GENERALIZATIONS OF A GAUSSIAN THEOREM

By Paul S. Dwyer

University of Michigan

1. Introduction and Summary. Plackett [1] has discussed the history and generalizations of the Gaussian theorem which states that least squares estimates are linear unbiased estimates with minimum variance. General forms of the theorem are due to Aitken [2], [3] and Rao [4], [5]. The essence of the proof for Aitken's general case consists in minimizing, simultaneously, certain quadratic forms involving linear combinations of the parameters. Plackett derived Aitken's result by using a matrix relation. The proof of the theorem follows quickly once the relation is established. A somewhat similar but simpler matrix relation is used by Rao ([4], page 10).

Aitken [2] and Rao [4], [5] obtain minimum variance with the use of Lagrange multipliers. Unless one has a method of working with matrices of derivatives it seems necessary to differentiate with respect to the many scalars constituting the matrices and to assemble the results in desired matrix form. Authors frequently give only the assembled results ([4], page 10, [5], page 17, [6], page 83).

The question arises as to whether it is possible to use the logically preferable matrix derivative methods of minimization. It is shown below that the use of matrices of partial derivatives [7] leads logically to the solution without the necessity of changing to and from scalar notation, or without the necessity of establishing some relation which implicitly contains the solution. Matrix derivative methods seem to be preferable methods for undertaking solutions of problems of simultaneous matrix minimization with side conditions for the same reason that derivative methods are preferable to the use of some (unknown) relation in solving problems of minimization involving scalars. They may also be used in establishing the relation which may then be verified without their use.

The paper includes generalizations of the results of Aitken [2], [3], Rao [4], [5], and David and Neyman [8]. It gives a general formula for simultaneous unbiased estimators of linear functions of parameters when the parameters are subject to linear restrictions and shows how the results are applicable to special cases. It provides formulas for the variance matrix of these estimators. It generalizes a matrix relation used by Plackett [1]. It uses the matrix square root transformation in establishing the general result for the variance of (weighted) residuals when there may be linear restrictions on the parameters. It provides a generalization of a formula of David and Neyman [8] in estimating the variance matrix of the unbiased linear estimators.

2. The least squares solution. The (inconsistent) observational equations are

$$A\theta = x \tag{2.1}$$

Received February 18, 1957; revised September 17, 1957.

and the true linear regression is given by

$$\mathcal{E}(x) = A\theta, \tag{2.2}$$

where the values of x, A, and θ are real. We set

$$A\theta - x = \epsilon \tag{2.3}$$

so that

$$\mathcal{E}(\epsilon) = 0. \tag{2.4}$$

In determining the least squares regression we have $\theta(s \times 1)$ as the vector of unknown parameters, $x(n \times 1)$ as the vector of measurements of the variable of regression, $\epsilon(n \times 1)$ as the vector of errors and $A(n \times s)$ as the matrix of measurements of the regressed variables. We take $s < n$ and A of rank s. Further, under the usual regression condition of fixed A,

$$\begin{aligned} V &= \mathcal{E}(xx^T) - \mathcal{E}(x)\mathcal{E}(x^T) = \mathcal{E}(\epsilon\epsilon^T) = V^T \\ &= \text{var } (x) = \text{var } \epsilon \end{aligned} \tag{2.5}$$

is the dispersion matrix of x and ϵ. We limit our discussion to the case where V is positive definite. A common dimensionless generalization of the least squares concept uses weights for the observations with $W = V^{-1}$ and leads to

$$Q = \epsilon^T V^{-1} \epsilon = (A\theta - x)^T V^{-1} (A\theta - x) \tag{2.6}$$

as the form to be minimized. The value of θ which minimizes (2.6) is known to be

$$\theta^* = (A^T V^{-1} A)^{-1} A^T V^{-1} x. \tag{2.7}$$

This result can be derived using symbolic matrix derivatives ([7], page 524). We have successively

$$Q = \theta^T A^T V^{-1} A\theta - \theta^T A^T V^{-1} x - x^T V^{-1} A\theta + x^T V^{-1} x, \tag{2.8}$$

