The Annals of Mathematical Statistics
1972, Vol. 43, No. 6, 2024-2030

THE MIXED EFFECTS MODEL AND SIMULTANEOUS DIAGONALIZATION OF SYMMETRIC MATRICES

BY ROBERT HULTQUIST AND ERWIN M. ATZINGER

The Pennsylvania State University

This paper presents results obtained in efforts to generalize ideas reported during the past 15 years on the subject of variance components. The paper applies the result of Newcomb (1960) to the mixed effects model

$$Y = \sum_{j=1}^{t} X_j \tau_j + \sum_{k=1}^{h} Z_k b_k$$

where the τ_j are fixed parameters and the b_k are random vectors distributed normally with covariance matrix $\sigma_k^2 I$. The restrictions made on the model are far less severe than those imposed by other authors. Through a nonsingular transformation of the data vector Y, minimal sufficient statistics are obtained. Theorems are presented which give conditions under which minimum variance unbiased estimates exist and these estimates are displayed. Properties of the model and the estimates are discussed in both the complete and noncomplete density cases. Perhaps the most important contribution relates to the simplicity with which the theoretical methods treat the general variance components situation.

1. Introduction. Many investigators have researched the area in statistics related to the concept of variance components. Today it still receives much attention. Efforts such as those of Kendall and Stuart (1960) have been directed toward unifying the theory. This paper consists of some results obtained in a further attempt to unify the theory and generalize concepts.

We consider the mixed effects model of the form

$$Y = \sum_{j=1}^{t} X_j \tau_j + \sum_{k=1}^{h} Z_k b_k = X\tau + \sum_{k=1}^{h} Z_k b_k , \tag{1}$$

where $\tau' = [\tau_1, \tau_2, \cdots, \tau_t]$ is a vector of t components and b_k, $k = 1, \cdots, h$ are random vectors. Models discussed in some earlier papers such as [9], [10], [12], [20], were special cases of this model. The obtaining of a set of minimal sufficient statistics was a principal objective of the papers.

Most easily handled is the situation where all $X_i X_i'$ and $Z_k Z_k'$ commute in pairs, that is where

$$X_j X_j' X_i X_i' = X_i X_i' X_j X_j' ; \qquad i, j = 1, 2, \cdots, t ;$$
$$X_j X_j' Z_k Z_k' = Z_k Z_k' X_j X_j' ; \qquad j = 1, 2, \cdots, t;\ k = 1, 2, \cdots, h; \text{ and}$$
$$Z_l Z_l' Z_k Z_k' = Z_k Z_k' Z_l Z_l' ; \qquad k, l = 1, 2, \cdots, h .$$

Some results have been reported for the non-commutative case.

The model

$$Y = j\mu + Z_1 b_1 + Z_2 b_2 + I b_3 ,$$

where μ is a scalar and j is a vector of ones, was studied by Weeks and Graybill

Received September 3, 1970; revised April 7, 1972.

(1962). They obtained a set of minimal sufficient statistics in the case where $Z_1 Z_1'$ and $Z_2 Z_2'$ do not necessarily commute.

The model

$$Y = X\tau + Z_1 b_1 + I b_2 ,$$

was studied by Hultquist and Graybill (1965). In that study XX' and $Z_1 Z_1'$ did not necessarily commute.

As is well known, when symmetric matrices A_i commute in pairs, there exists an orthogonal matrix P such that $PA_i P'$ are simultaneously diagonal. What is not so well known is the fact presented by Newcomb (1960) that for nonnegative matrices B_1 and B_2 there always exists a nonsingular matrix R such that both $RB_1 R'$ and $RB_2 R'$ are diagonal. Indeed there exist pairwise non-commutative matrices $B_1, B_2, \cdots, B_h$, $h > 2$ for which there is a nonsingular matrix R such that each $RB_i R'$ is diagonal. This paper concerns itself with the class of models where there exists a nonsingular matrix R independent of all parameters and such that $RZ_k Z_k' R'$ is diagonal for all $k = 1, 2, \cdots, h$.

In another important way the model discussed in this paper is a generalization of the models discussed in other papers, where the matrix Z_h was always chosen to be the identity matrix. In this paper we relax that assumption.

We now describe more carefully the model which will be investigated. The $(n \times 1)$ vector of observations Y is writen

$$Y = X\tau + \sum_{k=1}^{h} Z_k b_k ,$$

with the following assumptions:

(a) X is an $(n \times t)$ matrix of known constants.

