

Some new results concerning estimation and prediction
in mixed models¹⁾

Shayle R. Searle²⁾

Biometrics Unit, Cornell University, Ithaca, N.Y., U.S.A.

and

Institut für Mathematik, Universität Augsburg

D-8900 Augsburg, Federal Republic of Germany

BU-873-M

April 1985

Abstract

Three contributions to estimation and prediction in mixed models of the analysis of variance are described. First is a proof that for a broad class of equal-subclass-numbers data the best linear unbiased estimator (BLUE) of an estimable function of fixed effects is, in all mixed models, the same as the familiar ordinary least squares estimator. Second, are explicit expressions for the BLUE of treatment means in randomized complete blocks with unequal numbers of observations on the treatments in each block: a special case is balanced incomplete blocks. Third, is demonstration that the BLU predictor of random effects in a mixed model is always of the class of predictors that is invariant to the fixed effects; and other properties of BLUP are illustrated.

1) Prepared for the Pacific Statistical Congress, May 20-24, 1985, in Auckland, New Zealand.

2) Supported by a U.S. Senior Scientist Award, at the University of Augsburg, from the Alexander von Humboldt-Stiftung, January-April, 1985 and 1986.

1. Introduction

Linear models associated with analyses of variance traditionally involve parameters that we want to estimate - either as they stand, or in terms of linear combinations of them. Point and/or interval estimates may be wanted - and hypotheses tests may also be of interest. We restrict attention to point estimation.

The simplest example is a between- and within-groups analysis of variance, where y_{ij} is the j 'th observation in the i 'th group, for a groups and n_i observations in the i 'th group so that $i = 1, 2, \dots, a$, and $j = 1, 2, \dots, n_i$. A familiar model is

$$E(y_{ij}) = \mu + \alpha_i \quad (1)$$

where E represents expectation over repeated sampling; μ is a general mean and α_i is the effect on the response variable of its being from the i 'th group. Then we define a random error term e_{ij} as

$$e_{ij} = y_{ij} - E(y_{ij}), \quad \text{with} \quad E(e_{ij}) = 0 \quad (2)$$

so giving

$$y_{ij} = \mu + \alpha_i + e_{ij}. \quad (3)$$

To the e_{ij} 's we also attribute the variance-covariance properties

$$v(e_{ij}) = \sigma^2 \quad \text{and} \quad \text{cov}(e_{ij}, e_{hk}) = 0, \quad (4)$$

for all i, j, h and k , except $i = h$ and $j = k$ together. Under these conditions μ and α_i are parameters that we are interested in estimating; they are called fixed effects and the model represented by (3) and (4) is called a fixed effects model.

Estimation of fixed effects is usually by means of ordinary least squares (OLSE) or best linear unbiased estimation (BLUE). In the case of fixed effects models these two methods yield the same estimators of estimable functions: e.g., for the model just described

$$\text{OLSE}(\alpha_i - \alpha_h) = \text{BLUE}(\alpha_i - \alpha_h) = \bar{y}_{i\cdot} - \bar{y}_{h\cdot}, \quad (5)$$

where

$$\bar{y}_{i\cdot} = \sum_{j=1}^{n_i} y_{ij} / n_i. \quad (6)$$

And use is often made of

$$\text{BLUE} \left[\frac{\sum_{i=1}^a n_i (\mu + \alpha_i)}{\sum_{i=1}^a n_i} \right] = \bar{y}_{..} \quad \text{and} \quad \text{BLUE} \left[\frac{\sum_{i=1}^a (\mu + \alpha_i)/a}{\sum_{i=1}^a 1/a} \right] = \sum_{i=1}^a \bar{y}_{i.}/a, \quad (7)$$

where

$$\bar{y}_{..} = \frac{\sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij}}{\sum_{i=1}^a n_i} = \frac{\sum_{i=1}^a n_i \bar{y}_{i.}}{\sum_{i=1}^a n_i}. \quad (8)$$

