UNIVARIATE DATA FOR MULTI-VARIABLE SITUATIONS: ESTIMATING VARIANCE COMPONENTS*

BU-411-M

S. R. Searle
Biometrics Unit, Cornell University, Ithaca, New York, U.S.A.

Abstract

An outline is given of variance components models, of estimation procedures associated with them and of the difficulties involved in the unbalanced data that are so often available from such models. Several specific unsolved problems are described.

* Prepared for a Research Seminar on Multivariate Statistical Inference, Dalhousie University, Halifax, Nova Scotia, Canada, March 23-25, 1972.

1. Introduction

Multivariate analysis means many things to many people. Traditionally it is usually thought of in terms of such things as the multi-normal distribution, factor analysis, discriminant analysis and so on. However, it is because of this traditional view of multivariate analysis that Dempster [1971] accuses many academic statisticians of having defined it too narrowly, "excluding even such obviously multivariate data types as factorial experiments, contingency tables and time series". By the same token variance components models also come within this wider purview of multivariate analysis and since they are so often by-passed even in traditional univariate analysis presentations it is opportune to outline such models here, to indicate their practical importance and to highlight some of the unsolved problems associated with them. In a rather special sense they are truly multivariate models — but with the peculiarity that available data are only univariate. Hence my title.

The classical form of a linear model is

$$\underline{y} = \underline{\underline{X}}\underline{\beta} + \underline{e} \tag{1}$$

where $\underline{y}$ is a vector of observations on a random variable Y, $\underline{\beta}$ is a vector of parameters to be estimated, $\underline{\underline{X}}$ is a matrix of known values and $\underline{e}$ is a vector of residual error terms. The elements of $\underline{\beta}$ are called constants or fixed effects, and are regression slopes, main effects, or interactions, depending on the context. In this way the formulation in (1) embraces all analysis of variance models, regression models and mixtures of the two, namely analysis of covariance models. In all these cases the elements of $\underline{\beta}$ are never envisaged as random variables. They are parameters to be estimated, generically referred to as fixed effects. In contrast, there are models in which some of the elements of $\underline{\beta}$ are random variables. Usually they are the effects corresponding, in traditional analysis of variance situations, to the levels of one or more factors (or interactions) and have therefore been called random <u>effects</u>. It is the variances of these random effects (random variables) that are the parameters of interest insofar as these elements of $\underline{\beta}$ are concerned.

Despite Yates' [1967] comment that "unfortunately after the war a new concept of fixed and random effects models was introduced" there is currently an increasing interest in these random effects models (sometimes also called variance components models), with numerous new results having been published in recent years including, one notes, two papers in last year's inaugural volume of the Journal of Multivariate Analysis. Many of these deal with the more difficult aspects of estimation that arise from unequal-subclass-number data: for example, the multiplicity of estimation procedures, the intractability of criteria for judging between them and the prescience of negative estimates of parameters (namely variances) that are by definition positive.

2. Examples and models

Variance components have had a long use in genetics. Suppose a male animal has many progeny, a Holstein bull for example, which, through the use of artificial insemination, has sired many cows. If x_{ij} is the milk yield of the j^{th} daughter of the i^{th} sire a suitable model is

$$x_{ij} = \mu + \alpha_i + e_{ij} \quad \text{for} \quad \begin{array}{l} i = 1,2,\cdots,a \\ j = 1,2,\cdots,n_i \end{array} \quad . \tag{2}$$

The parameter μ is a general mean, α_i is the effect due to the i^{th} sire (a random effect from a population of α's that has zero mean and variance σ_α^2) and the e_{ij}'s are the usual random error terms, uncorrelated with the α's, having zero mean and variance σ_e^2. Thus $\sigma_x^2 = \sigma_\alpha^2 + \sigma_e^2$ and the problem is to estimate σ_α^2 and σ_e^2. The sire's effect α_i on his progeny's milk yield represents a random half of the sire's genetic make-up, so that $\sigma_\alpha^2 = \frac{1}{4}\sigma_G^2$ where σ_G^2 is the (additive) genetic variance of milk yield. The ratio of this to σ_x^2, namely $h = \sigma_G^2/\sigma_x^2 = 4\sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_e^2)$ which, known as heritability, is of great use in animal breeding programs. For example, expected increases in yield arising from selecting a high-yielding fraction of animals to be parents of the next generation are proportional to h.

Another example is crossing 2 varieties of corn, using pollen from replicate males (tassels) of one variety on replicate females (silks) of the other variety, the sample of males and females used in each case being considered random samples from the varieties concerned. If x_{ijk} is the yield of the k^{th}

plant resulting from crossing the i^{th} male with the j^{th} female the model

$$x_{ijk} = \mu + m_i + f_j + (mf)_{ij} + e_{ijk} \tag{3}$$

is appropriate. Here m_i, f_j and $(mf)_{ij}$ are uncorrelated random variables with zero means and homoscedastic with variances σ_m^2, σ_f^2 and σ_{mf}^2 respectively. These variables are also uncorrelated with the error terms e_{ijk} which have zero mean and variance σ_e^2.

There are also models involving both random and fixed effects. A non-biological example given in Thompson [1963] is that of analyzing the muzzle velocity x_{ij} of the i^{th} shell fired from a gun using the j^{th} of several measuring instruments. Here we have

$$x_{ij} = \mu + s_i + m_j + e_{ij} \tag{4}$$

where s_i is the effect due to the i^{th} shell and m_j is the bias in the j^{th} measuring instrument. Since the shells used are a random sample of shells, the s_i are random effects, whereas the m_j are fixed effects.

