VARIANCE COMPONENTS ESTIMATION: A REVIEW, IN NOTE FORM

BU-651-M

S. R. Searle
Cornell University, Ithaca, New York

June, 1978

Abstract

Part I of this paper is a brief account of early work (1931-1953) on variance component estimation and of some recent uses and applications of variance components models. Parts II and III are a summary account, in note form, of methods currently available for estimating variance components, particularly from unbalanced data (having unequal numbers of observations in the subclasses): part II summarizes analysis of variance methods of estimation, and Part III deals with maximum-likelihood style methods. Part IV shows relationships between the different methods.

PART I. HISTORY, AND CURRENT USES

1.1. Variance components models

There is widespread familiarity with the traditional analysis of variance model such as that for the completely randomized design:

\[ y_{ij} = \mu + \alpha_i + e_{ij}. \] (1)

In this equation \( y_{ij} \) is the \( j \)'th observation on the \( i \)'th treatment, with \( \mu \) representing a general mean and \( \alpha_i \) the effect of the \( i \)'th "treatment". The expected value of \( y_{ij} \) is taken as \( E(y_{ij}) = \mu + \alpha_i \) for \( E \) representing expectation over repeated sampling. The term \( e_{ij} \) in (1) represents the difference \( y_{ij} - E(y_{ij}) \) and is usually taken as being a random variable (often called residual, or error, or both), with

Paper No. BU-651-M in the Biometrics Unit, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14853. This paper is an amalgamation of Searle [1977, 1978].
zero mean and variance $\sigma^2_e$ for all $i = 1, \ldots, a$ and $j = 1, \ldots, n$, the individual $e_{ij}$'s also being uncorrelated.

The parameters of interest in this model are the mean $\mu$ and the $\alpha_i$'s, the effects of the "treatments" on the yield; and one object of analyzing the data is that of estimating linear functions of $\mu$ and the $\alpha_i$'s. Best linear unbiased estimators of two functions of interest are, for example,

$$\hat{\mu} + \hat{\alpha}_i = \bar{y}_i = \frac{1}{n} \sum_{j=1}^{n} y_{ij}$$

and

$$\hat{\alpha}_i - \hat{\alpha}_j = \bar{y}_i - \bar{y}_j.$$

In the context of this model the $\mu$ and the $\alpha_i$'s are taken as being constants, albeit unknown and unknowable, but nevertheless fixed constants. As such, they are usually called fixed effects. They are deemed to be constants representing the effects of the different "treatments" being studied. The "treatments" are the things of particular interest, chosen by some investigator because of his interest in them: different diets fed to laboratory animals, farm livestock or to humans, different fertilizers given to a corn crop, different forage crops grown in the same region, different machines used in a manufacturing process, different drugs given for the same illness, and so on. The possibilities are legion - as are the varieties of models and their complexities, reaching far beyond those of (1).

Now consider the $\alpha_i$'s of equation (1) as being realized (but unobservable) values of a random variable having zero mean and variance $\sigma^2_\alpha$, with the $\alpha_i$'s being uncorrelated with each other and with the $e_{ij}$'s. [In this case $E(y_{ij}) = \mu$, and $e_{ij}$ is defined as $e_{ij} = y_{ij} - E(y_{ij} | \alpha_i)$ where $E(y_{ij} | \alpha_i)$ is a conditional expected value.] In this context the individual $\alpha_i$'s are no longer things of particular interest as they are in fixed effects models; those that occur in the data are deemed to be
just a random sample of $\alpha$'s selected from a population defined as having zero mean and variance $\sigma^2$. There is therefore little or no reason for estimating either the $\alpha_i$'s or differences between them; the parameter of interest so far as they are concerned is now $\sigma^2$. Because in this case (1) gives $\sigma^2 = \sigma^2 + \sigma^2$, the variances $\sigma^2$ and $\sigma^2$, being components of the variance of $y$, are called variance components. This use of (1) leads to the model being called a variance components model, and the $\alpha_i$'s are called random effects. Correspondingly, the model is sometimes called the random model.

Some models have both fixed effects and random effects, in which case the name mixed model is used. An example would be a randomized complete block design having model equation

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

where the $\alpha_i$'s are fixed effects representing treatments and the $\beta_j$'s are random effects representing blocks. The parameters of interest would be $\mu$ and the $\alpha_i$'s and the variance components $\sigma^2$ and $\sigma^2$.

I.2. History, 1931-1953

An excellent history of variance components, starting at 1856, is to be found in Anderson [1978]. Only a few salient dates are given here.

The basic principle for estimating variance components has been, and to a large extent still is, that of equating quadratic functions of the observations to their expected values. Obvious candidates for such functions are the sums of squares of the analysis of variance table. The first formal description of this procedure appears to be that of Tippett [1931, Secs. 6.1, 6.2, 10.3], followed by papers by Daniels [1939], whose interest was in weights of slubbings from the carding process in the woolen industry, and by Winsor and Clarke [1940], who analyzed catches of different species in successive hauls of plankton nets. These two papers were
published only a few months apart and it seems certain they were the results of independent work by the respective authors. Both papers give expected sums of squares for two or three different analyses of variance, Winsor and Clarke explicitly showing

\[ E \sum_{i=1}^{a} n(\bar{y}_{i} - \bar{y})^2 = (a - 1)(n\sigma^2 + \sigma_e^2) \]

and

\[ E \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2 = a(n - 1)\sigma_e^2 \]

for the random model based on (1). They then estimate \( \sigma^2 \alpha \) and \( \sigma^2_e \) by \( \hat{\sigma}^2_\alpha \) and \( \hat{\sigma}^2_e \), the solutions to the equations

\[ E \sum_{i=1}^{a} n(\bar{y}_{i} - \bar{y})^2 = (a - 1)(n\hat{\sigma}^2_\alpha + \hat{\sigma}^2_e) \]

and

\[ E \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2 = a(n - 1)\hat{\sigma}^2_e \].

