AN INVERSE MATRIX ADJUSTMENT ARISING IN DISCRIMINANT ANALYSIS

BY M. S. BARTLETT

University of Manchester, England

1. Introduction. The adjustment of an inverse matrix arising from the change of a single element, or of elements in a single row or column, in the original matrix has recently been discussed by Sherman and Morrison [1, 2]. In discriminant function analysis the adjustment due to the addition of a degenerate matrix of rank one to the original matrix has sometimes been required, and the method used by the writer is described in this note. It will be noticed that this case includes the cases considered by Sherman and Morrison.

2. General formula. The new square matrix can always be written in the form

$$\mathbf{B} = \mathbf{A} + \mathbf{u}\mathbf{v}', \tag{1}$$

where $\mathbf{u}$ is a column vector (single column matrix), and $\mathbf{v}'$ a row vector (dashes denote matrix transposes). We write formally

$$\begin{aligned} \mathbf{B}^{-1} &= (\mathbf{A} + \mathbf{u}\mathbf{v}')^{-1} = \mathbf{A}^{-1}(1 + \mathbf{u}\mathbf{v}'\mathbf{A}^{-1})^{-1} \\ &= \mathbf{A}^{-1}(1 - \mathbf{u}\mathbf{v}'\mathbf{A}^{-1} + \mathbf{u}\mathbf{v}'\mathbf{A}^{-1}\mathbf{u}\mathbf{v}'\mathbf{A}^{-1} - \cdots) \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{u}\cdot\mathbf{v}'\mathbf{A}^{-1}\{1 - \mathbf{v}'\mathbf{A}^{-1}\mathbf{u} + (\mathbf{v}'\mathbf{A}^{-1}\mathbf{u})^2 - \cdots\} \\ &= \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\cdot\mathbf{v}'\mathbf{A}^{-1}}{1 + \mathbf{v}'\mathbf{A}^{-1}\mathbf{u}}, \end{aligned} \tag{2}$$

which has the same simple structure as (1) and can be determined when $\mathbf{A}^{-1}$ is known. To check this formal result, we may easily verify that pre- or post-multiplication of the expression (2) by $\mathbf{B}$ gives the unit matrix.

3. Numerical example in discriminant analysis. The general regression relation between two sample matrices S_2 and S_1 may be written (Bartlett [3])

$$S_2 = C_{21}C_{11}^{-1}S_1 + S_{2.1}\,. \tag{3}$$

Here the n observations of any variable (measured if necessary from the general mean) comprise one row in the appropriate matrix, S_2 and S_1 representing respectively the dependent and independent variables. S_2S_1' is written $\mathbf{C}_{21}$ for convenience, and similarly for $\mathbf{C}_{11}$, $\mathbf{C}_{22}$; also $\mathbf{C}_{22.1} = S_{2.1}S_{2.1}'$. In discriminant analysis in its strict sense S_1 stands for a single dummy variable serving to isolate a group or other contrast between the proper random variables S_2. In that case the equation

$$\mathbf{C}_{22} = \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12} + \mathbf{C}_{22.1} \tag{4}$$

derived from (3) becomes of the form

$$C_{22} = zz' + C_{22.1}. \tag{5}$$

The discriminant function coefficients in Fisher's original discussion [4] of this type of analysis are proportional to the solution **a** of the equation

$$C_{22.1}a = z \tag{6}$$

(see Bartlett [3], p. 37), and hence are obtained as $C_{22.1}^{-1}z$, where $C_{22.1}$ is the matrix of 'sums of squares and products' within groups. But in tests of significance of **a** it is convenient (see, for example, Bartlett [5], §5) to make use of the 'inverted regression relation' (first noted by Fisher [4], p. 184)

$$S_1 = C_{12}C_{22}^{-1}S_2 + S_{1.2}, \tag{7}$$

giving discriminant function coefficients $b = C_{22}^{-1}C_{21}$.

It is sometimes required to obtain the second (equivalent) form of solution involving C_{22}^{-1} from computations already available based on the first method of analysis involving $C_{22.1}^{-1}$. For example, in Fisher's original comparison of *Iris versicolor* and *Iris setosa* based on 50 observations, on each species, of the variables

$$x_1 = \text{sepal length},$$
$$x_2 = \text{sepal width},$$
$$x_3 = \text{petal length},$$
$$x_4 = \text{petal width},$$

he gives (p. 181) for $C_{22.1}^{-1}$ the (symmetric positive definite) matrix

(8)

	x_1	x_2	x_3	x_4
x_1	0.1187161			
x_2	−0.0668666	0.1452736		
x_3	−0.0816158	+0.0334101	0.2193614	
x_4	+0.0396350	−0.1107529	−0.2720206	0.8945506.