A model such as (4) is usually called a mixed model, involving as it does both fixed and random effects. But since in all models μ is a fixed effect and the error terms are random, all models can be considered as mixed. To distinguish the 2 kinds of effects a generalization of the fixed effects model (1) is

$$\underline{y} = \underline{X}\underline{\beta} + \underline{Z}\underline{u} + \underline{e} \tag{5}$$

for the mixed effects model. Here $\underline{\beta}$ is the vector of fixed effects (including coefficients of covariates if present), $\underline{u}$ is the vector of random effects and $\underline{e}$ is the vector of error terms. $\underline{X}$ and $\underline{Z}$ are matrices of known values corresponding to the incidence of the fixed and random effects respectively in $\underline{y}$. Properties generally attributed to the random variables in $\underline{u}$ and $\underline{e}$ are

$$E(\underline{u}) = \underline{0}, \quad E(\underline{e}) = \underline{0}, \quad E(\underline{u}\underline{e}') = \underline{0},$$

$$\text{var}(\underline{u}) = E(\underline{u}\underline{u}') = \underline{D} \quad \text{and} \quad \text{var}(\underline{e}) = E(\underline{e}\underline{e}') = \underline{R} \, .$$

Hence

$$E(\underline{y}) = \underline{X}\underline{\beta} \tag{6}$$

and

$$\mathrm{var}(\underline{y}) = \underline{Z}\underline{D}\underline{Z}' + \underline{R} \equiv \underline{V}\ , \text{ say.} \tag{7}$$

The problem is to estimate not only $\underline{\beta}$ but also $\underline{D}$ and $\underline{R}$. Usually $\underline{R}$ is taken as $\sigma_e^2\underline{I}$ and $\underline{D}$ is taken as a diagonal matrix. For example, in reformulating (2) in the form of (5), the vector $\underline{u}$ would be $\underline{u}' = \underline{\alpha}' = (\alpha_1\ \alpha_2\ \cdots\ \alpha_a)$ so that $\underline{D}$ would be $\underline{D} = \sigma_\alpha^2\underline{I}_a$. Similarly for (3), with a males, $i = 1,2,\cdots,a$ and b females, $j = 1,2,\cdots,b$, $\underline{D}$ would be

$$\underline{D} = \begin{bmatrix} \sigma_m^2\underline{I}_a & \underline{0} & \underline{0} \\ \underline{0} & \sigma_f^2\underline{I}_b & \underline{0} \\ \underline{0} & \underline{0} & \sigma_{mf}^2\underline{I}_{ab} \end{bmatrix} . \tag{8}$$

In general, we can specify $\underline{u}$ as

$$\underline{u}' = [\underline{u}_1'\ \ \underline{u}_2'\ \ \cdots\ \ \underline{u}_\theta'\ \ \cdots\ \ \underline{u}_K'] \tag{9}$$

and $\underline{Z}$ correspondingly as

$$\underline{Z} = [\underline{Z}_1\ \ \underline{Z}_2\ \ \cdots\ \ \underline{Z}_\theta\ \ \cdots\ \ \underline{Z}_K] \tag{10}$$

where $\underline{u}_\theta'$ is the vector of N_θ effects for the θ^{th} factor (main effects or interaction factor), for $\theta = 1,2,\cdots,K$. Customarily the elements of $\underline{u}_\theta$ are assumed to have zero mean, be uncorrelated and have uniform variance σ_θ^2, and to be uncorrelated with elements of all other $\underline{u}$'s; i.e., $E(\underline{u}_\theta) = 0$, $\mathrm{var}(\underline{u}_\theta) = \sigma_\theta^2\underline{I}_{N_\theta}$, $E(\underline{u}_\theta\underline{u}_\varphi') = \underline{0}$ for $\theta \neq \varphi$ and $E(\underline{u}_\theta\underline{e}') = \underline{0}$. Hence $\underline{D}$ is a diagonal matrix of the matrices $\sigma_\theta^2\underline{I}_{N_\theta}$ and so can be written as

$$\underline{D} = \sum_{\theta=1}^{K}{}^{+}\sigma_\theta^2\underline{I}_{N_\theta} \tag{11}$$

where Σ^+ denotes the operation of a direct (Kronecker) sum of matrices. An example is (8). Then $\underline{V}$ in (7) becomes, using $\underline{R} = \sigma_e^2\underline{I}$,

$$\underline{V} = \sum_{\theta=1}^{K} \sigma_\theta^2 \underline{Z}_\theta \underline{Z}_\theta' + \sigma_e^2 \underline{I} \tag{12}$$

which, by defining $\underline{Z}_{K+1} \equiv \underline{I}$ and $\sigma_{K+1}^2 \equiv \sigma_e^2$ can be further generalized as

$$\underline{V} = \sum_{\theta=1}^{K+1} \sigma_\theta^2 \underline{Z}_\theta \underline{Z}_\theta' \; . \tag{13}$$

The problem is to estimate the variance components σ_e^2 and σ_θ^2 for $\theta = 1,2,\cdots,K$.

Additional generality could be given to the model by assuming $\mathrm{var}(\underline{u}_\theta)$ to be $\underline{\Sigma}_{\theta\theta}$ say, rather than $\sigma_\theta^2 \underline{I}_{N_\theta}$, and more still by assuming $\mathrm{cov}(\underline{u}_\theta \underline{u}_\varphi')$ to be $\underline{\Sigma}_{\theta\varphi}$ say, rather than zero. Then $\underline{D}$ would be $\underline{D} = \{\underline{\Sigma}_{\theta\varphi}\}$ for $\theta,\varphi = 1,2,\cdots,K$. However, there are difficulties enough in estimating $\underline{D}$ when it has the simple form shown in (11) – i.e., estimating the σ^2's – so that (11) is the form usually assumed.

3. Available data

Situations for which random effects models are appropriate often yield, especially in biology and economics, what can be called "messy data". The data are often voluminous in extent and frequently include large samples of the random effects. However, these same data often stem from survey-like situations and seldon do they have a uniform number of observations in each sub-most subclass. Not only may the numbers vary greatly but many subclasses may be empty, having no observations at all – in some cases as many as 30% or more of the subclasses being empty. At first thought the prospects of having efficient estimation procedures for such data are gloomy, and indeed they are. But the pressing need by biologists, economists and others for variance components estimates that they can use reliably in their work is such that development of efficient estimators for their kinds of data is worth pursuing.

Dichotomizing data according to the number of observations in the subclasses, namely data having equal subclass numbers (which we call balanced data) or data having unequal subclass numbers (which we call unbalanced data) is pertinent to variance component estimation because in the one case estimation is easy and well documented and in the other it is difficult with unsolved problems. The easy case is balanced data; the difficult case is unbalanced data. We deal first, and briefly, with the easy case.

4. Estimation: balanced data

Consideration of variance components models in balanced data led to estimation methods based on the mean squares of classical analyses of variance. Expected values of these mean squares are linear functions of the variance components and equating them to observed values gives linear equations in the components, the solutions of which are taken as estimators. Suppose $\underline{m}$ and $\underline{\sigma}^2$ are vectors of analysis of variance mean squares and variance components respectively for some set of data. Writing the expected value of $\underline{m}$ as $\underline{P}\underline{\sigma}^2$ we have

$$E(\underline{m}) = \underline{P}\underline{\sigma}^2 \tag{14}$$

and the equations for deriving $\hat{\underline{\sigma}}^2$ as an estimator of $\underline{\sigma}^2$ are

$$\underline{m} = \underline{P}\hat{\underline{\sigma}}^2 \ . \tag{15}$$

For random models the elements of $\underline{m}$ are all the mean squares of the appropriate analysis of variance and in mixed models they are the mean squares whose expectations contain no fixed effects. In both cases (for balanced data) $\underline{P}$ is non-singular and the estimators are $\underline{P}^{-1}\hat{\underline{\sigma}}^2$.