$$\frac{\partial Q}{\partial \langle \theta \rangle} = J^T A^T V^{-1} A\theta + \theta^T A^T V^{-1} AJ - J^T A^T V^{-1} x - x^T V^{-1} AJ, \tag{2.9}$$

$$\frac{\partial \langle Q \rangle}{\partial \theta} = A^T V^{-1} A\theta K^T + A^T V^{-1} A\theta K - A^T V^{-1} x K^T - A^T V^{-1} x K, \tag{2.10}$$

and since Q is scalar, $K = K^T = 1$. Setting $\theta = \theta^*$ when $\dfrac{\partial \langle Q \rangle}{\partial \theta} = 0$, we get (2.7).

We note that θ^* is an unbiased estimate of θ because of (2.2) and (2.7).

3. Linear estimates with minimum variance. Now consider the k linear parametric functions

$$\phi = L\theta, \tag{3.1}$$

where $L = L(k \times s)$ is known. Then $\phi = \phi(k \times 1)$. We wish to find

$$\phi^* = L\theta^* \tag{3.2}$$

such that ϕ^* is an unbiased estimate of ϕ with minimum variance. This means that the diagonal terms of var ϕ (a matrix of order $k \times k$) should attain their minimum values simultaneously. Following Aitken we consider solutions of the form

$$\phi^* = Bx \tag{3.3}$$

and determine $B = B(k \times n)$. Rao [4] has shown that this homogeneous form is the general form. The relation

$$(BA - L)\theta = 0 \tag{3.4}$$

follows from (3.3), (2.2), and (3.1) in accordance with the requirement that ϕ^* be an unbiased estimate of ϕ.

Aitken [3] has shown using Lagrangian multipliers and Plackett [1] using a matrix relation that the value of θ^* in (3.2) which minimizes the diagonal term of var Q^* is identical with the θ^* resulting from least squares as given by (2.7). This Aitken theorem is a generalization of the Gaussian theorem that least squares linear estimators are unbiased with minimum variance.

Rao [5] further generalized the theorem with a consideration of linear restrictions on the parameters when $k = 1$. The argument is given below for the more general k. The preparation of the problem for minimization is similar to that of Rao in the special case with $k = 1$, though there are some modifications. The $u < s$ independent linear restrictions may be indicated by

$$g = R\theta \equiv 0, \tag{3.5}$$

where $R = R(u \times s)$ and $g = g(u \times 1)$, without loss of generality since any term not having some θ_i as a factor may be multiplied by $\theta_0 = 1$ and s replaced by $s' = s + 1$. We premultiply by the undetermined $D = D(k \times u)$ to get

$$DR\theta = Dg, \tag{3.6}$$

in which the matrix coefficient of θ has the same order as BA and L. Then the condition for unbiased estimation *and* the specific side conditions are incorporated in the matrix relation

$$(L - BA)\,\theta \equiv 0 \equiv DR\theta - Dg \tag{3.7}$$

so that the conditions for estimation can be written in the form

$$(L - BA - DR) = 0 \quad \text{and} \quad Dg = 0. \tag{3.8}$$

Specifically our purpose is the minimization of the diagonal terms of var ϕ^* subject to (3.8). Now

$$\text{var}\,\phi^* = \mathcal{E}(\phi^*\phi^{*T}) - \mathcal{E}(\phi^*)\mathcal{E}(\phi^{*T}) = B[\mathcal{E}(xx^T) - \mathcal{E}(x)\mathcal{E}(x^T)]B^T = BVB.^T \tag{3.9}$$

