C. R. Henderson, the Statistician; And His Contributions to Variance Components Estimation[1]

SHAYLE R. SEARLE
Biometrics Unit
Cornell University
Ithaca, NY 14853

ABSTRACT

C. R. Henderson's 1953 *Biometrics* paper "Estimation of Variance and Covariance Components" (25) is an outstanding landmark in the discipline of statistics. It sets out the very first ideas of how to estimate variance components from unbalanced (unequal subclass numbers) data in situations more complicated than the one-way classification (completely randomized design). As such it had three important, long-lasting impacts. First, it provided methods for actually using unbalanced data, even in large quantity, for estimating variance components. And this has played a tremendous role in population genetics and in animal breeding where the use of estimated variance components is vital to the application of selection theory and selection index techniques. Second, that 1953 paper stimulated numerous statisticians to become interested in random effects, mixed models, and variance components estimation, with such statistical greats as H. O. Hartley and C. R. Rao making contributions in the late 1960s and early 1970s. By then, improved methods of estimating variance components from unbalanced data had been developed, namely maximum likelihood (ML) and restricted maximum likelihood (REML). Once computing power had expanded to the point where these methods became feasible, Henderson made notable contributions to these methods, allied to his two great interests: animal breeding and feasible computing procedures. For both of these, his mixed model equations were a salient feature. Third, these methods reached a wide audience of geneticists and statisticians.

(Key words: Henderson, C. R.; statistics; variance components)

Abbreviation key: MME = mixed model equations.

THE HENDERSON LANDMARK

The early beginnings of estimating variance components date back to the work of two astronomers: Airy [1861 (1)] and Chauvenet [1863 (10)]. Modern-day work started with Fisher [1925 (17) and 1938 (18)], Cochran [1939 (11)], Daniels [1939 (15)], and Winsor and Clarke [1940 (54)]. In all of these papers we see the beginnings of what today is called the ANOVA method of estimating variance components, a method based on equating expected mean squares to their values computed from data. As an estimation method, it was given considerable impetus by the Anderson and Bancroft (3) book in which it is considered at some length. More details of this history are presented by Searle (50).

Most of this work had dealt with what are now called balanced data—data having equal numbers of observations in the subclasses. Cochran (11), Daniels (15) and Winsor and Clarke (54) had considered the case of unbalanced data (having unequal numbers of observations in the subclasses) from a one-way classification, and Ganguli (19) dealt comprehensively with unbalanced data from completely nested models (e.g., cows within herds and herds within states). But no one had tackled the difficult problem of estimating variance components from unbalanced data of cross-classified models (e.g., of milk production records of daughters of AI sires in different

Received August 30, 1990.
Accepted December 3, 1990.
[1]Paper BU-574 in the Biometrics Unit, College of Agriculture & Life Sciences, Cornell University, Ithaca, NY.

herds, where sires are cross-classified with herds, and, for a large group of herds, each sire has daughters in many herds and each herd has daughters of many sires). This was the statistical problem that Charles Roy Henderson addressed in 1953 in his classic paper in *Biometrics:* "Estimation of Variance and Covariance Components" (25). That paper was to be a landmark in the history of estimating variance components.

THE HENDERSON BACKGROUND

Here we had, in 1953, a statistical paper, couched in terms of dairy cows, that was later to be described by the Institute of Scientific Information as one of the most frequently cited papers in the scientific literature: in the decade from 1978 to 1988, 25 to 35 years after publication, it was listed by *Science Citation Index* as having been referenced 234 times, an average of twice a month! From whence, statistically, came its author?

Charles Henderson grew up in Iowa and received all his university education at Iowa State College (now University) in Ames. Following his M.S. degree in animal nutrition, he worked first with the Iowa Extension Service and then Ohio University, and in 1942 he joined the US Army, where he finished as Commanding Officer of an Army Medical Nutrition Laboratory in Chicago. It was during those army days that his statistical interest began, as evidenced by the 18 papers from 1944 to 1948 on which he is a coauthor. Almost all of these are concerned with such features as "effects of" and "observations on" various dietary limitations on young men.

Leaving the army in 1946, Chuck (as we all knew him) went back to Ames and pursued a Ph.D. degree under Jay L. Lush, Leroy Hazel, and Oscar Kempthorne. And it was in working on his thesis, "Estimation of Variances Due to General, Specific, and Maternal Combining Abilities Among Inbred Lines of Swine," that he came to the problem that concerned him, statistically, for the rest of his life. This was the question: how does one estimate variance components from large, cross-classified data sets that have unequal numbers of observations in the subclasses? At that time, the late 1940s and early 1950s, there was only the ANOVA method of estimation, based as it is on equating mean squares of an analysis of variance to their expected values under models that contain random effects; in general, mixed models. The result is a set of linear equations in the variance components, the solutions of which are taken as estimates of the variance components.

