The Annals of Statistics
1984, Vol. 12, No. 4, 1488–1499

ON ANALYSIS OF VARIANCE IN THE MIXED MODEL

BY K. G. BROWN

National Institute of Environmental Health Sciences

An *analysis of variance* (*ANOVA*) is defined to be a partition of the total sum of squares into independent terms which, when suitably scaled, are chi-squared variables. A partition of less than the total sum of squares, but with these properties, will often suffice and is referred to as a *partial ANOVA*. Conditions for an ANOVA, and for partial ANOVAs selected to contain only specific parameters, are given. Implications for estimation of variance components from an ANOVA are also discussed. These results are largely an extension of work by Graybill and Hultquist (1961).

With unbalanced data, conditions for an ANOVA and the number of terms in it both can depend on which effects in the model are fixed and which are random. This is not taken into account by those procedures for partitioning a sum of squares which distinguish between random and fixed effects only in the calculation of expected mean squares. Several examples are given.

1. Introduction. The term "analysis of variance" was introduced by Fisher (1918), who is responsible for its inception as a methodology and much of its development as practiced today. At the time of writing of Scheffé (1959), a general theory and treatment of the fixed effects model as elucidated by him was judged to be in "fairly permanent form," while the state of the art with respect to random and mixed models was less developed. Shortly thereafter Graybill and Hultquist (1961) presented a fairly general theory for the random effects model.

In the presence of random effects, the problem is one of determining when an ANOVA exists and then generating it; the expected mean squares dictate what inferences, by way of hypothesis testing and point or interval estimation, can be drawn. This contrasts with the fixed effects case in which hypotheses to be tested are generally decided in advance and the total sum of squares is partitioned accordingly. In this paper we attempt to combine the separate approaches to fixed and random effects models into a unified treatment suitable for study of the analysis of variance in the mixed model. Within such a framework, fixed effects and random effects models can be viewed as special cases.

The mixed model is described by

$$Y = X\beta + \varepsilon \tag{1.1}$$

where Y is a random n-vector, X is a given $n \times m$ matrix assumed (for convenience) to be of full column rank, β is an unknown m-vector of parameters. Further, the error term has the linear structure

$$\varepsilon = U_1\xi_1 + \cdots + U_k\xi_k$$

where U_i is a given $n \times k_i$ matrix, $U_k = I$, $\xi_i \sim N(0, I\sigma_i^2)$, and terms ξ_i and ξ_j are

Received June 1980; revised April 1983.

AMS 1980 *subject classifications*. Primary 62J10, secondary 62E10.

Key words and phrases. Analysis of variance, mixed model, chi-squared distribution, variance components.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*. www.jstor.org

independent, $i \neq j$. Then $V(\varepsilon) = \Sigma = \sum_{i=1}^{k} V_i\sigma_i^2$, where $V_i = U_iU_i'$. Letting $\sigma^2 = (\sigma_1^2, \cdots, \sigma_k^2)'$, the parameter space of σ^2 is defined to be

$$\Omega = \{\sigma^2: \sigma_i^2 \geq 0, i = 1, \cdots, k-1, \sigma_k^2 > 0\}.$$

The elements of X and U_i are not limited to zeroes and ones, and for most of what follows the covariance matrix of ε could be replaced by the more general linear structure in Anderson (1973).

Analysis of variance (ANOVA) is formally defined as a partition of $Y'Y$ into symmetric quadratic forms which

(a) are mutually independent,

(b) when suitably scaled are distributed as (possibly noncentral) chi-squared variables,

(c) have expectations which are different (beyond a known multiplicative constant) parametric functions.

Requirement (c) implies a minimum number of terms. If A is a known nonnegative definite matrix such that $Y'AY < Y'Y$ with probability one, then a partition of $Y'AY$ with the properties of an ANOVA (except for summing to $Y'AY$ instead of $Y'Y$) will be called a *partial ANOVA*.

2. Full and partial ANOVAs. Let P be the perpendicular projector onto $\mu(X)$, the range of X, and define $T = Y'Y$, $R = Y'PY$, $S = T - R$. When $m > 1$ it is generally of interest to partition β into some given set of subvectors, say $\beta = (\beta_1, \cdots, \beta_t)'$, and write $R = R_1 + \cdots + R_t$, where R_1 is the (least squares) sum of squares for β_1, and R_i, for $i > 1$, is the sum of squares for β_i adjusted for $\beta_1, \cdots, \beta_{i-1}$. Let $P_1, \cdots, P_t$ be the perpendicular projection matrices such that $R_i = Y'P_iY$, $i = 1, \cdots, t$, and $P = P_1 + \cdots + P_t$, $P_iP_j = 0$, $i \neq j$. It is well known that if $k = 1$ then $\{R_1, \cdots, R_t, S\}$ is an ANOVA.

If $k > 1$, then it is generally necessary to partition S further, say into quadratic forms denoted by $S_1, \cdots, S_s$, to form an ANOVA. If, for example, the objective is to test an hypothesis such as $\beta_t = 0$, then an F-statistic can be formulated if one of the S_i terms, say S_1, is such that the expected mean squares of R_t and S_1 are identical under the hypothesis. There is no guarantee that such a match is possible; however, it will be shown in Theorem 1 that $\{S_1, \cdots, S_s\}$ is unique up to order. Thus, the question of which partition of S is preferable does not arise, only whether one satisfying the ANOVA definition exists at all.