Unbiasedness is a well-evident property of estimators obtained from (14) and (15), e.g., Winsor and Clarke [1940]; but establishment of other properties has been relatively recent; e.g., minimum variance quadratic unbiasedness and, under normality assumptions, minimum variance unbiasedness, Graybill and Hultquist [1961]. Normality assumptions for the random effects also lead, as usual, to the analysis of variance sums of squares having χ^2-distributions (multiplied by constants) with the result that confidence intervals and test statistics for hypotheses about certain linear combinations of the components can be derived. However, the linear combinations of χ^2-variables that constitute the estimators have coefficients that involve the unknown components and so the distributions of the estimators are unknown. For example, for the model (2) with $n_i = n$ for all i

$$\hat{\sigma}^2_\alpha = \frac{n\sigma^2_\alpha + \sigma^2_e}{n(a-1)}\chi^2_{a-1} - \frac{\sigma^2_e}{an(n-1)}\chi^2_{a(n-1)} \tag{16}$$

where the χ^2_r symbol here denotes a variable distributed as χ^2 on r degrees of freedom. Furthermore, since the estimators involve differences between such variables, as in (16), their distributions involve sums of confluent hypergeometric functions as in Robinson [1965] and Wang [1967]. However, one characteristic of the estimators that can be derived is their variances, because the χ^2-variables of which they are linear combinations are independent. Unbiased estimators of these variances are also available, Ahrens [1965] (see also Searle [1971a]).

Maximum likelihood estimation using normality assumptions leads pro forma to almost the same estimators as given by the analysis of variance method summarized in (14) and (15). However, the estimators can be negative so they cannot truly be maximum likelihood estimators since these would stem from maximizing the likelihood over the parameter space which, for variance components, is non-negative. Herbach [1959], Thompson [1962] and Thompson and Moore [1963] discuss this problem.

Estimates obtained from (15) are, on some occasions, negative. This is clearly embarrassing because the corresponding parameters, being variances, are essentially positive. Many awkward moments arise between a consulting statistician and his client when explanation of this peculiarity is called for and found wanting. Various unsatisfactory alternatives are listed in Searle [1971a] but the need for developing essentially positive estimators remains.

5. Estimation: unbalanced data

The innocent looking difference between balanced and unbalanced data has widespread ramifications in the task of estimating variance components. It leads to a variety of methods of estimation, properties of which are mostly unknown (save for unbiasedness which is almost universally achieved). This, in combination with the largely empirical nature of the criteria used for deriving the estimation methods, makes it almost impossible to pass judgement on the different estimators. Furthermore, the algebra involved is horrendous ("algebraic heroics" are Hartley's [1967] words) and computing procedures are correspondingly difficult even for large computers; e.g., inverting matrices of order 1000 and greater. Nevertheless, the practical need for efficient estimators of variance components from unbalanced data is sufficiently compelling to pursue the problems involved.

5a. Basic development

The development of estimators has basically been that of a variety of quadratic forms in the observation vector $\underline{y}$ having expected values that are linear combinations of the variance components. Thus if $\underline{y}'\underline{Q}\underline{y}$ is one such quadratic form we know from (6) and (7), without any assumptions about the form of the distribution of $\underline{y}$, that

$$E(\underline{y}'\underline{Q}\underline{y}) = tr(\underline{Q}\underline{V}) + \underline{\beta}'\underline{X}'\underline{Q}\underline{X}\underline{\beta} \quad . \tag{17}$$

$\underline{Q}$ is therefore chosen so that

$$\underline{Q}\underline{X} = \underline{0} \; . \tag{18}$$

Then, analogous to (14) and (15), if $\underline{q}(\underline{y})$ is a vector of such quadratic forms with

$$E[\underline{q}(\underline{y})] = \underline{P}\underline{\sigma}^2 \; , \tag{19}$$

the equations for deriving $\hat{\underline{\sigma}}^2$ as an estimator of $\underline{\sigma}^2$ are

$$\underline{P}\hat{\underline{\sigma}}^2 = \underline{q}(\underline{y}) \; . \tag{20}$$

Now with balanced data there is an 'obvious' set of mean squares (quadratic forms) to use in $\underline{m}$ in the estimation procedure of (14) and (15), and the resulting estimators have been shown to have certain desirable properties in addition to unbiasedness. In contrast, with unbalanced data there is no single set of quadratic forms satisfying (18) that are 'obvious' for use in (19) and (20). A variety of suggestions have been made, some of them involving more equations in (19) than there are variance components. In this case equations (20) are over-identified, but provided $\underline{P}$ of (19) has full column rank, a 'least squares' solution can be obtained as

$$\hat{\underline{\sigma}}^2 = (\underline{P}'\underline{P})^{-1}\underline{P}'\underline{q}(\underline{y}) \; . \tag{21}$$

5b. Adaptations of analysis of variance

Until quite recently the quadratic forms $\underline{y}'\underline{Q}\underline{y}$ suggested for use in $\underline{q}(\underline{y})$ of (19) and (20) have been chosen by analogy with classical analysis of variance procedures. Three such analogies given in Henderson [1953] have received widespread use and attention. The first uses unbalanced data analogies of sums of squares used in analyses of variance of balanced data. For example, for the model (3) with $i = 1,2,\cdots,a$ and $j = 1,2,\cdots,b$ and $k = 1,2,\cdots,n_{ij}$, the interaction sum of squares for balanced data ($n_{ij} = n$ for all i and j) is

$$n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij.}-\bar{x}_{i..}-\bar{x}_{.j.}-\bar{x}_{...})^2 \equiv n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}^2_{ij.} - bn\sum_{i=1}^{a}\bar{x}^2_{i..} - an\sum_{j=1}^{b}\bar{x}^2_{.j.} + abn\bar{x}^2_{...} \; , \tag{22}$$

using familiar bar and subscript dot notation for means. Analogous to the right-hand side of this identity Henderson's [1953] first method suggests using for unbalanced data

$$\sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}\bar{x}^2_{ij.} - \sum_{i=1}^{a}n_{i.}\bar{x}^2_{i..} - \sum_{j=1}^{b}n_{.j}\bar{x}^2_{.j.} + n_{..}\bar{x}^2_{...} \; . \tag{23}$$

Although not a sum of squares (it is not a positive definite quadratic form), this and its counterparts for the error and 2 main effects sums of squares for

(3) do provide 4 elements for $\underline{q}(\underline{y})$ in (19) and so yield estimators from (20).

