

BEST ESTIMATION OF VARIANCE COMPONENTS FROM BALANCED DATA,
WITH ARBITRARY KURTOSIS

BU-691-M

by

March, 1982

R. D. Anderson^{1/}, H. V. Henderson^{2/}, F. Pukelsheim^{3/} and S. R. Searle^{4/}

Abstract

Properties are well known for analysis of variance estimators of variance components obtained from balanced data under assumptions either of normality or of zero kurtosis. We show here that even with non-zero kurtosis, these estimators still have uniformly minimum variance among all unbiased, translation invariant, quadratic estimators.

Examples of balanced data models with succinct matrix representations are given. An algorithm is presented for deriving from \underline{XX}' the matrix $\underline{M} = \underline{I} - \underline{XX}^+$, where \underline{X} is the incidence matrix for the fixed effects and \underline{X}^+ denotes its Moore-Penrose inverse. The algorithm involves only the Kronecker product operation and requires no explicit calculation of generalized inverses.

1. Introduction

Variance components estimators obtained from balanced data (having equal numbers of observations in the subclasses) by equating analysis of variance mean squares to their expected values are known as ANOVA (analysis of variance) estimators. Seely [1971] proposes a comprehensive theory for optimality properties

-
- ^{1/} Sheep Husbandry Department, Massey University, Palmerston North, New Zealand.
^{2/} Biometrics Section, Ruakura Agricultural Research Centre, Hamilton, New Zealand.
^{3/} Institut für Mathematische Stochastik der Universität, Freiburg im Breisgau, Federal Republic of Germany.
^{4/} Biometrics Unit, Cornell University, Ithaca, New York.

of ANOVA estimators, introducing the notion of a quadratic subspace of symmetric matrices. For ANOVA models with balanced data his results show that the ANOVA estimators are the same as MINQUE [Rao, 1971] and are UMVUIQ (uniformly minimum variance unbiased location-invariant quadratic) when fourth moments are the same as under normality, i.e.,

$$\text{ANOVA} = \text{UMVUIQ}, \text{ under zero kurtosis .} \quad (1.1)$$

Our main result is that for ANOVA models with balanced data the uniformly minimum variance property extends beyond normality and zero kurtosis to distributions with non-zero kurtosis:

$$\text{ANOVA} = \text{UMVUIQ}, \text{ under arbitrary kurtosis .} \quad (1.2)$$

This extends partial results earlier obtained by Hsu [1938, p. 100] and Atiquallah [1962, p. 85] for the residual error variance in fixed effects models, by Graybill [1954], Graybill and Wortham [1956], and Graybill and Hultquist [1961, Theorem 7] for random effects models satisfying certain assumptions, and by Tan [1979] for nested classifications with random models.

A special case of the zero kurtosis of (1.1) is the assumption of normality. Under this assumption, restriction to quadratic estimators is irrelevant:

$$\text{ANOVA} = \text{UMVUI}, \text{ under normality .} \quad (1.3)$$

In this case, restricted maximum likelihood (REML) estimators (e.g., see Patterson and Thompson [1971], Corbeil and Searle [1976] and Searle [1979a]) also coincide with UNVUI estimators (see Anderson [1978] and Pukelsheim and Styan [1979]), so that

$$\text{ANOVA} = \text{REML}, \text{ under normality ,} \quad (1.4)$$

as noted by Harville [1977, p. 325].

Section 2 starts with several examples of ANOVA models with balanced data and discusses succinct matrix representation of them. For calculating $\underline{\underline{M}} = \underline{\underline{I}} - \underline{\underline{X}}(\underline{\underline{X}}'\underline{\underline{X}})^{-}\underline{\underline{X}}'$, where $\underline{\underline{X}}$ is the incidence matrix for the fixed effects, this representation provides (in Section 3) an algorithm that does not explicitly need any generalized inverse $(\underline{\underline{X}}'\underline{\underline{X}})^{-}$ of $\underline{\underline{X}}'\underline{\underline{X}}$; this leads to a detailed knowledge of the structure of $\underline{\underline{M}}$ necessary to establish (1.2). Section 3 also discusses translation invariance, and in Section 4 we review the normal and zero-kurtosis case. And the condition of Hsu's and Atiqullah's results, that a certain matrix have equal diagonal elements, is reflected by a scalar invariance in our proof of (1.2) in Section 5.

Not only is our result more general than results presently available but our method of proof differs from that of earlier proofs, which explicitly involve calculating the variance that is to be minimized. Rather, we embed the problem in a general linear model for variance components and then view it as a special case of the question "When is least-squares estimation also minimum variance?". This approach is essentially due to Seely [1970] and has also been applied to similar problems by Drygas [1980], Kleffe [1977] and Brown [1978].

2. Notation and Examples

2.1. The general linear model

A representation of the general linear model that includes fixed and random effects is

$$\underline{\underline{y}} = \underline{\underline{X}}\underline{\underline{\beta}} + \underline{\underline{Z}}\underline{\underline{u}} = \underline{\underline{X}}\underline{\underline{\beta}} + \sum_{i=1}^c \underline{\underline{Z}}_i \underline{\underline{u}}_i \quad (2.1)$$

where $\underline{\underline{y}}$ is an $N \times 1$ vector of observations, $\underline{\underline{\beta}}$ is a vector of p parameters, $\underline{\underline{X}}$ is a known $N \times p$ incidence matrix, $\underline{\underline{u}}_i$ is a vector of q_i random effects and $\underline{\underline{Z}}_i$ is a known $N \times q_i$ incidence matrix, with $\underline{\underline{u}}_c = \underline{\underline{\epsilon}}$ being the vector of residual error

terms and $\underline{Z}_c = \underline{I}_N$. Partitioned as rows of sub-matrices and sub-vectors, respectively, $\underline{Z} = \{\underline{Z}_i\}$ and $\underline{u}' = \{\underline{u}'_i\}$, the means, variances and covariances of the \underline{u}_i -vectors are defined by $E(\underline{u}_i) = \underline{0}$, $\text{var}(\underline{u}_i) = E(\underline{u}_i \underline{u}'_i) = \sigma_{ii}^2 \underline{I}_{q_i}$ and $\text{cov}(\underline{u}_i, \underline{u}'_j) = E(\underline{u}_i \underline{u}'_j) = \underline{0}$ for $i \neq j = 1, \dots, c$. With these definitions, the dispersion (variance-covariance) matrix of \underline{y} is

$$\underline{V} = \sum_{i=1}^c \sigma_{ii}^2 \underline{V}_i \quad \text{with} \quad \underline{V}_i = \underline{Z}_i \underline{Z}'_i. \quad (2.2)$$

Each vector of random effects, \underline{u}_i , has elements u_{i1}, \dots, u_{iq_i} which are assumed to be independent, all with the same coefficient of kurtosis γ_i :

$$E(u_{ij}^4) = (\gamma_i + 3)\sigma_{ii}^4, \quad \text{for } j = 1, \dots, q_i. \quad (2.3)$$

In saying "zero kurtosis" in (1.1), we mean $\gamma_1 = \dots = \gamma_c = 0$. This occurs, for example, when each \underline{u}_i follows a normal distribution. Non-zero kurtosis in this context was first considered by Hsu [1938], so that (2.1) and (2.3) are called Hsu's model by Pukelsheim [1977].