(b) τ is a $(t \times 1)$ vector of functionally independent unknown parameters.

(c) b_k $(k = 1, 2, \cdots, h)$ are vectors distributed normally with mean matrix ϕ (a vector of zeros) and covariance matrix $\sigma_k^2 I$.

(d) All components of all vectors b_k are stochastically independent.

(e) σ_k^2 $(k = 1, 2, \cdots, h)$ are functionally independent. That is we assume no known relationship among them.

(f) Each σ_k^2 is functionally independent of the elements of τ.

(g) The rank of the matrix $[Z_1, Z_2, \cdots, Z_h]$ is equal to n.

2. Estimability. Perhaps the most natural first question to consider is the question of estimability of the parameters. (A function of parameters is said to be estimable if there exists an unbiased estimate of the function of the parameters.) In this section, necessary and sufficient conditions for estimability are presented in terms of the design matrices X_i and Z_k. The results obtained by the authors are special cases of results obtained by Seely (1969) and hence no proofs are given. Theorems 1 and 2 are presented because the statements are in a different framework and notation and of a different flavor from those presented by Seely.

THEOREM 1. *Necessary and sufficient conditions for σ_s^2; $s = 1, \cdots, h$ to be estimable by a quadratic form $Y'C_s Y$ are:*

(a) *The set of matrices* $\{Z_k Z_k' : k = 1, \cdots, h\}$ *is a linearly independent set and*

(b) $Z_k Z_k'$ $(k = 1, \cdots, h)$ *is not a linear combination of the matrices* $\{X_i X_j' + X_j X_i' : i \leqq j = 1, \cdots, t\}$.

THEOREM 2. *A necessary and sufficient condition for each* σ_s^2 *to be estimable by a quadratic form* $Y'C_s Y$ *and each product* $\tau_i \tau_j$ $(i \leqq j = 1, \cdots, t)$ *to be estimable by a quadratic form* $Y'B_{ij}Y$ *is that the set of matrices* $\{Z_k Z_k' : k = 1, \cdots, h\} \cup \{X_i X_j' + X_j X_i' : i \leqq j = 1, \cdots, t\}$ *be linearly independent.*

3. General theory. The theoretical development in this section exhibits the use of a nonsingular transformation of the observation vector Y in order to obtain mutually independent unbiased sufficient statistics for the parameters. The joint density of Y can be written

$$f(Y) = Ce^{-Q/2}$$

where $Q = (Y - X\tau)'V^{-1}(Y - X\tau)$ and $V = \sum_{k=1}^{h} \sigma_k^2 Z_k Z_k'$.

We consider the fairly broad class of designs where there exists a nonsingular matrix R such that $RZ_k Z_k' R'$ is diagonal for all $k = 1, \cdots, h$. The transformed covariance matrix is

$$RVR' = \sum_{k=1}^{h} \sigma_k^2 RZ_k Z_k' R'$$

which is diagonal. Without loss of generality the matrix R can be chosen in such a way that the like elements on the diagonal of RVR' are grouped and such that no diagonal element is a multiple of another diagonal element. Let ζ_i, $i = 1, \cdots, s$ be the s distinct diagonal elements of RVR', then the quadratic form Q can be written

$$Q = \sum_{i=1}^{s} \frac{1}{\zeta_i}(R_i Y - R_i X\tau)'(R_i Y - R_i X\tau) .$$

The dimension of R_i is $(m_i \times n)$ where m_i is the multiplicity of ζ_i.

Let q_i denote the rank of the $(m_i \times t)$ matrix $R_i X$. For each i we can now select an $(m_i \times m_i)$ orthogonal matrix (functionally independent of τ).

$$G_i = \begin{bmatrix} G_i^{(1)} \\ G_i^{(2)} \end{bmatrix}$$

with G_i partitioned such that $G_i^{(1)}$ is $(q_i \times m_i)$ and $G_i^{(2)} R_i X\tau = \phi$. In the case where $q_i = 0$, $G_i^{(1)}$ does not exist and when $q_i = m_i$ then $G_i^{(2)}$ does not exist. We employ the matrices G_i in the following way.