Daniels mentions Tippett but not R. A. Fisher, in deriving expected values of sums of squares, whereas Winsor and Clarke describe their derivation of (2) as being "a straightforward extension of the suggestions of R. A. Fisher in his 'Statistical Methods for Research Workers' [Sec.] 40." Presumably this is the Seventh Edition, published in 1938, wherein Sec. 40 is the section dealing with the intraclass correlation, exactly as does the same section, unchanged, in the Twelfth Edition of 1954. The important suggestion of Fisher's in table 39 which, although he makes no explicit mention of expectation whatever, contains exactly the results (2). Nor does he give any serious attention to estimating \( \sigma^2_\alpha \) beyond saying "we may make estimates of the values A and B [\( \sigma^2_\alpha \) and \( \sigma^2_e \)], or in other words we may analyze the variance into the portions contributed by the two causes". Both Daniels and Winsor and Clarke use the expectation notation and are concerned with estimating \( \sigma^2_\alpha \) and \( \sigma^2_e \).
At about the same time as the Daniels and Winsor and Clarke papers were published (the latter in what, even at that time, must have been somewhat of an obscure journal for statisticians), Snedecor's third edition [1940] became available with, as far as I can see, no reference to variance components at all. Page 205 contains discussion of estimating the intra-class correlation as \( \Lambda/(\Lambda + B) \), just as does Fisher [1938]. The nearest thing to characterizing \( \Lambda \) as a variance component is the description that "\( \Lambda \) is the same for all ... samples - it is the common element, analogous to covariance." And that is, of course, the case: the covariance between \( y_{ij} \) and \( y_{ij'} \), for \( j \neq j' \) is \( \sigma^2_{\alpha} \).

Winsor and Clarke not only use (2) and (3), for balanced data, but they also derive the expectations

\[
E \sum_{i=1}^{a} n_{i} (\bar{y}_{i} - \bar{y})^2 = (n - \sum_{i=1}^{a} n_{i}/n_{i}\sigma^2 + (a - 1)\sigma^2_{e}
\]

and

\[
E \sum_{i=1}^{a} \sum_{j=1}^{n_{i}} (y_{ij} - \bar{y}_{i})^2 = (n - a)\sigma^2_{e},
\]

for unbalanced data, something which Daniels [1939] does not address himself to. Interestingly enough, Snedecor [Third Edition, 1940] touches obliquely on this subject in Example 10.21 (p. 205) where, in referring to unbalanced data of Table 10.8, he asks the question "Why can't you calculate intra-class correlation accurately" for such data? Winsor and Clarke's results (4) would show that you could. Needless to say, that example does not appear in the sixth edition, Snedecor and Cochran [1967].

Although Daniels [1939] and Winsor and Clarke [1940] represent both sides of the Atlantic, it appears that major developments in variance components estimation subsequently took place mainly in the U. S. A. An exception to this was Ganguli...
[1941], dealing with nested classifications, and then came Crump [1946] concerned with randomized complete blocks, and Satterthwaite [1946] dealing with approximate sampling distributions of estimated variance components. This was followed by Eisenhart [1947] who put a firm foundation to the distinction between the fixed effects model (Model I) and the random effects model (Model II), a distinction which Yates [1967] later took great exception to. Sampling variances of estimators obtainable from (k) were given in Hammersley [1949] for arbitrary distributional properties, and in a doctoral thesis by Crump [1947] for normality assumptions. These results from the thesis were included in Crump's [1951] review paper, but not those concerned with maximum likelihood estimation, a topic which reasserted itself in Hartley and Rao [1967] and has been actively pursued ever since, with no immediate end in sight (e.g., Harville [1977]).

Anderson and Bancroft [1952] is the first book with extensive treatment of variance components, its final four chapters being devoted to the topic entirely. This really set the subject on a firm footing, and well and truly laid out the procedure of equating analysis of variance sums of squares to their expectations as a method of estimating variance components. The book deals very thoroughly with estimation from unbalanced data for both mixed and random models; it also deals with unbalanced data for nested classifications and, after considering incomplete blocks designs, it poses a number of pertinent research problems, many of which have still not been answered satisfactorily. In all, the book is a milestone in variance components estimation.

Active interest in estimation from balanced data continued well into the 1960's, by which time several optimum properties of the estimators had been established (see references in Searle [1971], for example, particularly those by Graybill and co-workers). Estimation from unbalanced data in crossed classifications,
and mixtures of crossed and nested classifications, with mixed or random models, got its prime start from Henderson [1953]. Active interest in unbalanced data has continued unabated to the present day and has not subsided yet.

I.3. Uses and applications

Most statistical methods are developed in response to the demands of practical problems. Variance components estimation is no exception. The first papers, by Daniels [1939] and Winsor and Clarke [1940], dealt with woolen industry and with plankton net data, respectively. Crump [1946] was interested in Drosophila egg production and he also refers to a variety of other applications of variance components: enumeration sampling, cereal experiments, swine breeding (three papers), corn breeding and soil sampling. Papers by Hazel and Terrill [e.g., 1945] on sheep breeding could be added to the list. Clearly, by the mid-40's, animal and plant breeders were making considerable use of variance components. The Anderson and Bancroft [1952] book also contains references to numerous uses of variance components in subject-matter disciplines: industrial experimentation, corn trials, psychological testing, sample surveys (wheat fields, soybean trials and forest nurseries), the sampling of baled wool, studies of egg production and hog prices, and analyses of the efficacy of measuring instruments.

Thus we see that, like many statistical techniques founded in practical problems, the early literature of variance components estimation contains plentiful reference to those problems. In contrast, today's statistical literature deals very largely with just the mathematics of statistical methodology, with much less space being devoted to practical problems and accompanying data than was the case twenty and more years ago. This is certainly true of variance component estimation no less than it is of other topics in statistics. Nevertheless, literature of the subject-matter disciplines (and occasionally of statistics) continues to bear witness
to the uses for variance components estimates. In addition to the frequent and
established uses in genetics and animal breeding such as estimating genetic
parameters (e.g., Becker et al. [1977]) and using them in prediction (Searle [1974])
there are now uses in a variety of other disciplines. Closely allied to the animal
breeder's parameter of repeatability is the psychologist's and educationalist's
measure of reliability of a test instrument, namely $\sigma^2_{\text{respondent}}/(\sigma^2_{\text{respondent}} +
\sigma^2_{\text{residual}})$, as, for example, in Alwin [1976]. Geneticists and others who use the
experimental design of the diallel cross (originating in genetics) also make great
use of variance components — Randall [1976] provides a comprehensive review — and
so do those designing sample surveys. Analyses of trajectory and orbital data in
rocket flight testing have been based on the mixed model version of variance com-
ponents models (e.g., Bush [1971]) and so have analyses of data from clinical trials
involving several clinics (Chakravorti and Grizzle [1975]). Kalman filtering
techniques of engineering, as described by Duncan and Horn [1972], also utilize
mixed model theory, as noted by Harville [1977]. And economists nowadays make very
wide use of mixed models in combining cross-section with time series data (e.g.,
Houthakker et al. [1974]), referring to their models as error components models.
Variance components estimation continues, therefore, to be a technique that is
quite widely used in data analysis, as well as receiving considerable attention on
its theoretical side.
I.4. References for Part I


Addendum


PART II. ESTIMATING VARIANCE COMPONENTS FROM UNBALANCED DATA IN MIXED MODELS: ANALYSIS OF VARIANCE METHODS

A summary of methods — in note form.