Example. Suppose in a given year there are n daughters of each of b bulls in a sire-proving experiment, with

$$y_{ij} = \mu + s_i + e_{ij} \qquad [1]$$

being the record of daughter j of sire i. Then the between-sires and within-sires analysis of variance table is based on Table 1, where $\bar{y}_{i.} = \sum_{j=1}^{n} y_{ij}/n$ is the mean record of the n daughters

TABLE 1. Sums of squares (SS) and mean squares of n daughters from each of b sires.

Source of variation	Degrees of freedom	Sum of squares	Mean square
Between sires	$b - 1$	$SSB = \sum_{i=1}^{b}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})^2$	$MSB = \frac{SSB}{b-1}$
Within sires	$b(n - 1)$	$SSW = \sum_{i=1}^{b}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$	$MSW = \frac{SSW}{b(n-1)}$
Total	$bn - 1$	$SST = \sum_{i=1}^{b}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$	

TABLE 2. Number of daughter records in each herd for 60 AI sires used in 800 herds.

	Herd								
Sire	1	2	3	4	5	6	...	800	Total
1	–	1	–	–	2	3	...	1	436
2	2	–	1	1	–	1	...	–	281
3	1	–	2	–	1	2	...	2	264
⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮	⋮
60	–	1	–	–	–	–			40
Total	65	48	51	47	32	61	...	24	9862

of sire i, and $\bar{y}_{..} = \sum_{i=1}^{b} \sum_{j=1}^{n} y_{ij}/bn$ is the mean of the records of all n daughters of all b sires. Then, with sires considered random, the expected mean squares (MS) in the usual variance component form of Equation [1] are, using σ_s^2 and σ_e^2 for sire and error variance, respectively

$$E(MSB) = n\sigma_s^2 + \sigma_e^2 \text{ and } E(MSW) = \sigma_e^2. \quad [2]$$

Equating mean squares to their expected values as a method of estimating variance components gives the equations

$$MSB = n\hat{\sigma}_s^2 + \hat{\sigma}_e^2 \text{ and } MSW = \hat{\sigma}_e^2, [3]$$

and these yield the ANOVA estimates

$$\hat{\sigma}_s^2 = \frac{1}{n}(MSA - MSE) \text{ and } \hat{\sigma}_e^2 = MSE \quad [4]$$

This is the ANOVA method of estimating variance components at its simplest: for the one-way classification with balanced data. By the time of Henderson's 1953 paper, this method of estimation was well established for balanced data. What Henderson did, for unbalanced data from models that could have any number of crossed factors, was to provide suggestions for the sums of squares that could be used in those models in the manner that SSB and SSW are used in [2], [3], and [4].

HENDERSON'S THREE METHODS

The philosophy of Henderson (25) was to adapt the ANOVA method of estimating variance components so as to have methodology available for the kind of analysis that was being considered in the 1950s, a year-by-herd-by-sire analysis with years as fixed effects and herd and sires as random effects. Thus, the distribution of 1 yr's records might have the appearance of Table 2. What Henderson gave us was three different sets of mean squares (or sums of squares, these two being effectively equivalent for the purpose at hand) that could be used in the basic ANOVA method of estimation, namely

equate calculated MS to E(MS) [5]

or

equate calculated SS to E(SS). [6]

The three different sets of sums of squares came to be known as Henderson's Methods 1, 2, and 3. In truth, they are not three different methods but three different applications of the basic ANOVA method as specified in [5] and [6]. In brief, the three methods are as follows.

Method 1

Use sums of squares analogous to the traditional sums of squares of the analysis of variance of balanced data. In fact, some of the terms used are actually not sums of squares; they are quadratic forms of the data that can be negative. For example, the analogy of

$$\sum_{i=1}^{a}\sum_{j=1}^{b} n(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 = \Sigma_i\Sigma_j n\bar{y}_{ij.}^2 - \Sigma_i bn\bar{y}_{i..}^2 - \Sigma_j an\bar{y}_{.j.}^2 + abn\bar{y}_{...}^2 \quad [7]$$

for balanced data becomes, for unbalanced data,

$$\Sigma_i\Sigma_j n_{ij}\bar{y}_{ij.}^2 - \Sigma_i n_{i.}\bar{y}_{i..}^2 - \Sigma_j n_{.j}\bar{y}_{.j.}^2 + n_{..}\bar{y}_{...}^2, \quad [8]$$

which is not a sum of squares. It can be negative. Thus, for $a = b = 2$ and data

2	4
10	4,40

expression [8] is –22. [9]

Method 2

Adjust the data to take account of fixed effects estimated as if the random effects were fixed; and then use Method 1.

