Bot. Bull. Academia Sinica **14**: 75-82 (1973)

ON CANONICAL REDUCTION OF GENERAL MULTIVARIATE LINEAR HYPOTHESIS

TENG-SHAN WENG

Mead Johnson Research Center
Evansville, Indiana, U. S. A.

(Received for publication February 7, 1973)

Abstract

This article discusses the problem of reducing the general multivariate linear hypothesis to a canonical form and proves the existence of an *ad hoc* orthogonal transformation for the more general case where the design matrix of the underlying linear model need not be of full rank.

Let Y be an $N\times p$ random matrix such that the row vectors thereof are stochastically independent, each following a p-variate normal distribution with common dispersion matrix $\Phi(p\times p)$ which is positive-definite. Suppose further that Y has the expectation

$$\mathcal{E}Y = A\Theta\,,$$

where A is an $N\times k$ known design matrix of rank r, $0<r\leq k<N$, and Θ is a $k\times p$ matrix of unknown parameters. To avoid triviality, we shall impose the condition that $p\geq 2$ and $r+p\leq N$.

The *general multivariate linear hypothesis* is defined in terms of two linear subspaces Ω and ω of an N-dimensional vector space having the respective dimensions r and $r-q$ $(0\leq r-q<r)$ with $\omega\subset\Omega$. In general, the column vectors of $\mathcal{E}Y$ are all assumed to lie in Ω *a priori* and the hypothesis in question specifies that they all lie in ω. In practice, the above problem is customarily represented by the hypothesis

$$\mathcal{H} : B_1\Theta = 0\,,$$

where B_1 is a $q\times k$ known matrix (usually called a *hypothesis matrix*) of rank q such that the row vectors thereof all lie in the row space of the matrix $A'A$, A' being the transpose of A. This latter condition implies that there exists a $q\times k$ matrix Λ_1 such that $B_1=\Lambda_1 A'A$ and is imposed thereupon to guarantee the estimability of the set of linear parametric functions represented by the matrix $B_1\Theta$. Under the general assumption that column vectors of Y all lie in Ω, the normal equation derivable by the method of least squares is $A'A\Theta = A'Y$. Let

$$\hat{\Theta} = (A'A)^{-}A'Y$$

be any solution to the normal equation, where $(A'A)^{-}$ denotes a generalized inverse of $A'A$ (*cf.* Rao, 1966). Then, $B_1\hat{\Theta}$ is invariant to $\hat{\Theta}$ and is the best linear unbiased estimator of $B_1\Theta$, and the usual tests (see the next paragraph) of the linear hypothesis $\mathscr{H}: B_1\Theta = O$ are based on two matrix functions of Y, namely,

$$H = (B_1\hat{\Theta})'\{B_1(A'A)^{-}B_1'\}^{-1}B_1\hat{\Theta}$$

and

$$E = Y'Y - \hat{\Theta}'A'Y,$$

where $\{B_1(A'A)^{-}B_1'\}^{-1}$ denotes the ordinary inverse of $B_1(A'A)^{-}B_1'$. The matrices H and E are called *hypothesis sum of products* (*S.P.*) *matrix* and *error S.P. matrix*, respectively.

About a dozen of test statistics have been proposed for testing the hypothesis $\mathscr{H}: B_1\Theta = O$. Among them, Wilks' U (Wilks, 1932 and Hsu, 1940), Lawley's V (Lawley, 1938) or Hotelling's T_0^2 (Hotelling, 1947 and 1951), and Roy's λ (Roy, 1954) are most often used in practical applications. Based on these test statistics, the respective level α rejection regions ($0<\alpha<1$) for the hypothesis $\mathscr{H}: B_1\Theta = O$ are given by

$$\text{Wilks' } U: U = \frac{|E|}{|H+E|} < u, \qquad (1)$$

$$\text{Lawley's } V \text{ or Hotelling's } T_0^2: V = tr\,(HE^{-1}) = \frac{T_0^2}{N-r} > v, \qquad (2)$$

and

$$\text{Roy's } \lambda: \lambda = Ch_{\max}(HE^{-1}) > \lambda_0, \qquad (3)$$

where "$|\ |$" denotes the determinant, "tr" the trace, and "$Ch_{\max}$" the maximum characteristic root of a matrix. The constants u, v and λ_0 are determined by the equations

$$\mathscr{P}\{U < u \mid \mathscr{H}\} = \mathscr{P}\{V > v \mid \mathscr{H}\} = \mathscr{P}\{\lambda > \lambda_0 \mid \mathscr{H}\} = \alpha.$$