We take S_1 as a pseudo-variate with value $+\frac{1}{2}$ for one species and $-\frac{1}{2}$ for the other, so that $C_{11} = 25$, and C_{21} is the column vector of differences in means multiplied by 25, and $z' = C_{21}/\sqrt{25}$. From (5) and (2) the inverse of C_{22} is

$$C_{22.1}^{-1} - \frac{C_{22.1}^{-1}z \cdot z'C_{22.1}^{-1}}{1 + z'C_{22.1}^{-1}z},$$

or from (6),

$$C_{22.1}^{-1} - \frac{aa'}{1 + a'z}. \tag{9}$$

Fisher actually gives the solution of (6) with **z** replaced by the vector of mean differences, so that, in terms of his solution **c**, where

$$c = a/5 = \begin{pmatrix} -0.0311511 \\ -0.1839075 \\ +0.2221044 \\ +0.3147370 \end{pmatrix}, \tag{10}$$

we find that (9) becomes

$$C_{22.1}^{-1} - 0.9146\ \mathbf{cc}'. \tag{11}$$

Hence we obtain C_{22}^{-1} (without having to re-work it from C_{22}) as

(12)

	x_1	x_2	x_3	x_4
x_1	0.11783			
x_2	−0.07211	0.11434		
x_3	−0.07529	+0.07077	0.17424	
x_4	+0.04860	−0.05781	−0.33595	0.80395.

With this matrix we can complete the formal regression analysis of S_1, giving for **b** and its 'standard errors'

$$\begin{array}{rcl} -0.02847 & \pm & 0.03368 \\ -0.16808 & \pm & 0.03318 \\ +0.20298 & \pm & 0.04095 \\ +0.28764 & \pm & 0.08798. \end{array} \tag{13}$$

The solution **b** we know to be a multiple of the solution **c** (as may be verified to within 2 in the fourth decimal place), but we also see from (12) that the first variable is not contributing to the discrimination and might be omitted. The corresponding analysis of variance of S_1 (c.f. Fisher's Table VII) gives

(14)

	D.F.	S.S.	M.S.
between species x_2, x_3, x_4	3	24.0785	
between species x_1 (partial)	1	0.0069	
within species	95	0.9146	0.011088
Total	99	25.0000	

so that the square of the multiple correlation coefficient is only reduced from 0.96342 to 0.96314 by the omission of x_1. It should be noticed that the multiplier 0.9146 in (11) is the 'within species' entry in (14).

4. Theoretical example in discriminant analysis. The formula (2) is also theoretically useful in deriving the discriminant function by 'size and shape' suggested by Penrose [6]. It is known that for multivariate normal variables **x** with constant variance matrix **V** the ideal discriminant function for contrasting two groups has coefficients $\mathbf{d}'\mathbf{V}^{-1}$, where **d** is the column vector of true differences in means of the two groups. It is now assumed that after standardization of each variable to unit variance we can write

$$\mathbf{V} = \begin{bmatrix} 1 & \rho & \rho & \cdots \\ \rho & 1 & \rho & \cdots \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \end{bmatrix} = (1-\rho)\mathbf{I} + \rho\mathbf{ww}', \tag{15}$$

where **I** is the unit matrix and **w** a column vector with unit components. Applying formula (2), we find the inverse matrix

$$\mathbf{V}^{-1} = \frac{\mathbf{I}}{1-\rho} - \frac{\rho}{1-\rho}\frac{\mathbf{w}\mathbf{w}'}{1+\rho(p-1)}, \tag{16}$$