Method 3

Use the sums of squares from the classical method of fitting constants that has long been available as a method of estimating fixed effects from unbalanced data from fixed effects models. These sums of squares used to be used in a hypothesis testing framework, too, although it has since been shown [e.g. Searle (49)] that the hypotheses being so tested do, in fact, differ from those which they were once thought to be.

Each of these methods has it merits and demerits. Method 1 is easy to calculate, even without modern computers, but it can be used only for random models. It cannot be used for mixed models. Adaptating the method to mixed models by ignoring the fixed effects or assuming them to be random yields biased estimators of the variance components. Method 2 was, in my opinion, initially difficult to understand, especially the manner of making the adjustments for the fixed effects. More serious, though, was the fact that Method 2 cannot be used for any model that includes interactions between fixed effects factors and random effects factors. Trying to do so yields biased estimators. Method 3 is easy to understand, but with large data sets, its calculation can require inverting matrices of large order. And for many situations it is a method that is not uniquely defined. For example, in the sires-by-herds case of Table 2, does one use the (reductions in) sum of squares,

R(herds, adjusted for years) and
R(sires, adjusted for years and herds), [10]

or does one use

R(sires, adjusted for years) and
R(herds, adjusted for years and sires)? [11]

These two possibilities are different and they do not yield the same estimates. Each of them is a legitimate Method 3 procedure, but Method 3 gives no indication as to which is preferable—or even if one is preferable to the other, or preferable to other pairs of sums of squares, such as, for example, the second R( ) in each of [10] and [11].

INITIAL FOLLOW UP: PROPERTIES OF THE HENDERSON METHODS

All three of the Henderson methods give unbiased estimators of variance components. This is to be expected, because all of them are applications of the most general form of ANOVA estimation, which can be typified as $E(q) = P\sigma^2$ giving $\hat{\sigma}^2 = P^{-1}q$ where q is a vector of quadratic forms of the data and σ^2 the vector of components to be estimated. Hence, $E(\hat{\sigma}^2) = P^{-1}E(q) = P^{-1}P\sigma^2 = \sigma^2$. What the Henderson methods do is provide suggestions for the elements of q.

The three methods also have one deficiency: there is nothing in their methodology that prevents the occurrence of negative estimates. Yet all of them simplify, for balanced data, to be the same: to yield the ANOVA estimators of variance components, which, for balanced data, are minimum variance unbiased (MVU) under normality assumptions [Graybill and Wortham (21)] and even without normality are minimum variance quadratic unbiased (MVQU) [Graybill and Hultquist (20)].

Henderson's development of his three methods of estimating variance components soon led to their being used extensively in many different applications, especially in animal breeding, as dairy scientists well know. But one particular sentence in that 1953 paper fired the imagination of a number of statisticians: "The relative sizes of the sampling variances of estimates obtained by the methods are not known" [Henderson, (25), p. 227]. This led to a small flood of papers that produced expressions for calculating sampling variances of estimated variance components for a number of special cases. Searle [1956 (45)], who extended Crump [1951 (13)], Searle [1958 (46), 1961 (47)], Mahamunulu [1963 (32)], Low [1964 (31)], Hirotsu [1966 (30)], Blischke [1966 (6), 1968 (7)], and Rhode and Tallis [1969 (42)] are among those who made contributions. To users of variance components it was great to have estimation methods available for unbalanced data (for that is what Henderson had provided), but for statisticians it was essential to know something about the sampling distribution of each resulting estimator or at least the sampling variance. Unfortunately, even on assuming normality of the random effects, there has been no progress at all in establishing closed-form expressions for the sampling distributions, and even trying to derive expressions for sampling variances turned out to be very messy algebraically. Those expressions are quadratic functions of the unknown variance components with coefficients that are incredibly miserable functions of the numbers of observations in the subclasses of the data. For example, for the one-way classification, the daughters-within-sires illustration of Table 1, only with differing numbers of daughters from the sires, the sampling variance of the estimated sire variance component is, for $N = \Sigma_i n_i$.

$$\mathrm{var}(\hat{\sigma}_s^2) = \frac{2N}{N^2 - \Sigma n_i^2}\left[\frac{N(N-1)(a-1)}{(N-a)\;(N^2-\Sigma n_i^2)}\,\sigma_e^4 + 2\sigma_s^2\sigma_e^2 + \frac{N\Sigma n_i^2 + (\Sigma n_i^2)^2 - 2N\Sigma n_i^3}{N(N^2-\Sigma n_i^2)}\,\sigma_s^4\right].$$

This expression is not amenable to analysis; i.e., one cannot, no matter what the values of σ_s^2 and σ_e^2, study the behavior of $\mathrm{var}(\hat{\sigma}_s^2)$ for variation in the n_i values.