In contrast to fixed effects models are mixed models, where some parameters of a fixed effects model instead of being treated as fixed constants to be estimated are taken as being random variables. For example, in (1) the α_i s would be taken as random variables. Then (2) and (3) are replaced by

$$E(y_{ij}) = \mu \quad (9)$$

and

$$e_{ij} = y_{ij} - E(y_{ij} | \alpha_i) = y_{ij} - \mu - \alpha_i \quad \text{and so} \quad E(e_{ij}) = 0. \quad (10)$$

The random variables $\alpha_1, \dots, \alpha_a$ are called random effects; and to them are attributed the following properties:

$$E(\alpha_i) = 0, \quad v(\alpha_i) = \sigma_\alpha^2 \quad (11)$$

$$\text{cov}(\alpha_i, \alpha_h) = 0 \quad \text{and} \quad \text{cov}(\alpha_i, e_{hk}) = 0 \quad (12)$$

for all i, h and k , save that $\text{cov}(\alpha_i, \alpha_i) = v(\alpha_i) = \sigma_\alpha^2$. This model, represented by (3), (4), and (9)-(12), is an example of what is called a mixed model because it consists of a mixture of fixed effects (just one in this case, μ), and random effects, the α_i s. Traditionally models in which all terms save μ are random are called random models - but since they do include μ as a fixed effect they can also be called mixed models. The variances in mixed models such as σ^2 and σ_α^2 of (4) and (11) are called variance components; in this case $\sigma_y^2 = \sigma_\alpha^2 + \sigma^2$. It is in this context that models where μ is the only fixed effect are also referred to as variance components models.

The change from a fixed effects model to a mixed model does, in many situations, bring to the problem of estimating fixed effects complications that do not occur with fixed effects models. For example, for the mixed model that uses (9)-(12), the OLSE and BLUE of μ are not the same:

$$\text{OLSE}(\mu) = \bar{y}_{..} \quad (13)$$

and

$$\text{BLUE}(\mu) = \frac{\sum_{i=1}^a \frac{n_i}{\sigma^2 + n_i \sigma_\alpha^2} \bar{y}_{i.}}{\sum_{i=1}^a \frac{n_i}{\sigma^2 + n_i \sigma_\alpha^2}} \quad (14)$$

Not only are these estimators not equal, but also $\bar{y}_{..}$ as an estimator is playing a different role in the mixed model from its role in the fixed effects model. Here, in (13), $\bar{y}_{..}$ in the mixed model is the OLSE of μ . In (7), $\bar{y}_{..}$ in the fixed effects model is the BLUE of a mean of the $(\mu + \alpha_i)$ -terms weighted by the n_i 's.

One of the most noticeable features of (14) as an estimator of the fixed effect μ is that it involves the variance components σ^2 and σ_α^2 of the random effects. Thus, in order to calculate an estimate of μ from (14) we need estimates of these variance components. This raises the whole question of how to estimate variance components; it is a large topic, but is not the subject of this paper.

Another feature of (14) is that when every group has the same number of observations, i.e., $n_i = n \forall i$, then (14) reduces to $\text{BLUE}(\mu) = \bar{y}_{..}$, i.e., for balanced data (equal subclass numbers data), the BLUE and OLSE of μ are equal.

This is an important result and one which generalizes as follows.

Theorem. In a broad, definable class of balanced data, for every mixed model the BLUE of an estimable function of fixed effects is the same as its OLSE.

The theorem is important not only for its broad application but also for the simplicity it brings to the calculation of BLUEs in a broad class of balanced data. As an illustration, consider data from a 2-factor experiment with a customary model equation

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \quad (15)$$

In the fixed effects version of this model, with balanced data,

$$\text{BLUE}(\alpha_i - \alpha_h) = \bar{y}_{i..} - \bar{y}_{h..} \quad (16)$$

where $\bar{y}_{i..}$ is the mean of the data in the i 'th level of the α -factor. Now consider a mixed model version of (15), where the β_j 's and γ_{ij} 's, or just the