ANOVA$(\beta_1, \cdots, \beta_t, \sigma^2)$ will refer to a set of terms $\{R_1, \cdots, R_t, S_1, \cdots, S_s\}$, where $\sum_{i=1}^{s} S_i = S$, which form an ANOVA. The first part of the following theorem is essentially an extension of Theorem 6 of [9] to the mixed model.

THEOREM 1. *An* ANOVA$(\beta_1, \cdots, \beta_t, \sigma^2)$ *exists iff* a) $P_1, \cdots, P_t, V_1, \cdots, V_k$ *commute, and* b) *for each* $i = 1, \cdots, t$, *the nonzero characteristic roots of* $P_i\Sigma P_i$ *(which may depend on* σ^2*) are all the same. An* ANOVA$(\beta_1, \cdots, \beta_t, \sigma^2)$ *is unique up to order for any given choice of* $t, R_1, \cdots, R_t$.

REMARK. Theorem 1 is about ANOVAs that can be formulated by first partitioning $Y'Y$ into R and S ($Y'PY$ and $Y'(I-P)Y$), and then partitioning R and S further. If P does not commute with each V_i (the n.s. condition for a BLUE of an estimable parametric function), then R and S are not independent and Theorem 1 is not applicable for any choice of $R_1, \cdots, R_t$. If P does commute with each V_i, and $V_1, \cdots, V_t$ commute as well, then there is a unique partition of S (not dependent on the value of t) with the independent and chi-squared properties. Assuming one is successful to this point, then one only needs to check if $R_1, \cdots, R_t$ form an ANOVA for $Y'PY$. If we drop the requirement that an ANOVA be formed by partitioning R and S, then the commutativity of $V_1, \cdots, V_k$ is n.s. for there to be at least one ANOVA (of unspecified form) of $Y'Y$ [5].

PROOF (*Sufficiency*). Let $\delta(\sigma^2)_i$ be the nonzero characteristic root of $P_i\Sigma P_i$ with multiplicity $m_i = \text{rank}(P_i\Sigma P_i) = \text{rank}(P_i)$, $i = 1, \cdots, t$. By definition, $P = P_1 + \cdots + P_t$ and the P_i are disjoint perpendicular projection matrices. Let L_i be an $n \times m_i$ matrix such that $L_i'L_i = I$ and

$$P_i\Sigma P_i L_i = \delta(\sigma^2)_i L_i, \quad i = 1, \cdots, t. \tag{2.1}$$

Multiplying on the left by P_i gives $P_i\Sigma P_i L_i = \delta(\sigma^2)_i P_i L_i = \delta(\sigma^2)_i L_i$, which implies that $P_i L_i = L_i$. It follows that $P_i = L_i L_i'$, $i = 1, \cdots, t$. By commutativity, $P_i\Sigma P_i = \Sigma P_i P_i = \Sigma P_i$. Multiply (2.1) on the right by L_i' to get

$$\Sigma P_i = \delta(\sigma^2)_i P_i. \tag{2.2}$$

Letting $Q = I - P$ gives

$$\Sigma = \Sigma P + \Sigma Q = \Sigma P_1 + \cdots + \Sigma P_t + \Sigma Q = \delta(\sigma^2)_1 P_1 + \cdots + \delta(\sigma^2)_t P_t + \Sigma Q.$$

By commutativity, $V_i Q = Q V_i Q$, which is symmetric, and $V_1 Q, \cdots, V_k Q$ commute. This implies there exists a matrix K such that $K' = K^{-1}$ and $KV_iQK' = D_i$ is diagonal, $i = 1, \cdots, k$ (Theorem J of [8]). Then $K\Sigma QK' = \sum_{i=1}^k D_i\sigma_i^2$. Let $\delta(\sigma^2)_{t+1}, \cdots, \delta(\sigma^2)_u$ denote the different parametric functions on the diagonal of $\sum_{i=1}^k D_i\sigma_i^2$ (excluding the function of multiplicity m which is identically zero) and let $m_{t+1}, \cdots, m_u$ be their respective multiplicities. (Note that $\delta(\sigma^2)_i$ and $\delta(\sigma^2)_j$ may be identical parametric functions if $i, j \le t$ or $i \le t$ and $j > t$, but not if $i, j > t$.) Let L_i be the $n \times m_i$ submatrix of K' consisting of characteristic vectors associated with $\delta(\sigma^2)_i$, $i = t+1, \cdots, u$. Then

$$\Sigma Q = L_{t+1}L_{t+1}'\delta(\sigma^2)_{t+1} + \cdots + L_u L_u'\delta(\sigma^2)_u.$$