These results slowly led to the realization that it is a practical impossibility to compare the three Henderson methods in terms of sampling variances (or mean squared error) of estimators. The mathematical expressions for these variances and covariances are simply too intractable to permit any general statements comparing one method with another, either mathematically or numerically. The few numerical studies that were done [Bainbridge, 1963 (5); Bush and Anderson, 1963 (9); Anderson and Crump, 1967 (4); and Anderson, 1975 (2)] are mostly for small hypothetical data layouts that give little or no relevant information for the kind of large data sets that animal breeders deal with so often.

Along with the flurry of papers on sampling variances, there also came papers on a variety of other features of the Henderson methods and associated ideas. Cunningham and Henderson (14), Zelen (55), and Thompson (53), for example, suggested extensions of Method 3; Searle (48) provided proof that Method 2 cannot be used on models having interactions between fixed and random effects; and Henderson et al. (27) showed how Method 2 can be computed and that it has an essential invariance property [in contrast to Searle (49), who wrongly asserted that it lacked such a property].

An interesting by-product of Henderson's contributions and influence in all this work is that many animal breeders, particularly those in academia, realized the necessity of having a knowledge of matrix algebra as a tool both for understanding the difficulties involved in estimating variance components from their data and for developing methods that can handle those difficulties. Henderson, from the late 1950s onward, was a steady user of matrix methods, and it seems to me that it was he who single-handedly promoted their use for animal breeding problems.

BLUP AND HENDERSON'S MME

One of Chuck's most important contributions not only for animal breeders but also for statisticians was in a paper (26) with Oscar Kempthorne and others. That paper was written in three sections, each with its own author or authors. The importance of the paper is largely in Henderson's section in which the estimation of variance components is not the emphasis but rather the derivation of the BLUP of random effects. Equations now known as the mixed model equations (MME), derived by Henderson, are also given in that paper. Before discussing these and their connection to BLUP it is interesting to give a little history.

During his Ph.D. training, Chuck had taken from Alexander Mood a statistics course wherein he first came upon the following exercise that later appeared in Mood (33, p. 164, exercise 23). With small changes it is also to be found in Mood and Graybill (34, p. 195, exercise 32), and in Mood, Graybill and Boes (35, p. 370, exercise 52):

> 23. Suppose intelligence quotients for students in a particular age group are normally distributed about a mean of 100 with standard deviation 15. The IQ, say x_1, of a particular student is to be estimated by a test on which he scores 130. It is further given that test scores are normally distributed about the true IQ as a mean with standard deviation 5. What is the maximum-likelihood estimate of the student's IQ? (The answer is not 130).

This exercise, with its tantalizing last sentence, played a prominent role in initially motivating C. R. Henderson in his lifelong contributions to the problem of estimating genetic merit of livestock. It is the classic prediction problem of predicting the unobservable realized value of a random effect that is part of a mixed model.

As the forerunner of BLUP, it is, as we know now, a procedure for estimating a conditional mean, the mean of the random effect of interest, given the data. In the sire-proving context, BLUP is the estimator of $E(s_i|y)$, where s_i is a sire effect and y is the vector of data. In 1950, Henderson (24) gave a paper on this topic at a meeting of the Institute of Mathematical Statistics and had referred to the technique as "estimation" of random effects, a phrase that, Chuck once told me, went down like a lead balloon. Statisticians simply do not estimate random variables. And that, along with a paper by Cochran (12), is presumably and appropriately what led to the name "predictor" rather than "estimator".

Brief details of the BLUP and MME ideas are as follows, given in the matrix notation that Henderson used so often. Begin with the model equation for a data vector y:

$$y = X\beta + Zu + e.$$

β is a vector of fixed effects, **u** is a vector of random effects, X and Z usually incidence matrices (elements of 0 or 1), and **e** is a residual error vector. Then with

$$\text{var}(\mathbf{u}) = \mathbf{D},\ \text{var}(\mathbf{e}) = \mathbf{R},\ \text{and}\ \text{cov}(\mathbf{u},\mathbf{e}) = \mathbf{0},$$