2.2. ANOVA models with balanced data

Searle and Henderson [1979] develop properties of \underline{V} , for balanced data, from noting that each $\underline{Z}_i \underline{Z}'_i$ of (2.2) can be expressed as a Kronecker product of \underline{I} - and \underline{J} -matrices where, in general, \underline{I}_a is the $a \times a$ identity matrix, and \underline{J}_a is the $a \times a$ matrix with every element unity. Similarly, \underline{X} and each \underline{Z}_i of (2.1) can be expressed in terms of Kronecker products of \underline{I} -matrices and $\underline{1}$ -vectors where $\underline{1}_a$ is the $a \times 1$ vector with every element unity. Then $\underline{J}_a = \underline{1}_a \underline{1}'_a$. We also define the symmetric idempotent (projection) matrices $\bar{\underline{J}}_a = \underline{J}_a/a$ and $\underline{K}_a = \underline{I}_a - \bar{\underline{J}}_a$. Products of \underline{J}_a , $\bar{\underline{J}}_a$ and \underline{K}_a commute, and $\underline{K}_a \bar{\underline{J}}_a = \underline{K}_a \underline{J}_a = \underline{0}$. For \otimes being the Kronecker product operator, the equality $\underline{T}_a \otimes \underline{T}_n = \underline{T}_{an}$ holds for \underline{T} representing \underline{I} , $\underline{1}$, \underline{J} or $\bar{\underline{J}}$ but not \underline{K} :

$$\underline{K}_{an} = \underline{K}_a \otimes \underline{K}_n + \underline{K}_a \otimes \bar{\underline{J}}_n + \bar{\underline{J}}_a \otimes \underline{K}_n \neq \underline{K}_a \otimes \underline{K}_n. \quad (2.4)$$

The reader will recognize that $\underline{\underline{K}}_a$ is a special case of the projection matrix associated with the general incidence matrix $\underline{\underline{X}}$ of the model (2.1):

$$\underline{\underline{M}} = \underline{\underline{I}} - \underline{\underline{X}}(\underline{\underline{X}}'\underline{\underline{X}})^-\underline{\underline{X}}' = \underline{\underline{I}} - \underline{\underline{X}}\underline{\underline{X}}^+ \quad (2.5)$$

where $(\underline{\underline{X}}'\underline{\underline{X}})^-$ is a generalized inverse of $\underline{\underline{X}}'\underline{\underline{X}}$, and $\underline{\underline{X}}^+$ is the Moore-Penrose inverse of $\underline{\underline{X}}$ (e.g., Searle [1971, p. 20]). $\underline{\underline{M}}$ is symmetric, idempotent and, for given $\underline{\underline{X}}$, invariant to the choice of $(\underline{\underline{X}}'\underline{\underline{X}})^-$, and $\underline{\underline{M}}\underline{\underline{X}} = \underline{\underline{0}}$.

2.3. Examples: three-way classifications

Numerous examples could be considered; we offer three that illustrate a variety of features. In each of them, μ is a general mean and ϵ -terms are random errors with zero mean and dispersion matrix $\sigma_\epsilon^2 \underline{\underline{I}}_N$. Other terms are defined as random or fixed. Variances are denoted by σ^2 with subscript corresponding to the term concerned. The subscript ranges are $i = 1, \dots, a$, $j = 1, \dots, b$, $k = 1, \dots, c$ and $\ell = 1, \dots, n$.

(1) A mixed effects model

Consider a 3-way classification with two of the three possible first-order interactions and without the second-order interaction:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\beta\gamma)_{jk} + \epsilon_{ijkl}.$$

With α 's, β 's, and $(\alpha\beta)$'s fixed, but γ 's and $(\beta\gamma)$'s random:

$$\underline{\underline{X}} = [\underline{\underline{1}}_a \otimes \underline{\underline{1}}_b \otimes \underline{\underline{1}}_c \otimes \underline{\underline{1}}_n \mid \underline{\underline{I}}_a \otimes \underline{\underline{1}}_b \otimes \underline{\underline{1}}_c \otimes \underline{\underline{1}}_n \mid \underline{\underline{1}}_a \otimes \underline{\underline{I}}_b \otimes \underline{\underline{1}}_c \otimes \underline{\underline{1}}_n \mid \underline{\underline{I}}_a \otimes \underline{\underline{I}}_b \otimes \underline{\underline{1}}_c \otimes \underline{\underline{1}}_n]$$

$$\underline{\underline{Z}}_1 = \underline{\underline{1}}_a \otimes \underline{\underline{1}}_b \otimes \underline{\underline{I}}_c \otimes \underline{\underline{1}}_n, \quad \underline{\underline{Z}}_2 = \underline{\underline{1}}_a \otimes \underline{\underline{I}}_b \otimes \underline{\underline{I}}_c \otimes \underline{\underline{1}}_n, \quad \underline{\underline{Z}}_3 = \underline{\underline{I}}_{abcn}$$

$$\underline{\underline{V}} = \sigma_\gamma^2 (\underline{\underline{J}}_a \otimes \underline{\underline{J}}_b \otimes \underline{\underline{I}}_c \otimes \underline{\underline{J}}_n) + \sigma_{\alpha\gamma}^2 (\underline{\underline{J}}_a \otimes \underline{\underline{I}}_b \otimes \underline{\underline{I}}_c \otimes \underline{\underline{J}}_n) + \sigma_\epsilon^2 (\underline{\underline{I}}_a \otimes \underline{\underline{I}}_b \otimes \underline{\underline{I}}_c \otimes \underline{\underline{I}}_n).$$

(2) A fixed effects model with a nested factor

Consider the case of having γ_{jk} 's nested within the β_j 's:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{jk} + (\alpha\gamma)_{ijk} + \epsilon_{ijkl} .$$

The \underline{X} matrix is

$$\underline{X} = [\underline{1}_{\underline{a}} \otimes \underline{1}_{\underline{b}} \otimes \underline{1}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{I}_{\underline{a}} \otimes \underline{1}_{\underline{b}} \otimes \underline{1}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{1}_{\underline{a}} \otimes \underline{I}_{\underline{b}} \otimes \underline{1}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{I}_{\underline{a}} \otimes \underline{I}_{\underline{b}} \otimes \underline{1}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{1}_{\underline{a}} \otimes \underline{1}_{\underline{b}} \otimes \underline{I}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{1}_{\underline{a}} \otimes \underline{1}_{\underline{b}} \otimes \underline{I}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{I}_{\underline{a}} \otimes \underline{1}_{\underline{b}} \otimes \underline{I}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{I}_{\underline{a}} \otimes \underline{I}_{\underline{b}} \otimes \underline{I}_{\underline{c}} \otimes \underline{1}_{\underline{n}}] .$$