$$Q = \sum_{i=1}^{s} \frac{1}{\zeta_i}(R_i Y - R_i X\tau)'G_i'G_i(R_i Y - R_i X\tau)$$

$$Q = \sum_{i=1}^{s} \frac{1}{\zeta_i}(G_i R_i Y - G_i R_i X\tau)'(G_i R_i Y - G_i R_i X\tau)$$

$$Q = \sum_{i=1}^{s} \frac{1}{\zeta_i} \begin{bmatrix} G_i^{(1)} R_i Y - G_i^{(1)} R_i X\tau \\ G_i^{(2)} R_i Y \end{bmatrix}' \begin{bmatrix} G_i^{(1)} R_i Y - G_i^{(1)} R_i X\tau \\ G_i^{(2)} R_i Y \end{bmatrix}$$

$$Q = \sum_{i=1}^{s} \frac{1}{\zeta_i}\{Y'R_i'G_i^{(1)\prime}G_i^{(1)}R_i Y + \tau'X'R_i'G_i^{(1)\prime}G_i^{(1)}R_i X\tau - 2\tau'X'R_i'G_i^{(1)\prime}G_i^{(1)}R_i Y + Y'R_i'G_i^{(2)\prime}G_i^{(2)}R_i Y\}.$$

Denote $\tau' X' R_i' G_i^{(1)'}$ by θ_i' and if they exist denote $G_i^{(1)} R_i Y$ by U_i, and $Y' R_i' G_i^{(2)'} G_i^{(2)} R_i Y$ by v_i. In this notation the quadratic form can be written

$$Q = \sum_{i=1}^{s} \frac{1}{\zeta_i} \{U_i' U_i + \theta_i' \theta_i - 2\theta_i' U_i + v_i\} . \tag{2}$$

Let $q. = \sum_{i=1}^{s} q_i$, let $r_i = 1$ if $q_i \neq m_i$, $r_i = 0$ if $q_i = m_i$ and let $r. = \sum_{i=1}^{s} r_i =$ no. of $q_i \neq m_i$. Equation (2) then exhibits sufficient statistics which when put into vector form has dimension $r. + q.$.

THEOREM 3. *The $r. + q.$ dimensional vector statistic consisting of the elements U_{ij} and v_i, is jointly sufficient for the $t + h$ parameters τ and σ_k^2 $(k = 1, \cdots, h)$. (U_{ij} denotes the* j*th element of U_i).*

Furthermore we have

THEOREM 4. *The $r. + q.$ statistics U_{ij} and v_i are stochastically independent with $U_{ij} \sim N(\theta_{ij}, \zeta_i)$ and $v_i/\zeta_i \sim \chi^2(m_i - q_i)$. ($\theta_{ij}$ denotes the* j*th element of θ_i.)*

It readily follows, from the minimality criterion of Lehmann and Scheffé (1950) that if the set $\{\zeta_i^{-1}: i = 1, 2, \cdots, s\}$ is a linearly independent set of distinct diagonal elements of $(RVR')^{-1}$ then the $r. + q.$ dimensional sufficient statistic is minimal sufficient. This fact provides additional incentive to investigate the nature of the distinct diagonal elements of RVR'.

LEMMA. *If $\delta_i = \sum_{k=1}^{h} c_{ik} \sigma_k^2$; $i = 1, 2, \cdots, s$ where the σ_k^2; $k = 1, 2, \cdots, h$ are functionally independent then the reciprocals δ_i^{-1} are linearly independent.*

PROOF. We prove this lemma by the method of contradiction. Assume $\sum_{i=1}^{h} a_i/\delta_i = 0$ with not all $a_i = 0$. Then

$$\sum_{i=1}^{h} \left\{ \frac{a_i}{\sum_{k=1}^{h} c_{ik} \sigma_k^2} \right\} = 0 .$$

This means that the σ_k^2 are functionally dependent which contradicts the hypothesis and proves the lemma.