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

Then the BLUE of $X\beta$ is

$$\text{BLUE}(X\beta) = X(X'V^{-1}X)^{-}X'V^{-1}y\ , \quad [12]$$

where $(X'V^{-1}X)^{-}$ is defined by $X'V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}X = X'V^{-1}X$. And then the BLUP of **u** is

$$\text{BLUP}(\mathbf{u}) = \mathbf{D}Z'V^{-1}[y - \text{BLUE}(X\beta)]. \quad [13]$$

Henderson's development was that BLUE($X\beta$) and BLUP(**u**) can both be obtained more easily than in Equations [12] and [13], from the one set of equations

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + D^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}, \quad [14]$$

which he called the MME. Then

$$X\hat{\beta} = \text{BLUE}(X\beta) \quad [15]$$

and

$$\tilde{u} = \text{BLUP}(\mathbf{u}). \quad [16]$$

Equations [14], the MME, have several important features. First, for obtaining $X\hat{\beta}$ they are easier to calculate than [12]: V^{-1} in [12] has order N, whereas [14] has order equal to the number of fixed and random effects occurring in the data—usually much less than N. Moreover, $\mathbf{R}^{-1}$ and $\mathbf{D}^{-1}$ in Equation [14] are usually easy to calculate, because R is commonly just

$\sigma_e^2\mathbf{I}$, and $\mathbf{D}$ is often diagonal. Second, Equations [14] yield $\hat{\beta}$ and $\tilde{\mathbf{u}}$ simultaneously. And then there are the two other properties of $\tilde{\mathbf{u}}$ that are so useful to animal breeders: one is the good ranking property of $\tilde{\mathbf{u}}$, that in ranking sires by their corresponding values in $\tilde{\mathbf{u}}$ one is maximizing the probability of correctly ranking the sires, as shown by Portnoy (37); and the other useful feature of $\hat{\beta}$ and $\tilde{\mathbf{u}}$ is that provided $\mathbf{k}'\beta$ is estimable (i.e., $\mathbf{k}' = \mathbf{t}'\mathbf{X}$ for some $\mathbf{t}'$), then the best estimator of $\mathbf{k}'\beta + \mathbf{p}'\mathbf{u}$ is $\mathbf{k}'\hat{\beta} + \mathbf{p}'\tilde{\mathbf{u}}$. And finally, of course, through their reliance on the variance components of whatever mixed model is being used, the MME provide solid motivation for wanting to estimate variance components.

These then, are Henderson's main specific contributions to statistics: timely methods in the early 1950s for estimating variance components from unbalanced data and development of the BLUP procedure and affiliated methodology. These contributions are substantial. No one before Henderson had really tackled the variance component problem; and his BLUP is equivalent to, and came long before, procedures in statistics in the framework of Bayes, Stein, and shrinkage estimation that are now so widely recognized. In this context, Henderson, the statistician, through his intense interest in animal breeding problems, was years ahead of his time.

MAXIMUM LIKELIHOOD ESTIMATION

Henderson's methods not only inspired other statisticians to study properties of those methods (e.g., sampling variances, as already discussed), but they also motivated others, in my opinion, to develop further methods of estimating variance components. The most important of these, by Hartley and J.N.K. Rao (22), has been the application of the long-established method of maximum likelihood—a method of estimation that, in almost all circumstances, yields estimators with a variety of optimal statistical properties: consistency, sufficiency, and asymptotic normality with asymptotic sampling variances being available. It is therefore a very attractive method of estimation. Yet, apart from Crump (13), it was not applied to the variance component problem until Hartley and J.N.K. Rao (22), their success undoubtedly being due to their matrix formulation of the problem. With the model equation

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e} \qquad [17]$$

they partitioned $\mathbf{u}$ into vectors $\mathbf{u}_i$ of order q_i, one for each random factor (main effect, interaction or nested factor) in the data, with $\text{var}(\mathbf{u}_i) = \sigma_i^2\mathbf{I}_{q_i}$. And $\mathbf{Z}$ is partitioned conformably with $\mathbf{u}$, to give

$$\mathbf{Z}\mathbf{u} = \sum_{i=1}^{r}\mathbf{Z}_i\mathbf{u}_i$$

for r random factors. Moreover, on defining

$$\mathbf{u}_0 = \mathbf{e} \text{ and } \mathbf{Z}_0 = \mathbf{I} \qquad [18]$$

the model equation becomes

$$\mathbf{y} = \mathbf{X}\beta + \sum_{i=0}^{r}\mathbf{Z}_i\mathbf{u}_i. \qquad [19]$$

Then, on assuming normality of all the random effects and residual error term, and defining

