C.R. HENDERSON THE STATISTICIAN; AND HIS CONTRIBUTIONS TO VARIANCE COMPONENTS ESTIMATION

Shayle R. Searle

BU-1085-MA

June 1990

C.R. HENDERSON THE STATISTICIAN; AND HIS CONTRIBUTIONS TO VARIANCE COMPONENTS ESTIMATION*

Shayle R. Searle
Biometrics Unit
N.Y. State College of Agriculture and Life Sciences
Cornell University, Ithaca, N.Y.

ABSTRACT

C. R. Henderson's 1953 *Biometrics* paper "Estimation of variance and covariance components" 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 1-way classification, or completely randomized design. As such it had two important, long-lasting impacts. First, it provided methods for actually using unbalanced data, even in large quantity, for estimating variance components. And this 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 1960's and early 1970's. 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 "Chuck", in his later years, 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.

*An invited paper for the Symposium "The Legacy of C.R. Henderson" at the 85'th Annual Meeting of the American Dairy Science Association, Raleigh, North Carolina, June 25-26, 1990.

1. THE HENDERSON LANDMARK

The early beginnings of estimating variance components date back to the work of two astronomers: Airy (1861) and Chauvenet (1863). Modern-day work started with Fisher (1925), Fisher (1938), Cochran (1939), Daniels (1939) and Winsor and Clarke (1940). In all of these papers we see the beginnings of what today is called the ANOVA (analysis of variance) method of estimating variance components, based on equating expected mean squares to their values computed from data. And as an estimation method it was given considerable impetus by the Anderson and Bancroft (1952) book wherein it is considered at some length. More details of this history are to be found in Searle (1989).

Most of this work had dealt with what are now called balanced data – data having equal numbers of observations in the subclasses. Cochran (1939), Daniels (1939) and Winsor and Clarke (1940) had considered the case of unbalanced data (having unequal numbers of observations in the subclasses) from a 1-way classification, and Ganguli (1941) dealt comprehensively with unbalanced data from completely nested models, e.g., cows within herds and, say, 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 A.I. sires in different herds – where sires are crossed 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 Ray Henderson (1911-1989) addressed in his classic 1953 paper in *Biometrics* "Estimation of variance and covariance components." It was to be a landmark in the history of estimating variance components.

2. 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: e.g., in the decade of 1978-1988, namely 25-35 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 had all his university education at Iowa State College (now University) in Ames. Following his M.S. in animal nutrition he worked first with the Iowa Extension Service and then Ohio University, and in 1942 joined the U.S. Army where he finished up 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 eighteen papers of 1944-1948 on which he is a co-author. Number three of those, for example, is "Certain statistical considerations in patch testing", published in the *Journal of Investigative Dermatology*; and almost all of the others 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 Lush, Lenoy 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 1940's and early 1950's, 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

$$y_{ij} = \mu + s_i + e_{ij} \tag{1}$$

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 of sire i, and $\bar{y}_{..} = \sum_{i=1}^{b} \sum_{j=1}^{n} y_{ij}/bn$ is the mean record of all n daughters of all b sires.

Table 1. Sums of squares 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	$SST = \sum_{i=1}^{b}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2$	

Then, with sires considered random, the expected mean squares in the usual variance component form of (1) are

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

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

$$MSB = n\hat{\sigma}_s^2 + \hat{\sigma}_e^2 \quad \text{and} \quad MSW = \hat{\sigma}_e^2 \; , \tag{3}$$

and these yield the ANOVA estimates

$$\hat{\sigma}_s^2 = \tfrac{1}{n}(MSA\text{-}MSE) \quad \text{and} \quad \hat{\sigma}_e^2 = MSE \; . \tag{4}$$

This is the ANOVA method of estimating variance components at its simplest: for the 1-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).

3. HENDERSON'S THREE METHODS

The philosophy of Henderson (1953) 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 1950's, a year-by-herd-by-sire analysis with years as fixed effects and herd and sires as random effects. Thus, one year's records might have the appearance of Table 2.

Table 2. Number of daughter records in each herd for 60 A.I. sires used in 800 herds.