We can then use

$$\psi = BVB^T + 2(L - BA - DR)\,\Lambda + 2Dg\mu^T, \tag{3.10}$$

where $\psi = \psi\ (k \times k)$, $\Lambda = \Lambda\ (s \times k)$, and $\mu = \mu\ (k \times 1)$ and differentiate with respect to B and D. We have

$$\frac{\partial\psi}{\partial\langle B\rangle} = JVB^T + BVJ^T - 2JA\Lambda,$$

$$\frac{\partial\langle\psi\rangle}{\partial B} = KBV + K^TBV - 2K\Lambda^TA^T,$$

so that the critical value, for each and every diagonal term, is given by

$$BV = \Lambda^TA^T. \tag{3.11}$$

Again

$$\frac{\partial\psi}{\partial\langle D\rangle} = -2JR\Lambda + 2Jg\mu^T,$$

$$\frac{\partial\langle\psi\rangle}{\partial D} = -2K\Lambda^TR^T + 2K\mu g^T,$$

$$\frac{\partial\psi_{ii}}{\partial D} = -2K_{ii}\Lambda^TR^T + 2K_{ii}\mu g^T,$$

so that, for each and every diagonal term

$$\Lambda^TR^T = \mu g^T,$$

so by (3.5),

$$\Lambda^TR^T = 0. \tag{3.12}$$

From (3.11) we get

$$B = \Lambda^TA^TV^{-1}. \tag{3.13}$$

Substituting in the first equation of (3.8), we arrive at

$$\Lambda^TA^TV^{-1}A + DR = L. \tag{3.14}$$

This equation and (3.12), for the special case with $k = 1$, were derived and emphasized by Rao [5], [17].

We next derive an estimate of ϕ in terms of Λ^T and θ^* for general k. We just multiply (3.14) by θ^* and use $R\theta^* = 0$ to get

$$\Lambda^TA^TV^{-1}A\theta^* = \phi^*. \tag{3.15}$$

The corresponding estimate in terms of Λ^T and x is

$$\phi^* = Bx = \Lambda^TA^TV^{-1}x. \tag{3.16}$$

It follows that θ^* satisfies

$$\Lambda^T A^T V^{-1} A \theta^* = \Lambda^T A^T V^{-1} x. \tag{3.17}$$

Equations (3.17) and (3.12) may be considered to be basic relations in θ^* and Λ^T.

4. The general Gaussian theorem. We next demonstrate the general Gaussian theorem that the value of θ^* obtained by least squares is consistent with that of (3.17) and (3.12). We note first that θ^* in the general solution is an $s \times 1$ vector and that the general solution is obtained by premultiplying θ^* by the fixed $k \times s$ matrix L. The general theorem is established by proving the typical case with $k = 1$ so that L, B, D, and Λ are vectors and (3.17) becomes

$$\lambda^T A^T V^{-1} A \theta^* = \lambda^T A^T V^{-1} x, \tag{4.1}$$

where λ^T is $\lambda^T(1 \times s)$. Also (3.12) becomes

$$\lambda^T R^T = 0. \tag{4.2}$$

Now we wish to minimize the scalar $Q = \epsilon^T V^{-1} \epsilon$, subject to the restriction conditions. Then

$$Q' = (A\theta - x)^T V^{-1}(A\theta - x) + 2(l - bA - dR)\lambda + 2\gamma R\theta. \tag{4.3}$$

Differentiation with respect to θ and d leads to the "normal" equations

$$A^T V^{-1} A \theta^* - A^T V^{-1} x + R^T \gamma^T = 0, \tag{4.4}$$

$$\lambda^T R^T = 0. \tag{4.5}$$

Premultiplying (4.4) by λ^T and substituting (4.5), we get

$$\lambda^T A^T V^{-1} A \theta^* = \lambda^T A^T V^{-1} x. \tag{4.6}$$

Since (4.6) and (4.5) are identical with (4.1) and (4.2), the λ's and θ's must be the same, so the general Gaussian theorem is true.

This solution, which is similar to that of Rao, is satisfactory in proving the generalized Gaussian theorem but it is not satisfactory in that it does not provide an explicit value of θ^* (only implicit relations involving the vector parameter λ) nor does it give an explicit expression for the unbiased linear estimator having minimum variance. These are provided in the sections following.

One further remark should be made before leaving these results on least squares. The Eqs. (4.6) and (4.5) may be considered to be the normal equations of a general least squares problem expressed in terms of the vector parameter λ. Comparison of (4.6) with (2.7) shows that these normal equations can be obtained from the normal equations of the problem with no restrictions by premultiplication by λ^T where λ^T is subject to the conditions $\lambda^T R^T = 0$.