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} \qquad [20]$$

the maximum likelihood equations for the variance components can be adapted [e.g., Searle (51)] from the Hartley and Rao (22) equations to be written succinctly as [see Searle (52, Sec. 7.1a) for notation]

$$\left\{_m\text{tr}(\tilde{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{V}}^{-1}\mathbf{Z}_j\mathbf{Z}_j')\right\}_{i,j=0}^{r} \left\{_c\tilde{\sigma}_j^2\right\}_{j=0}^{r} = \left\{_c\mathbf{y}'\tilde{\mathbf{P}}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{P}}\mathbf{y}\right\}_{i=0}^{r}. \qquad [21]$$

These equations may not be the easiest form for obtaining a solution, but they are well suited to considering what is involved in calculating maximum likelihood estimates. They have to be solved for the $\tilde{\sigma}_j^2$ terms; and these occur in $\tilde{\mathbf{V}}^{-1}$, or course, and also, from [20], in $\tilde{\mathbf{P}}$. So the equations are very nonlinear and in general do not yield closed-form solutions. Nevertheless, the asymptotic sampling dispersion matrix of the variance component estimators is known and is

$$\text{var}(\tilde{\sigma}^2) \doteq \left\{_m\text{tr}(\mathbf{V}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{V}^{-1}\mathbf{Z}_j\mathbf{Z}_j')\right\}_{i,j=0}^{r} . \qquad [22]$$

A variant of maximum likelihood is restricted (or, in Europe, residual) maximum likelihood (REML), that is, maximum likelihood for linear functions of the data that do not contain the fixed effects, the "error contrasts", as Harville [23] calls them. Use of REML results in equations

$$\left\{{}_m\mathrm{tr}(\hat{P}Z_iZ_i'\hat{P}Z_jZ_j')\right\}_{i,j=0}^{r}\left\{{}_c\hat{\sigma}_j^2\right\}_{j=0}^{r} = \left\{{}_c y'\hat{P}Z_iZ_i'\hat{P}y\right\}_{i=0}^{r} \qquad [23]$$

with asymptotic dispersion matrix

$$\mathrm{var}(\hat{\sigma}^2) \doteq \left\{{}_m\mathrm{tr}(PZ_iZ_i'PZ_jZ_j')\right\}_{i,j=0}^{r} \cdot \qquad [24]$$

Whichever one uses, maximum likelihood or REML, the equations for the estimates have to be solved numerically, usually by iteration, although with the availability of very large computers, methods of simply evaluating the logarithm of the likelihood are coming to be investigated. Whatever is done, writing the necessary computer program is no easy task. Among the difficulties are the following: 1) using an optimum form of the equation; 2) developing an optimum iteration technique; 3) using sparse matrices of very large order; 4) having no intermediate values of $\tilde{V}$ be singular; 5) being sure that choice of starting values does not affect the final solution; 6) reaching a global, not local, maximum for the likelihood; 7) being certain that convergence is achieved; 8) achieving $\tilde{\sigma}_0^2 > 0$ and $\tilde{\sigma}_i^2 \geq 0$ for $i = 1,\ldots,r$; and 9) having $\tilde{V}$ be non-negative definite.

Despite these difficulties, along with having to assume normality, the maximum likelihood approach is definitely coming to be accepted as an optimal approach to estimating variance components. Its difficult computing problems are getting to be solved both by the rapid development of hardware and software and by the attention that computer scientists and numerical analysts are giving to the solution of nonlinear equations. Furthermore, maximum likelihood came on the scene just as some statistical dissatisfaction with the Henderson methods was peaking—lack of distribution properties and the inability to informatively compare methods. And although it is my personal opinion that, to begin with, Chuck Henderson was not convinced that maximum likelihood was going to become acceptable, he had soon started working with maximum likelihood, and after 1977 was publishing comparative studies involving maximum likelihood and REML. [See, for example, Henderson and Quaas, 1977 (28); Rothschild, Henderson, and Quaas, 1978 (43); and Rothschild and Henderson, 1979 (44).] These studies also included comparisons with MINQUE estimation method of Rao (38, 39, 40, 41), which requires using a set of prior values of the components as part of the estimation process. The MINQUE estimates are then functions of those prior values. Brown [8], however, shows that iterating MINQUE, which is computationally equivalent to REML, gives estimators that are asymptotically normal.

At about the same time as this comparative work, Henderson (29) wrote a technical report in which he showed how maximum likelihood calculations can be achieved using elements of his MME. This was a remarkable piece of work, allied as it was to both Patterson and Thompson (36) in their original development of REML and to the later Dempster et al. (16) in their description of what is now called the expectation-maximization (EM) algorithm. Henderson's approach, by way of his MME, avoided some of the computing problems listed earlier; notably, as shown by Harville (23), it always gives positive values for the estimated components.