Sire	Herd								Total
	1	2	3	4	5	6	...	800	
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

What Henderson gave us was three different sets of mean squares (or sums of squares – effectively equivalent for the purpose at hand) that could be used in the basic ANOVA method of estimation, namely

$$\text{equate calculated MS to E(MS)} \tag{5}$$

or

$$\text{equate calculated SS to E(SS)} \ . \tag{6}$$

The three different sets of sums of squares came to be known as Henderson's Methods I, II and III. 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 I. 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

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

for balanced data becomes, for unbalanced data,

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

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

$$\begin{matrix} 2 & 4 \\ 10 & 4,40 \end{matrix} \quad \text{expression (8) is } -22 \ . \tag{9}$$

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

Method III. 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, 1971) that the hypotheses being so tested do, in fact, differ from those which they were once thought to be.

Each of these methods has its merits and demerits. Method I 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 through ignoring the fixed effects, or assuming them to be random yields biased estimators of the variance components. Method II 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 II 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 III 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 III procedure, but Method III gives no indication as to which is preferable – nor 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).

4. 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, since all of them are just applications of the most general form of ANOVA estimation, which can be typified as $E(\mathbf{q}) = \mathbf{P}\boldsymbol{\sigma}^2$ giving $\hat{\boldsymbol{\sigma}}^2 = \mathbf{P}^{-1}\mathbf{q}$ where $\mathbf{q}$ is a vector of quadratic forms of the data and $\boldsymbol{\sigma}^2$ the vector of components to be estimated. Hence $E(\hat{\boldsymbol{\sigma}}^2) = \mathbf{P}^{-1}E(\mathbf{q}) = \mathbf{P}^{-1}\mathbf{P}\boldsymbol{\sigma}^2 = \boldsymbol{\sigma}^2$. What the Henderson methods do is provide suggestions for the elements of $\mathbf{q}$.

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 thing, namely to yielding the ANOVA estimators of variance components which, for balanced data, are minimum variance unbiased (MVU) under normality assumptions, Graybill and Wortham (1956), and even without normality are minimum variance quadratic unbiased, MVQU, Graybill and Hultquist (1961).

Henderson's development of his three methods of estimating variance components soon led to their being used extensively in many different applications – and 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 namely "The relative sizes of the sampling variances of estimates obtained by the methods are not known." (Henderson, 1953, 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), which extended Crump (1951), Searle (1958, 1961), Mahamunulu (1963), Low (1964), Hirotsu (1966), Blischke (1966, 1968) and Rhode and Tallis (1969) 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 the resulting estimators, 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 1-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

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

This expression is not amenable to analysis; i.e., one cannot, no matter what the values of σ_s^2 and σ_e^2 are, study the behavior of $\operatorname{var}(\hat{\sigma}_s^2)$ 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 error mean squares) 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 (e.g. Bainbridge, 1963, Bush and Anderson, 1963, Anderson and Crump, 1967, and Anderson, 1975) are mostly for small hypothetical data layouts that give little or no relevant information for the kind of large data sets that animal breeders so often deal with.

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 (1968), Zelen (1968), and Thompson (1969), for example, suggest extensions of Method III, Searle (1968) provides proof that Method II cannot be used on models that have interactions between fixed and random effects, and Henderson, Searle and Schaeffer (1974) show how Method II can be computed and that it has an essential invariance property (in contrast to Searle, 1971, 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 1950's onwards 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.

5. BLUP AND HENDERSON'S MMEs

One of Chuck's most important contributions not only for animal breeders but equally so for statisticians was in a paper with Oscar Kempthorne and others, Henderson ***et al*** (1959). The paper has an interesting format because it is written in three sections, each with its own author(s). The importance of the paper is largely in Henderson's section where the estimation of variance components is not the emphasis but rather the derivation of the best linear unbiased predictor (BLUP) of random effects. Equations now known as the mixed model equations (MMEs), 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 a statistics course from Alexander Mood wherein he first came upon the following exercise that later appeared in Mood (1950, p. 164, exercise 23). With small changes it is also to be found in Graybill (1963, p. 195, exercise 32), and in Mood, Graybill and Boes (1974, 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 dairy cattle. 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 had given 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, Henderson (1950), a phrase that, Chuck once told me, went down like a lead balloon. Statisticians simply do not estimate random variables. And that, along with Cochran (1951), 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:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e} .$$

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

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

$$\mathbf{V} = \text{var}(\mathbf{y}) = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R} .$$

Then the best linear unbiased estimator of $\mathbf{X}\beta$ is

$$\text{BLUE}(\mathbf{X}\beta) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\ , \tag{12}$$

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

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

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

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

which he called the mixed model equations.