Let T be an orthogonal transformation such that, after having applied T to the matrix Y, the row vectors of the transformed matrix

$$TY = Y^*, \quad \text{say}, \qquad (4)$$

are again stochastically independent and distributed according to a p-variate normal probability density function with common dispersion matrix Φ. If the transformation T can operate on Y in such a way that

$$\mathcal{E}Y_i^* = M_i, \quad i = 1, 2; \quad \mathcal{E}Y_3^* = O,$$

where Y_i^* $(i=1,2,3)$ are submatrices of Y defined by the partition

$$Y^* = \begin{bmatrix} Y_1^* \\ Y_2^* \\ Y_3^* \end{bmatrix} \begin{matrix} q\times p \\ (r-q)\times p \\ (N-r)\times p \end{matrix}$$

and M_i $(i=1,2)$ are linearly dependent on the row vectors of Θ, then the problem of testing the hypothesis $\mathcal{H}: B_1\Theta = O$ is equivalent to testing the hypothesis

$$\mathcal{H}_0 : M_1 = O,$$

and we say that *the original hypothesis is reduced to a canonical form.* In fact, through the execution of such an orthogonal transformation, the original test can be resolved into a test that is concerned only with some mean vector of a multivariate normal distribution, which is obviously much simpler to work with. The existence of an orthogonal transformation as such has been proved for the case where the design matrix A is of full rank, *i.e.*, *rank* $A=k$ (*cf.* Hsu, 1941). The purpose of this article is to supply another proof for the more general case as described in the beginning by constructing an *ad hoc* orthogonal matrix. This will be done with the help of a couple of useful lemmas, as follows:

LEMMA 1. If the matrices $A(N\times k)$ and $B_1(q\times k)$ are, respectively, of rank r and q $(0<q\leq r\leq k<N)$ and $B_1 = A_1A'A$ for some matrix $A_1(q\times k)$, then there exists a matrix $A_2\{(r-q)\times k\}$ such that

$$A_1 A' A A_2' = O \tag{5}$$

and

$$rank \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} A' A = r. \tag{6}$$

PROOF. Let

$$\mathcal{N}(A'A) = \{y : A'Ay = 0\}$$

and

$$\mathcal{N}(A_1A'A) = \{y : A_1A'Ay = 0\}.$$

Here $\mathcal{N}(A'A)$ and $\mathcal{N}(A_1A'A)$ represent the null spaces of $A'A$ and $A_1A'A$, respectively. It is obvious that

$$\mathcal{N}(A'A) \subset \mathcal{N}(A_1A'A). \tag{7}$$

Since, by hypothesis,

$$rank\ A_1A'A = q \tag{8}$$

and

$$rank\ A'A = rank\ A = r,$$

it follows that (*cf.* Birkhoff and MacLane, 1965, p. 213)

$$dim\ \mathscr{N}(A'A) = k - rank\ A'A = k - r = n, \quad \text{say},$$

$$dim\ \mathscr{N}(A_1 A'A) = k - rank\ A_1 A'A = k - q = m, \quad \text{say},$$

and [*cf.* (7)]

$$0 \leq m - n = r - q .$$

Let $\mathscr{B}=\{b_1, \ldots, b_m\}$ be a basis of $\mathscr{N}(A_1A'A)$. Then, without loss of generality, we may let $\mathscr{B}_1=\{b_1, \ldots, b_n\}\subset\mathscr{B}$ be a basis of $\mathscr{N}(A'A)$ and denote by $\mathscr{B}_2=\{b_{n+1}, \ldots, b_m\}$ the remaining basis vectors in $\mathscr{B}$. Since $\mathscr{B}_2\subset\mathscr{N}(A_1A'A)$, we must have

$$A_1 A'A\, b_i = 0, \qquad i = n+1, \ldots, m. \tag{9}$$

Let $A_2'\{k\times(r-q)\}$ be a matrix consisting of the $r-q$ basis vectors in $\mathscr{B}_2$. It then follows from (9) that

$$A_1 A'A A_2' = O ,$$

which proves (5).