AN OVERVIEW

A list of C. R. Henderson's publications contains 233 items: 1 book, 1 thesis, 48 abstracts, and 183 papers. In a rough and subjective classification of these papers, I labeled 94 of them as primarily statistical, with 41 of those being concerned with variance components. With this record, it is clear that although dairy scientists will always think of Chuck as one of them (and quite rightly so), he was also a major player in statistics. His statistical contributions were, of course, in the specialized fields of linear models and variance components, resulting from his strong and lifelong interest in using statistics for the improvement of dairy farming. A dazzling lighthouse of his work was the 1953 *Biometrics* paper (25). It had three major effects. First, it showed us that there were indeed ways of estimating variance components from unbalanced data. Second, it motivated a number of statisticians to look at those methods and investigate their properties, and in so doing, to generate interest in the mixed model. Third, and possibly most important, this encouraged Chuck, I think, to continue promoting the mixed model to his colleagues and his students until his ideas reached a statistical audience far far wider than just the users of the mixed model and variance components. In the 1950s and early 1960s, interest in the mixed model was primarily among biometricians connected with genetics. (I remember in 1957 being at a 6-wk National Science Foundation seminar on analysis of variance that included a number of established statisticians who were publishing analysis of variance papers in the *Annals of Mathematical Statistics*. Several of them had never heard of random effects!) Yet by 1967 and in the years since, the mixed model has become a hot topic in numerous statistical journals. Much of this interest stems, I'm sure, from the man who got to the problems first: Charles Roy Henderson.

ACKNOWLEDGMENTS

Preparation of this paper was supported, in part, by the Department of Mathematics and Statistics at the University of Auckland, New Zealand, where the author was on leave from Cornell, February to May 1990.

REFERENCES

1 Airy, G. B. 1861. On the algebraical and numerical theory of errors of observations and the combinations of observations. MacMillan, London, Engl.

2 Anderson, R. L. 1975. Designs and estimators for variance components. Page 1 *in* Statistical design and linear models. J. N. Srivastava, ed. North Holland, Amsterdam, Neth.

3 Anderson, R. L., and T. A. Bancroft. 1952. Statistical theory in research. McGraw-Hill, New York, NY.

4 Anderson, R. L., and P. P. Crump. 1967. Comparisons of designs and estimation procedures for estimating parameters in a two-stage nested process. Technometrics 9:499.

5 Bainbridge, T. R. 1963. Staggered nested designs for estimating variance components. Page 93 *in* Am. Soc. Qual. Control Annu. Conf. Trans.

6 Blischke, W. R. 1966. Variances of estimates of variance components in a three-way classification. Biometrics 22:553.

7 Blischke, W. R. 1968. Variances of moment estimators of variance components in the unbalanced r-way classification. Biometrics 24:527.

8 Brown, K. G. 1976. Asymptotic behavior of MINQUE-typed estimators of variance components. Ann. Stat. 4: 746.

9 Bush, N., and R. L. Anderson. 1963. A comparison of three different procedures for estimating variance components. Technometrics 5:421.

10 Chauvenet, W. 1863. A manual of spherical and practical astronomy, 2: theory and use of astronomical instruments. Lippincott, Philadelphia, PA.

11 Cochran, W. G. 1939. The use of analysis of variance in enumeration by sampling. J. Am. Stat. Assoc. 34: 492.

12 Cochran, W. G. 1951. Improvement by means of selection. Page 499 *in* Proc. 2nd Berkeley Symp.

13 Crump, S. L. 1951. The present status of variance components analysis. Biometrics 7:1.

14 Cunningham, E. P., and C. R. Henderson. 1968. An iterative procedure for estimating fixed effects and variance components in mixed model situations. Biometrics 24:13.

15 Daniels, H. E. 1939. The estimation of components of variance. J. R. Stat. Soc.(Suppl.)6:186.

16 Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B 39:1.

17 Fisher, R. A. 1925, 1938. Statistical methods for research workers. 1st and 7th ed. Oliver and Boyd, Edinburgh and London.

18 Fisher, R. A. 1935. Discussion of Neyman et al. (1935). J. R. Stat. Soc. Ser. B 2:154.

19 Ganguli, M. 1941. A note on nested sampling. Sankhyā 5:449.

20 Graybill, F. A., and R. A. Hultquist. 1961. Theorems concerning Eisenhart's model II. Ann. Math. Stat. 32: 261.

21 Graybill, F. A., and A. W. Wortham. 1956. A note on uniformly best unbiased estimators for variance components. J. Am. Stat. Assoc. 51:266.

