

SOME USES OF GENERALIZED INVERSE MATRICES
IN STATISTICS*

S. R. Searle
Biometrics Unit, Cornell University

BU-229-M

November, 1966

ABSTRACT

This paper summarizes properties of generalized inverse matrices and some of their uses in statistics, particularly those relating to linear models. Neither proofs nor numerical examples are given.

* Paper prepared for the Department of Statistics Colloquium, The University of Connecticut, Storrs, December 8, 1966.

SOME USES OF GENERALIZED INVERSE MATRICES
IN STATISTICS*

S. R. Searle

Biometrics Unit, Cornell University

BU-229-M

November, 1966

Definition

Penrose (1955) showed that for any non-null matrix A , square or rectangular, there exists a unique matrix G satisfying the four conditions:

(i) $AGA = A$

(ii) $GAG = G$

(iii) GA is symmetric

and

(iv) AG is symmetric .

He called G the generalized inverse of A , and showed its use in solving systems of linear equations, namely that if equations $Ax = y$ are consistent then $x = Gy + (GA - I)z$ is a solution to x for any arbitrary vector z .

Several adaptations of the generalized inverse concept have been used by other workers. Rao (1955) and Greville (1957) for example, refer to a variant of G as a 'pseudo inverse', Wilkinson (1958) calls it an 'effective inverse' and Zelen and co-workers [Goldman and Zelen (1964), for example] use the term "weak generalized inverse" for a matrix G that satisfies condition (i), (ii) and (iii).

Concerning the solution of linear equations, it has been shown by Rao (1962) that only condition (i) is needed. Thus consistent equations $Ax = y$

* Paper prepared for the Department of Statistics Colloquium, The University of Connecticut, Storrs, December 8, 1966.

have the solution $x = Gy + (GA - I)z$ for any G satisfying $AGA = A$, z being any arbitrary vector of the same order as x . Indeed the condition $AGA = A$ is both necessary and sufficient for the solution of $Ax = y$ to be expressible in this form. This is perhaps the most important consequence of any matrix G that satisfies the first of the four conditions (i) - (iv). And since matrices which satisfy both (i) and any others of (ii), (iii) and (iv), or all of them, are just subsets of those that satisfy only (i), the definition 'generalized inverse' is here given to those matrices satisfying (i) alone. This is also the definition used in Searle (1966).

Thus if, for some matrix A , G is such that $AGA = A$, then G is said to be the generalized inverse of A . And furthermore, solutions to consistent equations $Ax = y$ are $x = Gy + (GA - I)z$ for any arbitrary vector z of the same order as x .

Notice that when A^{-1} exists it equals G , and the equations $Ax = y$ have the sole, familiar solution $x = A^{-1}y$. Thus A^{-1} , when it exists, is just a special case of G ; hence the two problems of solving $Ax = y$ when A is non-singular and when it is singular need not be distinguished. In all cases the solution can be taken as $x = Gy + (GA - I)z$.

Computation

The matrix G that satisfies (i) - (iv) is difficult to compute. But one that satisfies (i) and (ii) is easily derived. Suppose that

$$PAQ = \Delta = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}$$

is the diagonal form of A , where P and Q are non-singular and D is diagonal of order $r = r(A)$, the rank of A . Then, on defining

$$\Delta^{-1} = \begin{bmatrix} D_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

a matrix G satisfying (i) and (ii) is

$$G = Q\Delta^{-1}P.$$

Rao (1962) gives a somewhat similar method for calculating a G that satisfies (i) but not (ii).

Whether or not any of the conditions (ii) - (iv) are satisfied it is clear that many matrices G satisfy (i). Nonetheless, for each of them, we also have

$$H = GA, \text{ idempotent}$$

$$r(H) = r,$$

and
$$r(H - I_q) = r - q$$

where q is the order of x.