(3) A fixed effects model with a nested factor and a missing interaction

Now consider having δ_{ijk} 's nested within (α_i, β_j) cells but with the $(\alpha\beta)_{ij}$ interactions absent:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \delta_{ijk} + \epsilon_{ijkl} .$$

The matrix \underline{X} is

$$\underline{X} = [\underline{1}_{\underline{a}} \otimes \underline{1}_{\underline{b}} \otimes \underline{1}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{I}_{\underline{a}} \otimes \underline{1}_{\underline{b}} \otimes \underline{1}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{1}_{\underline{a}} \otimes \underline{I}_{\underline{b}} \otimes \underline{1}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{I}_{\underline{a}} \otimes \underline{I}_{\underline{b}} \otimes \underline{1}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{1}_{\underline{a}} \otimes \underline{1}_{\underline{b}} \otimes \underline{I}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{1}_{\underline{a}} \otimes \underline{1}_{\underline{b}} \otimes \underline{I}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{I}_{\underline{a}} \otimes \underline{1}_{\underline{b}} \otimes \underline{I}_{\underline{c}} \otimes \underline{1}_{\underline{n}} ; \underline{I}_{\underline{a}} \otimes \underline{I}_{\underline{b}} \otimes \underline{I}_{\underline{c}} \otimes \underline{1}_{\underline{n}}] .$$

2.4. Comments on a general m-way classification model with balanced data

Suppose we write a general m-way classification model as $\underline{y} = \sum_{j=1}^m \underline{G}_j \underline{\theta}_j + \underline{\epsilon}$ where $\underline{\theta}_j$ represents in turn $\mu, \alpha, \beta, \gamma, \dots, (\alpha\beta), (\alpha\gamma), \dots$, appropriate to any particular model. Then evidently \underline{G}_j is a Kronecker product of $m + 1$ matrices each of which is either an \underline{I} -matrix or a $\underline{1}$ -vector. To be more specific, suppose N_1, \dots, N_m are the numbers of levels of the m (main or nested) effects $\alpha, \beta, \gamma, \dots$, and n is the number of observations in each of the smallest subclasses. Then \underline{G}_j can be written, similar to Seifert [1979, p. 238], as

$$\underline{G}_j = \underline{F}_{N_1} \otimes \dots \otimes \underline{F}_{N_m} \otimes \underline{1}_n , \quad (2.6)$$

where, for $r = 1, \dots, m$, the matrix \underline{F}_{N_r} equals \underline{I}_{N_r} if the r 'th effect is

represented in θ_j and equals 1_{N_r} otherwise. Thus, in example (1), the $(\beta\gamma)$'s are represented by θ_6 for which $G_6 = 1_a \otimes I_b \otimes I_c \otimes 1_n$. In all models the Kronecker product of all 1 's (which is a 1 -vector) appears as G_1 for μ ; and the Kronecker product of all I 's (itself an I) corresponds to the error term ϵ .

The distinction between interaction effects and nested effects in this description is that a term which, from the nature of its G could be either, is an interaction if and only if all the corresponding main effects are present (whether they are effects due to crossed classifications, as is usual, or to nested classifications); otherwise it is a nested effect. This is evident, for instance, for the G -matrix $1_a \otimes I_b \otimes I_c \otimes 1_n$ which in example (1) is associated with the $(\beta\gamma)$ interaction terms. In contrast, if the γ_k 's were omitted from example (1) the effect in question would be nested within the β_j 's, as in example 2 of Seifert [1979].

The classifying of an effect as fixed or random determines whether its G_j becomes, in terms of the general linear model $y = X\beta + \sum_i Z_i u_i$ in (2.1), a submatrix of X or is a Z_i . For a θ_j representing fixed effects, G_j becomes a submatrix of X , whereas for a θ_j representing random effects its G_j becomes a Z_i . In any case, the product $G_j G_j'$ appears as a term either in XX' or in $V = \sum_i \sigma_i^2 Z_i Z_i'$, and from (2.6) is a Kronecker product of I 's and J 's. Indeed, both XX' and $V = \sum_i \sigma_i^2 Z_i Z_i'$ are sums of $G_j G_j'$ matrices, as dealt with by Searle and Henderson [1979].

3. Translation Invariance

Estimation of variance components is usually confined to estimators that are based on quadratic functions of the observation vector y of the form $y' A y$ with A being symmetric. Since the ANOVA table is set up after an initial least squares fit for β , ANOVA estimators of variance components have the property of depending on the observations y only through the residual statistic $My = y - X(X'X)^{-1}X'y$.

Seely [1971, p. 717] points out that $\underline{\underline{M}}_y$ is a maximal translation-invariant statistic, whence every estimator for variance components which depends on $\underline{\underline{y}}$ only through $\underline{\underline{M}}_y$ is called a (translation) invariant estimator.

The invariance requirement plays an ambiguous role in the theory. For a variance component estimator $\underline{\underline{y}}'\underline{\underline{A}}\underline{\underline{y}}$ there are at least three situations when invariance arises automatically: (i) When $\underline{\underline{A}}$ is non-negative definite, then unbiasedness necessitates invariance (see Atiqullah [1962, Lemma 2]). (ii) When the variance of $\underline{\underline{y}}'\underline{\underline{A}}\underline{\underline{y}}$ is assumed not to depend on $\underline{\underline{\beta}}$, then this necessitates invariance (see Hsu [1938, page 95] and Drygas [1972]). (iii) When Seely's [1971, p. 715] set of assumptions for the existence of a uniformly minimum variance unbiased estimator are satisfied, then that estimator is invariant. There are, however, instances when invariance does not arise automatically. In these cases the invariance requirement means a genuine restriction from the larger class of all unbiased quadratic estimators to the proper subclass of those estimators which, in addition, are invariant.