5. The explicit form of the estimator. It appears that no one has provided the explicit form for ϕ^* or for θ^*. Post multiplication of (3.14) by $(A^T V^{-1} A)^{-1} R^T$

followed by application of (3.12) eliminates Λ^T with the resulting

(5.1) $$DR(A^TV^{-1}A)^{-1}R^T = L(A^TV^{-1}A)^{-1}R^T.$$

Now since $R(A^TV^{-1}A)^{-1}R^T$ is of order and rank u, we can write

(5.2) $$D = L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1}.$$

The value of Λ^T is then from (3.14)

(5.2) $$\Lambda^T = L(A^TV^{-1}A)^{-1} - L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}$$

and from (3.13),

(5.4) $$B = L(A^TV^{-1}A)^{-1}A^TV^{-1} - L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1}$$

so that

(5.5) $$\phi^* = L(A^TV^{-1}A)^{-1}A^TV^{-1}x - L(A^TV^{-1}A)^{-1}R^T [R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1}x$$

is the linear unbiased estimator having minimum variance, and

(5.6) $$\theta^* = (A^TV^{-1}A)^{-1}A^TV^{-1}x - (A^TV^{-1}A)^{-1}R^T [R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1}x,$$

and θ^* is the explicit solution of the normal equations. Rao did not give an explicit answer even for the case $k = 1$, since he did not derive an explicit formula for λ^T. The argument above covers the Rao case with L and λ vectors. Thus (5.5) and (5.6) hold with L a vector. As is pointed out above, the θ^* which results from least squares and from minimum variance is independent of L.

The results above are also general enough to include the Aitken results. These can be obtained formally from the above results by using the convention that $R^T[R(A^TV^{-1}A)R^T]^{-1}$ is 0 when $R = 0$, the formal equivalent of $u = 0$ side conditions. Thus the last terms drop from (5.5) and (5.6) for the Aitken problem.

The above results also generalize those of David and Neyman [8] who placed specifications on the dispersion matrix V. They defined V to be a diagonal matrix with

(5.7) $$v_{ii} = \frac{\sigma^2}{P_{ii}}, \quad \text{where} \quad P_{ii} = \frac{\sigma^2}{\sigma_i^2}.$$

The formula for ϕ^* then becomes

(5.8) $$\phi^* = L(A^TPA)^{-1}A^TPx - L(A^TPA)^{-1}R^T[R(A^TPA)^{-1}R^T]^{-1}R(A^TPA)^{-1}A^TPx.$$

Now B is ϕ^* with $x = I$, and θ^* is ϕ^* with $L = I$.

If $P_{ii} = \sigma P'_{ii}$ with $P'_{ii} = 1/\sigma_i^2$, we have

$$\phi^* = L(A^TP'A)^{-1}A^TP'x - L(A^TP'A)^{-1}R^T[R(A^TP'A)^{-1}R^T]^{-1}R(A^TP'A)^{-1}A^TP'x. \tag{5.9}$$

Then dropping the side conditions on the parameters we get

$$B = L(A^TPA)^{-1}A^TP = L(A^TP'A)^{-1}A^TP'. \tag{5.10}$$

When L is restricted to a vector, this is the David-Neyman result in matrix form.

When $V = I, L = I$ and $R = 0$ we have the common case of unweighted least squares regression

$$\phi^* = \theta^* = (A^TA)^{-1}A^Tx$$

and

$$B = (A^TA)^{-1}A^T. \tag{5.11}$$

The general results are immediately applicable to a variety of special cases involving specifications on V, specifications on L, and specifications on R. separately or in combinations.