Linear Models

The equation of the general linear model can be written as

$$y = X\beta + e$$

where y is an N x 1 vector of responses, X is an N x p matrix of known coefficients, β is a p x 1 vector of parameters and e is an N x 1 vector of randomly distributed error terms having zero mean and variance-covariance matrix $\sigma_e^2 I$. (This variance-covariance matrix could be $\sigma_e^2 A$ in the general case situations, but our discussion here is confined to $\sigma_e^2 I$.) When the vector β represents fixed effects (Model I of Eisenhart, 1947) the normal equations resulting from the least squares procedure for estimating β are

$$X'Xb = X'y.$$

(a) Estimable functions. The linear function $q'\beta$ of β can be defined as estimable if there exists some vector t' such that $t'y$ has expectation $q'\beta$. And when $q'\beta$ is estimable its best linear unbiased (BLU) estimator is $q'b$ where b is any solution to $X'Xb = X'y$. Solutions to this can be expressed as

$$b = GX'y + (H - I)z$$

where G is a generalized inverse of $X'X$, $H = GX'X$ and z is arbitrary. In particular, taking z as null gives the specific solution

$$b_0 = GX'y.$$

By noting that $E(b_0) = GX'X\beta = H\beta$ it is clear that b_0 is the BLU estimator of $H\beta$.

In general, then, $q'b_0 = q'GX'y$ is the BLU estimator of the estimable function $q'\beta$, invariant to which generalized inverse of $X'X$ is used for G .

(b) What functions are estimable? It is well-known [e.g. Graybill (1961), Theorem 11.3] that the necessary and sufficient condition for $q'\beta$ to be estimable is that there be a solution for u to the equation $X'Xu = q$. A more helpful form of this condition is that $q'\beta$ is estimable for q' of the form $q' = q'H$. For, if $q'\beta$ is estimable, $q' = u'X'X$ and then

$$q'H = u'X'XGX'X = u'X'X = q' ;$$

and if q' does equal $q'H$ then

$$q'\beta = q'GX'\beta = q'GX'E(y) = E[(q'GX')y]$$

and so, by definition, $q'\beta$ is estimable.

Now note that $q' = q'H$ is true whenever

$$q' = w'H$$

no matter what the vector w' is. Furthermore, because $r(H) = r$ there are only r linearly independent vectors $q' = w'H$, and hence only r linearly independent estimable functions. And the BLU estimator of $q'\beta = w'H\beta$ is

$$q'b = q'b_0 = w'HGX'y = w'GX'XGX'y$$

and because, for any generalized inverse of $X'X$ it can be shown that $X'XGX' = X'$, we have

$$q'b = w'GX'y = w'b_0 .$$

Thus for any arbitrary vector w' , $w'H\beta$ is estimable, its BLU estimator is $w'b_0$, and there are such linearly independent functions. This is the result indicated in Searle (1965 and 1966).

(c) Variances. The variance of b is

$$\text{var}(b) = G\sigma^2 ;$$

and that of the BLU estimator of an estimable function $w'H\beta$ is

$$\text{var}(w'b_0) = w'Gw\sigma^2 .$$

(d) Residual sums of squares. The residual sum of squares after fitting the model $y = X\beta + e$ is

$$SSR_0 = (y - Xb)'(y - Xb)$$

which, because $b = GX'y + (H - I)z$ and $XH = X$, reduces to

$$SSR_0 = y'(I - XGX')y .$$

As one would expect, XGX' can be shown invariant to the choice of G .

The expected value of SSR_0 is

$$\begin{aligned} E(SSR_0) &= \text{tr}\{(I - XGX')E(yy')\} \\ &= \text{tr}\{(I - XGX')(X\beta\beta'X + \sigma^2I)\} \\ &= \text{tr}(X - XGX'X)(\beta\beta'X) + \sigma^2\text{tr}(I - GX'X) \\ &= (N - r)\sigma^2 . \end{aligned}$$

(e) Tests of hypotheses. The general linear hypothesis $Q'\beta = m$ can be tested by fitting the model $y = X\beta + e$ subject to the hypothesis $Q'\beta = m$. It can be shown (Searle, 1965) that in doing this the residual sum of squares, SSR_H , is

$$SSR_H = SSR_0 + (Q'b_0 - m)'(Q'GQ)^{-1}(Q'b_0 - m)$$

and hence the F-test of the hypothesis is

$$F = \frac{(Q'b_0 - m)'(Q'GQ)^{-1}(Q'b_0 - m)}{s} \cdot \frac{N - r}{SSR_0} ,$$

s being the number of rows in Q . Clearly $Q'b_0$ must be invariant to G for F to be meaningful, i.e. $Q'\beta$ must consist of s estimable functions. It will be found that under these conditions $(Q'GQ)^{-1}$ always exists.