γ_{ij} s, are random effects. The theorem then shows that in this mixed model, (16) still holds. Indeed more than that: if $\mu + \alpha_i$ is the only fixed effects part of a mixed model then no matter how many random effects there are, the BLUE of $\alpha_i - \alpha_h$ still has the form of (16), namely the difference between the data means for the i'th and h'th levels of the α -factors; e.g., for

$$y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \eta_k + \delta_l + \tau_{jkl} + e_{ijklm},$$

where, on the right-hand side, everything except $\mu + \alpha_i$ is a random effect,

$$\text{BLUE}(\alpha_i - \alpha_h) = \bar{y}_{i\dots} - \bar{y}_{h\dots}.$$

This is the importance of the theorem. The broad class of balanced data to which the theorem applies is described and proof of the theorem is indicated, in Section 2.

The BLUE procedure for fixed effects in mixed models with unbalanced data does not generally lead to specific formulae for those estimators, such as in (14), for example. Section 3 gives the general formula, and shows two cases where specific expressions have been obtained. Both cases are randomized blocks experiments with block effects random: one is unequal numbers of observations on the treatments in the blocks, and the other is a special case of this, balanced incomplete blocks.

A feature of mixed models that is different from fixed effects models is that of predicting the realized but unobservable value of the random effects occurring in the data. For example, with the mixed model (3), (4) and (9) - (12), the best, linear, unbiased predictor (BLUP) of α_i is

$$\text{BLUP}(\alpha_i) = \frac{n_i \sigma_\alpha^2}{\sigma^2 + n_i \sigma_\alpha^2} (\bar{y}_{i\cdot} - \hat{\mu}) \quad (17)$$

where $\hat{\mu} = \text{BLUE}(\mu)$ of (14). The general BLUP procedure, of which (17) is an example, is summarized in Section 4, together with its properties and several different ways in which it can be derived.

2. BLUEs from balanced data with mixed models

2.1 Models with balanced data

We write the model equation for a vector of observations y as

$$y = X\beta + Zu = \sum_{d=1}^f X_d \beta_d + \sum_{q=1}^r Z_q u_q \quad (18)$$

where β and u are vectors of fixed and random effects, respectively. Each β_d and u_q is a subvector of β and u respectively, with corresponding incidence matrices X_d and Z_q that are sub-matrices of X and Z . The last u_q vector, namely u_r , is the vector of error terms, with $Z_r = I_N$ where N is the total number of observations, the order of y . Each β_d (and u_q , except u_r) is a vector of the effects corresponding to the levels of a fixed (random) effects factor, either a main effect or interaction factor. For the random effects vectors the variance-covariance properties are extensions of (11) and (12): $E(u_q) = 0$, $\text{var}(u_q) = \sigma_q^2 I$ of appropriate order, and $\text{cov}(u_q, u_{q'}) = 0 \forall q \neq q'$. Hence, defining V as $\text{var}(Y)$, we have

$$V = \text{var}(Y) = \sum_{q=1}^r \sigma_q^2 Z_q Z_q' \quad (19)$$

Suppose there are m main effect factors. The class of balanced data we deal with is where each "cell" of the data defined by one level of every main effects factor contains n observations, the same for every cell. Then each X_d and each Z_q of (18) is a Kronecker product (KP) of $m + 1$ matrices that are each an identity matrix I or a summing vector $\mathbf{1}$, of appropriate order. Special cases are $X_1 = \mathbf{1}_N$ and $Z_r = I_N$, but otherwise in the KP for each X_d and Z_q the $(m + 1)$ 'th matrix is $\mathbf{1}_n$ and for $t = 1, \dots, m$, the t 'th matrix corresponds to the main effects factor and is I when that factor is instrumental in defining the factor corresponding to X_d or Z_q ; otherwise it is $\mathbf{1}$. Details and examples of this formulation of balanced data are shown in Searle (1984). It is the same formulation as used by Smith and Hocking (1978), Searle and Henderson (1979), Seifert (1979), Khuri (1981) and Anderson et al (1984) for specific cases.