A quadratic estimator $\underline{\underline{y}}'\underline{\underline{A}}\underline{\underline{y}}$ is invariant, if and only if it remains unaltered when $\underline{\underline{y}}$ is replaced by $\underline{\underline{y}} - \underline{\underline{X}}\underline{\underline{\delta}}$ for every $\underline{\underline{\delta}}$, and for this, a necessary and sufficient condition is $\underline{\underline{A}}\underline{\underline{X}} = \underline{\underline{0}}$ (see Rao [1971, p. 267]). Since it is easily verified that $\underline{\underline{A}}\underline{\underline{X}} = \underline{\underline{0}}$ if and only if $\underline{\underline{A}} = \underline{\underline{M}}\underline{\underline{M}}'$, it follows that $\underline{\underline{y}}'\underline{\underline{A}}\underline{\underline{y}}$ is invariant if and only if it has the form $\underline{\underline{y}}'\underline{\underline{M}}\underline{\underline{M}}'\underline{\underline{y}}$. A general form of $\underline{\underline{M}}$ in ANOVA models with balanced data can be derived by first writing $\underline{\underline{I}}_{\underline{\underline{N}}}$ as the Kronecker product $\underline{\underline{I}}_{\underline{\underline{N}}_1} \otimes \cdots \otimes \underline{\underline{I}}_{\underline{\underline{N}}_m} \otimes \underline{\underline{I}}_{\underline{\underline{N}}}$ and replacing each $\underline{\underline{I}}_{\underline{\underline{N}}_r}$ by $\underline{\underline{J}}_{\underline{\underline{N}}_r} + \underline{\underline{K}}_{\underline{\underline{N}}_r}$. This gives $\underline{\underline{I}}_{\underline{\underline{N}}}$ as a sum of 2^{m+1} terms. Each term is a Kronecker product of $\underline{\underline{J}}$'s and $\underline{\underline{K}}$'s and hence a projection matrix. The product of any two of these terms is null because it involves the product of a $\underline{\underline{J}}$ and a $\underline{\underline{K}}$ which is null, and so the representation thus obtained corresponds to the orthogonal partitioning traditional to the ANOVA of balanced data. As a consequence, $\underline{\underline{M}}$ is a sum of Kronecker products of $\underline{\underline{I}}$'s, $\underline{\underline{J}}$'s and $\underline{\underline{K}}$'s. This is the property of $\underline{\underline{M}}$ required in Section 5 for our proof concerning non-zero kurtosis.

For certain ANOVA models with balanced data we suggest the following algorithm for deriving $\underline{\underline{M}}$ of (2.5) without having to calculate a generalized inverse $(\underline{\underline{X}}'\underline{\underline{X}})^-$ or the Moore-Penrose inverse $\underline{\underline{X}}^+$. First recall that $\underline{\underline{X}}'\underline{\underline{X}}$ is a sum of terms $\underline{\underline{G}}_j\underline{\underline{G}}_j'$ that, from (2.6) are each Kronecker products of $\underline{\underline{I}}$'s and $\underline{\underline{J}}$'s, as described at the end of Section 2. Then the algorithm is as follows.

- (i) In $\underline{\underline{G}}_j\underline{\underline{G}}_j'$ corresponding to each main effect or interaction factor, replace each $\underline{\underline{J}}_{\underline{\underline{N}}_r}$ by $\bar{\underline{\underline{J}}}_{\underline{\underline{N}}_r}$, and each $\underline{\underline{I}}_{\underline{\underline{N}}_r}$ by $\underline{\underline{K}}_{\underline{\underline{N}}_r}$.
- (ii) In $\underline{\underline{G}}_j\underline{\underline{G}}_j'$ corresponding to a factor (represented by $\underline{\underline{\theta}}_j$) nested in some other (main effect or interaction) factor which is specifically represented in the model as $\underline{\underline{\theta}}_k$ for $k \neq j$, proceed as in (i) except leave unchanged all those $\underline{\underline{I}}$'s for the factors within which $\underline{\underline{\theta}}_j$ is nested.

Then $\underline{\underline{X}}\underline{\underline{X}}^+$ is the sum of the changed $\underline{\underline{G}}_j\underline{\underline{G}}_j'$ terms, and $\underline{\underline{M}} = \underline{\underline{I}} - \underline{\underline{X}}\underline{\underline{X}}^+$.

We use the examples of Section 2 to illustrate the algorithm. For example

(1),

$$\underline{\underline{X}}\underline{\underline{X}}' = \underline{\underline{J}}_{\underline{\underline{a}}} \otimes \underline{\underline{J}}_{\underline{\underline{b}}} \otimes \underline{\underline{J}}_{\underline{\underline{c}}} \otimes \underline{\underline{J}}_{\underline{\underline{n}}} + \underline{\underline{I}}_{\underline{\underline{a}}} \otimes \underline{\underline{J}}_{\underline{\underline{b}}} \otimes \underline{\underline{J}}_{\underline{\underline{c}}} \otimes \underline{\underline{J}}_{\underline{\underline{n}}} + \underline{\underline{J}}_{\underline{\underline{a}}} \otimes \underline{\underline{I}}_{\underline{\underline{b}}} \otimes \underline{\underline{J}}_{\underline{\underline{c}}} \otimes \underline{\underline{J}}_{\underline{\underline{n}}} + \underline{\underline{I}}_{\underline{\underline{a}}} \otimes \underline{\underline{I}}_{\underline{\underline{b}}} \otimes \underline{\underline{J}}_{\underline{\underline{c}}} \otimes \underline{\underline{J}}_{\underline{\underline{n}}}$$

and the algorithm converts this to

$$\begin{aligned} \underline{\underline{X}}\underline{\underline{X}}^+ &= \bar{\underline{\underline{J}}}_{\underline{\underline{a}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{b}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{c}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{n}}} + \underline{\underline{K}}_{\underline{\underline{a}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{b}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{c}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{n}}} + \bar{\underline{\underline{J}}}_{\underline{\underline{a}}} \otimes \underline{\underline{K}}_{\underline{\underline{b}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{c}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{n}}} + \underline{\underline{K}}_{\underline{\underline{a}}} \otimes \underline{\underline{K}}_{\underline{\underline{b}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{c}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{n}}} \\ &= (\bar{\underline{\underline{J}}}_{\underline{\underline{a}}} + \underline{\underline{K}}_{\underline{\underline{a}}}) \otimes (\bar{\underline{\underline{J}}}_{\underline{\underline{b}}} + \underline{\underline{K}}_{\underline{\underline{b}}}) \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{c}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{n}}} = \underline{\underline{I}}_{\underline{\underline{ab}}} \otimes \bar{\underline{\underline{J}}}_{\underline{\underline{cn}}} . \end{aligned}$$