Then

$$\mathbf{X}\hat{\beta} = \text{BLUE}(\mathbf{X}\beta) \tag{15}$$

and

$$\tilde{\mathbf{u}} = \text{BLUP}(\mathbf{u})\ . \tag{16}$$

Equations (14), the mixed model equations, 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 equations (14) have order equal to the number of fixed and random effects occurring in the data – usually much less than N. Moreover R^{-1} and D^{-1} in (14) are usually easy to calculate, because R is commonly just $\sigma_e^2 I$, and D is often diagonal. Second, equations (14) yield $\hat{\beta}$ and $\tilde{u}$ simultaneously. And then there are the two other properties of $\tilde{u}$ that are so useful to animal breeders: one is the good ranking property of $\tilde{u}$, that in ranking sires by their corresponding values in $\tilde{u}$ one is maximizing the probability of correctly ranking the sires, as in Portnoy (1982); and the other useful feature of $\hat{\beta}$ and $\tilde{u}$ is that provided $k'\beta$ is estimable (i.e., $k' = t'X$ for some t'), then the best estimator of $k'\beta + p'u$ is $k'\hat{\beta} + p'\tilde{u}$. And finally, of course, through their reliance on the variance components of whatever mixed model is being used, the MMEs provide solid motivation for wanting to estimate variance components.

These then, are Henderson's main specific contributions to statistics: timely methods in the early 1950's for estimating variance components from unbalanced data; and development of the BLUP procedure and affiliated methodology. These contributions are substantial. No one before Chuck 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.

6. 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 (1967), has been the application of the long-established method of maximum likelihood – a method of estimation which, in almost all circumstances, yields estimators that have 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 (1951), it was not applied to the variance component problem until Hartley and Rao (1967), their success undoubtedly being due to their matrix formulation of the problem. With the model equation

$$y = X\beta + Zu + e \tag{17}$$

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

$$Zu = \sum_{i=1}^{r} Z_i u_i$$

for r random factors. Moreover, on defining

$$e = u_0 \quad \text{and} \quad I = Z_0 \tag{18}$$

the model equation becomes

$$y = X\beta + \sum_{i=0}^{r} Z_i u_i \, . \tag{19}$$

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

$$P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-} X'V^{-1} \tag{20}$$

the maximum likelihood equations for the variance components can be adapted (e.g., Searle, 1979) from the Hartley - Rao equations to be written succinctly as (see Searle, 1987, Sec. 7.1a for notation)

$$\left\{_m \ \mathrm{tr}(\tilde{V}^{-1}Z_iZ_i'\tilde{V}^{-1}Z_jZ_j')\right\}_{i,j=0}^{r} \left\{_c \ \tilde{\sigma}_j^2\right\}_{j=0}^{r} = \left\{_c \ y'\tilde{P}Z_iZ_i'\tilde{P}y\right\}_{i=0}^{r} \, . \tag{21}$$

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 $\tilde{V}^{-1}$, of course, and also, from (20), in $\tilde{P}$. So the equations are very non-linear, and in general do not yield closed-form solutions. Nevertheless, the asymptotic sampling dispersion matrix of the variance component estimators is known and is

$$\mathrm{var}(\tilde{\sigma}^2) \doteqdot \left\{_m \ \mathrm{tr}(V^{-1} Z_i Z_i'V^{-1}Z_jZ_j')\right\}_{i,\,j=0}^{r} \, . \tag{22}$$

A variant of ML 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, "error contrasts" as Harville (1977) calls them. This results in equations

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

with asymptotic dispersion matrix

$$\mathrm{var}(\hat{\sigma}) \doteqdot \left\{_m \ \mathrm{tr}(PZ_iZ_i' \ PZ_jZ_j')\right\}_{i,j=0}^{r} \, . \tag{24}$$

Whichever one uses, ML 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:

(i) Using an optimum form of the equation.

(ii) Developing an optimum iteration technique.

(iii) Using sparse matrices of very large order.

(iv) Having no intermediate values of $\tilde{V}$ be singular.

(v) Being sure that choice of starting values does not effect the final solution.

(vi) Reaching a global, not local, maximum for the likelihood.

(vii) Being certain that convergence is achieved.

(viii) Achieving $\tilde{\sigma}_0^2 > 0$ and $\tilde{\sigma}_i^2 \geq 0$ for $i=1,\cdots,r$.