2.2 Estimation

With the model of (18) and (19), the OLSE estimator of the estimable function $\lambda'X\beta$ is

$$\text{OLSE}(\lambda'X\beta) = \lambda'X(X'X)^{-1}X'y \quad (20)$$

where $(X'X)^{-}$ is a generalized inverse of $X'X$. And the BLUE is

$$\text{BLUE}(\lambda'X\beta) = \lambda'X(X'V^{-1}X)^{-}X'V^{-1}y, \quad (21)$$

where V is assumed to be positive definite.

Zyskind (1967) has shown that (20) and (21) are equal if and only if there exists some Q such that

$$VX = XQ. \quad (22)$$

It is this result which provides the theorem that precedes (15). For the model (18) and (19) with balanced data as just described, a Q does exist such that (22) is satisfied. The proof consists of displaying an appropriate matrix Q . The essence of doing this is that, from (19), VX is a row of matrices of the form $\sum_{q=1}^r \sigma_q^2 Z_q Z_q' X_d$. Since each Z_q and X_d is a KP, so is $Z_q Z_q' X_d$; and the nature of the matrices in the KP that is $Z_q Z_q' X_d$ is such that that KP can be written as $X_d M_{qd}$ where M_{qd} can be prescribed. This leads to expressing VX as XQ for Q being a block diagonal matrix of matrices $\sum_{q=1}^r \sigma_q^2 M_{qd}$ for $d = 1, \dots, f$; thus (22) is satisfied and the theorem is proved.

3. Randomized blocks with unbalanced data

In contrast to balanced data, where BLUE and OLSE are the same, the use of mixed models with unbalanced data demands estimating estimable functions of fixed effects by the BLUE of (21). This is how the variance-covariance structure of the random effects gets taken into account in estimating the fixed effects. OLSE ignores this structure, and thus effectively ignores the random effects. Unfortunately, of course, with unbalanced data $\text{BLUE}(\lambda'X\beta)$ is seldom a nice analytic function. Usually one has to first estimate the variance components that make up the elements of V , replace those components in V by their estimates and so obtain \hat{V} ; and then calculate

$$\text{BLUE}(\lambda'X\beta)_{\hat{V}} = \lambda'X(X'\hat{V}^{-1}X)^{-}X'\hat{V}^{-1}y. \quad (23)$$

Fortunately, Kackar and Harville (1981, 1984) have found that for a wide range of estimators of the variance components (23) is unbiased for $\lambda'X\beta$; and also, the approximate variance of it can be calculated.

One situation where $\text{BLUE}(\lambda'X\beta)$ is tractable is randomized blocks with

blocks random and n_{ij} observations on treatment i in block j , for $n_{ij} \geq 0$. We use the model equation for a treatments and b blocks,

$$y_{ijk} = \mu_i + \beta_j + e_{ijk} \quad (24)$$

for the k 'th observation on the i 'th treatment in block j , where μ_i is a treatment effect, β_j is a random block effect, and e_{ijk} a random error term, with $i = 1, \dots, a$, $j = 1, \dots, b$ and $k = 1, \dots, n_{ij}$. The customary variance component structure attributed to the β_j s and e_{ijk} s is, with $E(e_{ijk}) = 0$ and $E(\beta_j) = 0$,

$$v(e_{ijk}) = \sigma_e^2 \quad \text{and} \quad v(\beta_j) = \sigma_\beta^2 \quad (25)$$

for all i, j and k ; and all covariances are taken as zero. We define

$$Z_i = \left(\begin{array}{c} b \\ (+) \sum_{j=1}^b 1_{n_{ij}} \end{array} \right) \quad \text{and} \quad Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_a \end{bmatrix}$$

where $(+)$ represents the direct sum operator adapted so that every Z_i has b columns and for every $n_{ij} \neq 0$, $1_{n_{ij}}$ is always in column j of Z_i , for $j = 1, \dots, b$. For example, with $b = 3$ and $n_{i1} = 4$, $n_{i2} = 0$ and $n_{i3} = 5$