Example (2) has the γ -factor nested within the β -factor. Keeping this in mind, the algorithm converts

$$\begin{aligned} \underline{\underline{X}}\underline{\underline{X}}' &= \underline{\underline{J}}_{\underline{\underline{a}}} \otimes \underline{\underline{J}}_{\underline{\underline{b}}} \otimes \underline{\underline{J}}_{\underline{\underline{c}}} \otimes \underline{\underline{J}}_{\underline{\underline{n}}} + \underline{\underline{I}}_{\underline{\underline{a}}} \otimes \underline{\underline{J}}_{\underline{\underline{b}}} \otimes \underline{\underline{J}}_{\underline{\underline{c}}} \otimes \underline{\underline{J}}_{\underline{\underline{n}}} + \underline{\underline{J}}_{\underline{\underline{a}}} \otimes \underline{\underline{I}}_{\underline{\underline{b}}} \otimes \underline{\underline{J}}_{\underline{\underline{c}}} \otimes \underline{\underline{J}}_{\underline{\underline{n}}} \\ &\quad + \underline{\underline{I}}_{\underline{\underline{a}}} \otimes \underline{\underline{I}}_{\underline{\underline{b}}} \otimes \underline{\underline{J}}_{\underline{\underline{c}}} \otimes \underline{\underline{J}}_{\underline{\underline{n}}} + \underline{\underline{J}}_{\underline{\underline{a}}} \otimes \underline{\underline{I}}_{\underline{\underline{b}}} \otimes \underline{\underline{I}}_{\underline{\underline{c}}} \otimes \underline{\underline{J}}_{\underline{\underline{n}}} + \underline{\underline{I}}_{\underline{\underline{a}}} \otimes \underline{\underline{I}}_{\underline{\underline{b}}} \otimes \underline{\underline{I}}_{\underline{\underline{c}}} \otimes \underline{\underline{J}}_{\underline{\underline{n}}} \end{aligned}$$

into

$$\begin{aligned}
 \bar{X}\bar{X}^+ &= \bar{J}_{\bar{a}} \otimes \bar{J}_{\bar{b}} \otimes \bar{J}_{\bar{c}} \otimes \bar{J}_{\bar{n}} + \bar{K}_{\bar{a}} \otimes \bar{J}_{\bar{b}} \otimes \bar{J}_{\bar{c}} \otimes \bar{J}_{\bar{n}} + \bar{J}_{\bar{a}} \otimes \bar{K}_{\bar{b}} \otimes \bar{J}_{\bar{c}} \otimes \bar{J}_{\bar{n}} \\
 &\quad + \bar{K}_{\bar{a}} \otimes \bar{K}_{\bar{b}} \otimes \bar{J}_{\bar{c}} \otimes \bar{J}_{\bar{n}} + \bar{J}_{\bar{a}} \otimes \bar{I}_{\bar{b}} \otimes \bar{K}_{\bar{c}} \otimes \bar{J}_{\bar{n}} + \bar{K}_{\bar{a}} \otimes \bar{I}_{\bar{b}} \otimes \bar{K}_{\bar{c}} \otimes \bar{J}_{\bar{n}} \\
 &= \bar{I}_{\bar{abc}} \otimes \bar{J}_{\bar{n}} .
 \end{aligned}$$

Example (3) is not covered by the algorithm because the δ_{ijk} effects are nested within the (α_i, β_j) -cells for which the model contains no specific (interaction) effects. The latter are needed in the model in order for part (ii) of the algorithm to be effective; if $(\alpha\beta)_{ij}$ effects were part of the model they would act as θ_k of the algorithm. The inability of the algorithm to deal with a model of this nature is of no consequence because models with factors nested within factors that are represented in the model in less than a fully parameterized manner are unlikely to be met with in practice.

The algorithm utilizes structural features inherent in any model, arising out of its statistical meaning: e.g., interactions cannot be fixed effects if their associated main effects are random, nor can fixed effects be nested within random effects. Formalization of these features and rigorous proof of the algorithm would, as Cornfield and Tukey [1956] so rightly say in a situation of similar repetitive algebraic complexity, involve "systematic algebra [which] can take us deep into the forest of notation. But the detailed manipulation will, sooner or later, blot out any understanding we may have started with."

4. Estimation Under Zero Kurtosis

Minimum variance properties of analysis of variance (ANOVA) estimators of variance components from balanced data are summarized in (1.1). The variances of these estimators do, of course, depend on $\text{var}(\bar{y}) = \bar{V} = \sum_{i=1}^c \sigma_i^2 \bar{V}_i$ of (2.2). More than that, existence of UMVUQ estimators of the σ_i^2 's comes from \bar{V} and the \bar{V}_i 's having a certain structure. When the kurtosis is zero, as defined following (2.3),

it is sufficient that the structure for the \underline{V}_i 's be that they form a quadratic subspace of symmetric matrices, as defined by Seely [1971]. To be precise, let

$$\mathcal{B} = \left\{ \sum_{i=1}^c t_i \underline{V}_i \mid t_1, \dots, t_c \in \mathcal{R} \right\}, \quad (4.1)$$

be the set of all matrices that are linear combinations $\sum_{i=1}^c t_i \underline{V}_i$ of the \underline{V}_i 's, for the t_i 's being any real scalars. Then \mathcal{B} is defined by Seely [1971] as a quadratic subspace of symmetric matrices when every member B of \mathcal{B} has B^2 also in \mathcal{B} .

Seely's [1971, p. 715] results on uniform minimum variance unbiased estimation are established on the basis of two assumptions:

- (a) that \mathcal{B} is a quadratic subspace of symmetric matrices,
- and
- (b) that matrices \underline{H}_i exist such that $\underline{V}_i \underline{X} = \underline{X} \underline{H}_i$ for $i = 1, \dots, c$.

These assumptions certainly hold in a fixed effects model, as in Atiquallah [1962], wherein the only \underline{V} matrix is \underline{I}_N . They also hold for the random effects model in Theorem 7 of Graybill and Hultquist [1961], since their requirement that an analysis of variance exist leads to Seely's assumption (a), while their assumption (iv) is Seely's assumption (b). Since Seely [1971, p. 717] shows that his assumptions (a) and (b) necessitate translation-invariance of the resulting estimator, neither Atiquallah [1962] nor Graybill and Hultquist [1961] need a restriction to invariant quadratic estimates.

In general, however, an ANOVA model with balanced data does not necessarily satisfy Seely's assumption (a) for the same kind of reasons that Seely's [1971, p. 719] example of the balanced incomplete block design does not, and as further evidenced in example 1 of Kleffe and Pincus [1974, p. 53]. Another demonstration that \mathcal{B} is not always a quadratic subspace is given by Searle and Henderson [1979] for the 2-way crossed classification where both the inverse \underline{V}^{-1} and \underline{V}^2 include a term in \underline{J}_N whereas \underline{V} itself does not. But the \underline{V}_i 's of \underline{V} , together with \underline{J}_N do

form a quadratic subspace and \underline{V} is a member of it. Indeed, there are typically two distinct situations:

- (1) For some models (e.g., crossed classification models having no nested factors), the \underline{V}_i do not define a quadratic subspace. This is because, by the crossed nature of the factors, there is a product of two \underline{V}_i 's that yields \underline{J}_N , and \underline{J}_N has to be included in \mathcal{B} .
- (2) For other models (e.g., completely nested models, and mixed models having random factors that are, within themselves, effectively nested) the \underline{V}_i 's define a quadratic subspace and no product $\underline{V}_i \underline{V}_j$ yields \underline{J}_N , and so there is no need to include \underline{J}_N .

