last copy

ESTIMABILITY, ESTIMATES, AND VARIANCES FOR LINEAR MODELS

David M. Allen*

BU-529-M October 1974

SUMMARY

In the linear model $Y = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$ interest often centers on inference about $\theta = \underset{\sim}{\ell}'\underset{\sim}{\beta}$. Procedures for inference require a criterion for estimability, an estimate, and the estimated variance of the estimator. This paper presents formulas for these quantities that are fast, accurate, and have low storage requirements. Derivations of the formulas and a numerical example are also presented.

* David M. Allen is an Associate Professor in the Department of Statistics, University of Kentucky. This manuscript was prepared while the author was on leave at the Biometrics Unit, Cornell University, Ithaca, N. Y. 14850. The research was supported by a grant from Mead Johnson and Company, Evansville, Ind. The author is grateful for the inspiration provided by Mead Johnson employees and for the suggestions of his colleagues at Cornell.

ESTIMABILITY, ESTIMATES, AND VARIANCES FOR LINEAR MODELS

David M. Allen*

1. INTRODUCTION

A matrix representation of the linear model is $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ where $\underline{Y}$ is an $n \times 1$ observation vector, X is an $n \times p$ matrix of known constants of rank r, β is a $p \times 1$ vector of unknown parameters, and $\underline{\epsilon}$ is an unobserved random normal vector with mean $\underline{0}$ and covariance matrix $\sigma^2 I$ where σ^2 is positive and unknown. For a $(p \times 1)$ vector of constants $\underline{\ell}$, $\theta = \underline{\ell}'\underline{\beta}$ is said to be estimable if and only if $\underline{\ell}$ is in the vector space spanned by the rows of X . Our statistical objective is to determine if θ is estimable and, if so, to make inferences about θ . Statistics used for this purpose are given by Searle [5]. The following is a summary of these statistics.

The parameter θ is estimable if and only if

$$(I - (X'X)(X'X)^-)\underline{\ell} = \underline{0} , \tag{1.1}$$

where $(X'X)^-$ is any generalized inverse of $X'X$. If θ is estimable, then its best linear unbiased estimator is

$$\hat{\theta} = \underline{\ell}'(X'X)^-X'\underline{Y} \quad . \tag{1.2}$$

---

* David M. Allen is an Associate Professor in the Department of Statistics, University of Kentucky. This manuscript was prepared while the author was on leave at the Biometrics Unit, Cornell University, Ithaca, N. Y. 14850. The research was supported by a grant from Mead Johnson and Company, Evansville, Ind. The author is grateful for the inspiration provided by Mead Johnson employees and for the suggestions of his colleagues at Cornell.

The coefficient of σ^2 in the variance of $\hat{\theta}$ is

$$\underset{\sim}{\ell}'(X'X)^{-}\underset{\sim}{\ell} \tag{1.3}$$

and the error sum of squares is

$$\underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{Y}'X(X'X)^{-}X'\underset{\sim}{Y} \ . \tag{1.4}$$

Most linear models computations are done by computer. Formulas for computation should be evaluated with respect to operation count (speed), accuracy, and storage requirements. Direct application of textbook formulas is not desirable for several reasons. Any method where X'X is formed is potentially hazardous from the standpoint of numerical accuracy. Computation of $(X'X)^{-}$ is relatively slow. If repeated values of $\underset{\sim}{\ell}$ are to be processed, then both $(X'X)(X'X)^{-}$ and $(X'X)^{-}$ need to be stored.

Milliken [3] gives an alternative to (1.1). The supposed advantage of his method is that only one number need be checked. However, checking the norm of the left hand side of (1.1) has the same property. Milliken's computational method requires high storage. He does not give an operation count.

The objective of this paper is to present formulas for (1.1) through (1.4) that are fast, accurate, and require low storage. Derivations of these formulas are also presented.

2. DEFINITIONS AND A PRELIMINARY RESULT

Definition 2.1. For any $k \times k$ matrix A with real elements a_{ij} let $\mathcal{L}_A = \{i \mid i=1,2,\cdots,k;\ a_{ii} = 0\}$ and let I_A be a diagonal matrix whose i^{th} diagonal element is 1 if $i \in \mathcal{L}_A$ and 0 otherwise.

Definition 2.2. A $k \times k$ matrix A is said to be layer triangular if it is upper triangular, has no negative diagonal elements, and $a_{ij} = 0$ for all i, j such that $i \in \mathcal{L}_A$.

For any $k \times k$ layer triangular matrix A, let A^+ denote $(A + I_A)^{-1}$. A property of A^+ is

$$AA^+ = I - I_A \quad . \tag{2.1}$$

To prove (2.1) observe that $I_A A = 0$, $I_A^2 = I_A$, and thus $A = (I - I_A)(A - I_A)$. Multiply from the right by A^+ to obtain (2.1).