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

23 Harville, D. A. 1977. Maximum-likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Assoc. 72:320.

24 Henderson, C. R. 1950. Estimation of genetic parameters. Ann. Math. Stat. 21:309.

25 Henderson, C. R. 1953. Estimation of variance and covariance components. Biometrics 9:226.

26 Henderson, C. R., O. Kempthorne, S. R. Searle, and C. N. Von Krosigk. 1959. Estimation of environmental and genetic trends from records subject to culling. Biometrics 15:192.

27 Henderson, C. R., S. R. Searle, and L. R. Schaeffer. 1974. The invariance and calculation of Method 2 for estimating variance components. Biometrics 30:583.

28 Henderson, C. R., and R. L. Quaas. 1977. A comparison of Method 3 and maximum likelihood estimators of heritabilities and genetic correlation. J. Anim. Sci. 45(Suppl. 1):22.(Abstr.)

29 Henderson, C. R. 1973. Maximum likelihood estimation of variance components. Tech. Rep. Dep. Anim. Sci., Cornell Univ., Ithaca, NY.

30 Hirotsu, C. 1966. Estimating variance components in a two-way layout with unequal numbers of observation. Rep. Stat. Appl., Union Jpn. Sci. Eng. (JUSE) 13:29.

31 Low, L. Y. 1964. Sampling variances of estimates of

components of variance from a non-orthogonal two-way classification. Biometrika 51:491.

32 Mahamunulu, D. M. 1963. Sampling variances of the estimates of variance components in the unbalanced three-way nested classification. Ann. Math. Stat. 34: 521.

33 Mood, A. M. 1950. Introduction to the theory of statistics. McGraw-Hill, New York, NY.

34 Mood, A. M., and F. A. Graybill. 1963. Introduction to the theory of statistics. 2nd ed. McGraw-Hill, New York, NY.

35 Mood, A. M., F. A. Graybill, and D. C. Boes. 1974. Introduction to the theory of statistics. 3rd ed. McGraw-Hill, New York, NY.

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

37 Portnoy, S. 1982. Maximizing the probability of correctly ordering random variables using linear predictors. J. Multivariate Anal. 12:256.

38 Rao, C. R. 1970. Estimation of heteroscedastic variances in linear models. J. Am. Stat. Assoc. 65:161.

39 Rao, C. R. 1971. Estimation of variance and covariance components—MINQUE theory. J. Multivariate Anal. 1:257.

40 Rao, C. R. 1971. Minimum variance quadratic unbiased estimation of variance components. J. Multivariate Anal. 1:445.

41 Rao, C. R. 1972. Estimation of variance and covariance components in linear models. J. Am. Stat. Assoc. 67:112.

42 Rohde, D. A., and G. M. Tallis. 1969. Exact first- and second-order moments of estimates of components of covariance. Biometrika 56:517.

43 Rothschild, M. F., C. R. Henderson, and R. L. Quaas. 1978. Comparison of maximum likelihood and Henderson's Method I estimates of variances and covariances for two traits when selection is present. J. Dairy Sci. (Suppl. 1):91.(Abstr.)

44 Rothschild, M. F., and C. R. Henderson. 1979. Maximum likelihood estimation of population parameters of first and second lactation milk records. J. Dairy Sci. 62:603.

45 Searle, S. R. 1956. Matrix methods in components of variance and covariance analysis. Ann. Math. Stat. 27: 737.

46 Searle, S. R. 1958. Sampling variance of estimates of components of variance. Ann. Math. Stat. 29:167.

47 Searle, S. R. 1961. Variance components in the unbalanced two-way nested classification. Ann. Math. Stat. 32:1161.

48 Searle, S. R. 1968. Another look at Henderson's method of estimating variance components. Biometrics 24:749.

49 Searle, S. R. 1971. Linear models. Wiley, New York, NY.

50 Searle, S. R. 1989. Variance components—some history and a summary account of estimation methods. J. Anim. Breeding Genet. 106:1.

51 Searle, S. R. 1979. Notes on variance component estimation. Tech. Rep. BU-673-M, Biometrics Unit, Cornell Univ., Ithaca, NY.

52 Searle, S. R. 1987. Linear models for unbalanced data. Wiley, New York, NY.

53 Thompson, R. 1969. Iterative estimation of variance components for non-orthogonal data. Biometrics 25: 767.

54 Winsor, C. P., and G. L. Clarke. 1940. Statistical study of variation in the catch of plankton nets. Sears Found. J. Marine Res. 3:1.

55 Zelen, M. 1968. Discussion of Searle (1968). Biometrics 24:779.