II.0. Introduction

Confine attention to the 2-way crossed classification model:

\[ y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \]

\[ i = 1 \ldots a \quad j = 1 \ldots b \quad k = 1 \ldots n_{ij} \quad \Sigma n_{ij} = N \]

Fixed effects model

Balanced data

All \( n_{ij} = n \): the analysis of variance is familiar.

\[
\begin{align*}
\text{Mean} & \quad 1 \quad \text{abn}
\quad \ldots \\
\text{Rows} & \quad a-1 \quad \text{SSA} = \Sigma bny_{i..}^2 - \text{abn}y^2
\quad \ldots \\
\text{Columns} & \quad b-1 \quad \text{SSB} = \Sigma any_{..j}^2 - \text{abn}y^2
\quad \ldots \\
\text{Interaction} & \quad (a-1)(b-1) \quad \text{SSAB} = \Sigma \Sigma n_{ij}^2 - \Sigma bny_{i..}^2 - \Sigma any_{..j}^2 + \text{abn}y^2
\quad \ldots \\
\text{Residual} & \quad ab(n-1) \quad \text{SSE} = \Sigma \Sigma \Sigma y_{ijk}^2 - \Sigma \Sigma n_{ij}^2
\quad \ldots \\
\text{Total} & \quad \text{abn} \quad \Sigma \Sigma \Sigma y_{ijk}^2
\end{align*}
\]

Unbalanced data: \( s \) cells containing data.

2 partitionings of sums of squares.

\[
\begin{array}{ccc}
\text{Rows before columns} & \text{OR} & \text{Columns before rows} \\
R(\mu) & 1 & R(\mu) & 1 \\
R(\alpha|\mu) & a-1 & R(\beta|\mu) & b-1 \\
R(\beta|\mu,\alpha) & b-1 & R(\alpha|\mu,\beta) & a-1 \\
R(\gamma|\mu,\alpha,\beta) & s-a-b+1 & R(\gamma|\mu,\alpha,\beta) & s-a-b+1 \\
\text{SSE} & N-s & \text{SSE} & N-s \\
\text{Total} & N & \text{Total} & N
\end{array}
\]
Mixed model:
\[ \beta_j's \text{ remain as fixed effects} \quad \alpha_i's \text{ random} \quad \gamma_{ij}'s \text{ random} \]
\[ E(\alpha_i) = 0 \quad E(\gamma_{ij}) = 0 \]
\[ \text{var}(\alpha) = \sigma_{\alpha}^2 \quad \text{var}(\gamma) = \sigma_{\gamma}^2 \]
\[ s = ab \text{ for balanced data} \]

Want to estimate: \( \mu, \beta, \sigma_{\alpha}^2, \sigma_{\gamma}^2 \) and \( \sigma_e^2 \)

Balanced data:

Use part of analysis of variance table for fixed effects model
\[ E(SSA) = (a-1)(b-1)(\text{var}(\alpha) + \text{var}(\gamma) + \sigma_e^2) \]
\[ E(SSAB) = (a-1)(b-1)(\text{var}(\gamma) + \sigma_e^2) \]
\[ E(SSE) = ab(n-1)\sigma_e^2 \]

Estimators
\[ SSA = (a-1)(b-1)(\text{var}(\alpha) + \text{var}(\gamma) + \sigma_e^2) \]
\[ SSAB = (a-1)(b-1)(\text{var}(\gamma) + \sigma_e^2) \]
\[ SSE = ab(n-1)\sigma_e^2 \]

Properties of estimators: minimum variance quadratic unbiased
[Graybill and Wortham [1956])

under normality, minimum variance unbiased
[Graybill and Hultquist [1961])

Unbalanced data:

Variety of methods available, several based on same principle as preceding:

Develop \( q \) as a vector of quadratic forms in \( \gamma \).

Derive \( E(q) \); each element will be a linear combination of variance components, elements of \( \sigma^2 \).
\[ E(q) = C\sigma^2 \text{ for some } \Sigma \]

Estimation: \( \sigma^2 = \Sigma^{-1}q \)

Question: What quadratics are used as elements for \( q \) ?

This method uses quadratic forms analogous to sums of squares of balanced data ANOVA, e.g.,

$$SSA^* = \sum_{i} \tilde{y}_{i.}^2 - N \bar{y}_{..}^2$$

$$SSAB^* = \sum_{ij} \tilde{y}_{ij}^2 - \sum_{i} \tilde{y}_{i.}^2 - \sum_{j} \tilde{y}_{.j}^2 + N \bar{y}_{..}^2$$

Note: SSAB* is not positive definite; it is not a sum of squares.

Estimation: equate SS*'s to expectations

Properties: easy to compute
unbiased for random models-
sampling variances available for 1, 2 and 3-way classifications
not unbiased for mixed models, because the fixed effects, \( \beta_{ij} \)'s, occur in \( E(SS^*) \)'s.

History: Henderson [1953]: described method
Searle [1971a]: collected details for 1-, 2- and 3-way classifications, including sampling variances of estimates.

Designed to overcome biasedness of Method 1 for mixed models.

Retains relative ease of computing.

**Principle:** "Correct" data for fixed effects.

Use Method 1 on corrected data.

Make slight adjustments.

\[ y = X\beta + Zu + e \]

Use normal equations as if \( u \) were fixed:

\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & Z'Z
\end{bmatrix}
\begin{bmatrix}
\beta^0 \\
u^0
\end{bmatrix} =
\begin{bmatrix}
X'y \\
Z'y
\end{bmatrix}
\]

Correct for \( \beta \): use a \( \beta^0 \) such that

\[ z = y - X\beta^0 = \mu^1 + Zu + Ke, \text{ for some } K. \]

Use Method 1 on \( z \) just as if it were \( y \) without fixed effects.

**Adjustments:** To coefficients of \( \sigma^2_e \) in E(SS's), to account for \( K \).

**Condition:** No interactions, fixed-by-random.

**History:**
- Henderson [1953]: first described, but not clearly.
- Searle [1968]: generalized, clarified, decried as not invariant.
- Henderson, Searle and Schaeffer [1974]: established invariance, and described computing procedure.