Although the estimability of $Q'\beta$ is a sufficient condition for the existence of $(Q'GQ)^{-1}$ it is not a necessary one. Thus it is possible to calculate an F value as given above for some hypotheses $Q'\beta = m$ where $Q'\beta$ is not estimable. The calculated F is then a test of the hypothesis $Q'H\beta = m$.

(f) Restricted models. There is no need at all, when using generalized inverses, to introduce the "customary constraints" or "usual restraints" that are often added to the normal equations in order to solve them. For example, in balanced models involving k treatment effects τ_i , $i = 1, 2, \dots, k$, the restraint $\sum_{i=1}^k \tau_i = 0$ is often used. This kind of expression simplifies many normal equations, especially in cases of equal subclass numbers, and it can be logically rationalized along the lines of "making the mean treatment effect zero" or "considering the treatment effects measured from their mean". But such restraints have no use in unequal subclass analyses, and their presence often confuses the student. This, because different restraints give different solutions but, of course, not different BLU estimators of estimable functions, a fact not as readily presentable as with generalized inverses.

Despite what has just been said, the elements of a model sometimes are restricted. Thus, confining ourselves to linear restrictions, the elements of β in the model $y = X\beta + e$ might be subject to the restriction $P'\beta = \alpha$. If $P'\beta$ is not estimable the only effect on the model is on the solution to the normal equations. It becomes $b^* = b_0 + (H - I)z^*$ where z^* is given by $P'(H - I)z^* = \alpha - P'b_0$. Thus solutions b^* are simply a subset of the solutions b and there is no other effect on the estimation processes.

But when $P'\beta$ is estimable the solutions are

$$b^* = b_0 - GP(P'GP)^{-1}(P'b_0 - \alpha) + (H - I)z.$$

All functions that were estimable in the unrestricted model are still estimable, although of course they can be re-interpreted in light of the restrictions. Thus the estimable function $Q'\beta$ becomes equivalent to $Q'\beta + \Lambda(P'\beta - \alpha)$ for any matrix Λ of order $s \times k$, Q' having order $s \times p$ and P' being $k \times p$. The residual sum of squares in the restricted model is

$$SSR_{0*} = SSR_0 + (P'b_0 - \alpha)'(P'GP)^{-1}(P'b_0 - \alpha)$$

with expected value $[n - (r - k)]\sigma^2$. To test the hypothesis $Q'\beta = m$ in this restricted model the residual sum of squares for fitting the model $y = X\beta + e$ subject to both the restriction $P'\beta = \alpha$ and the hypothesis $Q'\beta = m$ is

$$SSR_{H*} = SSR_0 + (T'b_0 - \tau)'(T'GT)^{-1}(T'b_0 - \tau)$$

where

$$T = \begin{bmatrix} P' \\ Q' \end{bmatrix} \quad \text{and} \quad \tau = \begin{bmatrix} \alpha \\ m \end{bmatrix} .$$

Under the null hypothesis, $E(SSR_{H^*}) = [n - (r - s - k)]\sigma^2$ and the F-value for testing the hypothesis in this case is

$$F = \frac{SSR_{H^*} - SSR_{O^*}}{s} \cdot \frac{n - (r - k)}{SSR_{O^*}} .$$

This result and many of the others given above are developed in detail in Searle (1966a).

Variance Components

For a suitable partitioning of X and β the model $y = X\beta + e$ can be written as $y = X_1\beta_1 + X_2\beta_2 + \dots + X_k\beta_k + e$ where $\beta_1, \beta_2, \dots, \beta_k$ can represent the vectors of effects due to different levels of k classifications, one of which may be the 1-level class of the general mean. In this form the model is well suited to variance component estimation by the method (expounded in Henderson, 1953) of equating observed values of quadratic forms of the observations to their expected values. In the completely random model (model II of Eisenhart, 1947) unbiased estimators of variance components are obtained by this method. But in the mixed model, if fixed effects are treated as random, the method gives unbiased estimators. Under certain conditions this bias can be removed by estimating the fixed effects (treating the random effects as fixed, for so doing), correcting the data in accord with the resulting estimates, and then proceeding as if there were no longer any fixed effects in the model. This is Method 2 of Henderson (1953). Its use demands solving normal equations to estimate the fixed effects. Since the solutions to these equations for the random effects are not needed the equations are best partitioned and solutions for the fixed effects derived from using a generalized inverse of a partitioned matrix, a result given by Rohde (1965). This result is as follows:

$$\begin{bmatrix} A & B \\ B' & C \end{bmatrix} \text{ has a generalized inverse } \begin{bmatrix} R^- & -R^-BC^- \\ -C^-B'R^- & C^- + C^-B'R^-BC^- \end{bmatrix}$$

where $R = A - BC\bar{B}$, with $R\bar{}$ and $C\bar{}$ being generalized inverses of R and C respectively.