The second Henderson method, described by Searle [1968] in matrix terminology, is intended for mixed models like (5). Balanced data present no difficulties in estimating variance components in mixed models because the analysis of variance sums of squares for the random effects factors have expected values free of the fixed effects. But with unbalanced data there is need for eliminating these fixed effects. An apparently easy procedure is to first estimate the fixed effects as

$$\tilde{\underline{\beta}} = \underline{L}\underline{y} \tag{24}$$

say, and then estimate the variance components from $\underline{y}$ corrected for $\tilde{\underline{\beta}}$ in the form

$$\underline{z} = \underline{y} - \underline{X}\tilde{\underline{\beta}} \ , \tag{25}$$

for which the model is, from (5),

$$\underline{z} = (\underline{X} - \underline{XLX})\underline{\beta} + (\underline{Z} - \underline{XLZ})\underline{u} + (\underline{I} - \underline{XL})\underline{e} \ . \tag{26}$$

Henderson's Method 2 chooses $\underline{L}$ to reduce this to a random model that is, apart from the error terms, directly suited to his first method. However, the choice of $\underline{L}$ is not unique and it necessitates preclusion of models for $\underline{y}$ that contain interactions between fixed and main effects, Searle [1968]. These are rather severe limitations.

A recent use of (24) and (25) is made by Wallace and Hussain [1969, sec. 5.4] for eliminating covariates from a mixed model. In using $(\underline{X}'\underline{X})^{-1}$ for $\underline{L}$ they certainly eliminate $\underline{\beta}$ from (26), but their then estimating variance components from familiar analysis of variance mean squares of the z's (their data is balanced) predicates the assumption that the model for $\underline{z}$ is $\mu^*\underline{1} + \underline{Zu} + \underline{e}$. This is incorrect, as is evident from (26).

The third Henderson [1953] method uses the reductions in sums of squares calculated when fitting constants. Suppose $R(\underline{\beta}_1, \underline{\beta}_2)$ and $R(\underline{\beta}_1)$ are the reductions in sum of squares for fitting

$$y = \underline{X}_1\underline{\beta}_1 + \underline{X}_2\underline{\beta}_2 + \underline{e} \ , \tag{27}$$

and $\underline{y} = \underline{X}_1\underline{\beta}_1 + \underline{\epsilon}$ respectively. Then the expectation under the model (27) of

$$R(\underline{\beta}_2|\underline{\beta}_1) = R(\underline{\beta}_1,\underline{\beta}_2) - R(\underline{\beta}_1)$$

is

$$E\, R(\underline{\beta}_2|\underline{\beta}_1) = tr\{\underline{X}_2'[\underline{I} - \underline{X}_1(\underline{X}_1'\underline{X}_1)^{-}\underline{X}_1']\underline{X}_2\, E(\underline{\beta}_2\, \underline{\beta}_2')\} + \sigma_e^2[r(\underline{X}_1 \quad \underline{X}_2) - r(\underline{X}_1)] \,. \quad (28)$$

The importance of this result is that $E\, R(\underline{\beta}_2|\underline{\beta}_1)$ contains no terms in $\underline{\beta}_1$. Therefore for the mixed model $\underline{y} = \underline{X}\underline{\beta} + \underline{Z}\underline{u} + \underline{e}$ of (5), reductions in sums of squares of the form $R(\underline{B}_2|\underline{\beta}_1)$ can be derived having expected values free of the fixed effects so long as $\underline{\beta}_1$ always contains $\underline{\beta}$. For example using (9), (10) and (11) with $K = 2$ the model is

$$\underline{y} = \underline{X}\underline{\beta} + \underline{X}_1\underline{u}_1 + \underline{Z}_2\underline{u}_2 + \underline{e} \,, \quad (29)$$

and expected values of $R(\underline{u}_1|\underline{\beta},\underline{u}_2)$ and $R(\underline{u}_2|\beta,\underline{u}_1)$ will by (28) be linear functions of σ_1^2,σ_e^2 and σ_2^2,σ_e^2 respectively, with $E[\underline{y}'\underline{y} - R(\underline{\beta},\underline{u}_1,\underline{u}_2)]$ being a multiple of σ_e^2 in the usual way. Note that $E[R(\underline{u}_1,\underline{u}_2|\underline{\beta})]$ will be a linear function of σ_1^2, σ_2^2 and σ_e^2 so that there are 4 elements of $\underline{q}(\underline{y})$ for (19) and (20) with only 3 variance components to estimate. This is the problem of over-identifiability referred to earlier. It is also discussed in Searle [1971a,b] for the model (3) and in Mount and Searle [1972] for a covariance model.

5c. Maximum likelihood

On the basis of normality assumptions Hartley and Rao [1967] consider maximum likelihood estimation for mixed models. This involves equations that are extremely complicated in the estimators: using $\underline{V}$ of (12) they are

$$\underline{X}'\underline{\tilde{V}}^{-1}\underline{X}\underline{\tilde{\beta}} = \underline{X}'\underline{\tilde{V}}^{-1}\underline{y} \,,$$

$$(\underline{y} - \underline{X}\underline{\tilde{\beta}})'\underline{\tilde{V}}^{-1}(\underline{y} - \underline{X}\underline{\tilde{\beta}}) = N \,,$$

and

$$tr(\underline{\tilde{V}}^{-1}\underline{Z}_\theta\underline{Z}_\theta') = (\underline{y} - \underline{X}\underline{\tilde{\beta}})'\underline{\tilde{V}}^{-1}\underline{Z}_\theta\underline{Z}_\theta'\underline{\tilde{V}}^{-1}(\underline{y} - \underline{X}\underline{\tilde{\beta}}) \,. \quad (30)$$

Numerical solution by the method of steepest ascent is suggested. However, the computing procedures are neither easy, nor yet widely available. Nor are they attractive for the typically large-sized data set of variance components models, because $\underline{V}^{-1}$ has order equal to the number of observations. And with unbalanced data $\underline{V}$ is in no sense a patterned matrix and permits of no easy analytical inverse.