Use $R(\cdot)$'s of fitting constants for fixed effects models

$$E R(\alpha, \gamma | \mu, \beta) = c_1 \sigma^2 + c_1 \gamma^2 + (s-b) \sigma^2_e$$

$$E R(\gamma | \mu, \alpha, \beta) = c_2 \gamma^2 + (s-a-b+1) \sigma^2_e$$

$$E \text{SSE} = (N-s) \sigma^2_e$$

or, if no interaction

$$E R(\alpha | \mu, \beta) = c_1 \sigma^2 + (a-1) \sigma^2_e$$

$$E \text{SSE} = (N-a-b+1) \sigma^2_e$$

Properties: Unbiased.

Reduces to ANOVA for balanced data.

Difficulties: Can be difficult to compute (i.e., inverting large matrices).

Not uniquely defined: can have more equations than variance components, e.g., for 2-way random model, can use

$$R(\alpha | \mu) \quad R(\beta | \mu) \quad R(\beta | \mu, \alpha)$$

$$R(\beta | \mu, \alpha) \quad \text{OR} \quad R(\alpha | \mu, \beta) \quad \text{OR} \quad R(\alpha | \mu, \beta)$$

$$R(\gamma | \mu, \alpha, \beta) \quad R(\gamma | \mu, \alpha, \beta) \quad R(\gamma | \mu, \alpha, \beta)$$

$$\text{SSE} \quad \text{SSE} \quad \text{SSE}$$

$$y'y - N \bar{y}^2 \quad y'y - N \bar{y}^2$$

History: Henderson [1953]: described method.
Rohde and Tallis [1969]: give general expressions for sampling variances.
Searle [1971a, b]: discusses in detail.
II.4. A general linear model

**Equation of the model**

The general linear model can be represented as

\[ y = X\alpha + Zb + e = X\alpha + \sum_{i=1}^{\text{q}} Z_i b_i + e \]  \hfill (1)

where

- \( y_{\text{nxl}} \) is a vector of \( N \) observations,
- \( X_{\text{nxp}} \) is a known matrix, of rank \( p^* \leq p < N \),
- \( \alpha_{\text{pxl}} \) is a vector of \( p \) fixed effects parameters,
- \( Z_{\text{nxq}} \) is a known matrix,
- \( b_{\text{qxl}} \) is a vector of random effects, and
- \( e_{\text{nxl}} \) is a vector of residual errors.

The second equality in (1) comes from the partitioning

\[ b' = [b'_{1} b'_{2} \cdots b'_{c}] \quad \text{and} \quad Z = [Z_{1} Z_{2} \cdots Z_{c}] \]  \hfill (2)

where \( b_i \) has order \( q_i \times 1 \) and is the vector of \( q_i \) effects corresponding to \( q_i \) levels of the \( i'\text{th} \) random factor (main effect or interaction factor) in the model, with

\[ q = \sum_{i=1}^{c} q_i. \]

**Distributional properties**

\[ E(b_i) = 0, \quad E(e) = 0, \quad \text{and} \quad E(y) = X\alpha \]  \hfill (3)

\[ \text{var}(b_i) = \sigma^2 I_{q_i}, \quad \text{cov}(b_i, b_j') = 0, \quad i \neq j, \]  \hfill (4)

\[ \text{var}(b) = D = \text{diag}[\sigma^2 I_{q_1} \cdots \sigma^2 I_{q_c}] \]

\[ \text{var}(e) = R, \quad \text{cov}(b_i e') = 0, \]  \hfill (5)

\[ \text{var}(y) = V = ZDZ' + R. \]

It is customary to have \( R = \sigma^2 I. \)
II.5. MME: Henderson's mixed model equations

In the mixed linear model (1), equations designed for estimating the fixed effects $\alpha$ and for predicting the random effects $b$ also have uses in calculating estimates of variance components. They have computational uses, but provide no new estimation procedure.

The generalized least squares (GLS, or Aitken) equations for $\alpha$ are $X'V^{-1}X\hat{\alpha} = X'V^{-1}y$. If $b$, instead of representing random effects, were to represent fixed effects the normal equations for $\alpha$ and that $b$ would be

$$
\begin{bmatrix}
X'R^{-1}X & X'R^{-1}Z \\
Z'R^{-1}X & Z'R^{-1}Z
\end{bmatrix}
\begin{bmatrix}
\alpha \\
b
\end{bmatrix}
= 
\begin{bmatrix}
X'R^{-1}y \\
Z'R^{-1}y
\end{bmatrix}.
$$

(6)

The MME's are (6) adapted by adding $D^{-1}$ to $Z'R^{-1}Z$:

$$
\begin{bmatrix}
X'R^{-1}X & X'R^{-1}Z \\
Z'R^{-1}X & Z'R^{-1}Z + D^{-1}
\end{bmatrix}
\begin{bmatrix}
\alpha \\
b
\end{bmatrix}
= 
\begin{bmatrix}
X'R^{-1}y \\
Z'R^{-1}y
\end{bmatrix}.
$$

(7)

These have also been written by Harville [1975, 1977] as

$$
\begin{bmatrix}
X'R^{-1}X & X'R^{-1}ZD \\
Z'R^{-1}X & Z'R^{-1}ZD + I
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\nu
\end{bmatrix}
= 
\begin{bmatrix}
X'R^{-1}y \\
Z'R^{-1}y
\end{bmatrix}.
$$

(8)

with $\bar{b} = D\nu$.

A special case: For the special case of just a single random factor, i.e., $c = 1$, $b = b_1$, $D = \sigma_1^2 I$ and $R = \sigma^2 I$,

the equations (7) become
\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & D[n_1] + (\sigma^2/\sigma_1^2)I
\end{bmatrix}
\begin{bmatrix}
\hat{\beta} \\
\hat{\gamma}
\end{bmatrix}
= \begin{bmatrix}
X'y \\
Z'y
\end{bmatrix}.
\] (7a)

Comments

(i) \(\hat{\alpha}\) of (7) and (8) equals the Aitken \(\hat{\alpha}\).

(ii) \(\hat{\alpha}\) is easier to compute than \(\hat{\alpha}\).

(iii) \(\tilde{\beta}\) of (7) is the BLUP (best linear unbiased predictor) of \(\tilde{\beta}\).

(iv) \(\tilde{\beta} = DZ'\tilde{V}^{-1}(y - X\tilde{\alpha})\), and with normality assumptions this is the ML estimator of \(E(\tilde{b}|y)\). In genetics, \(\tilde{\beta}\) is used for estimating genetic values.

(v) In equations (7) the partitioned matrix on the l.h.s. is symmetric and p.s.d. whereas that in (8) is not.

(vi) Equations (8) can be used with non-singular \(D\) whereas (7) cannot.