In contrast to $\underline{V} = \text{var}(\underline{y})$ consider the variance of \underline{My} from which the translation invariant quadratic form $\underline{y}' \underline{M} \underline{My}$ of Section 3 is formed:

$$\text{var}(\underline{My}) = \underline{MVM} = \sum_{i=1}^c \sigma_i^2 \underline{MV}_i \underline{M} . \quad (4.2)$$

The analogous form of \mathcal{B} for matrices $\underline{MV}_i \underline{M}$ is then

$$\mathcal{B}_M = \left\{ \sum_{i=1}^c t_i \underline{MV}_i \underline{M} \mid t_1, \dots, t_c \in \mathcal{R} \right\} . \quad (4.3)$$

Concerning \mathcal{B}_M , Theorem 6 of Kleffe and Pincus [1974, p. 52] shows that in any linear model the quadratic subspace property that is not always evident in \underline{V} is needed only of \mathcal{B}_M . For balanced data this is always the case, i.e., \mathcal{B}_M defines a quadratic subspace, resulting from the fact that \underline{M} and the \underline{V}_i 's are all linear combinations of Kronecker products of \underline{I} 's, \underline{J} 's and \underline{K} 's. No matrix such as \underline{J}_N ever has to be included with the $\underline{MV}_i \underline{M}$'s. This is so because \underline{MJ}_N is null. [Note that the analogue of Seely's assumption (b) is trivially satisfied, since \underline{My} has expectation zero.]

Theorems 1 and 3 of Seely [1971] assert that for balanced data with zero kurtosis there exists an unbiased invariant quadratic estimator of the variance components which has uniformly minimum variance in its class (UMVUIQ). Under normality this estimator retains the UMV property among all unbiased invariant estimators, whether they are quadratic or not (UMVUI). We now show that this estimator also coincides with the ANOVA estimator, thus justifying (1.1) and (1.3).

4.2. The derived, dispersion-mean model

The general linear model for \underline{y} is $\underline{y} = \underline{X}\underline{\beta} + \underline{Z}\underline{u}$ of (2.1). For \underline{M} of (2.5) the model for \underline{My} is $\underline{My} = \underline{MZu}$. Seely [1970, 1971], Pukelsheim [1976, 1977, 1979], Brown [1978, 1979] and Anderson [1978, 1979] show how this model for \underline{My} can be converted into a linear model for σ^2 . To this end we set out some notation.

For a matrix \underline{A} we denote by $\text{vec}\underline{A}$ the vector formed by stacking the columns of \underline{A} one under the other to form a single column vector. Its history, properties, and uses in statistics are extensively reviewed in Henderson and Searle [1979]. An important result connecting vec with Kronecker products is

$$\text{vec}(\underline{ABC}) = (\underline{C}' \otimes \underline{A})\text{vec}\underline{B} ; \quad (4.4)$$

and an important matrix is $\underline{I}_{N,N}$ defined, for \underline{A} of order $N \times N$, by

$$\text{vec}(\underline{A}_{N \times N}) = \underline{I}_{N,N} \text{vec}(\underline{A}') , \quad (4.5)$$

as discussed in Henderson and Searle [1979, 1981], where it is called a vec permutation matrix. We also need $\underline{q} = \Sigma \underline{q}_i$, for \underline{q}_i of (2.3), and introduce the $q^2 \times q$ matrix

$$\underline{D}_q = [\underline{e}_1 \otimes \underline{e}_1 \mid \cdots \mid \underline{e}_q \otimes \underline{e}_q] , \quad (4.6)$$

where \underline{e}_i is the i 'th column of \underline{I}_q .

Using these terms we then define

$$\underline{y} = \underline{M}\underline{y} \otimes \underline{M}\underline{y} \quad (4.7)$$

$$\underline{x} = [\text{vec}(\underline{M}\underline{V}_1\underline{M}) \mid \cdots \mid \text{vec}(\underline{M}\underline{V}_c\underline{M})] \quad (4.8)$$

$$\underline{W}_1 = (\underline{M} \otimes \underline{M})(\underline{V} \otimes \underline{V})(\underline{I}_{N^2} + \underline{I}_{N,N})(\underline{M} \otimes \underline{M}) \quad (4.9)$$

$$\underline{\Delta} = \bigoplus_{i=1}^c \sigma_i^2 \underline{Y}_i \underline{I}_{q_i} = \text{block diagonal}(\sigma_1^2 \underline{Y}_1 \underline{I}_{q_1}, \cdots, \sigma_c^2 \underline{Y}_c \underline{I}_{q_c}) \quad (4.10)$$

$$\underline{W}_2 = (\underline{M} \otimes \underline{M})(\underline{Z} \otimes \underline{Z}) \underline{D} \underline{\Delta} \underline{D}' (\underline{Z} \otimes \underline{Z})' (\underline{M} \otimes \underline{M}) . \quad (4.11)$$

Then the derived, or dispersion-mean, model for σ^2 can be expressed as

$$E(\underline{y}) = \underline{x}\sigma^2 \quad \text{and} \quad \text{var}(\underline{y}) = \underline{W}_1 + \underline{W}_2 . \quad (4.12)$$

Ordinary least squares on (4.12) leads, as is not hard to show, to equations

$$\left\{ \text{tr}(\underline{M}\underline{V}_i \underline{M}\underline{V}_j) \right\} \hat{\sigma}^2 = \left\{ \underline{y}' \underline{M}\underline{V}_i \underline{M}\underline{y} \right\} \quad (4.13)$$

for $i = 1, \cdots, c$. Then, since \mathcal{B}_M of (4.3) is a quadratic subspace of symmetric matrices, the result of Seely [1971] discussed in the first paragraph of this section shows that the estimators $\hat{\sigma}^2$ of (4.13) are UMVUQ – and because they are also translation invariant they are thus UMVUIQ; and, under normality, they are UMVUI. Furthermore, because in ANOVA models with balanced data, ANOVA estimators have these same properties, as discussed in Section 1, the estimators in (4.13) are the ANOVA estimators.

(4.13) is also the generalized least squares solution for (4.12) when $\sigma_c^2 = \sigma_\epsilon^2 = 1$ (or more generally, any value whatever) and $\sigma_1^2 = \cdots = \sigma_{c-1}^2 = \gamma_1 = \cdots = \gamma_c = 0$. For this reason, equations (4.13) have recently been called MIVQUE-0 by Goodnight [1979] and MINQUE0 by Searle [1979a,b].