$$Z_i = \begin{bmatrix} 1_4 & 0 & 0 \\ 0 & 0 & 1_5 \end{bmatrix}.$$

Then with the y_{ijk} s being in lexicon order in y ,

$$V = \text{var}(y) = \sigma_e^2 Z Z' + \sigma_\beta^2 I. \quad (26)$$

For the μ_i s of (24) define $\mu = [\mu_1 \dots \mu_a]'$. With $X = \bigoplus_{i=1}^a 1_{n_i}$, a regular direct sum, which has full column rank, we then seek

$$\text{BLUE}(\mu) = (X'V^{-1}X)^{-1}X'V^{-1}y. \quad (27)$$

V^{-1} is obtained from (27) by the "tearing" algorithm (e.g. Bonney and Kissing, 1984; Searle, 1982, p.261, and, for a review, Henderson and Searle, 1981) that, in general,

$$(D - CA^{-1}B)^{-1} = D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}.$$

Using

$$c_j = [n_{1j} \quad n_{2j} \quad \dots \quad n_{aj}]',$$

and

$$D\{n_{i.}\} = \text{diagonal matrix of } n_{1.}, n_{2.}, \dots, n_{a.},$$

we find that (27) reduces to

$$\text{BLUE}(\mu) = \left[D\{n_{i.}\} - \sum_{j=1}^b \frac{\sigma_{\beta}^2}{\sigma_e^2 + n_{.j}\sigma_{\beta}^2} c_j c_j' \right]^{-1} \left\{ y_{i..} - \frac{n_{i.} \sigma_{\beta}^2 y_{.j.}}{\sigma_e^2 + n_{.j}\sigma_{\beta}^2} \right\}_{i=1}^{i=a} \quad (28)$$

The variance-covariance matrix of this estimator is σ_e^2 multiplying the inverted matrix shown in (28).

The occurrence of σ_e^2 and σ_{β}^2 in (28) is at once apparent. Estimates of them must be used to calculate an estimate of μ from (28). As mentioned following (23), Kackar and Harville (1981,1984) indicate that under wide conditions that estimate of μ will be unbiased, and approximations to its sampling variance can be calculated.

Various simplifications of (28) are easily derived:

- (i) When $n_{ij} = n$ (corresponding to balanced data), $\text{BLUE}(\mu_i) = \bar{y}_{i..}$.
- (ii) When $\sigma_{\beta}^2 = 0$ (corresponding to the model $y_{ijk} = \mu + \alpha_i + e_{ijk}$), $\text{BLUE}(\mu_i) = \bar{y}_{i..}$.
- (iii) When $\sigma_{\beta}^2 \rightarrow \infty$ (corresponding to a fixed effects no-interaction model), $\text{BLUE}(\mu_i - \mu_h)$ is the standard result - e.g. Searle (1971, Chapter 7).
- (iv) For balanced incomplete blocks, with a treatments, b blocks, k treatments (one observation each) per block, r blocks containing each particular treatment, and each treatment pair occurring in λ blocks, (28) simplifies so that with

$$\rho = \sigma_e^2 / \sigma_{\beta}^2,$$

$$\text{BLUE}(\hat{\mu}_i) = \frac{r(\rho+k)}{r\rho+\lambda a} \left[\bar{y}_{i.} - \frac{k}{\rho+k} \bar{y}_{i(j)} + \frac{\lambda a}{r(\rho+k)} \bar{y}_{..} \right] \quad (29)$$

where $\bar{y}_{i(j\cdot)}$ is the mean of the block means for the r blocks that contain treatment i . (29) is consistent with the results given in Scheffé (1959, pp. 161-175). Details of deriving (29) are available in Searle (1985).