(vii) Define \(\tilde{T}^* = (Z'R^{-1}ZD + I)^{-1} = (Z'D/\sigma^2 + I)^{-1} = \{T^*_{ij}\} \text{ for } i,j = 1, \cdots, c.\) (9)

(viii) Define the lower right sub-matrix of the generalized inverse of the partitioned matrix on the l.h.s. of (8) as \(\tilde{T}^{-1}\). Then

\[
\tilde{T} = [I + Z'[R^{-1} - R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1}]ZD]^{-1} = \{T_{ij}\} \text{ for } i,j = 1, \cdots, c.\) (10)

(ix) \(\tilde{T}^*\) and \(\tilde{T}\) can be used in computing ML, REML, and I-MINQUE estimators of \(\tilde{\sigma}^2\).

History: Henderson et al. [1959] and Henderson [1963]: developed MME.
Lindley and Smith [1972]: equations arise in Bayesian setting.
Henderson [1973]: MME related to MINQUE and ML.
Harville [1976]: renewed attention.
II.6. Thompson's iterative method

Models with only 1 random factor; e.g., 2-way classification without interaction

\[ y_{ijk} = \mu + \alpha_i + b_j + e_{ijk} \quad i = 1, \ldots, p \quad j = 1, \ldots, q \]

\[ y = X\alpha + Zb + e \]

**Fitting constants method**: Based on

\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & Z'Z
\end{bmatrix}
\begin{bmatrix}
\alpha^0 \\
b^0
\end{bmatrix}
= \begin{bmatrix}
X'y \\
Z'y
\end{bmatrix}
\text{ and } R(\mu, \alpha, b) = (\alpha^0', b^0') \begin{bmatrix}
X'y \\
Z'y
\end{bmatrix}
\]

\[
\hat{\sigma}^2_0 = \frac{y'y - R(\mu, \alpha, b)}{N - p^2 - q_1 + 1}
\hat{\sigma}^2_1 = \frac{R(b'\mu, \alpha) - (q_1 - 1)\hat{\sigma}^2_e}{N - \Sigma \Sigma n_i / n_j}
\]

Cunningham and Henderson [1968]: Use mixed model equations (7a) to calculate

\[
R^*(\mu, \alpha, b) = \begin{bmatrix}
\alpha^* \\
b^*
\end{bmatrix}
\begin{bmatrix}
X'y \\
Z'y
\end{bmatrix}
\]

and get

\[
\hat{\sigma}^2_0 = \frac{y'y - R^*(\mu, \alpha, b)}{N - p^2 - q_1 + 1}
\hat{\sigma}^2_1 = \frac{R^*(\mu, \alpha, b) - R(\mu, \alpha) - (q_1 - 1)\hat{\sigma}^2_e}{N + p^2 (\hat{\sigma}^2_0 / \hat{\sigma}^2_1) - \Sigma \Sigma n_i / n_j}
\]

then iterate on \( \hat{\sigma}^2_0 / \hat{\sigma}^2_1 \).

Thompson's [1969] method: Located error in expectations of Cunningham and Henderson; correction yields

\[
\hat{\sigma}^2_0 = \frac{y'y - R^*(\mu, \alpha, b)}{N - q_1}
\hat{\sigma}^2_1 = \frac{R^*(\mu, \alpha, b) - R(\mu, \alpha)}{N - \Sigma \Sigma n_i / n_j}
\]

iterate on \( \hat{\sigma}^2_0 / \hat{\sigma}^2_1 \).

Computing formulae:

2-way, no interaction: Searle [1973].

2-way, with interaction (i.e., 2 random factors): Corbeil and Searle [1973].
PART III. ESTIMATING VARIANCE COMPONENTS FROM UNBALANCED DATA IN MIXED MODELS:
MAXIMUM-LIKELIHOOD STYLE METHODS

III.0. Introduction

For unbalanced data (having unequal numbers of observations in the subclasses), Henderson [1953] is a landmark paper with its three methods of estimation based on the principle of equating quadratic forms to their expected values. Succeeding years saw expansion and explanation of these methods together with exploration of their properties, but there were no really new developments until Hartley and Rao [1967] described maximum likelihood (ML) procedures - based, as is so often the case, on normality assumptions. Since then there has been a whole host of new methods, not only ML, but REML, MINQUE, I-MINQUE, and MIVQUE - and doubtless some other alphabetic horrors also. In addition, there are peripheral topics tangential to computing techniques - such as Henderson's MME's (mixed model equations) and the Dispersion-mean model suggested by Pukelsheim [1976b]. As foundation for all this there is a large corpus of matrix algebra, there are numerous notations that look sufficiently alike to add the traditional amount of confusion and, hanging like a thunder cloud over everything, are numerical and computing problems involved with very large data sets, sparse matrices, and the solving of non-linear equations subject to non-linear (non-negativity) constraints.

Obviously, a whole tome would be needed to deal thoroughly with all three aspects of the subject: description of each method, details of the underlying algebra, and the computing algorithms. Attention is confined here to just description of the methods in note form. The prime purpose is to give, in summary form, the basic rationale and methodology for each of the estimation procedures considered; and to do this with a unifying notation, and to show relationships between the methods. There are voluminous details of underlying algebra and equivalent expressions for each method, which will all be available in Searle and Quaas [1978]; and computing algorithms are left to others.
Ancillary notation for the general model

\[ R = \sigma^2 I_N \quad \text{and} \quad H = V/\sigma^2 \]

\[ b_0 = e, \quad q_0 = N \quad \text{and} \quad Z_0 = I_N \]

\[ \sigma^2 = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_c^2 \end{bmatrix} \]

\[ \sigma^2 = \begin{bmatrix} \sigma_0^2 & \sigma_1^2 & \cdots & \sigma_c^2 \end{bmatrix} = \begin{bmatrix} \sigma_0^2 & \sigma^2 \end{bmatrix} \]

\[ b' = \begin{bmatrix} b_0' & b_1' & \cdots & b_c' \end{bmatrix} = \begin{bmatrix} b_0' & b' \end{bmatrix} \]

\[ Z = \begin{bmatrix} Z_0 & Z_1 & \cdots & Z_c \end{bmatrix} = \begin{bmatrix} Z_0 & Z \end{bmatrix} \]

\[ \hat{D} = \text{diag}(\sigma_0^2 q_0, \sigma_1^2 q_1, \ldots, \sigma_c^2 q_c) = [\sigma^2 I_N, D] \]

\[ y^* = X \alpha + Z b \]

\[ V = Z D Z^{\dagger} + R = \tilde{Z} \tilde{D} \tilde{Z}^{\dagger} \]

\[ P = V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1} \]
III.1. ML: Maximum Likelihood