6. The dispersion matrix of solutions. The dispersion matrix of solutions is var $\phi^* = BVB^T$. Using the value of B in (5.4), we get

$$\text{var}(\phi^*) = BVB^T = L(A^TV^{-1}A)^{-1}L^T - L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1}R(A^TV^{-1}A)^{-1}L^T. \tag{6.1}$$

When $k = 1$, this is an explicit result for the Rao problem. When there are no side conditions we have the Aitken result

$$\text{var}(\phi^*) = L(A^TV^{-1}A)^{-1}L^T. \tag{6.2}$$

When the values of x are uncorrelated with $v_{ii} = \sigma^2/P_{ii}$, (6.1) and (6.2) become

$$\text{var}(\phi^*) = L(A^TPA)^{-1}L^T\sigma^2 - L(A^TPA)^{-1}R^T[R(A^TPA)^{-1}R^T]^{-1}R(A^TPA)^{-1}L^T\sigma^2 \tag{6.3}$$

and

$$\text{var}(\phi^*) = L(A^TPA)^{-1}L^T\sigma^2. \tag{6.4}$$

When in addition the variables have a common variance σ^2, $\sigma_i^2 = \sigma^2$ and $P = I$. The Eqs. (6.3) and (6.4) appear with $(A^TA)^{-1}$ replacing $(A^TPA)^{-1}$.

If $\phi = \theta$, the above formulas appear with $L = I$. The simple case in which there are no side conditions, $\phi = \theta$, with variables uncorrelated but with equal variances gives

$$\text{var}(\theta^*) = (A^TA)^{-1}\sigma^2, \tag{6.5}$$

which is the formula for the dispersion matrix of regression coefficients in a common model.

7. Use of a matrix relation. The results (6.1) and (5.4) enable us to write a relation involving the value of B which gives the value of BVB^T having minimum diagnonal terms and the resulting matrix. In order to write this relation in compact form we use

$$C = (A^TV^{-1}A)^{-1} - (A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1}R(A^TV^{-1}A)^{-1}, \tag{7.1}$$

which is Λ with $L = I$ to get

$$BVB^T = LCL^T + (B - LCA^TV^{-1})V(B - LCA^TV^{-1})^T. \tag{7.2}$$

The relation used by Plackett ([1], page 459) is a special case of this relation with the terms involving R deleted. Then $C = (A^TV^{-1}A)^{-1}$. Plackett's relation may be considered to be a generalization of the relation used by Gauss in establishing the theorem. Once the relation is established we see at once that the diagonal terms of BVB^T are minimized for general B when

$$B = LCA^TV^{-1} \tag{7.3}$$

as indicated in (5.4) and that the minimum values of the diagonal terms of the dispersion matrix are the diagonal terms of

$$BVB^T = LCL^T \tag{7.4}$$

as given in (6.1).

Once this general relation (7.2) is proposed, it may be verified by direct expansion. Then the whole solution of the problem of the minimization of the diagonal terms of the dispersion matrix of the estimators is immediately available as indicated by Plackett. If the relation is not known, and it has not been known previously for the general problem, it can be established with the use of matrix derivatives as shown above.

The various special cases of the general matrix relation result from the application of specified conditions to V, L, and R.

8. The variance of the residuals. Returning to the problem of least squares, we call $\mathcal{E}(\epsilon^TV^{-1}\epsilon)$ the variance of the (weighted) residuals. Then ϵ can be written

$$\epsilon = (ACA^TV^{-1} - I)x, \tag{8.1}$$

where C is given by (7.1), and ACA^TV^{-1}, and hence $ACA^TV^{-1} - I$, are idempotent. Hence

$$\begin{aligned} \epsilon^TV^{-1}\epsilon &= x^T(ACA^TV^{-1} - I)^TV^{-1}(ACA^TV^{-1} - I)x \\ &= x^TV^{-1}x - x^TV^{-1}ACA^TV^{-1}x. \end{aligned} \tag{8.2}$$

There is no loss in generality, for purpose of derivation, in assuming that x in (8.1) and (8.2) is a deviate with var $(x) = \mathcal{E}(xx^T) = V$.