4. Prediction of random effects

In the model equation (18) separate out $u_r = e$ from u , and re-define Z and u accordingly so as to rewrite (18) as

$$y = X\beta + Zu + e \quad (30)$$

where β and u are vectors of fixed and random effects, respectively: to u and e we attribute the mean and variance structure

$$E \begin{bmatrix} u \\ e \end{bmatrix} = 0, \quad \text{var} \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} D & 0 \\ 0 & R \end{bmatrix}, \quad (31)$$

so that

$$V = \text{var}(y) = ZDZ' + R. \quad (32)$$

Then, as is well known, the best, linear unbiased predictor of the random effects, $\text{BLUP}(u)$, is

$$\tilde{u} = \text{BLUP}(u) = DZ'V^{-1}(y - X\beta^0) \quad (33)$$

where

$$X\beta^0 = \text{BLUE}(X\beta) = X(X'V^{-1}X)^{-1}X'V^{-1}y. \quad (34)$$

Well-known properties of \tilde{u}_i , an element of $\tilde{u} = \text{BLUP}(u)$ of (33), are as follows.

- (a) \tilde{u}_i is linear in the observations.
- (b) \tilde{u}_i is unbiased for $E(u_i)$; i.e., $E(\tilde{u}_i) = E(u_i)$.
- (c) Of all linear functions that are unbiased for $E(u_i)$ none has smaller error mean square than \tilde{u}_i ; i.e., $E(\tilde{u}_i - u_i)^2$ is minimum for \tilde{u}_i of (33).
- (d) \tilde{u}_i and u_i have maximum correlation.
- (e) The probability that the ranking of \tilde{u}_i and \tilde{u}_h is the same as that of u_i and u_h is maximized.

In addition to the preceding properties there are at least seven ways of deriving \tilde{u} of (33), each of which contribute to our understanding of it. They are summarized as follows.

4.1 Mixed model equations

\tilde{u} was first derived, by Henderson et al (1959), from maximizing the joint distribution of u and y , assuming normality and the mean and variance properties of (31). This led to what are called the mixed model equations

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}z \\ z'R^{-1}X & z'R^{-1}z+D^{-1} \end{bmatrix} \begin{bmatrix} \beta^0 \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ z'R^{-1}y \end{bmatrix}, \quad (35)$$

solutions to which give precisely \tilde{u} and $X\beta^0$ of (33) and (34). Thus \tilde{u} maximizes a density function - not a likelihood.

4.2 Estimated conditional mean

On the basis of normality, for D and R of (31) assumed known, \tilde{u} is the maximum likelihood estimator of a conditional mean $\tilde{u} = E(u|y)$.

4.3 Best, linear, unbiased prediction

In Henderson (1963,1973) \tilde{u} is derived as the best, linear, unbiased predictor of u , the name by which it has come to be known; i.e., let $\lambda_i'y$ be the predictor of u_i and choose λ_i so that $E(\lambda_i'y) = E(u_i)$ and $E(\lambda_i'y - u_i)^2$ is minimized. The result is \tilde{u} of (33).

4.4 Bayes estimation

Lindley and Smith (1972), using normality for y and assuming normal priors for the parameters β and u , derive a Bayes estimator of u from the mean of the posterior distribution of the parameters. As examples, Lindley and Smith's elementary case, (1), is our (14) with every $n_i = 1$; and α^* of their (20) is our (17) with every $n_i = b$.

4.5 A weighted mean, in the Bayes context

Lindley and Smith (1972) also suggest that \tilde{u} is a weighted mean obtainable from two estimators of $[\beta' u']$. One is the BLUE that can be gotten from y given β and u ; and the other is the expected value of $[\beta' u']$ in their prior distribution. Combining these means weighted by the inverses of their dispersion matrices leads to a Bayes estimator of u , which is $\tilde{u} = \text{BLUP}(u)$ of (33). Dempfle (1977) shows details specific to deriving the mixed model equations (35).

4.6 Regression on corrected records

\tilde{u} is also obtainable as a regression estimator from records adjusted for the fixed effects: define

$$w = y - X\beta^0 \quad \text{and} \quad u^* = \text{cov}(u, w') [\text{var}(w)]^{-1} w .$$

This is suggested by Bulmer (1980). Gianola and Goffinet (1982) point out that $\text{var}(w)$ is singular and its inverse must be replaced by a generalized inverse in u^* ; and then $u^* = \tilde{u}$.