Invoking this lemma we then have according to the result of Lehmann and Scheffé (1950), the following

THEOREM 5. *The $r. + q.$ dimensional statistic consisting of the U_{ij} and the v_i is minimal sufficient for the $t + h$ parameters τ and σ_k^2 $(k = 1, \cdots, h)$.*

In the search for minimum variance unbiased estimates, the next logical step is to determine when the joint density for the minimal sufficient statistic is complete. In a straightforward manner the application of a completeness lemma introduced by Herbach (1959) and generalized by Imhof (1960), establishes

THEOREM 6. *When $q. = t$ and $r. = h$ then the joint density, of the minimal sufficient statistic, with components U_{ij} and v_i, is complete.*

In application of Theorem 6, the question arises as to whether q_i (equal to the rank of $R_i X$) is dependent on the particular choice of the diagonalizing transformation R. Possibly, by a more judicious choice of R a complete statistic could be obtained in certain situations which, for the particular choice of R, was incomplete. It is shown however by Atzinger (1970) that q_i is in fact unique for all R and hence only one choice of the diagonalizing transformation need be examined to determine whether the conditions of Theorem 6 are satisfied.

4. Applications and illustrations. The following example illustrates how the theoretical concepts of Section 3 are used to obtain best estimates. Consider the five observations

$$\begin{aligned} y_1 &= \tau + b_{11} + b_{21} + b_{31} \\ y_2 &= \tau + b_{11} + b_{21} + b_{32} \\ y_3 &= \tau + b_{11} + b_{22} + b_{31} \\ y_4 &= \tau + b_{12} + b_{22} + b_{32} \\ y_5 &= \tau + b_{12} + b_{31} \end{aligned}$$

where τ is fixed and $b_{1j} \sim N(0, \sigma_1^2)$ and $b_{2j} \sim N(0, \sigma_2^2)$ and $b_{3j} \sim N(0, \sigma_3^2)$ for $j = 1, 2$. This can be represented by the model

$$Y = X\tau + Z_1 b_1 + Z_2 b_2 + Z_3 b_3$$

where

$$X' = [1 \quad 1 \quad 1 \quad 1], \qquad Z_1' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \qquad Z_2' = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix},$$

$$Z_3' = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad b_i' = [b_{i1} \quad b_{i2}] \qquad (i = 1, 2, 3).$$

By applying Theorem 1, it is seen that σ_i^2 $(i = 1, 2, 3)$ is estimable. Thus, it appears practical to continue the search for minimum variance unbiased estimates in this case.

Clearly, $A_i = Z_i Z_i'$ $(i = 1, 2, 3)$ do not commute in pairs. Their structure, however, is such that they can be simultaneously diagonalized by the following matrix:

$$R = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ -2 & 2 & 1 & -1 & 2 \\ 1 & 0 & -1 & 0 & 0 \\ 3 & -2 & -1 & 2 & -2 \end{bmatrix}.$$

The diagonalized covariance matrix is then

$$RVR' = \operatorname{diag}(2\sigma_1^2, 2\sigma_3^2, 2\sigma_1^2 + 2\sigma_3^2, 2\sigma_2^2, 2\sigma_2^2)$$

where, $\zeta_1 = 2\sigma_1^2$, $\zeta_2 = 2\sigma_3^2$, $\zeta_3 = 2(\sigma_1^2 + \sigma_3^2)$, and $\zeta_4 = 2\sigma_2^2$ are the $s = 4$ distinct diagonal elements.

Corresponding to each ζ_i are the following values of m_i and q_i,

$$m_1 = 1, \quad m_2 = 1, \quad m_3 = 1, \quad m_4 = 2$$
$$q_1 = 0, \quad q_2 = 0, \quad q_3 = 1, \quad q_4 = 0$$

and hence, $q. = 1$ and $r. = 3$. The second transformation, G is then chosen as

$$G_1 = G_1^{(2)} = 1,$$
$$G_2 = G_2^{(2)} = 1,$$
$$G_3 = G_3^{(1)} = 1,$$
$$G_4 = G_4^{(2)} = \begin{bmatrix} 2^{-\frac{1}{2}} & 2^{-\frac{1}{2}} \\ 2^{-\frac{1}{2}} & -2^{-\frac{1}{2}} \end{bmatrix}.$$

Having obtained G, the following statistics are formed.

$$U_3 = G_3^{(1)}R_1 Y = -2y_1 + 2y_2 + y_3 - y_4 + 2y_5,$$
$$v_1 = Y'R_1'G_1^{(2)'}G_1^{(2)}R_1 Y = (y_1 - y_2 - y_3 + y_4)^2,$$
$$v_2 = Y'R_2'G_2^{(2)'}G_2^{(2)}R_2 Y = (y_1 - y_2)^2,$$

and

$$v_4 = Y'R_4'G_4^{(2)'}G_4^{(2)}R_4 Y = 2\{(2y_1 - y_2 - y_3 + y_4 - y_5)^2 + (-y_1 + y_2 - y_4 + y_5)^2\}.$$