Rationale

Assume normality of random effects with moments as in (3), (4) and (5), and maximize logarithm of likelihood of $y$:

$$L_y = \frac{1}{2} N \log 2\pi - \frac{1}{2} \log |V| - \frac{1}{2} (y - \bar{X}\alpha)'V^{-1}(y - \bar{X}\alpha)$$  \hspace{1cm} (11)

Method

Maximizing $L_y$ with respect to $\alpha$ and to the $\sigma^2$'s implicit in $V$ leads to equations in $\alpha$ and $V$:

$$X'\tilde{V}^{-1}\tilde{X}\alpha = X'\tilde{V}^{-1}y$$  \hspace{1cm} (12)

and

$$\text{tr}(\tilde{V}^{-1}Z_iZ_i') = (y - \bar{X}\alpha)'\tilde{V}^{-1}Z_iZ_i'(y - \bar{X}\alpha)$$  \hspace{1cm} (13)

for $i = 0, 1, \ldots, c$. We use the notation $\hat{\alpha}$ for a solution to (12), because (12) is the same as the Aitken equations with $\tilde{V}$ in place of $V$. Solutions to (12) and (13) restricted by $\sigma^2_0 > 0$ and $\sigma^2_i > 0$ for $i = 1, \ldots, c$ are ML estimators.

Equivalent equations, given by Hartley and Rao [1967] are

$$X'\tilde{H}^{-1}\tilde{X}\alpha = X'\tilde{H}^{-1}y$$  \hspace{1cm} (14)

$$\sigma^2_0 = (y - \bar{X}\alpha)'\tilde{H}^{-1}(y - \bar{X}\alpha)/N$$  \hspace{1cm} (15)

$$\text{tr}(\tilde{H}^{-1}Z_iZ_i') = (y - \bar{X}\alpha)'\tilde{H}^{-1}Z_iZ_i'(y - \bar{X}\alpha)/\sigma^2_0$$  \hspace{1cm} (16)

for $i = 1, 2, \ldots, c$. An alternative form of (13) is

$$\{\text{tr}(\tilde{V}^{-1}Z_iZ_i'\tilde{V}^{-1}Z_jZ_j')\}^{\tilde{2}} = \{y'\tilde{Z}_i\tilde{Z}_j'\tilde{V}y\}$$  \hspace{1cm} (17)

III.2. REML: Restricted Maximum Likelihood

Rationale

Assume normality, as with ML, but maximize only that portion of the log likelihood which is invariant to E(y). This is equivalent to maximizing

$$L_1 = \text{constant} - \frac{1}{2} \log |V| - \frac{1}{2} \log |X^k V^{-1} X^k| - \frac{1}{2} y' P y$$  \hspace{1cm} (18)

for \(X^k\) being any \(p^{1'}\) (rank of \(X\)) linearly independent columns of \(X\).

Method

Maximizing \(L_1\) with respect to the \(\sigma^2\)'s implicit in \(V\) (and hence in \(P\), also) leads to equations

$$\text{tr}(\hat{P}Z_i Z_i') = y' \hat{P}Z_i Z_i' \hat{P} y \text{ for } i = 0, \ldots, c$$  \hspace{1cm} (19)

equivalent to

$$\{\text{tr}(\hat{P}Z_i Z_j Z_j')\}_{i,j} = \{y' \hat{P}Z_i Z_i' \hat{P} y\} \text{ for } i, j = 0, 1, \ldots, c.$$  \hspace{1cm} (20)

Define \(^\hat{\alpha}\) by

$$X' V^{-1} \hat{\alpha} = X' V^{-1} y.$$  \hspace{1cm}

Then equation (19) leads to

$$\sigma^2 = (y - X\hat{\alpha})' \hat{H}^{-1} (y - X\hat{\alpha}) / (N - p^k).$$  \hspace{1cm} (21)

Solutions for \(^\hat{\sigma}_0^2\) and \(^\hat{\sigma}_i^2\), restricted to \(^\hat{\sigma}_0^2 > 0\) and \(^\hat{\sigma}_i^2 \geq 0\) for \(i = 1, \ldots, c\) are REML estimators.
An equivalent procedure

Maximize the log likelihood of $K'y$ where $K'$ is any matrix of full row rank

such that $K'X = 0$:

$$L(K'y) = \frac{1}{2}(N - p^*)\log 2\pi - \frac{1}{2}\log |K'VK| - \frac{1}{2}y'K(K'VK)^{-1}K'y. \quad (22)$$

Comments

(i) Elements of $K'y$ are called "error contrasts", Harville [1977].

(ii) $K(K'VK)^{-1}K' = P$.

(iii) $L(K'y) - L_1$ = constant not dependent on $\sigma$ or $\sigma^2$.

(iv) Corbeil and Searle [1976a] use a special form of $K'$: define the fixed effects part of the model as the vector of cell means of those sub-most cells of the fixed effects factors that contain data, $k$ such cells, say, with $n_1$, $\ldots$, $n_k$ observations. Then delete the $n_1$'th, $(n_1 + n_2)$'th, $\ldots$, $(n_1 + n_2 + \cdots + n_k)$'th rows from $I_N - \text{diag}\{J_{n_1}, \ldots, J_{n_k}\}$, and the matrix remaining is a $K'$. It is called $T$ by Searle and Corbeil.

History: Thompson [1962]: initial idea.

Patterson and Thompson [1971]: unbalanced data, largely b.i.b. designs.

Harville [1974]: shows results for any error contrasts.

Hocking and Kutner [1975]: simulate results on b.i.b. designs.

Harville [1975, 1977]: gives $L_1$ and makes comprehensive review.

Corbeil and Searle [1976a]: use a particular $K$ (T in their notation).

Corbeil and Searle [1976b]: give analytic comparisons of REML and ML for balanced data, and numeric comparisons for unbalanced data.
III.3. MINQUE: Minimum Norm Quadratic Unbiased Estimation

Rationale

Estimate \( p' \hat{\sigma}^2 \), a linear function of the \( \sigma^2 \)'s by a quadratic form \( y' Ay \), choosing \( A \) so that the estimator \( y' Ay \)

(i) is translation invariant: \( \alpha - \alpha + 5 \) does not alter \( y' Ay, = AX = 0. \)

(ii) is unbiased: \( E(y' Ay) = p' \sigma^2 \).

