

Deferment of Computation in the Method of Least Squares

By Irving H. Siegel

This note calls attention to a variant formulation of the least-squares adjustment procedure [1] that should prove increasingly attractive as electronic data-processing equipment becomes ever more widely available. It indicates that the statistician need not concern himself with the two rituals featured in traditional textbook presentations—namely, the construction and solution of the conventional “normal equations.” By simply arranging the original data (together with 1’s and 0’s) in a particular pattern and without engaging in any arithmetic operations, he (or an aide) may instantly form the essential matrix system to be solved. This system contains no processed elements—no sums or products—yet is algebraically equivalent to the usual normal equations. Thus, model design, data compilation, and explicit statement of the simultaneous equations to be solved are entirely divorcible, in linear regression analysis, from computation, which may be deferred and delegated to specialists as a unified task.

To illustrate the variant approach, we consider the classical linear regression problem in which n observed values of only one variate are deemed subject to (random) error. This linear model has the matrix form

$$(1) \quad Y = Xb + e,$$

where Y is the $(n \times 1)$ vector of observations subject to error, X is the given $(n \times k)$ matrix of explanatory coefficients ($n > k$), b is the $(k \times 1)$ vector of parameters to be estimated, and e is the $(n \times 1)$ vector of residuals with zero expectation. (One of the columns of the X matrix consists entirely of 1’s when a constant intercept is included in the model design.)

When the n observations are unweighted, the least-squares principle requires minimization of the scalar sum of n squared residuals, $S = e'e = (Y - Xb)' \cdot (Y - Xb)$. The accent here denotes transposition. Minimizing this expression with respect to b , we obtain the adjustment condition

$$(2a) \quad X'e = 0,$$

which is more often written in the form of a composite matrix normal equation,

$$(2b) \quad X'Xb = X'Y.$$

It is a significant fact that (1) and (2a), which state the initial observation equations and the additional least-squares constraints, may be incorporated at once into a supermatrix system corresponding to (2b) but containing no processed elements. Thus, literally, all of the arithmetic operations associated with regression analysis may be shifted to the programmer and to other computer-associated personnel for execution in the manner they deem best. Specifically, (1) and (2a) may be combined to form the basic matrix equation

$$(3) \quad \begin{bmatrix} I & X \\ X' & 0 \end{bmatrix} \begin{bmatrix} e \\ b \end{bmatrix} = \begin{bmatrix} Y \\ 0 \end{bmatrix}.$$

Here, I is the $(n \times n)$ identity submatrix that premultiplies the $(n \times 1)$ error subvector, e , upon expansion. All the other "packing" included in (3) consists of zero elements. The system does not need to be solved for e (except, perhaps, as a check). Other compatible arrangements of the submatrices and subvectors are feasible without any change in the results.

The deferment of computation is achieved, of course, at the expense of compactness. The leftmost matrix of (3) is square, symmetric, and nonsingular, like the corresponding matrix for conventional normal equations, and its determinant also has the same numerical value as the conventional least-squares determinant (i.e., $X'X$, the product taken along the secondary diagonal with sign ignored). It is of greater dimensionality, however, incorporating explicit information for the coefficients of two sets of unknown elements—errors in addition to parameters. Since the coefficients assigned to the errors in (3) by the least-squares principle (X') are given immediately by the transposition of the submatrix associated with the original observations (X), no new data are required for completion of the array of unprocessed elements.

When the n observations represented by (1) are unequally weighted, the least-squares criterion requires minimization of $S_w = e'We = (Y - Xb)'W(Y - Xb)$. Accordingly, for the adjustment condition we obtain

$$(4a) \quad X'We = 0,$$

or, if we prefer, the matrix version of the conventional weighted normal equations:

$$(4b) \quad X'WXb = X'WY.$$

Equations (1) and (4a) may easily be incorporated into a supermatrix system that is equivalent to (4b). If all preliminary processing of the data is to be avoided, however, it is necessary first to expand (4a) into two equations. Accordingly, the resulting supermatrix system without processed elements has three tiers:

$$(5) \quad \begin{bmatrix} 0 & I & X \\ I & W & 0 \\ X' & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ e \\ b \end{bmatrix} = \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix}.$$

Here, W is the diagonal $(n \times n)$ submatrix of given weights, and λ refers to n dummy variables (resembling Lagrange multipliers) that do not need to be determined explicitly.

The leftmost term of (5) has a simple memorable structure. It is triangular with respect to its submatrices and is symmetric with reference to its principal diagonal. Its determinant has the numerical value of the product along the secondary diagonal, i.e., $X'WX$. The same product occurs in (4b) and is also the value of the determinant of the smaller square matrix with processed elements that is typically derived for the weighted regression case.

In the concentrated computation phase, inversion of the turgid squares matrices in (3) and (5) may be tackled directly. Despite their size, they need not tax modern computing equipment; and certain features, such as symmetry, may contribute to efficient handling. At worst, the work may be programmed to entail the construction and solution of conventional normal equations. In any case, it is clear that all

the arithmetic processing involved in a large least-squares problem is assignable to specialists in one package.

8312 Bryant Drive
Bethesda, Maryland

1. I. H. SIEGEL, "Least-squares 'without normal equations,'" *Trans. New York Acad. Sci. Ser. II*, v. 24, pp. 362-371.