Define $C_i = L_i L_i'$, $i = 1, \cdots, u$. (Note that $C_i = P_i$, $i = 1, \cdots, t$). Thus

$$\Sigma = C_1\delta(\sigma^2)_1 + \cdots + C_u\ \delta(\sigma^2)_u, \tag{2.3}$$

where C_i has rank m_i, $i = 1, \cdots, u$. By construction, $C_1, \cdots, C_u$ are symmetric, idempotent, and disjoint, and $\sum_{i=1}^u C_i = I$. Multiplying (2.3) on the right by C_i gives $\Sigma C_i = C_i\delta(\sigma^2)_i$, $i = 1, \cdots, u$. It follows from Theorem 1 of [1] that $Y'C_1Y, \cdots, Y'C_uY$ are independent, and $Y'C_iY\delta(\sigma^2)_i^{-1}$ has a chi-squared distribution with m_i degrees of freedom and noncentrality parameter

$$\lambda_i = \beta' X' C_i X \beta \delta(\sigma^2)_i^{-1}, \; i = 1, \cdots, u.$$

For $i > t$, $C_i P = 0$ which implies $\lambda_i = 0$. Thus $\{R_1, \cdots, R_t, S_1, \cdots, S_s\}$ is an ANOVA$(\beta_1, \cdots, \beta_t, \sigma^2)$ where $R_i = Y'C_iY$, $i = 1, \cdots, t$ and $S_j = Y'C_{j+t}Y$, $j = 1, \cdots, s$, $s = u - t$.

(*Necessity*). Let $\{R_1, \cdots, R_t, S_1, \cdots, S_s\}$ be an ANOVA$(\beta_1, \cdots, \beta_t, \sigma^2)$. For $i = 1, \cdots, t$, define $C_i = P_i$; for $i = t + 1, \cdots, u$, where $u = t + s$, let C_i be a symmetric matrix such that $Y'C_iY = S_{i-t}$. Let $\delta(\sigma^2)_i$, $i = 1, \cdots, u$, be scale factors such that $\delta(\sigma^2)_i^{-1} Y'C_iY$ has a chi-squared distribution for all possible values of $\sigma^2 \in \Omega$. Theorem 9.2.1(i) of [17] and the nonsingularity of Σ imply

$$C_i \Sigma C_i = C_i \delta(\sigma^2)_i, \quad i = 1, \cdots, u. \tag{2.4}$$

By Theorem 9.4.1(d) of [13],

$$C_i \Sigma C_j = 0, \quad i \neq j, \quad i, j = 1, \cdots, u. \tag{2.5}$$

Identities (2.4) and (2.5) hold for all possible values of $\sigma^2 \in \Omega$, which implies they hold if $\Sigma = I$. Thus $C_iC_j = 0$, $i \neq j$. Multiply $I = C_1 + \cdots + C_u$ on the left by C_i to get $C_i = C_i^2$, $i = 1, \cdots, u$. Thus $C_1, \cdots, C_u$ are disjoint perpendicular projection matrices.

Identity (2.4) implies that $\delta(\sigma^2)_i$ is a nonzero characteristic root of $C_i \Sigma C_i$ and the columns of C_i are characteristic vectors associated with it. The multiplicity of $\delta(\sigma^2)_i$ is $\text{rank}(C_i) = \text{rank}(C_i \Sigma C_i)$, so the remaining characteristic roots of $C_i \Sigma C_i$ are zero, and this is true for all possible values of $\sigma^2 \in \Omega$. This establishes condition (b).

Using (2.4) and (2.5), $\sum_{i=1}^{u} \sum_{j=1}^{u} C_i \Sigma C_j =$

$$\Sigma = C_i \delta(\sigma^2)_1 + \cdots + C_u \delta(\sigma^2)_u. \tag{2.6}$$

Thus $V_1 \sigma_1^2 + \cdots + V_k \sigma_k^2 = C_1 \delta(\sigma^2)_1 + \cdots + C_u \delta(\sigma^2)_u$, for all possible values of $\sigma^2 \in \Omega$, which implies each V_i is a linear combination of $C_1, \cdots, C_u$. Let $\mathscr{V}$ be the linear space spanned by $C_1, \cdots, C_u$. Since $C_1, \cdots, C_u$ are idempotent and disjoint, $\mathscr{V}$ is commutative ($A, B \in \mathscr{V} \Rightarrow AB = BA$). The matrices $P_1, \cdots, P_t$, $V_1, \cdots, V_k$ are contained in $\mathscr{V}$ (recall $P_i = C_i$, $i = 1, \cdots, t$), so they commute, which establishes (a).

(*Uniqueness*). Let $\{R_1, \cdots, R_t, S_1, \cdots, S_s\}$ be an ANOVA$(\beta_1, \cdots, \beta_t, \sigma^2)$ with C_i and $\delta(\sigma^2)_i$ defined as in the necessity part of the proof. For $i = 1, \cdots, t$, $C_i = P_i$ is given and $Y'C_iY\delta(\sigma^2)_i^{-1}$ has a chi-squared distribution which implies that the (unknown) value of $\delta(\sigma^2)_i$ is unique. Thus $\Delta = \sum_{i=1}^{t} C_i \delta(\sigma^2)_i$ is uniquely determined. Using (2.6) gives

$$\Sigma - \Delta = C_{t+1} \delta(\sigma^2)_{t+1} + \cdots + C_u \delta(\sigma^2)_u, \tag{2.7}$$

