UNCORRELATED LINEAR TRANSFORMATIONS OF RESIDUALS IN MULTIPLE REGRESSION

BU-216-M D. S. Robson April, 1966

Abstract

If G is a generalized inverse of $X'X = (X_1'X_1 + X_2'X_2)$ and $\mathcal{G}$ is a generalized inverse of $X_1'X_1$ then

$$Y_1 - X_1\mathcal{G}X_1'Y_1 \equiv Y_1 - X_1GX'Y + X_1GX_2'K(Y_2 - X_2GX'Y)$$

where K is a generalized inverse of $I - X_2GX_2'$. Further, if $Y = X\beta + \epsilon$ and $E(\epsilon\epsilon') = \sigma^2 I$ then $E(Y_1 - X_1\mathcal{G}X_1'Y)(Y_2 - X_2GX'Y) = \phi$.

Biometrics Unit, Plant Breeding Department, Cornell University.

UNCORRELATED LINEAR TRANSFORMATIONS OF RESIDUALS IN MULTIPLE REGRESSION

BU-216-M D. S. Robson April, 1966

In an earlier note (BU-210-M) it was shown that if the regression model

$$\underset{n\times 1}{Y} = \begin{pmatrix} Y_1 \\ \hline Y_2 \end{pmatrix} = \underset{n\times p}{\begin{pmatrix} X_1 \\ \hline X_2 \end{pmatrix}} \underset{p\times 1}{\beta} + \underset{n\times 1}{\begin{pmatrix} \epsilon_1 \\ \hline \epsilon_2 \end{pmatrix}} , \quad E(\epsilon\epsilon') = \underset{n\times n}{I}\sigma^2$$

is fitted separately to the first k observations Y_1 and to all n observations Y then the least squares residuals in the first instance

$$\underset{k\times 1}{f_1} = \underset{k\times 1}{Y_1} - X_1\mathcal{G}X_1'Y_1 \quad , \quad X_1'X_1\mathcal{G}X_1'X_1 = X_1'X_1$$

are uncorrelated with the n-k least squares residuals

$$e_2 = Y_2 - X_2GX'Y \quad , \quad X'XGX'X = X'X$$

obtained after fitting the model to all n observations. The purpose of the present note is to show that

(1) $$f_1 = e_1 + X_1GX_2'Ke_2$$

for all K such that

(2) $$(I - X_2GX_2')K(I - X_2GX_2') = I - X_2GX_2'$$

Rewriting (1) in terms of Y_1 and Y_2 we get

(3) $$e_1 + X_1GX_2'Ke_2 = (I - X_1GX_1')Y_1 - X_1GX_2'Y_2 + X_1GX_2'K(I - X_2GX_2')Y_2 - X_1GX_2'KX_2GX_1'Y_1$$

The coefficient of Y_2 in (3) is null since

(4) $$[X_1GX_2'K(I - X_2GX_2') - X_1GX_2']'[X_1GX_2'K(I - X_2GX_2') - X_1GX_2']$$
$$= (I - X_2GX_2')K'X_2G'X_1'X_1GX_2'K(I - X_2GX_2') - X_2GX_1'X_1GX_2'K(I - X_2GX_2')$$
$$- (I - X_2GX_2')K'X_2G'X_1'X_1GX_2' + X_2G'X_1'X_1GX_2'$$

and from the relation $XGX'X = X$ we have

(5) $$X_2G'X_1'X_1 = (I - X_2G'X_2')X_2$$

and from the invariance of XGX' then

$$X_2G'X_1'X_1 = (I - X_2GX_2')X_2$$

Upon substituting this experssion into (4) and utilizing the relation (2) we see that the right hand side of (4) is null, and hence

$$X_1GX_2'K(I - X_2GX_2') = X_1GX_2'$$

Equation (3) thus reduces to

$$e_1 + X_1GX_2'Ke_2 = [I - X_1(G + GX_2'KX_2G)X_1']Y_1$$

and since

$$f_1 = (I - X_1GX_1')Y_1$$

then to establish (1) we must show that $G + GX_2'KX_2G$ is a generalized inverse of $X_1'X_1$; i.e., that

(6) $$X_1'X_1(G + GX_2'KX_2G)X_1'X_1 = X_1'X_1$$

As in (5) we may write

$$X_1'X_1GX_1'X_1 = X_1'X_1 - X_2'X_2GX_1'X_1$$

so that (6) reduces to

(7) $$X_1'X_1GX_2'KX_2GX_1'X_1 = X_2'X_2GX_1'X_1$$

From (5), the left hand side of (7) may be written as

$$X_2'(I - X_2GX_2')K(I - X_2GX_2')X_2 = X_2'(I - X_2GX_2')X_2$$

and applying (5) to the right hand side of (7) gives the same result, thus confirming (6) and hence also (1).