Although explicit maximum likelihood estimators cannot be obtained their large sample variances can. In fact their large sample variance-covariance matrix is

$$[\mathrm{var}(\underline{\tilde{\sigma}}^2)]^{-1} = \left\{\tfrac{1}{2}\ \mathrm{tr}\left(\underline{V}^{-1} \frac{\partial \underline{V}}{\partial\sigma_\theta^2}\, \underline{V}^{-1} \frac{\partial \underline{V}}{\partial\sigma_\varphi^2}\right)\right\} \quad \text{for } \theta,\varphi = 1,2,\cdots,K+1\ . \tag{31}$$

This is derived in Searle [1970] where explicit elements of the matrix on the right-hand-side are obtained for the 2-way nested classification, those for the 3-way nested classification being given in Rudan and Searle [1971]. Unfortunately all attempts at obtaining $\underline{V}^{-1}$ analytically for the 2-way crossed classification have failed, see Rudan and Searle [1971a]. Deriving this inverse for use in (31) remains an unsolved problem.

The computing difficulties associated with numerical inversion of matrices of large order that arise with $\underline{V}^{-1}$ in the maximum likelihood method would also occur in trying to use numerical methods for obtaining sampling variances from (31). Similar difficulties can arise in (28) where $(\underline{X}_1'\,\underline{X}_1)^-$ can be large; however, its order is only the number of effects in $\underline{\beta}_1$ of (27), which is usually far less than the number of observations, the order of $\underline{V}$. Furthermore, in at least one case of widespread application the 2-way crossed classification, explicit computing procedures for (28) are available, e.g., Searle [1971b, pp. 483-4].

5d. Minimization criteria

Several new quadratic forms for use in $\underline{q}(\underline{y})$ of (19) and (20) have recently been suggested, derived by setting up estimation criteria that seem appropriate. Rao [1971a], in suggesting that earlier methods other than maximum likelihood are "ad hoc" has developed, Rao [1970, 1971a], what he calls the MINQUE method, a method of minimum norm quadratic unbiased estimation. This entails establishing $\underline{y}'\underline{Q}\underline{y}$ as an estimator of $\Sigma\, p_i\sigma_i^2$ by choosing $\underline{Q}$ so that, as in (18), $\underline{Q}\underline{X} = \underline{0}$ and, in terms of (13), $\mathrm{tr}(\underline{Q} \sum_{\theta=1}^{K+1} \underline{Z}_\theta \underline{Z}'_\theta)^2$ is a minimum. Writing

$$\underline{W} = \sum_{\theta=1}^{K+1} \underline{W}_\theta \quad \text{for} \quad \underline{W}_\theta = \underline{Z}_\theta \underline{Z}'_\theta$$

this means minimizing $\mathrm{tr}(\underline{QW})^2$. Rao [1971a] gives a variety of theorems useful to this kind of minimization problem, the results for this particular case being that $\hat{\underline{\sigma}}^2$ is derived from

$$\underline{S}\,\hat{\underline{\sigma}}^2 = \underline{q}(\underline{y})$$

where

$$\underline{S} = \{\mathrm{tr}[\underline{RW}_\theta \underline{R}\, \underline{W}_{\theta'}]\} \quad \text{for} \quad \theta, \theta' = 1, 2, \cdots, K+1$$

$$\underline{q}(\underline{y}) = \{\underline{y}'\underline{RW}_\theta \underline{Ry}\} \qquad \text{for} \quad \theta = 1, \cdots, K+1$$

and

$$\underline{R} = \underline{W}^{-1}[\underline{I} - \underline{X}(\underline{X}'\underline{W}^{-1}\underline{X})^{-}\underline{X}'\underline{W}^{-1}] \quad .$$

In Rao [1971b] this development is extended to minimum variance estimation, MIVQUE, and minimum mean square estimation, MIMSQUE.

Whilst these solutions to the problem are to be applauded, they appear to have two strikes against them insofar as practical usage is concerned. First, $\underline{W}^{-1}$ is a matrix of order equal to the number of observations, as is $\underline{V}^{-1}$; and second, to quote Rao [1971b],

> "In all the formulas for estimating $\Sigma\, p_i \sigma_i^2$ the true σ_i^2 appear. In practice we use a priori values or a given set of values at which a minimum is sought."

In addition to these difficulties they also have the deficiencies of other estimators, namely that they are not necessarily non-negative and their distributions are unknown.

Estimators similar to those of Rao have also been suggested by LaMotte [1970] who calls his procedure QUESOM, quadratic unbiased estimation orthogonal to the mean. And Townsend and Searle [1971] have developed explicit expressions for the BQUE's of σ_α^2 and σ_e^2 in the random model $y_{ij} = \alpha_i + e_{ij}$ for unbalanced data, a BQUE being a best (in the minimum variance sense) quadratic unbiased estimator.

6. Some specific problems

The particulars are now given of some specific unfinished problems, both small and large.

6.a. Variances of binomial probabilities

Sometimes the probability parameter p of the binomial distribution can be considered a random variable; e.g., the conception rate for each bull of a population of bulls available for use in artificial breeding; or hatchability rate of a hen's eggs in poultry. The analysis of data on such variables is often an analysis of variance of the appropriate (0,1) variable representing success and failure. This analysis is tantamount to a weighted analysis of variance of the $\hat{p}$'s corresponding to the p's. A simple relationship exists between the variance components of the (0,1) variables and those of the population of p's, even for unbalanced data. However, as indicated in Gates and Searle [1971], unweighted analyses of variance calculations can also yield unbiassed estimators of the variance components of the p's. Although the two methods are the same for balanced data they are not for unbalanced data and in this case investigation is needed into their relative efficiency. Assuming a beta distribution for the p's may also yield distribution properties for the estimators, at least for balanced data.