Details of the use of this in the mixed model variance component estimation situation are given in Searle (1966b).

Distribution of Quadratic Forms

Graybill (1961), in his fourth chapter, gives several theorems relating to the distribution of quadratic forms. His Theorem 4.9 states that if x is a vector of normally distributed random variables having vector of means μ and variance-covariance matrix V , then the quadratic form $x'Ax$ is distributed as a non-central chi-square with q degrees of freedom and non-centrality parameter $\frac{1}{2}\mu'A\mu$ if and only if AV is idempotent, q being the rank of A . Implicit in the proof of the theorem is the non-singularity of the variance-covariance matrix V .

The case of V being non-singular, of rank q , can also be considered. We then have a theorem: if A is a generalized inverse of V , $x'Ax$ has the non-central chi-square distribution with q degrees of freedom and non-centrality parameter $\frac{1}{2}\mu'A\mu$. Notice that this condition is sufficient, but not necessary. Rao (1962) has a lemma indicating both necessity and sufficiency, but only the sufficiency can be upheld. This error is acknowledged in Rao (1965). The theorem referred to is a particular result of a more general theorem given in Rohde et al (1966): if x is a vector of variables that are normally distributed with vector of means μ and singular variance-covariance matrix V , of rank q , then a necessary and sufficient condition that the quadratic form $x'Ax$ has a non-central χ^2 -distribution with q degrees of freedom and non-centrality parameter $\frac{1}{2}\mu'A\mu$ is that

$$k\mu'A(VA)^{k-1}\mu + \text{tr}(VA)^k = k\mu'A\mu + q$$

for all positive integers k .

References

- Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics* 3, 1.
- Graybill, Franklin A. (1961). "An Introduction to Linear Statistical Models". Vol I. McGraw-Hill, New York.
- Goldman, A. J. and Zelen, M. (1964). Weak generalized inverses and minimum variance linear unbiased estimators. *J. Res. Nat. Bur. Stand.* 68B, 151-172.
- Greville, T. N. E. (1957). The pseudo inverse of a rectangular or singular matrix and its application to the solution of systems of linear equations. *S.I.A.M. Newsletter*, 5, 3-6.
- Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics* 9, 226.
- Penrose, R. A. (1955). A generalized inverse for matrices. *Proc. Camb. Phil. Soc.* 51, 406-413.
- Rao, C. R. (1955). Analysis of dispersion for multiple classified data with unequal numbers of cells. *Sankhya*, 15, 253-280.
- Rao, C. R. (1962). A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *J.R.S.S. (B)*, 24, 152-158.
- Rao, C. R. (1965). Generalized inverse for matrices and its applications in mathematical statistics, in "Festschrift for J. Neyman, Research Papers in Statistics" pp. 263-279, Ed. F. N. David, Wiley and Sons, N.Y.
- Rohde, Charles A. (1965). Generalized inverses of partitioned matrices. *J. Soc. Ind. Appl. Math.* 13, 1033-1035.
- Rohde, Charles A., Urquhart, N. S. and Searle, S. R. (1966). Quadratic forms in singular normal variables. (Submitted to *Ann. Math. Stat.*)
- Searle, S. R. (1965). Additional results concerning estimable functions and generalized inverse matrices. *J.R.S.S. (B)*, 27, 486-490.
- Searle, S. R. (1966). "Matrix Algebra for the Biological Sciences". Wiley, New York.
- Searle, S. R. (1966a). Estimable functions and testable hypotheses in linear models. Paper No. BU-213-M, Mimeo Series, Biometrics Unit, Cornell Univ.
- Searle, S. R. (1966b). A matrix reformulation of Henderson's three methods of estimating variance components. Paper No. BU-225-M, Mimeo Series, Biometrics Unit, Cornell University.
- Wilkinson, G. N. (1958). Estimation of missing values for the analysis of incomplete data. *Biometrics*, 14, 257-286.