(ix) 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 non-linear equations. Furthermore, ML 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 ML was going to become acceptable, he had soon started working with ML, and after 1977 was publishing comparative studies involving ML and REML. (See, for example Henderson and Quaas, 1977, and Rothschild and Henderson and Quaas, 1978, and Rothschild and Henderson, 1979). These studies also included comparisons with the MINQUE estimation method of Rao (1970, 1971a,b and 1972) which requires using a set of prior values of the components as part of the estimation process. MINQUE estimates are then functions of those prior values. Brown (1976), 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 wrote an unpublished paper, Henderson (1973) in which he shows how ML calculations can be achieved using elements of his MMEs, (14). This was a remarkable piece of work, allied as it was to both Patterson and Thompson (1971) in their original development of REML and to the later Dempster *et al* (1977) in their description of what is now called the EM algorithm. Henderson's approach, by way of his mixed model equations, avoided some of the computing problems listed earlier, notably that, as shown by Harville (1977), it always gives positive values for the estimated components.

7. AN OVERVIEW

A list of C.R. Henderson's publications contains 233 items, being one book, one thesis, 48 abstracts and 183 papers. In a rough and subjective classification of these papers I labelled 94 of them as primarily statistical, and 41 of those as 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 (Henderson, 1953). 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. This encouraged Chuck, I believe, to continue promoting the mixed model, to his colleagues and his students, until (and as the 1953 paper's third, and possibly greatest, effect) his ideas reached a statistical audience far far wider than just the users of the mixed model and variance components. In the 1950's and early 1960's interest in the mixed model was primarily among biometricians connected with genetics. (I remember in 1957 being at a 6-week NSF 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. And much of this interest stems, I'm sure, from the man who got to the problems first: Charles Ray Henderson.

Acknowledgement

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-May, 1990.

REFERENCES

Airy, G.B. (1861) *On the Algebraical and Numerical Theory of Errors of Observations and the Combinations of Observations.* London, MacMillan.

Anderson, R.L. (1975) Designs and estimators for variance components. In: *Statistical Design and Linear Models*, Ed. J.N. Srivastava, 1-30. Amsterdam, North Holland.

Anderson, R.L. and Bancroft, T.A. (1952) *Statistical Theory in Research.* McGraw-Hill, New York.

Anderson, R.L. and Crump, P.P. (1967) Comparisons of designs and estimation procedures for estimating parameters in a two-stage nested process. *Technometrics* 9, 499-416.

Bainbridge, T.R. (1963) Staggered nested designs for estimating variance components. *American Society for Quality Control Annual Conference Transactions* 93-103.

Blischke, W.R. (1966) Variances of estimates of variance components in a three-way classification. *Biometrics* 22, 553-565.

Blischke, W.R. (1968) Variances of moment estimators of variance components in the unbalanced r-way classification. *Biometrics* 24, 527-540.

Brown, K.G. (1976) Asymptotic behavior of MINQUE-typed estimators of variance components. *Annals of Statistics* 4, 746-754.

Bush, N. and Anderson, R.L. (1963) A comparison of three different procedures for estimating variance components. *Technometrics* 5, 421-440.

Chauvenet, W. (1863) *A Manual of Spherical and Practical Astronomy, 2: Theory and Use of Astronomical Instruments.* Lippincott, Philadelphia.

Cochran, W.G. (1939) The use of analysis of variance in enumeration by sampling. *Journal of the American Statistical Association* 34, 492-510.

Cochran, W.G. (1951) Improvement by means of selection *Proceedings Second Berkeley Symposium*, 499-470.

Crump, S.L. (1951) The present status of variance components analysis. *Biometrics* 7, 1-16.

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

Daniels, H.E. (1939) The estimation of components of variance. *Journal of the Royal Statistical Society, Supplement.* 6, 186-197.

Dempster, A.P., Laird, N.M. and Rubin D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* (B) 39, 1-38.

Fisher, R.A. (1925, 1938) *Statistical Methods for Research Workers, 1st and 7th editions.* Oliver and Boyd, Edinburgh and London.

Fisher, R.A. (1935) Discussion of Neyman *et al.* (1935). *Journal of the Royal Statistical Society*, Series B, 2, 154-155.

Ganguli, M. (1941) A note on nested sampling. *Sankhyā* 5, 449-452.

Graybill, F.A. and Hultquist, R.A. (1961) Theorems concerning Eisenhart's Model II. *Annals of Mathematical Statistics* 32, 261-269.