6.b. Models with covariates

The coefficients of covariates in a covariance model are nothing more than fixed effects and can be handled in accord with (27) and (28); i.e., so long as $\underline{X}_1$ of (27) always includes the covariates, (28) will yield variance components estimators unencumbered by the covariates. For example, consider the model

$$y_{ij} = \mu + \sum_{t=1}^{c} \beta_t x_{tij} + u_{1i} + e_{ij}$$

which can be written as

$$\underset{\sim}{y} = \mu \underset{\sim}{1} + \underset{\sim}{X}_1 \underset{\sim}{\beta}_1 + \underset{\sim}{Z}_1 \underset{\sim}{u}_1 + e \ .$$

Then for n_i observations in the i^{th} level of the random effects $u_{11}, \cdots, u_{1a}$ we have $\underset{\sim}{Z}_1 = \sum_{i=1}^{a^+} \underset{\sim}{1}_{n_i}$ where $\underset{\sim}{1}_{n_i}$ is a vector of n_i unities. The variance component to be estimated, in addition to σ_e^2, is σ_1^2 corresponding to the random effects of $\underset{\sim}{u}_1$. Since the model is that of the 1-way classification with covariance, the estimator of $\hat{\sigma}_e^2$ is

$$\hat{\sigma}_e^2 = \left[\sum_{i=1}^{a} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 - \underset{\sim}{w}' \underset{\sim}{W}_1^{-1} \underset{\sim}{w} \right] / \ (N - a - c)$$

where $\underset{\sim}{w}$ is the vector of within-group sums of products of the covariates with the y's and $\underset{\sim}{W}_1$ is the matrix of within-group sums of squares and products of the covariates. The estimator of σ_1^2 is obtained from using (28), with its $\underset{\sim}{X}_1$ and $\underset{\sim}{\beta}_1'$ now being $[\underset{\sim}{1} \quad \underset{\sim}{X}_1]$ and $[\mu' \quad \underset{\sim}{\beta}_1']$ and its $\underset{\sim}{Z}_2$ and $\underset{\sim}{u}_2$ now being $\underset{\sim}{Z}_1$ and $\underset{\sim}{u}_1$. The result is

$$\hat{\underset{\sim}{\sigma}}_1^2 = \frac{R(\underset{\sim}{u}_1 | \mu, \underset{\sim}{\beta}_1) - (a - 1)\hat{\sigma}_e^2}{N - \sum_{i=1}^{a} n_i^2/N - \mathrm{tr}(\underset{\sim}{T}^{-1} \underset{\sim}{U}_1 \underset{\sim}{U}_1')}$$

where

$$\underset{\sim}{T} = \left\{ \sum_{i=1}^{a} \sum_{j=1}^{n_i} (x_{tij} - \bar{x}_{t..})(x_{t'ij} - \bar{x}_{t'..}) \right\} \quad \text{for } t, t' = 1, \cdots, c$$

and

$$\underset{\sim}{U}_1 \underset{\sim}{U}_1' = \left\{ \sum_{i=1}^{a} n_i^2 (\bar{x}_{ti.} - \bar{x}_{t..})(\bar{x}_{t'i.} - \bar{x}_{t'..}) \right\} \quad \text{for } t, t' = 1, \cdots, c \ .$$

Mount and Searle [1972] derive these results. They also obtain results for the 2-way cross-classification with covariates, one observation per cell and both factors of the classification being random effects factors; i.e., the model (30) with $X = [\underset{\sim}{1} \quad \underset{\sim}{X}_1]$ and $\underset{\sim}{\beta} = [\mu \quad \underset{\sim}{\beta}_1]$ as above. Further extensions of this application of (28) to covariate models are needed.

6c. Components of covariance

The simplest form of dispersion matrix for the random effects of a model is $\sum_{\theta=1}^{K} \sigma_\theta^2 \underline{I}_{N_\theta}$ shown in (11); although simple it involves difficult estimation problems, as we have seen. A more general form, $\{\underline{\Sigma}_{\theta\varphi}\}$ for $\theta,\varphi = 1,\cdots,K$, is discussed following (13). Rao [1971a] calls this a covariance components model. But it is nothing more than a variance components model with covariances among the random effects. A components of covariance model is one for 2 (or more) observable variables having a covariance between them. It is this covariance whose components are of interest; and this is the components of covariance model that biologists have used for many years; e.g., it is the basis of procedures for estimating genetic correlations given in Hazel [1943]. Suppose, for example, we have observations on the staple length and crimp of the fleeces shorn from ewes sired by a variety of rams. If x_{ij} and y_{ij} are the 2 observations from the j^{th} ewe sired by the i^{th} ram, suitable random effects models might be

$$x_{ij} = \mu + \alpha_i + e_{ij}$$

and (32)

$$y_{ij} = \mu' + \alpha_i' + e_{ij}'$$

having variance components σ_α^2, σ_e^2 and $\sigma_{\alpha'}^2$, $\sigma_{e'}^2$, respectively. The components of covariance are the covariances $\sigma_{\alpha\alpha'}$ and $\sigma_{ee'}$ between α_i and α_i' and between e_{ij} and e_{ij}', with

$$\sigma_{xy} = \sigma_{\alpha\alpha'} + \sigma_{ee'} \; . \tag{33}$$

Estimation of components of covariance of the nature described for (33) is no more difficult than is estimation of components of variance. On all occasions, components of covariance estimators will be the same linear combinations of the same bilinear forms in $\underline{x}$ and $\underline{y}$ as variance components estimators are of quadratic forms in $\underline{x}$ and in $\underline{y}$. However, the need for bilinear forms can be by-passed by using quadratic forms in $\underline{x} + \underline{y}$ and relying on the identities

$$\sigma_{xy} \equiv \tfrac{1}{2}(\sigma_{x+y}^2 - \sigma_x^2 - \sigma_y^2)$$

and (34)

$$\underline{x}'\underline{Q}\underline{y} \equiv \tfrac{1}{2}[(\underline{x} + \underline{y})'\underline{Q}(\underline{x} + \underline{y}) - \underline{x}'\underline{Q}\underline{x} - \underline{y}'\underline{Q}\underline{y}] \; .$$

Thus for any data set in which the vectors of components of variance of x and y, $\underline{\sigma}_x^2$ and $\underline{\sigma}_y^2$ respectively, are estimated in accord with (20) by

$$\hat{\underline{\sigma}}_x^2 = \underline{P}^{-1}\underline{q}(\underline{x}) \quad \text{and} \quad \hat{\underline{\sigma}}_y^2 = \underline{P}^{-1}\underline{q}(\underline{y}) ,$$

the vector of components of covariance will be estimated by

$$\begin{aligned} \hat{\sigma}_{xy} &= \tfrac{1}{2}\underline{P}^{-1}[\underline{q}(\underline{x} + \underline{y}) - \underline{q}(\underline{x}) - \underline{q}(\underline{y})] \\ &= \tfrac{1}{2}(\hat{\underline{\sigma}}_{x+y}^2 - \hat{\underline{\sigma}}_x^2 - \hat{\underline{\sigma}}_y^2) . \end{aligned} \tag{35}$$

Hence components of covariance can be estimated directly from the estimated components of variance of the two variables concerned and of their sum.