The $q. + r. = 4$ statistics are complete, sufficient and minimal dimension. Their distributions are as follows:

$$U_3 \sim N(2\tau, 2\sigma_1^2 + 2\sigma_3^2), \qquad \frac{v_2}{2\sigma_1^2} \sim \chi^2(1), \qquad \frac{v_2}{2\sigma_3^2} \sim \chi^2(1),$$

and

$$\frac{v_4}{2\sigma_2^2} \sim \chi^2(2).$$

Thus, the estimates of the parameters τ, σ_1^2, σ_2^2 and σ_3^2 are:

$$\hat{\tau} = \frac{U_3}{2} = (-y_1 + y_2 + y_3/2 - y_4/2 + y_5),$$

$$\hat{\sigma}_1^2 = \frac{v_1}{2} = (y_1 - y_2 - y_3 + y_4)^2/2,$$

$$\hat{\sigma}_2^2 = \frac{v_4}{4} = \tfrac{1}{2}[(2y_1 - y_2 - y_3 + y_4 - y_5)^2 + (-y_1 + y_2 - y_4 + y_5)^2],$$

and

$$\hat{\sigma}_3^2 = \frac{v_2}{2} = (y_1 - y_2)^2/2.$$

These unbiased estimates, being based on a complete minimal sufficient set of statistics, are minimum variance unbiased.

REFERENCES

[1] ATZINGER, E. M. (1970). Mixed effects model estimation-optimal properties. Ph. D. dissertation, Pennsylvania State Univ.

[2] BASSON, R. (1965). On unbiased estimation in variance component models. Unpublished Ph. D. dissertation, Iowa State Univ.

[3] BIRKOFF, G. and MACLANE, S. (1963). *A Survey of Modern Algebra.* MacMillan, New York.

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

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

[6] GRAYBILL, F. A. and WORTHAM, A. W. (1956). A note on uniformly but unbiased estimators for variance components. *J. Amer. Statist. Assoc.* **51** 266–268.

[7] HENDERSON, C. R. (1953). Estimation of variance and covariance components. *Biometrics* **9** 226–252.

[8] HERBACH, L. H. (1959). Properties of model II-type analysis of variance tests, A: Optimum nature of the γ-test for model II in the balanced case. *Ann. Math. Statist.* **30** 939–959.

[9] HULTQUIST, R. A. (1959). Minimal sufficient statistics for Eisenhart's model III. Ph. D. dissertation, Oklahoma State Univ.

[10] HULTQUIST, R. A. and GRAYBILL, F. A. (1965). Minimal sufficient statistics for the two-way classification mixed model design. *J. Amer. Statist. Assoc.* **60** 182–192.

[11] IMHOF, J. P. (1960). A mixed model for the complete three-way layout with two random-effects factors. *Ann. Math. Statist.* **31** 906–928.

[12] KAPADIA, C. H. and WEEKS, D. L. (1963). Variance components in two-way classification models with interaction. *Biometrika* **50** 327–334.

[13] KENDALL, M. G. and STUART, A. (1966). *The Advanced Theory of Statistics*, **3**. *Design and Analysis, and Time Series.* Hafner, New York.

[14] LEHMANN, E. L. and SCHEFFE, H. (1950). Completeness, similar regions and unbiased estimation I. *Sankhyā* **10** 305–340.

[15] MITRA, S. K. and RAO, C. R. (1968). Simultaneous reduction of a pair of quadratic forms. *Sankhyā Ser. A.* **30** 313–323.

[16] NEWCOMB, R. W. (1960). On the simultaneous diagonalization of two semi-definite matrices. *Quart. Appl. Math.* **19** 144–146.

[17] SEELY, J. (1969). Estimation in finite dimensional vector spaces with application to the mixed linear model. Ph. D. dissertation, Iowa State Univ.

[18] THRALL, R. M. and TORNHEIM, L. (1957). *Vector Spaces and Matrices.* Wiley, New York, 189.

[19] WILKS, S. S. (1963). *Mathematical Statistics.* Wiley, New York.

[20] WEEKS, D. L. and GRAYBILL, F. A. (1962). A minimal sufficient statistic for a general class of designs. *Sankhyā Ser. A.* **24** 339–353.

DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENN. 16802