which implies that $\delta(\sigma^2)_{t+1}, \cdots, \delta(\sigma^2)_u$ are nonzero characteristic roots of $\Sigma - \Delta$. The definition of ANOVA implies that $\delta(\cdot)_{t+1}, \cdots, \delta(\cdot)_u$ are different functions defined on Ω, but the values of $\delta(\sigma^2)_{t+1}, \cdots, \delta(\sigma^2)_u$ may not be distinct. Let i be fixed with $i > t$ and define τ_i to be the vector space formed by taking the intersection of the eigenspaces of $\delta(h)_i$, as h ranges over Ω. Then $\tau_i = \mu(C_i + \sum_{j \in G_i} C_j)$ where $G_i = \{k: \delta(h)_i = \delta(h)_k$ for all $h \in \Omega\}$ and $\mu(\cdot)$ indicates range. As $\delta(\cdot)_{t+1}, \cdots, \delta(\cdot)_u$ are different functions defined on Ω, it follows that G_i is void. Thus $\tau_i = \mu(C_i)$, and since C_i is idempotent and symmetric, it follows that C_i is uniquely the perpendicular projector onto τ_i. □

The random effects model discussed by Graybill and Hultquist (1961) is obtained from (1.1) when $t = 1$, X is a column of ones, and β is a scalar generally denoted by μ. Their definition of ANOVA corresponds to our ANOVA(μ, σ^2) plus two additional requirements: s (the number of S_i terms) must equal k (the number of variance components); the scale factors to make $S_1, \cdots, S_s$ chi-squared variables must be different linear parametric functions. The first requirement has been dropped here because s is unique (Theorem 1) and thus will equal k automatically if possible. The second requirement is actually implicit here. The scale factor of S_i is uniquely $E(S_i)$, which is a linear parametric function of σ^2, and the definition of ANOVA implies that $E(S_i)$ are different parametric functions. The condition in our Theorem 1 on the characteristic roots of $P_i \Sigma P_i$ does not appear in their Theorem 6 because it is always satisfied when $m = 1$.

The conditions of Theorem 1 are very strong and one way in which they might not be satisfied is if R_i, for one or more values of i, violates the chi-squared or independence requirement of an ANOVA. For a problem such as the one above of testing H: $\beta_t = 0$, this is inconsequential for all R_i, $i \neq t$. Thus, it is helpful to know when a partial ANOVA consisting of only the terms $R_t, S_1, \cdots, S_s$ is possible. *ANOVA*(β_t, σ^2) will refer to a set of terms $\{R_t, S_1, \cdots, S_s\}$ where $\sum_{i=1}^{s} S_i = S$, which form a partial ANOVA. Let L be a matrix such that $L'L = I$ and $\mu(L) = \mu(P - P_t)^{\perp}$ where $\perp$ means orthogonal complement. Then $L'Y \sim N(L'X_t\beta_t, \sum_{i=1}^{k} L'V_iL\sigma_i^2)$, where X_t consists of the columns of X corresponding to β_t. Applying Theorem 1 to the model transformed by L' gives the following corollary.

COROLLARY 1. *An* ANOVA(β_t, σ^2) *exists iff* a) $P_t^*, L'V_1L, \cdots, L'V_kL$ *commute, and* b) *the nonzero characteristic roots of* $P_t^* L'\Sigma L P_t^*$ *are all the same parametric function of* σ^2, *where* P_t^* *is the perpendicular projector onto* $\mu(L'X_t)$. *An* ANOVA(β_t, σ^2) *is unique up to order for any given choice of* t, R_t.

In practice, Corollary 1 would be applied by taking $t = 2$ and defining β_1 and β_2 such that β_2 consists of the elements of β to be retained and β_1 contains elements to be eliminated (adjusted for) by the transformation L'. The ANOVA(β_2, σ^2) then consists of $R_2, S_1, \cdots, S_s$ where R_2 is the sum of squares for β_2 adjusted for β_1. By repeated applications of Corollary 1 in which β_2 is redefined as required, one may be able to construct some or all of the tests of interest on the fixed parameters in this piecemeal fashion.

For inference on variance components alone, the R_t terms is not required which allows a further weakening of the conditions. *ANOVA*(σ^2) will refer to a set of terms $\{S_1, \cdots, S_s\}$, where $\sum_{i=1}^{s} S_i = S$, which forms a partial ANOVA. Let W be a matrix such that $W'W = I$ and $\mu(W) = \mu(X)^{\perp}$, and define $Z = W'Y$,

$T_i = W'V_iW$, $i = 1, \cdots, k$. Applying Theorem 1 to model (1.1) transformed by W', i.e.

$$Z \sim N(0, \textstyle\sum_{i=1}^k T_i\sigma_i^2), \tag{2.8}$$

yields the following corollary. Since the fixed effects have been transformed out of the original model, uniqueness does not depend on t or otherwise on a partition of R.

COROLLARY 2. *An* ANOVA(σ^2) *exists iff* $T_1, \cdots, T_k$ *commute. An* ANOVA(σ^2) *is unique up to order.*

The necessity of Corollary 2 is the only part which is new to this paper. Olsen, Seely, and Birkes (1976) form a minimal sufficient statistic for a model with two variance components ($k = 2$) and show that it has the properties defined as an ANOVA(σ^2) here. LaMotte (1976), in treating the random one-way model, uses the same procedure, and notes that it can be applied whenever the matrices $T_1, \cdots, T_k$ can be simultaneously diagonalized. The condition for a diagonalizing transformation is the commutativity given in the corollary.