Investigation of properties of estimated components of covariance is also needed. Their variances, for example, on the basis of normality can be derived using

$$\text{cov}(\underline{x}'\underline{Q}\underline{x}, \underline{x}'\underline{Q}\underline{y}) = 2\ \text{tr}(\underline{Q}\underline{C})^2 + 4\underline{\mu}_x'\underline{Q}\underline{C}\underline{Q}\underline{\mu}_y \tag{36}$$

where $\underline{C}$ is the matrix of covariances between $\underline{x}$ and $\underline{y}$ and $\underline{\mu}_x = E(\underline{x})$ and $\underline{\mu}_y = E(\underline{y})$. Similarly

$$\text{cov}(\underline{x}'\underline{Q}\underline{x}, \underline{x}'\underline{Q}\underline{y}) = \text{tr}(\underline{Q}\underline{V}_x\underline{Q}\underline{C}) + \text{tr}(\underline{Q}\underline{V}_x)^2 + 2\underline{\mu}_x'\underline{Q}\underline{C}\underline{Q}\underline{\mu}_x + 2\underline{\mu}_x'\underline{Q}\underline{V}_x\underline{Q}\underline{\mu}_y , \tag{37}$$

where $\underline{V}_x$ and $\underline{V}_y$ are the variance - covariance matrices of $\underline{x}$ and $\underline{y}$ respectively; and

$$\text{var}(\underline{x}'\underline{Q}\underline{y}) = \text{tr}(\underline{Q}\underline{C})^2 + \text{tr}(\underline{Q}\underline{V}_y\underline{Q}\underline{V}_x) + 2\underline{\mu}_x'\underline{Q}\underline{C}\underline{Q}\underline{\mu}_y + \underline{\mu}_x'\underline{Q}\underline{V}_y\underline{Q}\underline{\mu}_x + \underline{\mu}_y'\underline{Q}\underline{V}_x\underline{Q}\underline{\mu}_y . \tag{38}$$

These expressions come from the general form of the covariance between any two bilinear forms in normal variables, e.g., Searle [1971b, p. 66].

6.d. Criteria for estimation

The various estimation procedures originating from analysis of variance calculations undoubtedly arose as a matter of convenience and because they seemed "obvious". The only known property of the resulting estimators is unbiasedness (and the χ^2-nature of $\hat{\sigma}_e^2$ under normality). This property is retained in the MINQUE, MIVQUE and BQUE procedures, and others are added.

Although these procedures stem from desirable criteria they do not overcome the problem of yielding negative estimators which are such an embarrassment. Furthermore, the property of unbiasedness itself merits questioning in the case of variance components estimators. This is so because with unbalanced data from random models the concept of repetitions of similarly structured data and associated repetitions of estimators is often not appropriate – more data, maybe, but not necessarily with the same pattern of unbalancedness. Replications of data cannot therefore be thought of as mere resamplings of the data already available. This situation appears to demand that consideration of expected values over repeated samplings should take into account the varying numbers of observations that arise from sample to sample. Also to be taken into account is the fact that random model data are often available in such large quantities that additional data may involve other populations. Investigation of these and other ideas is needed to develop estimators that have properties more in keeping with them than do those currently available. Some form of modal unbiasedness is one possibility that has been suggested, Searle [1968]; or estimators for which the probability of small deviations from true value is maximized rather than minimizing the probability of large deviations. Even though, as Eisenhart [1968] points out, this was the idea that Gauss rejected in favour of his minimum-mean-squared-error approach, it may not be inappropriate for variance components situations where data sets are often large and non-replicable in the usual sense.

A standard procedure for comparing different estimators is by means of their sampling variances. With this in view, expressions for the sampling variances of variance components estimators under normality assumptions have been derived in several palces; e.g., Searle [1956, 1958, 1961, 1970], Rohde and Tallis [1969], Searle and Rudan [1971b]. Further work is needed to derive from Rohde and Tallis explicit expressions for particular cases; and a great deal of work is needed in comparative studies, using these results.

6.e. Computing difficulties

Reference has already been made to some of the computing difficulties involved in calculating estimates from some of the estimators discussed. These largely revolve around the difficulty of inverting matrices of very large order, such as calculating $\underline{V}^{-1}$. In addition to these computing difficulties the resulting estimates are such that their distributional properties are mostly unknown, or known only in terms of the unknown variance components parameters. Furthermore these properties themselves can involve computing headaches. Two questions therefore arise: (i) Can we develop variance components estimators based on much simpler calculations than are needed now — such as using rank order statistics, perhaps? (ii) Can we numerically investigate the properties of present estimators in a manner which will yield definitive information about their behaviour for variations in the unknown variance components and variations in the numbers of observations in the subclasses? Investigation of these two questions surely seems worthwhile.

6.f. Defining unbalancedness

Referring to the values that the numbers of observations take on in a data set as an n-pattern, one of the preceding questions is to what extent does the n-pattern affect the properties of an estimator? Unfortunately the algebra of most properties usually involves the n-pattern in such a complicated way that the effect of different n-patterns cannot be studied analytically. For example, assuming normality in the model (2), the sampling variance of the analysis of variance estimator of σ_α^2 is

$$v(\hat{\sigma}_\alpha^2) = \frac{2\sigma_e^4 N^2(N-1)(a-1)}{(N^2-S_2)^2(N-a)} + \frac{4\sigma_e^2\sigma_\alpha^2 N}{N^2-S_2} + \frac{2\sigma_\alpha^4(N^2S_2+S_2^2-2NS_3)}{(N^2-S_2)^2} \tag{39}$$

where

$$N = \sum_{i=1}^{a} n_i\ , \quad S_2 = \sum_{i=1}^{a} n_i^2 \quad \text{and} \quad S_3 = \sum_{i=1}^{a} n_i^3\ .$$

The involvement of the n_i's in this expression is clearly such that investigating its behaviour for changes in the n_i's is out of the question. It seems, therefore, that analytical comparisons of estimators are likely to be quite intractable and recourse must be made to numerical studies.