For the Aitken problem, $C = (A^TV^{-1}A)^{-1}$ and we have

$$\epsilon^TV^{-1}\epsilon = x^TV^{-1}x - x^TV^{-1}A(A^TV^{-1}A)^{-1}A^TV^{-1}x. \tag{8.3}$$

To this we apply the triangular matrix square root transformation[1]

$$y = Wx \text{ with } W^TW = V^{-1}. \tag{8.4}$$

We then have

$$e^TV^{-1}e = y^Ty - y^TWA(A^TV^{-1}A)^{-1}A^TW^Ty \tag{8.5}$$

with

$$\text{var }(y) = E(Wxx^TW^T) = WVW^T = I_n \tag{8.6}$$

so that, using the trace

$$E(x^TV^{-1}x) = E(y^Ty) = n. \tag{8.7}$$

In order to find the expected value of the second term on the right in (8.5), we use the additional transformation

$$z = Sy \text{ with } S^TS = WA(A^TV^{-1}A)^{-1}A^TW^T, \tag{8.8}$$

where S is a triangular matrix. Since the rank of $WA(A^TV^{-1}A)^{-1}A^TW^T$ is s, S is of rank s, and there are $n - s$ rows identically zero. Then

$$e^TV^{-1}e = y^Ty - z^Tz, \tag{8.9}$$

and since

$$E(zz^T) = \begin{bmatrix} I_s & 0 \\ 0 & 0 \end{bmatrix}, \tag{8.10}$$

then

$$E(z^Tz) = s$$

and

$$E(e^TV^{-1}e) = E(y^Ty) - E(z^Tz) = n - s. \tag{8.11}$$

In the general problem with more complex C we have the additional quadratic form

$$x^TV^{-1}A(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^TR^T]^{-1}R(A^TV^{-1}A)^{-1}\cdot A^TV^{-1}x \tag{8.12}$$

whose matrix is of rank u. Application of (8.4) followed by application of

$$t = U_y\text{, where } U^TU = WA(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1} \cdot R(A^TV^{-1}A)^{-1}A^TW^T \tag{8.13}$$

[1] A triangular matrix square root, as applied to this problem, is a triangular matrix W defined by $W^TW = V^{-1}$. This should not be confused with the (non-triangular) algebraic matrix square root defined by $(V^{-\frac{1}{2}})^2 = V^{-1}$.

reduces this term to

$$t^T t \text{ with } E(t^T t) = u. \tag{8.14}$$

Then

$$E(e^T V^{-1} e) = n - s + u = n - (s - u). \tag{8.15}$$

This result is what one would expect. If the values of x were distributed normally, the positive definite quadratic form $\epsilon^T V^{-1} \epsilon$ would be distributed as χ^2 with $E(X^2) = n - (s - u)$ indicates the number of *independent* parameters.

This result is independent of k. In the Rao problem, $k = 1$, and the value of $E(\epsilon^T V^{-1} \epsilon)$ is $n - s + u$ as above. For the Aitken problem, $u = 0$, and the value is $n - s$. Where $V^{-1} = P/\sigma^2$ we have

$$E(\epsilon^T P \epsilon) = (n - s + u)\sigma^2 \tag{8.16}$$

and when $u = 0$, this is

$$E(\epsilon^T P \epsilon) = (n - s)\sigma^2 \tag{8.17}$$

as shown by David and Neyman for the case of uncorrelated variables ([8], pages 110–112). When $P = I$ this becomes

$$E(\epsilon^T \epsilon) = (n - s)\sigma^2 \tag{8.18}$$

as shown by Aitken using the properties of idempotent matrices ([3], page 139).

9. An estimator of the dispersion matrix of ϕ^*. David and Neyman [8] have provided an unbiased estimate of var ϕ^* for the case in which $V^{-1} = P/\sigma^2$, the x's are uncorrelated and L is a vector. A generalization related to the David-Neyman formula for the general problem is, for known V,

$$E^{-1} \operatorname{var}(\phi^*) = \frac{\epsilon^T V^{-1} \epsilon}{n - s + u} LCL^T, \tag{9.1}$$

since its expected value is the dispersion matrix of ϕ^*.