3. Estimation of σ^2 in an ANOVA. An ANOVA is a useful summary of the data from which tests and confidence intervals on fixed effects and variance components can be constructed based on the chi-squared and independence properties. An estimate of the variance components is also commonly constructed from an ANOVA table, when possible, by equating the sums of squares involving only the variance components, $S_1, \cdots, S_s$, to their expectations and solving for the unknown parameters. This *ANOVA estimator* of σ^2 (known as Henderson's Method I when reference is to the random effects model) is in a class $\mathscr{L}$ of estimators which are quadratic in Y, unbiased, and translation invariant (invariant of the transformation $Y \to Y + X\gamma$ for arbitrary fixed γ).

We will say that σ^2 is *estimable* if $\mathscr{L}$ is nonempty. The condition of estimability, given for example in [19], can be shown to be equivalent to linear independence of $T_1, \cdots, T_k$. This is, of course, the condition under which the variance components are identifiable in (2.8).

We note that $\{S_1, \cdots, S_s\}$, from either a full or partial ANOVA, is a) a minimal sufficient set of statistics for the family of distributions induced by W', $\{N(0, \sum_{i=1}^k T_i\sigma_i^2): \sigma^2 \in \Omega\}$, and is b) complete if and only if $s \le k$. Both a) and b) are straightforward extensions of the case when $k = 2$ [16, page 880]. All but the "only if" part of b) can be shown directly by arguments along the lines of [9]. The "only if" part was first established under the conditions of [9] by Basson (1965).

If $s < k$, then σ^2 is not estimable. This follows by first observing that $E(\mathbf{S}) = F\sigma^2$, where $\mathbf{S} = (S_1, \cdots, S_s)'$ and F is a known $s \times k$ matrix. It will be useful to form the matrix F in the following way. By Corollary 2, the matrices $T_1, \cdots, T_k$ commute so there exists a matrix K such that $K' = K^{-1}$ and $KT_iK' = D_i$ is diagonal, $i = 1, \cdots, k$. Let G be a matrix with the ith column identical to the diagonal vector of D_i. Then F consists of the distinct rows of G, up to order, and the linear space spanned by the columns of F is of the same dimension as ω, the linear space spanned by $T_1, \cdots, T_k$. Since F is $s \times k$, this common dimension is equal to the rank of F, which is no larger than $\min(s, k)$. Thus, if $s < k$ then the dimension of ω is less than k, which implies that $T_1, \cdots, T_k$ are not linearly independent and σ^2 is not estimable.

For the remainder of this section, it will be assumed that σ^2 is estimable, i.e. $\mathscr{L}$ is nonempty, which implies $s \geq k$. When $s = k$, the ANOVA estimator of σ^2, $\hat{\sigma}^2 = F^{-1}\mathbf{S}$, is defined and has minimum (uniformly smallest) variance in $\mathscr{L}$. That property is implied by the completeness in fact b) above. It also follows directly from Theorem 2.2 of [16] when $k = 2$, and the more general version of that theorem appearing in Kleffe and Pincus (1974), otherwise.

If $s > k$, the equations $\mathbf{S} = F\sigma^2$ are overdetermined (in general) so that the ANOVA estimator is undefined. But the class of estimators is not empty, by assumption, so this raises the question of whether $\mathscr{L}$ contains a minimum variance estimator when $\hat{\sigma}^2$ is undefined and, if so, how to construct it. Since completeness is not satisfied when $s > k$, one would suspect that such an estimator does not exist. This is in fact the case, as will be shown in Theorem 2. We will need the following result of Seely (1971) stated as a lemma here.

LEMMA. *Let $\mathscr{L}$ be nonempty. There is an estimator of $\mathscr{L}$ of uniformly smallest variance if and only if ω is a quadratic subspace ($A \in \omega \Rightarrow A^2 \in \omega$). (Jensen (1975) observes that a linear subspace is quadratic if and only if it is a Jordan Algebra.)*

THEOREM 2. *Let $\mathscr{L}$ be nonempty. When there is a full or partial ANOVA with $s > k$, there is no estimator in $\mathscr{L}$ of uniformly smallest variance.*

PROOF. Assume the contrary. By Corollary 2, ω is commutative and by the lemma it is also quadratic. By Lemma 6 of [19] there exists a basis $R_1, \cdots, R_k$ for ω such that $R_i^2 = R_i$ and $R_iR_j = 0$, $i \neq j$. Since $T_1, \cdots, T_k$ also form a basis for ω, there are k different parametric functions $\gamma(\sigma^2)_1, \cdots, \gamma(\sigma^2)_k$ such that $\sum_{i=1}^k T_i\sigma_i^2 = W'(\sum_{i=1}^k V_i\sigma_i^2)W =$

$$W'\Sigma W = R_1\gamma(\sigma^2)_1 + \cdots + R_k\gamma(\sigma^2)_k. \tag{4.1}$$