A necessary and sufficient condition under which ordinary least squares estimation of $\underline{t}'\underline{\beta}$ in the model $E(\underline{y}) = \underline{X}\underline{\beta}$, $\text{var}(\underline{y}) = \underline{V}$ is the same as best linear unbiased estimation is $\underline{V}\underline{X} = \underline{X}\underline{H}$, for some matrix \underline{H} . Zyskind [1967] and Seely and Zyskind [1971] give a broad presentation of this result which, under normality, leads to the quadratic subspace condition of Seely [1971]. While (1.1) and (1.3) thus follow from Seely's [1971] general theory, the $\underline{V}\underline{X} = \underline{X}\underline{H}$ requirement for the model (4.12) with zero kurtosis (i.e., with $\underline{\Delta}$ and \underline{W}_2 null) is $\underline{W}_1\underline{\mathcal{X}} = \underline{\mathcal{X}}\underline{H}_1$ for some \underline{H}_1 . Direct verification of this may be found in Anderson [1978]. Under normality, Anderson [1979] also establishes (1.4), i.e., equality of REML and UMVUI, as do Pukelsheim and Styan [1979]. We now turn to the case where the kurtosis need not be zero.

5. Estimation Under Non-zero Kurtosis

For ANOVA models with balanced data we now verify (1.2) by exhibiting a matrix \underline{H}_2 that satisfies

$$\underline{W}_2\underline{\mathcal{X}} = \underline{\mathcal{X}}\underline{H}_2, \quad (5.1)$$

where $\underline{\mathcal{X}}$ and \underline{W}_2 are defined in (4.8) and (4.11). Then, since $\underline{W}_1\underline{\mathcal{X}} = \underline{\mathcal{X}}\underline{H}_1$, we have $(\underline{W}_1 + \underline{W}_2)\underline{\mathcal{X}} = \underline{\mathcal{X}}\underline{H}$ for $\underline{H} = \underline{H}_1 + \underline{H}_2$, and so the condition for ordinary least squares estimation being the same as best linear unbiased estimation is satisfied for the non-zero kurtosis case. Theorem 4.5 of Pukelsheim [1977], Theorem 6 of Kleffe [1977] and Theorem 1.4 of Drygas [1980] point out the need for a matrix \underline{H}_2 ; we substantiate this by showing its existence for the non-zero kurtosis in ANOVA estimation from balanced data.

It suffices to show that for every $k = 1, \dots, c$ there exists some $c \times 1$ vector \underline{h}_k such that

$$\underline{W}_2 \text{vec}(\underline{MV}_k \underline{M}) = \underline{\mathcal{Z}} \underline{h}_k. \quad (5.2)$$

We now use (4.4) and (4.11), and also $\underline{D} \underline{\Delta} \underline{\Delta}' \text{vec}(\underline{A}) = \text{vec}(\underline{\Delta} \text{diag} \underline{A})$ from Pukelsheim [1977, p. 326], where $\text{diag} \underline{A}$ is the diagonal matrix with diagonal elements being those of \underline{A} . The left-hand side of (5.2) then becomes

$$\begin{aligned} \underline{W}_2 \text{vec}(\underline{MV}_k \underline{M}) &= (\underline{M} \otimes \underline{M})(\underline{Z} \otimes \underline{Z}) \underline{D} \underline{\Delta} \underline{\Delta}' \text{vec}(\underline{Z}' \underline{MV}_k \underline{MZ}) \\ &= (\underline{M} \otimes \underline{M})(\underline{Z} \otimes \underline{Z}) \text{vec}\{\underline{\Delta} \text{diag}(\underline{Z}' \underline{MV}_k \underline{MZ})\}. \end{aligned} \quad (5.3)$$

Now suppose there exist numbers $\lambda_{1k}, \dots, \lambda_{ck}$ such that

$$\text{diag}(\underline{Z}' \underline{MV}_k \underline{MZ}) = \bigoplus_{i=1}^c \lambda_{ik} \underline{I}_{q_i} \equiv \underline{\Lambda}, \text{ say.} \quad (5.4)$$

Then

$$\underline{W}_2 \text{vec}(\underline{MV}_k \underline{M}) = \text{vec}(\underline{MZ} \underline{\Lambda} \underline{Z}' \underline{M}) = \text{vec} \sum_{i=1}^c \sigma_i^4 \gamma_i \lambda_{ik} \underline{MZ}_i \underline{Z}_i' \underline{M} = \underline{\mathcal{Z}} \underline{h}_k, \quad (5.5)$$

with $\underline{h}_k = (\sigma_1^4 \gamma_1 \lambda_{1k}, \dots, \sigma_c^4 \gamma_c \lambda_{ck})'$. This establishes (5.2), and therefore (5.1).

Thus for (5.5) to hold, it remains to exhibit numbers λ_{ik} satisfying (5.4).

To this end define

$$\lambda_{ijk} = \underline{z}_{ij}' \underline{MZ}_i \underline{Z}_i' \underline{MZ}_i \underline{z}_{ij} \quad (5.6)$$

$$= \underline{z}_{ij}' (\underline{I} - \underline{XX}^+) \underline{Z}_i \underline{Z}_i' (\underline{I} - \underline{XX}^+) \underline{z}_{ij} \quad (5.7)$$

$$= \underline{z}_{ij}' \underline{Z}_i \underline{Z}_i' \underline{z}_{ij} - \underline{z}_{ij}' \underline{XX}^+ \underline{Z}_i \underline{Z}_i' \underline{z}_{ij} - \underline{z}_{ij}' \underline{Z}_i \underline{Z}_i' \underline{XX}^+ \underline{z}_{ij} + \underline{z}_{ij}' \underline{XX}^+ \underline{Z}_i \underline{Z}_i' \underline{XX}^+ \underline{z}_{ij}. \quad (5.8)$$

We show that λ_{ijk} does not depend on j . Partition \underline{Z}_i into its columns \underline{z}_{ij} , for $j = 1, \dots, q_i$.

First observe from Section 2.4 that any \underline{Z}_i is a Kronecker product (KP) of \underline{I} 's and $\underline{1}$'s of the form (2.6). Hence, through partitioning the \underline{I} 's into their columns, denoted as \underline{e} -vectors, each

$$\underline{z}_{ij} \text{ is a KP of } \underline{e}'\text{'s and } \underline{1}'\text{'s} . \quad (5.9)$$

Along with this, we also have from (2.10) and the algorithm in Section 3.1 that

$$\underline{Z}\underline{Z}' \text{ is a KP of } \underline{I}'\text{'s and } \underline{J}'\text{'s} \quad (5.10)$$

and

$$\underline{XX}^+ \text{ is a sum of KP's of } \underline{I}'\text{'s, } \underline{\bar{J}}'\text{'s and } \underline{K}'\text{'s} . \quad (5.11)$$

All KP's in (5.9), (5.10) and (5.11) are conformable, whereupon each term in (5.8) is a KP. We use \underline{e}_j to represent the dependence on j of an \underline{e} referred to in (5.9). Then any position t in the KP that is a term of (5.8) is a scalar that has, for some matrix \underline{Q}_t , one of the following forms:

$$\left. \begin{array}{l} \text{or} \\ \underline{e}'\underline{Q}_t\underline{e}_j = j\text{'th diagonal element of } \underline{Q}_t , \\ \\ \underline{e}'\underline{Q}_t\underline{1} = j\text{'th row sum of } \underline{Q}_t , \\ \text{or} \\ \underline{1}'\underline{Q}_t\underline{1} = \text{sum of all elements in } \underline{Q}_t . \end{array} \right\} \quad (5.12)$$

Therefore each term in (5.8) is a product of scalars like (5.12), where \underline{Q}_t is either the matrix in position t of $\underline{Z}\underline{Z}'$ or is a product of matrices in position t of $\underline{Z}\underline{Z}'$ and \underline{XX}^+ . Hence, from (5.10) and (5.11) and using Section 2.2,

$$\underline{Q}_t \text{ is either an } \underline{I}, \underline{J}, \underline{\bar{J}} \text{ or } \underline{K} . \quad (5.13)$$

Hence, from (5.13), the scalars in (5.12) are all independent of j and so this is also true of each term in (5.8), and thus of λ_{ijk} itself.

