can succeed in deciding on relevant parameter functions to employ in an identity such as (2), what statistics (such as sums of squares) are we to compute to estimate the parameter functions? For simple random and stratified sampling certain principles of UMV and maximum likelihood estimation have recently been obtained for the estimation of population moments by sample moments [Hartley and Rao (1968)] but the results are as yet somewhat limited. It is very unlikely therefore that the adherence to the finite population sampling premise will provide any 'guide lines' for an 'Analysis of Variance' of unbalanced data. Indeed, it is doubtful whether an ANOVA can be claimed to arise from any optimality principle of estimation theory with this premise.

(b) Additive component model

(i) With this model there is of course no problem of definition, as the only difficulty with this approach is in the justification of the postulated model. Once justified the model itself provides the definition of its elements.

(ii) As is well known, with unbalanced data from a completely fixed effects model the theory of linear estimation provides an entirely satisfactory estimation theory based on well established optimality principles. However, in the case of a mixed model for unbalanced data most of the published ANOVA procedures are based on 'reasonable' analogies to the fixed model case [see e.g. Henderson (1953), Searle and Henderson (1961 and 1967) and Searle (1968)]. In addition, Hartley and Rao (1967) have recently provided a maximum likelihood estimation procedure which, in the general case of unbalanced data does not lead to the computation of ANOVA mean squares. However maximum likelihood estimation does, of course, reproduce the familiar ANOVA estimators in the special case of balanced data. The acceptance of model (b) therefore provides perfect continuity of both concepts and analyses, for with it balanced data analyses arise as just special cases of the unbalanced data estimation procedure.

In view of the above we regard the difference between the two approaches as one between a parametric [case (b)] and non-parametric [case (a)] analysis of identical data. The parametric analysis is based on a very specific model (b), but if the postulation of the model is justified this method covers both the analysis of unbalanced and balanced data, in consistent fashion backed by well established principles of estimation. Of course, this parametric model (b) is undoubtedly postulated in many situations when it is not warranted. By

contrast the non-parametric approach (a) provides an ingenious solution free from any assumptions but at present restricted to balanced data analysis. Now the customary strategy in choosing between parametric and non-parametric methods is to prefer the former in situations when the parametric model is really justified and use the latter only if it is not. Yet, to our view regrettably, most textbooks on analysis of variance invariably recommend the non-parametric treatment of the mixed model for the special case of balanced data, despite the fact that most workers who analyze mixed model data (e.g. Searle and Henderson, 1960) retreat to the parametric treatment for the general case, with a consequential discontinuity in concepts and statistical inference procedures. These discontinuities can be avoided by adopting the parametric model (b). For this reason we prefer this model in situations when it can be justified. Since there are obviously situations in which the parametric model may not be justified, additional research is needed on the robustness of analysis (b) to departures from the parametric model.

4. Some Consequences of the Parametric Model

Consequences of the parametric model are vividly demonstrated by considering the analysis of variance, for balanced data, for the model specified in (1). As a mixed model we take the treatment effects τ_i as fixed and the block effects β_j as random with zero means and variances σ_β^2 ; and the interaction effects are also assumed random with zero means and variances $\sigma_{\tau\beta}^2$. All covariances are assumed zero.

Calculation of the sums of squares and mean squares in the analysis of variance for this situation is well-known [e.g. Steele and Torrie (1960), Table 8.5)]. Only the degrees of freedom and expected mean squares are there-

fore shown in Table 2.

Table 2. 2-way classification, mixed model,
treatment effects, τ_i , fixed

Source of variation	Degrees of freedom	Mean Square	Expected value of mean square
Treatments	$t - 1$	MST	$\frac{bn}{t - 1} \sum_{i=1}^t (\tau_i - \bar{\tau})^2 + n\sigma_{\tau\beta}^2 + \sigma_e^2$
Blocks	$b - 1$	MSB	$tn\sigma_{\beta}^2 + n\sigma_{\tau\beta}^{2*} + \sigma_e^2$
Interaction	$(t - 1)(b - 1)$	MSTB	$n\sigma_{\tau\beta}^2 + \sigma_e^2$
Error	$tb(n - 1)$	MSE	σ_e^2
Total	$tbn - 1$		

*This term is not present in the 'finite population' model.

With the additive component premise (b)--the parametric model--we will, in both the completely random model and the fixed model, obtain the same expected mean squares for this analysis of variance as are obtained with the finite population model (a) using the identity (2) and definitions (3). However, with the mixed model there is a difference: with the parametric model (b), the term in $\sigma_{\tau\beta}^2$ occurs in $E(\text{MSB})$ as shown in Table 2. With the finite population model (a), it does not. This difference between the two models is of material importance because the presence or absence of $\sigma_{\tau\beta}^2$ from $E(\text{MSB})$ determines the mean square (MSTB or MSE) which is used for testing the hypothesis $H_0 : \sigma_{\beta}^2 = 0$. This hypothesis can be of prime interest in

situations where the β -effects are random effects of more specific interest than that usually attached to the 'blocks' of the experimental block concept. With the finite population (non-parametric) model MSE is the appropriate error term, but with the additive component (parametric) model MSTB is the error term, as indicated in Table 2. For reasons already given, we prefer this latter case, the parametric model.

Extension to mixed models in general is clear. Rules of thumb for writing down expected values of mean squares in analyses of variance for balanced data in crossed and nested classifications have, for example, been given by Schultz (1955) and Henderson (1959). They include provision for first treating any model as if it were completely random and then both eliminating certain variance components from the expected mean squares if some effects are fixed, and changing others to be quadratic functions of the fixed effects themselves. In keeping with Table 2 we now see that no variance components get eliminated, they either remain or get changed to quadratic functions. For example, the rules as given by Henderson (1959) would eliminate $n\sigma_{\tau\beta}^2$ from $E(\text{MSB})$, and we have suggested that it should remain. Extended to the completely fixed model we would find that although the terms in $\sigma_{\tau\beta}^2$ would not be there as such, they would be changed into quadratic functions in the $(\tau\beta)_{ij}$'s.

Acknowledgment

Thanks go to Dr. C. E. Gates for bringing this problem to our notice.

References

- Hartley, H. O. and Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 54, 93-108.
- Hartley, H. O. and Rao, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika* 55, --. (in press)
- Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics* 9, 226-252.
- Henderson, C. R. (1959). Design and analysis of animal husbandry experiments. Chapter 1 of "Techniques and Procedures in Animal Production Research", American Society of Animal Production Monograph.
- Kempthorne, O. (1951). "Design and Analysis of Experiments", John Wiley & Sons, New York.
- Kempthorne, O. (1957). "An Introduction to Genetic Statistics", John Wiley & Sons, New York.
- Mood, A. M. (1950). "Introduction to the Theory of Statistics", McGraw-Hill, New York.
- Schultz, E. F., Jr. (1955). Rules of thumb for determining expectations of mean squares in analysis of variance. *Biometrics* 11, 123-135.
- Searle, S. R. and Henderson, C. R. (1960). Judging the effectiveness of age-correction factors. *J. Dairy Sci.* 43, 966-974.
- Searle, S. R. and Henderson, C. R. (1961 and 1967). Computing procedures for estimating components of variance in the two-way classification. *Biometrics* 17, 607-616; and Correction: *Biometrics* 23, 852, 1967.
- Searle, S. R. (1968). Another look at Henderson's methods of estimating variance components. *Biometrics* 24, --. (in press)
- Steele, R. G. D. and Torrie, J. H. (1960). "Principles and Procedures of Statistics", McGraw-Hill, New York.