At the point

$$\sigma_0^2 = (0, \cdots, 0, 1)' \in \Omega,\ W'\Sigma_0 W = W'IW = R_1\gamma(\sigma_0^2)_1 + \cdots + R_k\gamma(\sigma_0^2)_k.$$

Multiply on the right by R_i to get $R_i = R_i^2\gamma(\sigma_0^2)_i = R_i\gamma(\sigma_0^2)_i$ which implies $\gamma(\sigma_0^2)_i = 1$, $i = 1, \cdots, k$, and $R_1 + \cdots + R_k = I$. Theorem 1 of [1] is satisfied with respect to the transformed model (2.8), and it implies that

$$\{Y'R_1Y, \cdots, Y'R_kY\} = \{S_1, \cdots, S_s\}.$$

By the uniqueness part of Corollary 2, s must equal k, a contradiction. □

When there is a full or partial ANOVA it is immediately apparent from the number of terms whether or not $\mathscr{L}$ contains a minimum variance estimator, and it is simply constructed from the ANOVA terms when there is. If an ANOVA is to be calculated anyway, this is much easier than verifying that ω is a quadratic subspace and then calculating the estimate separately as well. However, ω can be quadratic without being commutative, in which case there is a minimum variance estimator in $\mathscr{L}$, but there is not an ANOVA. Such cases would appear to be uncommon.

It is not the intent of this paper to necessarily recommend the class $\mathscr{L}$ of estimators of variance components. Although it has been a focus of interest in some of the literature, it has been introduced here because it arises naturally in the discussion of the analysis of variance. The estimators in $\mathscr{L}$ are unbiased, but at the expense of producing negative values sometimes. Maximum likelihood and restricted maximum likelihood (REML) estimates are nonnegative but biased. When data are balanced, the REML and ANOVA estimates of σ^2 are identical provided the latter has no negative values. A review of these and other methods is given by Harville (1977).

4. Further remarks and examples. The construction of a full or partial ANOVA for a given example can be accomplished by simultaneously diagonalizing the matrices which are required to commute in Theorem 1, Corollary 1, or Corollary 2, depending on which is satisfied. The sufficiency part of the proof of Theorem 1 illustrates construction of a full ANOVA, and construction of a partial ANOVA follows in like manner after a suitable transformation on the model. For the examples in this paper, a computer program was constructed which would accept any design conforming to (1.1), check the conditions of Theorem 1 and its corollaries, and construct what is possible in the way of an ANOVA. Although completely general in principle, this approach is limited in practice by the size of n, k, and possibly t which space and computing time will allow. Of course, in well-behaved cases where explicit expressions for the terms of an ANOVA are known, or can be constructed, these can be used to advantage.

In practice, it is common to construct an analysis of variance table by treating the random effects as if they were fixed effects for the purpose of partitioning the total sum of squares, and then to take the random effects into account when calculating mean squares. If sums of squares are formulated by fitting effects sequentially, while adjusting for the preceding effects, then this results in a partition of R as described in Section 2 and a partition of S into exactly k terms (possibly less than k if there is confounding between random and fixed effects). In the well-behaved cases with so-called balanced data this is equivalent to partitioning S as described here. With unbalanced data, however, the two methods are generally different with $s > k$. In such a case, the partition of S into only k terms will not have the ANOVA properties defined here since a partition of S with these properties is unique (Corollary 2).

The simple unbalanced one-way random classification, which is well understood, will provide a transparent example where $s > k$, and will also illustrate the notion of a partial ANOVA. Also, LaMotte (1976) has treated this example and provides additional details. In our preceding notation, $t = 1$, $k = 2$, $P_1 = (1/n)$, where $n = n_1 + \cdots + n_p$, n_i is the number of observations in the ith cell, and the n_i are not all equal. The matrix V_1 is block diagonal with blocks $J_1, \cdots, J_p$, where J_i is an $n_i \times n_i$ matrix with every element equal to one. The matrices P_1 and V_1 do not commute unless the n_i are all equal, so by appeal to Theorem 1, an ANOVA with $Y'P_1Y$ as a term does not exist. However, if we translate the mean out of the model (which is almost never of interest anyway) by the transformation W', the condition of Corollary 2 is satisfied so that a partial ANOVA, an ANOVA(σ^2) in this case (or equivalently, a full ANOVA in the model after the W' transformation), exists and is unique. The value of s, the number of terms in it, is the number of different eigenvalues of $W'V_1W + I$, or equivalently, the number of different nonzero eigenvalues of $(I - P_1)V_1(I - P_1) + I$. It is easy to verify that $s > k(=2)$ since the n_i are not all equal. Thus it is clear that the method described above, which would partition S into only two terms, would not produce an ANOVA in this case by virtue of uniqueness.