4.7 Unbiased prediction

Finally, it is shown by Searle and Casella (1984) that any linear function of y that is unbiased for $E(u) = 0$ must have the form By where $BX = 0$. Consequently $B = K[I - X(X'TX)^{-1}X'T]$ for any positive definite matrix T and any matrix K . Choosing $K = I$ and $T = V^{-1}$ ensures minimum error mean square and gives \tilde{u} .

References

- Anderson, R.D., Henderson, H.V., Pukelsheim, F. and Searle, S.R. (1984). Best estimation of variance components from balanced data with arbitrary kurtosis. Mathematische Operationsforschung und Statistik, Series Statistics 15, 163-176.
- Bonney, G.E. and Kissling, G.E. (1984). On the inverse of a patterned covariance matrix. Journal of the American Statistical Association 79, 722-723.
- Bulmer, M.G. (1980). The Mathematical Theory of Quantitative Genetics. Oxford University Press.
- Dempfle, L. (1977). Relation entre BLUP (best linear unbiased prediction) et estimateurs Bayesiens. Annals de Genetique et de Selection Animale 9, 27-32.
- Gionola, D. and Goffinet, B. (1982). Sire evaluation with best linear unbiased predictors. Biometrics 38, 1085-1088.
- Henderson, C.R. (1963). Selection index and expected genetic advance. Statistical Genetics in Plant Breeding, 141-163. National Academy of Sciences, National Research Council publication No. 982, Washington D.C.
- Henderson, C.R. (1973). Sire evaluation and genetic trends. In Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush, 10-41. Champaign, Illinois: American Society of Animal Science and American Dairy Science Association.
- Henderson, C.R., Kempthorne, O., Searle, S.R. and von Krosigk, C.N. (1959). Estimation of environmental and genetic trends from records subject to culling. Biometrics 15, 192-218.
- Henderson, H.V. and Searle, S.R. (1981). On deriving the inverse of a sum of matrices. SIAM Review 23, 53-60.
- Kackar, R.N. and Harville, D.A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. Communications in Statistics - Theory and Methods, Ser. A, 10, 1249-1261.
- Kackar, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimation of fixed and random effects in mixed linear models. Journal of the American Statistical Assoc. 79, 853-861.
- Khuri, A.I. (1981). Anova sums of squares for balanced mixed-effects models revisited. Technical Report No. 162, University of Florida.
- Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for the linear model (with discussion). Journal of the Royal Statistical Society (B) 34, 1-42.

- Scheffé, H. (1959). The Analysis of Variance. John Wiley & Sons, New York.
- Searle, S.R. (1971). Linear Models. John Wiley & Sons, New York.
- Searle, S.R. (1982). Matrix Algebra Useful for Statistics. John Wiley & Sons, New York.
- Searle, S.R. (1984). Best linear unbiased estimation in mixed models of the analysis of variance. Technical Report BU-864-M, Biometrics Unit, Cornell University.
- Searle, S.R. (1985). Cell means formulations of mixed models. Technical Report, Biometrics Unit, Cornell University.
- Searle, S.R. and Casella, G. (1984). Alternative derivations of best linear unbiased prediction (BLUP) in the mixed model. Technical Report BU-861-M, Biometrics Unit, Cornell University.
- Searle, S.R. and Henderson, H.V. (1979). Dispersion matrices in variance components models. Journal of the American Statistical Assoc. 74, 465-470.
- Seifert, B. (1979). Optimal testing for fixed effects in general balanced mixed classification models. Mathematische Operationsforschung und Statistik, Series Statistics 10, 237-255.
- Smith, D.W. and Hocking, R.R. (1978). Maximum likelihood analysis of the mixed model: the balanced case. Communications in Statistics - Theory and Methods A7, 1253-1266.
- Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. Annals of Mathematical Statistics 38, 1092-1109.