Even if the behaviour of expressions like (39) in terms of the n_i's could be delineated it would be advantageous to be able to summarize an n-pattern in terms of some characteristic, such as a measure of unbalancedness. The behaviour of (39) could then be described in terms of this measure. The possibility of doing this may be remote, however, because preliminary indications are that even in the simplest of cases the effect of the n-pattern on properties of estimators is apparently itself a function of the variance components being estimated. The effects of unbalancedness therefore appear to differ according to the values of the true variance components. This suggests that unbalancedness may have to be defined in terms of the components being estimated, an unsatisfactory conclusion from the point of view of considering the effect of unbalancedness on estimation.

Numeric studies for comparing estimators also involve a problem concerning n-patterns. It is that of planning a set of n-patterns to be used, in conjunction with sets of parameter values. Deciding on the latter usually poses no great difficulty because only a small number of variance components are involved. But deciding on a set of n-patterns is a situation of having infinitely many choices; for example, in a 2-way crossed classification of rows and columns, the planning of a set of n-patterns demands answering such questions as how many rows, how many columns, how many empty cells, and how many observations in the cells that are not empty? The sky's the limit, a fact which makes it exceedingly difficult to plan a set of n-patterns that are sufficiently disparate to encompass an interesting range but which also differ in some logical manner in such a way that effects on the properties of the estimators might be apparent. One possibility is to draw samples of n_{ij}'s from some distribution, provided a useful, realistic and tractable distribution can be postulated. Even then, this course of action would provide little information on just exactly how it is that the characteristic of unbalancedness affects estimation procedures. Investigation on this problem is therefore badly needed.

References

1. Ahrens, H. [1965]. Standardfehler geschätzter Varianzkomponenten eines unbalanzierten Versuchplanes in r-stufiger hierarchischer Klassifikation. Monatsb. Deutsch. Akad. Wiss., Berlin, 7:89-94.

2. Dempster, A. P. [1971]. An overview of multivariate data analysis. J. Multivariate Analysis, 1:316-346.

3. Eisenhart, C. [1968]. Discussion of Searle [1968]. Biometrics, 24:784.

4. Gates, C. E. and Searle, S. R. [1971]. Estimating variance components of binomial frequencies. Paper BU-368-M in the Biometrics Unit, Cornell University, Ithaca, N. Y.

5. Graybill, F. A. and Hultquist, R. A. [1961]. Theorems concerning Eisenhart's Model II. Ann. Math. Statist., 32:261-269.

6. Hazel, L. N. [1943]. The genetic basis for constructing selection indexes. Genetics, 28:476-490.

7. Hartley, H. O. [1967]. Expectation, variances and covariances of ANOVA mean squares by 'synthesis'. Biometrics, 23:105-114. Correction 23:853.

8. Hartley, H. O. and Rao, J. N. K. [1967]. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika, 54:93-108.

9. Henderson, C. R. [1953]. Estimation of variance and covariance components. Biometrics, 9:226-252.

10. Herbach, L. H. [1959]. Properties of Model II type analysis of variance tests. Ann. Math. Statist., 30:939-959.

11. LaMotte, L. R. [1970]. A class of estimators of variance components. Technical Report No. 10, Dept. of Statistics, University of Kentucky, Lexington.

12. Mount, T. D. and Searle, S. R. [1972]. Estimating variance components in covariance models. Paper BU-403-M in the Biometrics Unit, Cornell University, Ithaca, N. Y. Submitted to Econometrica.

13. Rao, C. R. [1970]. Estimation of heteroscedastic variances in linear models. J. Amer. Statist. Assoc., 65:161-172.

14. Rao, C. R. [1971a]. Estimation of variance and covariance components — MINQUE theory. J. Multivariate Analysis, 1:257-275.

15. Rao, C. R. [1971b]. Minimum variance quadratic unbiased estimation of variance components. J. Multivariate Analysis, 1:445-456.

16. Robinson, J. [1965]. The distribution of a general quadratic form in normal variables. Aust. J. Statist., 7:110-114.

17. Rohde, C. A. and Tallis, G. M. [1969]. Exact first- and second-order moments of estimates of components of covariance. Biometrika, 56:517-526.

18. Rudan, J. W. and Searle, S. R. [1971a]. Attempts at inverting the variance-covariance matrix of the 2-way crossed classification, unbalanced data, random model. Paper BU-353-M in the Biometrics Unit, Cornell University, Ithaca, N. Y.

19. Rudan, J. W. and Searle, S. R. [1971b]. Large sample variances of maximum likelihood estimators of variance components in the 3-way nested classification, random model, with unbalanced data. Biometrics, 27:1087-1091.

20. Searle, S. R. [1956]. Matrix methods in variance and covariance components analysis. Ann. Math. Statist., 27:737-748.

21. Searle, S. R. [1958]. Sampling variances of estimates of components of variance. Ann. Math. Statist., 29:167-178.

22. Searle, S. R. [1961]. Variance components in the unbalanced 2-way nested classification. Ann. Math. Statist., 32:1161-1166.

23. Searle, S. R. [1968]. Another look at Henderson's methods of estimating variance components. Biometrics, 24:749-788.

24. Searle, S. R. [1970]. Large sample variances of maximum likelihood estimators of variance components. Biometrics, 26:505-524.

25. Searle, S. R. [1971a]. Topics in variance components estimation. Biometrics, 27:1-76.

26. Searle, S. R. [1971b]. Linear Models. Wiley, New York.

27. Thompson, W. A., Jr. [1962]. The problem of negative estimates of variance components. Ann. Math. Statist., 33:273-289.

28. Thompson, W. A., Jr. [1963]. Precision of simultaneous measurement procedures. J. Am. Statist. Assoc., 58:474-479.

29. Thompson, W. A., Jr. and Moore, J. R. [1963]. Nonnegative estimates of variance components. Technometrics, 5:441-450.

30. Townsend, E. C. and Searle, S. R. [1972]. Best quadratic unbiased estimation of variance components from unbalanced data in the 1-way classification. Biometrics, 27:643-657.

31. Wallace, T. D. and Hussain, A. [1969]. The use of error components models in combining time series with cross section data. Econometrica, 37:55-72.

32. Wang, Y. Y. [1967]. A comparison of several variance component estimators. Biometrika, 54:301-305.

33. Winsor, C. P. and Clarke, G. L. [1940]. A statistical study of variation in the catch of plankton nets. Sears Foundation J. Marine Res., 3:1.

34. Yates, F. [1967]. A fresh look at the basic principles of the design and analysis of experiments. Proc. 5th Berkeley Symp. Math. Statist. Prob. IV:777-790. L. Lecam and J. Neyman, Eds., Univ. of Calif. Press.