Continuing with the same example, a broader range of hypotheses can be tested using the ANOVA(σ^2) (or expanded ANOVA in LaMotte's terminology) than the customary analysis of variance adjusted for the mean (Graybill, 1961, page 353). The reason is simply that the ANOVA(σ^2) is a finer partition of the sum of squares S. To illustrate, consider the simplest case with $s > k$, namely $s = 3$, and let S_1, S_2, S_3 be the terms of the ANOVA(σ^2) with f_1, f_2, f_3 degrees of freedom respectively. The expected mean squares are of the form $a_1\sigma_b^2 + \sigma_w^2$, $a_2\sigma_b^2 + \sigma_w^2$, σ_w^2, which we will let correspond to S_1, S_2, S_3, respectively, where σ_b^2 and σ_w^2 are the between-block and within-block variances and a_1, a_2 are known constants. The customary ANOVA contains only $k(=2)$ terms, denoted here by SS_b and SS_w for sum of squares for between and within blocks. The correspondence is $SS_b = S_1 + S_2$, $SS_w = S_3$. Consider the null hypothesis $H_o: \sigma_b^2/\sigma_w^2$ ($\leq, \geq, =$) c, for given $c \geq o$. It is straightforward to construct a test statistic using $\{S_1, S_2, S_3\}$, namely $f_3(S_1/(a_1c + 1) + S_2/(a_2c + 1))/(f_1 + f_2)S_3$, which is distributed as $F(f_1 + f_2, f_3)$ if $\sigma_b^2/\sigma_w^2 = c$. For $c = o$, the statistic reduces to the usual F statistic obtained from $\{SS_b, SS_w\}$ for testing of $\sigma_b^2 = o$. However, for $c > o$, construction of an exact test from $\{SS_b, SS_w\}$ does not appear possible.

The following examples serve to further illustrate the preceding results and demonstrate that the utility of a design with respect to an analysis of variance can be sensitive to which effects are fixed and which are random.

EXAMPLE 1. Suppose the following two-way uncrossed classification without interaction is given, where the number in each cell indicates the number of observations. It is assumed that the conditions of model (1.1) are satisfied. Consider first the case where both row and column effects are fixed. If the usual assumptions of a nested classification are appropriate, $\alpha_1 + \alpha_2 = 0$, $\tau_1 + \tau_2 = 0$, $\tau_3 + \tau_4 = 0$, there is no difficulty forming an ANOVA. However, under the usual assumptions of a two-way classification, $\sum_{i=1}^{4} \tau_j = 0$ and $\alpha_1 + \alpha_2 = 0$, there is clearly confounding so that row and column effects cannot be completely separated in an ANOVA, that is, the sums of squares for row effects will have to be adjusted for columns or vice versa, so that there are two ANOVAs of interest corresponding to two choices for R_1, R_2, R_3.

τ_j

α_i	2	2	0	0
	0	0	2	2

FIG. 1. *Two-way layout of Example* 1.

If both row and column classifications are random, or if row effects are fixed and column effects are random, then there is no longer a distinction from the nested classification. In either case, S is partitioned into $s = k$ terms, and there is only one ANOVA of interest.

To make the example less familiar, suppose that a third row of observations is added introducing a third row effect α_3.

1	1	1	1

FIG. 2. *Row to be added to Figure* 1.

Consider the mixed case where row effects are fixed with $\alpha_1 + \alpha_2 + \alpha_3 = 0$, column effects are random with variance σ_τ^2, and the error variance is denoted by σ_e^2. Taking $R_1 = n\bar{y}^2$ and R_2 equal to the remaining part of the sum of squares for fixed effects defines P_1 and P_2. Neither Theorem 1 nor Corollary 1 is satisfied, but Corollary 2 is. The ANOVA(σ^2) has the following terms.

TABLE 1
ANOVA(σ^2) *for Example* 1 *with three rows, fixed row effects and random column effects*

Source	S.S.	d.f.	$E(M.S.)$
τ	$\{S_1$	1	$\sigma_\tau^2 + \sigma_e^2$
	S_2	2	$3\sigma_\tau^2 + \sigma_e^2$
error	S_3	6	σ_e^2
TOTAL	S		

As $s = 3 > k = 2$, there is not a minimum variance estimator of σ^2 in $\mathscr{L}$ (Theorem 2), although σ^2 is estimable. An F-statistic to test H: $\sigma_\tau^2 = 0$ is given by $2(S_1 + S_2)/S_3$, but no test is provided for the hypothesis of zero row effects since only an ANOVA(σ^2) was generated.

τ_j

	1	1		
α_i	1	1		
			1	1
			1	1

FIG. 3. *Two-way layout of Example* 2.

TABLE 2
ANOVA(μ, σ^2) *for Example* 2 *with random row and column effects*

Source	S.S.	d.f.	*E(M.S.)*
μ	R	1	$2\sigma_\alpha^2 + 2\sigma_\tau^2 + \sigma_e^2 + 8\mu^2$
α and τ	S_1	2	$2\sigma_\alpha^2 + \sigma_e^2$
	S_2	2	$2\sigma_\tau^2 + \sigma_e^2$
	S_3	1	$2\sigma_\alpha^2 + 2\sigma_\tau^2 + \sigma_e^2$
error	S_4	2	σ_e^2
TOTAL	T		

TABLE 3
ANOVA(σ^2) *for Example* 2 *with fixed row effects and random column effects*

Source	S.S.	d.f.	*E(M.S.)*
τ	S_1	2	$2\sigma_\tau^2 + \sigma_e^2$
error	S_2	2	σ_e^2
TOTAL	S		

EXAMPLE 2. Consider another two-way classification without interaction, satisfying the assumptions of model (1.1). When both row and column effects are random, with $\tau_j \sim N(0, \sigma_\tau^2)$, $\alpha_i \sim N(0, \sigma_\alpha^2)$ and error $e_i \sim N(0, \sigma_e^2)$, this example is given in [9] and is what Gaylor (1960) calls a BD2-8 design. Theorem 1 is satisfied and an ANOVA(μ, σ^2), where μ is a term for the mean, consists of the terms in Table 2.