(iii) minimizes the weighted norm: \[
\| D^2 (Z'AZ - \Delta) D^2 \|
\] (23)

where, for arbitrary weights \( w_i \) and for \( r_i = p_i/q_i \) for \( i = 0, 1, \ldots, c, \)

\[
\hat{b}_w = \text{diag} \{ w_i I_{q_i} \ldots w_c I_{q_c} \} \quad \text{and} \quad \hat{\Delta} = \text{diag} \{ r_i I_{q_i} \ldots r_c I_{q_c} \}. \quad (24)
\]

Were the elements of \( b \) actually known, a "natural estimator" of \( p' \sigma^2 \) would be

(Rao [1972])

\[
\sum_{i=0}^{c} \frac{p_i}{q_i} \frac{b_i^2}{q_i} = \hat{b}' \hat{A} \hat{b}.
\]

Method

Minimizing (23) can be shown equivalent to minimizing \( \text{tr}(AV_w)^2 \). (25)

This leads to solving the equations

\[
\{ \text{tr}(P \hat{Z}_i \hat{Z}_i' P \hat{Z}_j \hat{Z}_j') \hat{b}_w^2 \} \approx \{ y' P \hat{Z}_i \hat{Z}_i' P \hat{Z}_j \hat{Z}_j' y \} \quad \text{for } i, j = 0, \ldots, c, \quad (26)
\]

where

\[
V_w = \hat{Z}_i \hat{Z}_i' \quad \text{and} \quad P = V_w^{-1} - V_w^{-1}X'(V_w^{-1}X)X'V_w^{-1} \cdot (27)
\]

Comments

(i) No distributional properties assumed.

(ii) No iteration: for given \( w_i \), solve (26).

(iii) Special cases

(a) \( w_i = 1, \) for all \( i: V_w = \sum_{i=0}^{c} Z_i Z_i' \).

(b) \( w_0 = 1, w_i = 0 \) for \( i \geq 1: V_w = I, \) and \( P = I - X(X'X)^{-1}X' \).
III.4. I-MINQUE: Iterative MINQUE

Rationale

The occurrence of \( w \)'s in \( V \) of (27) is exactly the same as that of \( \sigma^2 \)'s in \( V \) of (9). This prompts the idea of using the MINQUE equations iteratively — using \( \hat{\sigma}^2 \) as the \( w \) for a succeeding set of equations, which are then solved for \( \hat{\sigma}^2 \) — and so on.

Method

Solve, iteratively, for \( i, j = 0, 1, \ldots, c \)

\[
\begin{align*}
\{\text{tr}(\hat{P}_L Z_i \hat{P}_L Z_j')\} \hat{\sigma}^2 &= \{y' \hat{P}_L Z_i \hat{P}_L y\} \\
\end{align*}
\]

(28)

III.5. MIVQUE: Minimum Variance Quadratic Unbiased Estimation

Rationale

The same as MINQUE, except that instead of minimizing a norm, one minimizes the variance of the estimator \( y' Ay \). Combined with translation invariance and unbiasedness this involves minimizing

\[
\nu(y' Ay) = 2\text{tr}(AV)^2 + \sum_{i=0}^{c} \gamma_i \sigma_i^4 \lambda_i
\]

(29)

for \( \lambda_i = \text{sum of squares of diagonal elements of } Z_i' A Z_i' \), and where \( \gamma_i \) is the kurtosis parameter for \( b_i \) (see, e.g., Anderson et al. [1977]). Under normality this is equivalent to

minimizing \( \text{tr}(AV)^2 \)

(30)

similar to (25).

Method

Solve the equations, for \( i, j = 0, 1, \ldots, c \)

\[
\begin{align*}
\{\text{tr}(\hat{P}_L Z_i \hat{P}_L Z_j')\} \hat{\sigma}^2 &= \{y' \hat{P}_L Z_i \hat{P}_L y\} \\
\end{align*}
\]

(31)

History:

III.6. Dispersion-Mean Models

Rationale

Estimate \( \sigma_i^2 \) unbiasedly by \( y_i'KAK'y_i \) for \( K'y_i \) being error contrasts. This leads to the following models.

The distribution of

\[
y = Ky \otimes Ky
\]

is

\[
y \sim (\psi^2, \Omega)
\]

where

\[
y = K'y \otimes K'y \quad \text{and} \quad \chi = (K' \otimes K')\psi \quad \text{with} \quad \psi = [\text{vec}(Z_1Z_1') \cdots \text{vec}(Z_cZ_c')].
\]

Also,

\[
\Omega = (K' \otimes K')F(K \otimes K)
\]

with

\[
F = (\psi \otimes \psi)(I_{\psi^2} + I(N,N)) + (\psi \otimes \psi)[\text{diag} \{ \text{vec} \{ \text{diag}(\gamma_0 \sigma_0^2 \mathcal{I} \cdots \gamma_c \sigma_c^2 \mathcal{I}) \} \}](\mathcal{Z}' \otimes \mathcal{Z}').
\]

An alternative model (Anderson [1978]) is

\[
y_1 = (y - x\alpha) \otimes (y - x\alpha)
\]

with

\[
y_1 \sim (\psi^2, F).
\]

Methods

Generalized least squares, GLS (replacing \( V \) by a pre-assigned \( \psi_0 \)), and ordinary least squares, OLS, on (33) and (36) lead to MINQUE and ML.

PART IV. RELATIONSHIPS AMONG METHODS

IV.1. General Results

(1) Henderson 1 = ANOVA (definition).

(2) Henderson 3 = Fitting constants (definition, but not unique).

(3) ML (REML) estimators are ML (REML) solutions subject to non-negativity conditions.

IV.2. Balanced Data

(4) ANOVA = Henderson 2 = Henderson 3 = REML = MINQUE (MIVQUE under normality).

(5) Some ML equations have no explicit solution. When solutions exist, some are ANOVA (differences occur in "degrees of freedom").