3. THE MAIN RESULT

Let $[X \mid \underline{Y}]$ denote X augmented on the right by $\underline{Y}$. There exists an orthonormal matrix R such that

$$R[X \mid \underline{Y}] = \left[\frac{T}{0} \right] \tag{3.1}$$

where T is $(p+1) \times (p+1)$ layer triangular. Such a matrix T is unique. Gentleman [1] gives a numerically stable method for obtaining T that has low storage requirements. Other methods are summarized by Lawson and Hanson [2].

Partition T as

$$T = \left[\begin{array}{c|c} T_1 & \underline{t}_2 \\ \hline 0 & t_3 \end{array} \right] , \tag{3.2}$$

where T_1 is $p \times p$. Since T was obtained from $[X \mid \underline{Y}]$ by an orthogonal transformation

$$T_1'T_1 = X'X \ , \tag{3.3}$$

$$T_1'\underline{t}_2 = X'\underline{Y} \ , \tag{3.4}$$

and

$$\underline{t}_2'\underline{t}_2 + t_3't_3 = \underline{Y}'\underline{Y} \ . \tag{3.5}$$

Theorem 3.1. Let T_1, $\underline{t}_2$, and t_3 be as given by (3.1) and (3.2). Let $\mathcal{L}_{T_1}$ and I_{T_1} be as defined in Definition 2.1. Let $\underline{\ell}$ be a p vector of constants, $\theta = \underline{\ell}'\underline{\beta}$, and $\underline{q}' = \underline{\ell}'T_1^+$ where $T_1^+ = (T_1 - I_{T_1})^{-1}$. The following statements are true.

(i) The parameter θ is estimable if and only if

$$\sum_{i \in \mathcal{L}_{T_1}} q_i^2 = 0 \ . \tag{3.6}$$

(ii) If θ is estimable then

$$\underline{\ell}'(X'X)^-X'\underline{Y} = \underline{\ell}'T_1^+\underline{t}_2 \tag{3.7}$$

and

$$\underline{\ell}'(X'X)^-\underline{\ell} = \sum_{i \notin \mathcal{L}_{T_1}} q_i^2 \ . \tag{3.8}$$

(iii) For t_3 as given in (3.2)

$$\underline{Y}'\underline{Y} - \underline{Y}'X(X'X)^-X'\underline{Y} = t_3^2 \ .$$

PROOF. First we consider (i). By 3.1 and 3.2 the vector space spanned by the rows of T_1 is the same as the space spanned by the rows of X . The rank of this space is r . The rank of the orthocomplement to the row space of X is p - r . The matrix $T_1^+I_{T_1}$ has p - r linearly independent columns. By (2.1) $T_1T_1^+I_{T_1} = 0$ and thus the nonzero columns of $T_1^+I_{T_1}$ are a basis of the orthocomplement to the row space of X . A vector $\underset{\sim}{\ell}$ is in the row space of X, and thus $\underset{\sim}{\ell}'\underset{\sim}{\beta}$ is estimable, if and only if $\underset{\sim}{\ell}'T_1^+ I_{\mathcal{T}_1} = \underset{\sim}{0}'$. This condition is satisfied if and only if

$$\sum_{i \notin \mathcal{T}_1} q_i^2 = \underset{\sim}{\ell}' T_1^+ I_{\mathcal{T}_1} T_1^{+'} \underset{\sim}{\ell} = 0 .$$

We now consider (ii) and (iii). Assume that (3.6) is satisfied. Using relationships (3.3) — (3.5), (2.1) and that $T_1^+ T_1^{+'}$ is a choice for $(X'X)^-$, substitute into the left hand sides of (3.7) — (3.9) and simplify.

4. DISCUSSION

Most numerical problems in linear model computations are associated with obtaining T rather than with inverting $T_1 - I_{\mathcal{T}_1}$. The references cited discuss the numerical properties of several algorithms. Computing T and all computations suggested by this paper may be done in approximately $(p+1)^2/2$ floating point storage locations. Conventional methods must store both $(X'X)^-$ and $X'X(X'X)^-$ and require about twice this amount of storage.

For the purpose of comparing algorithms with respect to speed, we consider an operation to be a multiplication or a division and an addition. The standard method for calculating $(X'X)^-$ is to obtain T_1, T_1^-, and then $(X'X)^- = T_1^- T_1^{-'}$ (for example see [4]). Since calculation of T_1 is common to both the standard method and the method proposed by this paper, we compare only the additional number of operations required. The asymptotic number of operations required for both T_1^+ and $T_1^+ \underset{\sim}{t}_2$ is $p^3/6$. For each different $\underset{\sim}{\ell}$ considered, (3.6), (3.7), and (3.8) require about $p^2/2$ operations. The asymptotic number of operations required for both $(X'X)^-$ and $X'X(X'X)^-$ is $4p^3/3$. For each different $\underset{\sim}{\ell}$ considered, (1.1), (1.2), and (1.3) require about $2p^2$ operations.

Other matrices can be shown to have the same properties as T_1^+ . However, T_1^+ has the practical feature that $i \in \mathcal{L}_{T_1}$ if and only if the i^{th} diagonal element of T_1^+ is -1 . For this reason there need not be any storage allocated for $\mathcal{L}_{T_1}$.