Again, $\mathscr{L}$ does not contain a minimum variance estimator since $s = 4 > k = 3$, although σ^2 is estimable. Two independent F-tests of the hypothesis H: $\sigma_\tau^2 = 0$ can be constructed using (S_1, S_3) and (S_2, S_4); similarly for H: $\sigma_\alpha^2 = 0$. An F-statistic for H: $\sigma_\alpha^2 = \sigma_\tau^2 = 0$ is given by $.4(S_1 + S_2 + S_3)/S_4$.

Now let the row effects be fixed and sum to zero, and the column effects remain random. Let P_1 and P_2 be the symmetric idempotent matrices such that $R_1 = Y'P_1Y$ is the sum of squares for the mean and $R_2 = Y'P_2Y$ is the sum of squares for fixed effects adjusted for the mean. For suitable parametrization, $\beta = (\beta_1, \beta_2)'$, where $\beta_1 = \mu$, the mean, and β_2 is a 3×1 vector which is all zeroes iff there is no row effect. An ANOVA(β_2, σ^2) is of interest, but it does not exist because $P_2^*\Sigma P_2^*$ has two different parametric functions as characteristic roots, $2\sigma_\tau^2 + \sigma_e^2$ and σ_e^2, instead of only one (Corollary 1). However, Corollary 2 is satisfied so an ANOVA(σ^2) is possible, with terms given in Table 3.

The ANOVA estimator is defined and is the minimum variance estimator in $\mathscr{L}$. The usual methods of inference on the variance components in an ANOVA with $s = k$ are applicable, but no statistic is produced for a test on the row effects.

Acknowledgement. Thanks are due to an associate editor and to referees for their careful reviews and suggestions which improved the paper. I am also grateful to Robert I. Jennrich for many helpful discussions and for his meticulous reading of an earlier draft which resulted in an improved version of the manuscript.

REFERENCES

[1] ALBERT, A. (1976). When is a sum of squares an analysis of variance? *Ann. Statist.* **4** 775–778.

[2] ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141.

[3] BASSON, P. R. (1965). On unbiased estimation in variance components models. Ph.D. thesis, Iowa State University.

[4] BOARDMAN, T. J. (1974). Confidence intervals for variance components in comparative Monte Carlo study. *Biometrics* **30** 251–262.

[5] BROWN, K. G. (1983). Sub-balanced data and the mixed analysis of variance. *J. Amer. Statist. Assoc.* **78** 162–167.

[6] FISHER, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edin.* **52** 399–433.

[7] GAYLOR, D. W. (1960). The construction and evolution of some designs for the estimation of parameters in random models. Ph.D. thesis, North Carolina State University. *Inst. Stat. Mimeo Ser.* **256**.

[8] GRAYBILL, F. A. and MARSAGLIA, G. (1951). Idempotent matrices and quadratic forms in the general linear hypothesis. *Ann. Math. Statist.* **28** 678–686.

[9] GRAYBILL, F. A. and HULTQUIST, R. A. (1961). Theorems concerning Eisenhart's model II. *Ann. Math. Statist.* **32** 261–269.

[10] GRAYBILL, F. A. (1961). *An Introduction to Linear Statistical Models, Vol.* 1. McGraw-Hill, New York.

[11] HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and related problems. *J. Amer. Statist. Assoc.* **72** 320–340.

[12] HULTQUIST, R. A. and GRAYBILL, F. A. (1956). Minimal sufficient statistics for the two-way classification mixed model design. *J. Amer. Statist. Assoc.* **60** 182–192.

[13] JENSEN, S. T. (1975). Covariance hypotheses which are linear in both the covariance and the inverse covariance. (Preprint). Institute of Mathematical Statistics, University of Copenhagen.

[14] KLEFFE, J. and PINCUS, R. (1976). Bayes and best quadratic unbiased estimators for parameters of the covariance matrix in a normal linear model. *Math. Operationsforsch. Statist.* **5** 43–67.

[15] LAMOTTE, L. R. (1976). Invariant quadratic estimators in the random one-way ANOVA model. *Biometrics* **32** 793–804.

[16] OLSEN, A., SEELY, J., and BIRKES, D. (1976). Invariant quadratic unbiased estimation for two variance components. *Ann. Statist.* **4** 878–890.

[17] RAO, C. R. and MITRA, S. K. (1971). *Generalized Inverse of Matrices and its Applications.* Wiley, New York.

[18] SCHEFFÉ, H. (1959). *The Analysis of Variance.* Wiley, New York.

[19] SEELY, J. F. (1971). Quadratic subspaces and completeness. *Ann. Math. Statist.* **43** 710–721.

BIOMETRY AND RISK ASSESSMENT PROGRAM
NATIONAL INSTITUTE OF ENVIRONMENTAL
HEALTH SCIENCES
P.O. BOX 12233
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709