IV.3. Unbalanced Data

(6) ML and REML

\[ \text{ML: } \{ \text{tr} \left( \tilde{V}^{-1} \tilde{Z}_i \tilde{Z}'_i \tilde{V}^{-1} \tilde{Z}_j \tilde{Z}'_j \right) \} \tilde{e}^2 = \{ y' \tilde{P}_i \tilde{Z}_i \tilde{P}_j \tilde{Z}_j \tilde{P}_y \} \]  \hspace{1cm} (17)

\[ \text{REML: } \{ \text{tr} \left( \tilde{P}_i \tilde{Z}_i \tilde{P}_j \tilde{Z}_j \tilde{P}_y \right) \} \tilde{e}^2 = \{ y' \tilde{P}_i \tilde{Z}_i \tilde{P}_j \tilde{Z}_j \tilde{P}_y \} \]  \hspace{1cm} (20)

Note: $\tilde{V}^{-1}$ in the l.h.s. for ML is replaced by $\tilde{P}$ in REML.

\[ \text{ML: } \tilde{e}_0^2 = (\tilde{y} - \tilde{X}\hat{\alpha})'N^{-1}(\tilde{y} - \tilde{X}\hat{\alpha})/N \]  \hspace{1cm} (15)

\[ \text{REML: } \tilde{e}_0^2 = (\tilde{y} - \tilde{X}\hat{\alpha})'N^{-1}(\tilde{y} - \tilde{X}\hat{\alpha})/(N - \hat{p}) \]  \hspace{1cm} (21)

Note: Degrees of freedom for fixed effects are taken account of in REML.
(7) **REML and MINQUE**

A first iterate from REML is a MINQUE.

If the same arbitrary value is used initially for \( \hat{\sigma}^2 \) in \( P \) in the REML equations (20) as is used for \( w \) in \( P \) in the MINQUE equations (26), the resulting first iterate from REML will be identical to the MINQUE.

(8) **REML, I-MINQUE, and MIVQUE under normality**

Estimators from REML, I-MINQUE, and MIVQUE under normality are the same.

Equations (20), (28) and (31) are identical.

(9) **ML and MME**

Using \( \hat{\alpha} \) and \( \hat{b} \) from the MME (7)

\[
\hat{\sigma}^2_0 = y^\prime (y - \hat{\alpha} \hat{x} - \hat{z} b) / N
\]

and with \( T^* \) of (9)

\[
\hat{\sigma}^2_1 (r+1) = \frac{\hat{b}_i \hat{b}_j (r) + \hat{\sigma}^2_1 \text{tr}(T^*_i (r))}{q_i}
\]

(37)

\[
\hat{\sigma}^2_1 (r+1) = \frac{\hat{b}_i \hat{b}_j (r)}{q_i - \text{tr}(T^*_i (r))}
\]

(38)

Note: Equation (38) always yields positive estimates.

(10) **REML and MME**

\[
\hat{\sigma}^2_0 = y^\prime (y - \hat{x} \hat{a} - \hat{z} \hat{b}) / (N - p^*)
\]

and

\[
\hat{\sigma}^2_1 (r+1) = \frac{\hat{b}_i \hat{b}_j (r) + \hat{\sigma}^2_1 \text{tr}(T_i (r))}{q_i}
\]

(39)

\[
\hat{\sigma}^2_1 (r+1) = \frac{\hat{b}_i \hat{b}_j (r)}{q_i - \text{tr}(T_i (r))}
\]

(40)

Note: Equation (40) always yields positive estimates.

Equations (39) and (40) are (37) and (38) with \( T_{ii} \) in place of \( T^*_i \).
Equations for REML, MINQUE, I-MINQUE and MIVQUE under normality, namely (20), (26), (28) and (31), respectively, all have the form

\[
\{\text{tr}(P_{ij} Z_{ij})\}_{ij}^{\bar{P}} = \{y' P_{ij} Z_{ij}^2 y\} \text{ for } i,j = 0, 1, \ldots, c.
\]

These are equivalent to

\[
\begin{bmatrix}
\text{tr}(P^{\bar{P}}) & \{\text{tr}(P_{ij}^2 Z_{ij}^2)\} \\
\{\text{tr}(P_{ij}^2 Z_{ij}^2)\} & \{\text{tr}(P_{ij}^2 Z_{ij}^2)\}
\end{bmatrix}
\begin{bmatrix}
\bar{P}^2 \\
\bar{P}^2
\end{bmatrix}
= \begin{bmatrix}
y' P_{ij}^2 y \\
y' P_{ij}^2 y
\end{bmatrix} \text{ for } i,j = 1, \ldots, c.
\]

The terms in this equation can be expressed as functions of sub-matrices of \(T_{ij} \) of (10) as follows:

\[
\text{tr}(P^{\bar{P}}) = \frac{N - r}{\sigma_0^2} + \frac{1}{2} \sum_{i=1}^{c} \sum_{j=1}^{c} \text{tr}(T_{ij} T_{ji})
\]

\[
\text{tr}(P_{ij}^2 Z_{ij}^2) = \frac{1}{\sigma_0^2 \sigma_i^2} \left[ \text{tr}(T_{ii}) - \sum_{j=1}^{c} \text{tr}(T_{ij} T_{ji}) \right] \text{ for } i = 1, \ldots, c
\]

\[
\text{tr}(P_{ij}^2 Z_{ij}^2) = \frac{q_i}{\sigma_0^2} - \frac{2}{\sigma_i^2} \text{tr}(T_{ii}) + \frac{1}{\sigma_i^2} \text{tr}(T_{ii})^2 \text{ for } i = 1, \ldots, c
\]

\[
\text{tr}(P_{ij}^2 Z_{ij}^2) = \frac{1}{\sigma_i^2 \sigma_j^2} \text{tr}(T_{ij} T_{ji}) \text{ for } i \neq j = 1, \ldots, c
\]

\[
y' P_{ij}^2 y = (y'y - \tilde{a}'X' y - \tilde{b}'Z' y)/\sigma_0^2 - \frac{1}{2} \sum_{i=1}^{c} \frac{b_i v_i}{\sigma_i^2}
\]

\[
y' P_{ij}^2 Z_{ij}^2 y = \tilde{b}_i v_i / \sigma_i^2 \text{ for } i = 1, \ldots, c.
\]
(12) Dispersion-Mean Models, MINQUE and ML

(i) GLS on (33), with \( \gamma_0 \) replacing \( \gamma \), gives MINQUE.

(ii) GLS on (36), with \( \gamma_0 \) replacing \( \gamma \) and with \( \alpha \) replaced by
\[
\hat{\alpha} = (X' \gamma_0^{-1} X)^{-1} X' \gamma_0^{-1} y
\]
gives, on iterating, ML.

(iii) OLS on (33) gives MINQUE with a \( \gamma_0 \) of \( \sigma^2 I \).

History:

Patterson and Thompson [1971]: first indication of the \( y'PZ\bar{Z}'Py \) result.

Henderson [1973]: extended results, MME's, ML and MINQUE.

LaMotte [1973]: indicated results for REML and MINQUE.

Schaeffer [1975]: published some details of MME's and MINQUE, with many misprints.

Harville [1975, 1977]: comprehensive review.

Searle and Quaas [1978]: complete details given.
REFERENCES
(for Parts II, III and IV)


Patterson, H. D. and Thompson, R. [1971]. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545-554.