5. NUMERICAL EXAMPLE

The preceding methods are illustrated using data from a two-way classification model without interaction. Both factors have two levels. There are four observations at each combination of levels. Let

$$[X \mid Y] = \left[\begin{array}{ccccc|c}
1 & 1 & 0 & 1 & 0 & 45 \\
1 & 1 & 0 & 1 & 0 & 96 \\
1 & 1 & 0 & 1 & 0 & 37 \\
1 & 1 & 0 & 1 & 0 & 64 \\
1 & 1 & 0 & 0 & 1 & 85 \\
1 & 1 & 0 & 0 & 1 & 92 \\
1 & 1 & 0 & 0 & 1 & 75 \\
1 & 1 & 0 & 0 & 1 & 39 \\
1 & 0 & 1 & 1 & 0 & 53 \\
1 & 0 & 1 & 1 & 0 & 21 \\
1 & 0 & 1 & 1 & 0 & 68 \\
1 & 0 & 1 & 1 & 0 & 45 \\
1 & 0 & 1 & 0 & 1 & 91 \\
1 & 0 & 1 & 0 & 1 & 75 \\
1 & 0 & 1 & 0 & 1 & 64 \\
1 & 0 & 1 & 0 & 1 & 34
\end{array}\right] ,$$

and

$$\beta = [\mu \quad \alpha_1 \quad \alpha_2 \quad \delta_1 \quad \delta_2]' \quad .$$

Thus

$$T = \left[\begin{array}{ccccc|c}
4 & 2 & 2 & 2 & 2 & 246 \\
 & 2 & -2 & 0 & 0 & 20.5 \\
 & & 0 & 0 & 0 & 0 \\
 & & & 2 & -2 & -31.5 \\
 & & & & 0 & 0 \\
\hline
 & & & & & 81.6670
\end{array}\right] ,$$

$$\mathcal{L}_{T_1} = \{3, 5\} ,$$

and

$$[T_1^+ \mid T_1^+ \underline{t}_2] = \left[\begin{array}{ccccc|c} \frac{1}{4} & -\frac{1}{4} & 1 & -\frac{1}{4} & 1 & 64.25 \\ & \frac{1}{2} & -1 & 0 & 0 & 10.25 \\ & & -1 & 0 & 0 & 0 \\ & & & \frac{1}{2} & -1 & -15.75 \\ & & & & -1 & 0 \end{array}\right] .$$

Suppose $\theta = \mu + \gamma_1 + \delta_2 = \underline{\ell}'\beta$ where $\underline{\ell}' = (1 \;\; 1 \;\; 0 \;\; 0 \;\; 1)$. Then $\underline{\ell}'T_1^+ = \underline{q}'$ $= (\frac{1}{4} \;\; \frac{1}{4} \;\; 0 \;\; -\frac{1}{4} \;\; 0)$. The parameter θ is estimable since

$$\sum_{i \in \mathcal{L}_{T_1}} q_i^2 = q_3^2 + q_5^2 = 0 .$$

The estimate is

$$\hat{\theta} = \underline{\ell}'T_1^+\underline{t}_2 = 64.25 + 10.25 + 0$$
$$= 74.50 .$$

The variance of $\hat{\theta}$ (apart from σ^2) is

$$\sum_{i \notin \mathcal{L}_{T_1}} q_i = q_1^2 + q_2^2 + q_4^2$$
$$= (\tfrac{1}{4})^2 + (\tfrac{1}{4})^2 + (-\tfrac{1}{4})^2$$
$$= 3/16 .$$

SSE is $(81.667)^2 = 6669.5$. Suppose $\theta = \gamma_1 + \delta_2 = \underline{\ell}'\underline{\beta}$ where $\underline{\ell}' = (0 \;\; 1 \;\; 0 \;\; 0 \;\; 1)$. Then $\underline{\ell}'T_1^+ = \underline{q}' = (0 \;\; \frac{1}{2} \;\; -1 \;\; 0 \;\; -1)$. The parameter θ is not estimable since

$$\sum_{i \in \mathcal{L}_{T_1}} q_i^2 = q_3^2 + q_5^2$$
$$= (-1)^2 + (-1)^2$$
$$\neq 0 .$$

REFERENCES

[1] GENTLEMAN, W. MORVEN (1973). Least squares computation by Givens transformations without square roots. J. Inst. Math. Applic. 12:329-336.

[2] LAWSON, CHARLES L., and HANSON, RICHARD J. (1974). Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs, New Jersey.

[3] MILLIKEN, GEORGE A. (1971). New criteria for estimability for linear models. Ann. Math. Statist. 42:1588-1594.

[4] SCHWARZ, H. R. (1973). Numerical Analysis of Symmetric Matrices. Prentice-Hall, Englewood Cliffs, New Jersey.

[5] SEARLE, S. R. (1971). Linear Models. Wiley, New York.