Graybill, F.A. and Wortham, A.W. (1956) A note on uniformly best unbiased estimators for variance components. *Journal of the American Statistical Association* 51, 266-268.

Hartley, H.O. and Rao, J.N.K. (1967) Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 54, 93-108.

Harville, D.A. (1977) Maximum-likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72, 320-340.

Henderson, C.R. (1950) Estimation of genetic parameters. *Annals of Mathematical Statistics* 21, 309 (abstract).

Henderson, C.R. (1953) Estimation of variance and covariance components. *Biometrics* 9, 226-252.

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, C.R., Searle, S.R. and Schaeffer, L.R. (1974) The invariance and calculation of Method 2 for estimating variance components. *Biometrics* 30, 583-588.

Henderson, C.R. and Quaas, R.L. (1977) A comparison of Method 3 and maximum likelihood estimators of heritabilities and genetic correlation. *Journal of Animal Science* (abstract), p. 22.

Henderson, C.R. (1973) Maximum likelihood estimation of variance components. Technical Report, Department of Animal Science, Cornell University.

Hirotsu, C. (1966) Estimating variance components in a two-way layout with unequal numbers of observation. *Reports of Statistical Applications, Union of Japanese Scientists and Engineers* (JUSE), 13, No. 2, 29-34.

Low, L.Y. (1964) Sampling variances of estimates of components of variance from a non-orthogonal two-way classification. *Biometrika* 51, 491-494.

Mahamunulu, D.M. (1963) Sampling variances of the estimates of variance components in the unbalanced three-way nested classification. *Annals of Mathematical Statistics* 34, 521-527.

Mood, A.M. (1950). *Introduction to the Theory of Statistics.* McGraw-Hill, New York.

Mood, A.M. and Graybill, F.A. (1963) *Introduction to the Theory of Statistics*, 2nd ed. McGraw-Hill, New York.

Mood, A.M., Graybill, F.A. and Boes, D.C. (1974) *Introduction to the Theory of Statistics*, 3rd ed. McGraw-Hill, New York.

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

Portnoy, S. (1982) Maximizing the probability of correctly ordering random variables using linear predictors. *J. Multivariate Analysis* 12, 256-269.

Rothschild, M.F, Henderson, C.R. and Quaas, R.L. (1978) Comparison of maximum likelihood and Henderson's Method I estimates of variances and covariances for two traits when selection is present. *Journal Dairy Science Supplement* 1, 91 (Abstract).

Rothschild, M.F. and Henderson, C.R. (1979) Maximum likelihood estimation of population parameters of first and second lactation milk records. *Journal Dairy Science* 62, 603.

Rao, C.R. (1970) Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association* 65, 161-172.

Rao, C.R. (1971a) Estimation of variance and covariance components – MINQUE theory. *Journal of Multivariate Analysis* 1, 257-275.

Rao, C.R. (1971b) Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis* 1, 445-446.

Rao, C.R. (1972) Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association* 67, 112-115.

Rohde, D.A. and Tallis, G.M. (1969) Exact first- and second-order moments of estimates of components of covariance. *Biometrika* 56, 517-525.

Searle, S.R. (1956) Matrix methods in components of variance and covariance analysis. *Annals of Mathematical Statistics* 27, 737-748.

Searle, S.R. (1958) Sampling variance of estimates of components of variance. *Annals of Mathematical Statistics* 29, 167-178.

Searle, S.R. (1961) Variance components in the unbalanced two-way nested classification. *Annals of Mathematical Statistics* 32, 1161-1166.

Searle, S.R. (1968) Another look at Henderson's methods of estimating variance components. *Biometrics* 24, 749-778

Searle, S.R. (1971) *Linear Models.* Wiley, New York.

Searle, S.R. (1989) Variance components – some history and a summary account of estimation methods. *Journal of Animal Breeding and Genetics* 106, 1-29.

Searle, S.R. (1979) Notes on variance component estimation. Technical Report BU-673-M, Biometrics Unit, Cornell University.

Searle, S.R. (1987) *Linear Models for Unbalanced Data.* Wiley, New York.

Thompson, R. (1969) Iterative estimation of variance components for non-orthogonal data. *Biometrics* 25, 767-773.

Winsor, C.P. and Clarke, G.L. (1940) Statistical study of variation in the catch of plankton nets. *Sears Foundation Journal of Marine Research* 3, 1-34.

Zelen, M. (1968) Discussion of Searle (1968). *Biometrics* 24, 779-780.