When V is known this formula is of little value since BVB^T can be computed and no estimation is necessary. However if V is not known, but P is, we have

$$\begin{aligned} E^{-1} \operatorname{var}(\phi^*) = {} & \frac{\epsilon^T P \epsilon}{n - s + u} \\ & L\{(A^T PA)^{-1} - (A^T PA)^{-1} R^T [R(A^T PA)^{-1} R^T]^{-1} R(A^T PA)^{-1}\} L^T. \end{aligned} \tag{9.2}$$

When $P = I$, the case of equal variances, we have the important

$$\begin{aligned} E^{-1} \operatorname{var}(\phi^*) = {} & \frac{\epsilon^T \epsilon}{n - s + u} L\{(A^T A)^{-1} \\ & - (A^T A)^{-1} R^T [R(A^T A)^{-1} R^T]^{-1} R(A^T A)^{-1}\} L^T. \end{aligned} \tag{9.3}$$

In the case of no side conditions we have

$$E^{-1}\operatorname{var}(\phi^*) = \frac{\epsilon^T P \epsilon}{n - s} L(A^T P A)^{-1} L^T. \tag{9.4}$$

Using the value of B in (5.10) we get

$$E^{-1}\operatorname{var}(\phi^*) = \frac{\epsilon^T P \epsilon}{n - s} B P^{-1} B^T. \tag{9.5}$$

If L is a vector, the estimate is a scalar. In the David-Neyman scalar notation, with the x's uncorrelated and B a row vector (λ) we have

$$\mu_1^2 = \frac{S_0}{n - s} \sum_{i=1}^{n} \frac{\lambda_i^2}{Pi}, \tag{9.6}$$

where $[\lambda_i] = \lambda = L(A^T P A)^{-1} A^T P = B$. Hence (9.2) gives the estimator matrix of var (ϕ^*) for a more general problem than does (9.6).

Appendix Showing Orders of Matrices and Conditions

Matrix	Order	Matrix	Order
X	$n \times 1$	ψ	$k \times k$
A	$n \times s$	Λ	$s \times k$
θ and θ^*	$s \times 1$	λ	$s \times 1$
ϵ	$n \times 1$	R	$u \times s$
V	$n \times n$	g	$u \times 1$
Q and Q'	1×1	D	$k \times u$
$A^T V^{-1} A$	$s \times s$	μ	$k \times 1$
L	$k \times s$	γ	$1 \times u$
ϕ and ϕ^*	$k \times 1$	$R(A^T V^{-1} A)^{-1} R^T$	$u \times u$
B	$k \times n$	C	$s \times s$
var ϕ^*	$k \times k$	P	$n \times n$
BVB^T	$k \times k$		

$u < s < n$, $u = 0$ gives Aitken problem, $k = 1$ gives Rao problem, $V^{-1} = P/\sigma^2$ gives David-Neyman condition.

REFERENCES

[1] R. L. Plackett, "A historical note on the method of least squares," *Biometrika*, Vol. 36 (1949), pp. 458–460.

[2] A. C. Aitken, "On least squares and linear combinations of observations," *Proc. Roy. Soc. Edin.*, Vol. 55 (1934), pp. 42–48.

[3] A. C. Aitken, "Studies in mathematics. IV. On linear approximations by least squares," *Proc. Roy. Soc. Edin.*, Vol. 62 (1945), pp. 138–146.

[4] C. R. Rao, "Generalization of Markoff's theorem and tests of linear hypotheses," *Sankhya*, Vol. 7 (1945), pp. 9–16.

[5] C. R. RAO, "Markoff's Theorem with linear restrictions on the parameters," *Sankhya*, Vol. 7 (1945), pp. 16–19.
[6] P. S. DWYER, "A matrix presentation of least squares and correlation theory," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 82–89.
[7] P. S. DWYER AND M. S. MACPHAIL, "Symbolic matrix derivatives," *Ann Math. Stat.*, Vol. 19 (1948), pp. 517–534.
[8] F. N. DAVID AND J. NEYMAN, "Extension of the Markoff Theorem on least squares," *Statistical Research Memoirs*, Vol. 2 (1938), pp. 105–116.