Acknowledgments

Multi-authorship is sometimes a cause for scepticism; but in this case it is very genuine. The four of us came together through the generosity of sponsors of

"Variance Components and Animal Breeding: A Conference in Honor of C. R. Henderson" held at Cornell University, Ithaca, New York, July 1979. We gratefully acknowledge this opportunity to have worked together at and following the Conference. We also benefited from discussions with Justus Seely, and from detailed remarks from a referee.

References

- Anderson, R. D. [1978]. Studies on the estimation of variance components. Ph.D. Thesis, Animal Science Department, Cornell University.
- Anderson, R. D. [1979]. Estimating variance components from balanced data: Optimum properties of REML solutions and MIVQUE estimators. Proceedings of Variance Components and Animal Breeding: A Conference in Honor of C. R. Henderson, 205-215. Animal Science Department, Cornell University.
- Atiqullah, M. [1962]. The estimation of residual variance in quadratically balanced least-squares problems and the robustness of the F-test. Biometrika, 49, 83-91.
- Brown, K. G. [1978]. Estimation of variance components using residuals. J. Amer. Statist. Assoc., 73, 141-146.
- Brown, K. G. [1979]. Estimation of variance components using residuals: Some empirical evidence. Proceedings of Variance Components and Animal Breeding: A Conference in Honor of C. R. Henderson, 139-154. Animal Science Department, Cornell University.
- Corbeil, R. R. and Searle, S. R. [1976]. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. Technometrics, 18, 31-38.
- Cornfield, J. and Tukey, J. W. [1956]. Average values of mean squares in factorials. Ann. Math. Statist., 27, 907-949.
- Drygas, H. [1972]. The estimation of residual variance in regression analysis. Math. Operationsforsch. u. Statist., 3, 373-388.
- Drygas, H. [1980]. Hsu's theorem in variance component models. Mathematical Statistics, Banach Center Publications, 6, PWN - Polish Scientific Publishers, Warsaw.
- Goodnight, J. H. [1979]. New features in GLIM and VARCOMP. Proceedings, Fourth Annual SAS Users' Group International Conference, SAS Institute, Raleigh, North Carolina.
- Graybill, F. A. [1954]. On quadratic estimates of variance components. Ann. Math. Statist., 25, 367-372.

- Graybill, F. A. and Hultquist, R. A. [1961]. Theorems concerning Eisenhart's Model II. Ann. Math. Statist., 32, 261-269.
- Graybill, F. A. and Wortham, A. W. [1956]. A note on uniformly best unbiased estimators for variance components. J. Amer. Statist. Assoc., 51, 266-268.
- Harville, D. A. [1977]. Maximum likelihood approaches to variance component estimation and related problems. J. Amer. Statist. Assoc., 72, 320-340.
- Henderson, H. V. and Searle, S. R. [1979]. Vec and vech operators, with some uses in Jacobians and multivariate statistics. Canadian J. of Statist., 7, 65-81.
- Henderson, H. V. and Searle, S. R. [1981]. The vec-permutation matrix, the vec operator and Kronecker products: A review. Linear and Multilinear Algebra, 9, 271-288.
- Hsu, P. L. [1938]. On the best unbiased quadratic estimate of the variance. Statist. Res. Mem., 2, 91-104.
- Kleffe, J. [1977]. Invariant methods for estimating variance components in mixed linear models. Math. Operationsforsch. Statist. Ser. Statist., 8, 233-250.
- Kleffe, J. and Pincus, R. [1974]. Bayes and best quadratic unbiased estimators for parameters of the covariance matrix in a normal linear model. Math. Operationsforsch. u. Statist., 5, 43-67.
- Patterson, H. D. and Thompson, R. [1971]. Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545-554.
- Pukelsheim, F. [1976]. Estimating variance components in linear models. J. Multivariate Anal., 6, 626-629.
- Pukelsheim, F. [1977]. On Hsu's model in regression analysis. Math. Operationsforsch. Statist. Ser. Statist., 8, 323-331.
- Pukelsheim, F. [1979]. Classes of linear models. Proceedings of Variance Components and Animal Breeding: A Conference in Honor of C. R. Henderson, 69-83. Animal Science Department, Cornell University.
- Pukelsheim, F. and Styan, G. P. H. [1979]. Nonnegative definiteness of the estimated dispersion matrix in a multivariate linear model. Bull. Acad. Polon. Sci. Sér. Sci. Math., XXVII, 327-330.
- Rao, C. R. [1971]. Estimation of variance and covariance components - MINQUE theory. J. Multivariate Anal., 1, 257-275.
- Searle, S. R. [1971]. Linear Models. John Wiley and Sons, New York.
- Searle, S. R. [1979a]. Notes on variance component estimation: A detailed account of maximum likelihood and kindred methodology. Paper BU-673-M in the Biometrics Unit, Cornell University.

- Searle, S. R. [1979b]. Maximum likelihood and minimum variance estimators of variance components. Proceedings of Variance Components and Animal Breeding: A Conference in Honor of C. R. Henderson, 59-68. Animal Science Department, Cornell University.
- Searle, S. R. and Henderson, H. V. [1979]. Dispersion matrices for variance components models. J. Amer. Statist. Assoc., 74, 465-470.
- Seely, J. [1970]. Linear spaces and unbiased estimation - application to the mixed model. Ann. Math. Statist., 41, 1735-1748.
- Seely, J. [1971]. Quadratic subspaces and completeness. Ann. Math. Statist., 42, 710-721.
- Seely, J. and Zyskind, G. [1971]. Linear spaces and minimum variance unbiased estimation. Ann. Math. Statist., 42, 691-703.
- Seifert, B. [1979]. Optimal testing for fixed effects in general balanced mixed classification models. Math. Operationsforsch. Statist. Ser. Statist., 10, 237-256.
- Tan, Waiyuan [1979]. On the quadratic estimation of covariance matrices in MANOVA random effect models. Statistica, 38, 449-458.
- Zyskind, G. [1967]. On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. Ann. Math. Statist., 38, 1092-1109.