where p is the number of variables. Hence

$$\begin{aligned} \mathbf{d}'\mathbf{V}^{-1} &= \frac{\mathbf{d}'}{1-\rho} - \frac{\rho}{1-\rho}\frac{(\mathbf{w}'\mathbf{d})\mathbf{w}'}{1+\rho(p-1)} \\ &= \frac{\mathbf{w}'\mathbf{d}}{p(1-\rho)}\left\{\left[\frac{p\mathbf{d}'}{\mathbf{w}'\mathbf{d}} - \mathbf{w}'\right] + \mathbf{w}'\left[1 - \frac{p\rho}{1+\rho(p-1)}\right]\right\} \\ &\propto \left[\frac{p\mathbf{d}'}{\mathbf{w}'\mathbf{d}} - \mathbf{w}'\right] + \mathbf{w}'\left[\frac{1-\rho}{1+\rho(p-1)}\right], \end{aligned} \tag{17}$$

where the two sets of coefficients in (17), $\mathbf{h}'$ and $\mathbf{g}'$ ($\propto \mathbf{w}'$), say (respectively), are arranged to give zero correlation between $\mathbf{g}'\mathbf{x}$ and $\mathbf{h}'\mathbf{x}$. This is checked by evaluating the covariance $E\{\mathbf{w}'\mathbf{y}\cdot\mathbf{h}'\mathbf{y}\}$, where E denotes expectation, and $\mathbf{y}$ the standardized vector deviate with variance matrix $E\{\mathbf{y}\mathbf{y}'\} = \mathbf{V}$. We have

$$\begin{aligned} E\{\mathbf{w}'\mathbf{y}\cdot\mathbf{h}'\mathbf{y}\} &= E\{\mathbf{w}'\mathbf{y}\mathbf{y}'\mathbf{h}\} = \mathbf{w}'\mathbf{V}\mathbf{h} = \mathbf{w}'[(1-\rho) + \rho\mathbf{w}\mathbf{w}']\left[\frac{p\mathbf{d}}{\mathbf{w}'\mathbf{d}} - \mathbf{w}\right] \\ &= p(1-\rho) + \rho p\mathbf{w}'\mathbf{w} - (1-\rho)\mathbf{w}'\mathbf{w} - \rho(\mathbf{w}'\mathbf{w})^2 = 0. \end{aligned}$$

In view of this zero correlation the best discriminant function is of the form

$$\frac{d_1}{v_1}y_1 + \frac{d_2}{v_2}y_2, \tag{18}$$

$$\text{where } y_1 = \mathbf{w}'\mathbf{x} \text{ (the 'size' variable)},$$
$$y_2 = \mathbf{h}'\mathbf{x} \text{ (the 'shape' variable)},$$

d_1 is the difference in means for y_1 and v_1 its variance, and similarly for y_2. Penrose has shown that even if **V** is not exactly of the homogeneous type (15), the above method often gives a very good discriminant function. Applying it to the numerical data referred to in section 3 above, for example, it will be found that we obtain estimates

(19)

	Size weighting ($d_1\mathbf{w}/v_1$)	Shape weighting ($d_2\mathbf{h}/v_2$)	Final weighting
x_1	1.4351	−2.3353	−0.9002
x_2	1.4351	−8.0664	−6.6313
x_3	1.4351	+5.9774	7.4125
x_4	1.4351	+4.4243	5.8594.

It should be noted that the final weightings in (19) correspond with formula (18), and differ slightly from those given by Penrose (Table 5), who makes allowance for the *observed* correlation between y_1 and y_2. This allowance seems somewhat illogical and in any case rather a refinement. Thus Penrose's coefficients give a squared multiple correlation coefficient of 0.96334, whereas those in (19) give 0.96329 (compared with the maximum given in Section 3 of 0.96342).

This method is much quicker than the exact method, but of course the full analysis, as has been indicated in Section 3, enables the most efficient yet economical discriminant function to be found.

REFERENCES

[1] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix", *Annals of Math. Stat.*, Vol. 21 (1950), p. 124.

[2] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix", abstract, *Annals of Math. Stat.*, Vol. 20 (1949), p. 621.

[3] M. S. Bartlett, "Further aspects of the theory of multiple regression", *Proc. Camb. Phil. Soc.*, Vol. 34 (1938), p. 33.

[4] R. A. Fisher, "The use of multiple measurement in taxonomic problems", *Annals of Eugenics*, Vol. 7 (1936), p. 179.

[5] M. S. Bartlett, "Multivariate analysis", *Jour. Roy. Stat. Soc., Suppl.*, Vol. 9 (1947), p. 176.

[6] S. Penrose, "Some notes on discrimination", *Annals of Eugenics*, Vol. 13 (1946–47), p. 228.