To prove (6), let us first observe that all the q row vectors of the matrix $A_1A'A$ are linearly independent [obvious from (8)]. We next show that all the $r-q$ row vectors of the matrix $A_2A'A$ are also linearly independent. So let

$$\beta_{n+1} A'A\, b_{n+1} + \ldots + \beta_m A'A\, b_m = 0 .$$

Then,

$$A'A\, (\beta_{n+1} b_{n+1} + \ldots + \beta_m b_m) = 0$$

and it follows that the vector $x_0=\beta_{n+1}b_{n+1}+\ldots+\beta_m b_m$ is a solution to the linear equation $A'Ax=0$ and hence $x_0 \in \mathscr{N}(A'A)$. But this is impossible unless $x_0=0$. Therefore, we must have

$$\beta_{n+1} = \ldots = \beta_m = 0 .$$

Now it is obvious that $A'Ab_{n+1}, \ldots, A'Ab_m$, the $r-q$ column vectors of the matrix $A'AA_2'$, are linearly independent and so are the $r-q$ row vectors of the matrix $A_2A'A$.

We now show that all the r row vectors of the matrix $\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} A'A$ are linearly independent. Suppose that

$$b_1' A_1 A'A + b_2' A_2 A'A = O', \tag{10}$$

where the vectors b_i $(i=1,2)$ are arbitrary. It follows from (5) and (10) that the sum of squares of components in the row vector $b_2'A_2A'$ is zero. In fact,

$$\begin{aligned} b_2' A_2 A'A A_2' b_2 &= b_1' A_1 A'A A_2' b_2 + b_2' A_2 A'A A_2' b_2 \\ &= (b_1' A_1 A'A + b_2' A_2 A'A) A_2' b_2 \\ &= 0 . \end{aligned}$$

Hence we must have $b_2'A_2A'=0'$ and it follows that

$$b_2' A_2 A'A = 0' . \tag{11}$$

Similarly, we can show that

$$b_1' A_1 A'A = 0' . \tag{12}$$

Since the q row vectors of $A_1A'A$ are linearly independent and so are the $r-q$ ones of $A_2A'A$, the relations (11) and (12) imply that

$$b_1' = 0' \qquad \text{and} \qquad b_2' = 0' .$$

We thus see that the r row vectors of the matrix $\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} A'A$ are all linearly independent. Hence (6) follows. Q. E. D.

Corollary. If we write $B_2=A_2A'A$ in Lemma 1, then the matrices $B_i(A'A)^-B_i'$ $(i=1,2)$ are positive-definite.

Proof. We observe that

$$B_1 (A'A)^- B_1' = A_1 A'A A_1'$$

and

$$rank\ B_1 = q .$$

By the rule for the rank of a product matrix,

$$rank\ B_1 \leq rank\ A_1 A' \leq min\ (q,\ N) = q$$

and

$$rank\ A_1 A' = rank\ A_1 A'A A_1'$$

it follows that

$$rank\ B_1 (A'A)^- B_1' = rank\ A_1 A' = q .$$

We thus see that $B_1(A'A)^-B_1'$ has full row rank and is necessarily positive-definite (*cf.* Searle, 1971, p. 36).

Using Lemma 1 and following the similar procedure, we can also show that $B_2(A'A)^-B_2'$ is positive-definite. Q. E. D.

For the formulation of the next lemma, we need the following definitions:

Definition 1. An $s\times t$ matrix S is said to be *semiorthogonal* if and only if $s<t$ and $SS'=I_s$, where I_s denotes the identity matrix of order s.

Definitjon 2. Let $S(s\times t)$, $s<t$, be a semiorthogonal matrix. A $(t-s)\times t$ semiorthogonal matrix R is called an *orthogonal completion* of S if and only if the $t\times t$ matrix $\begin{bmatrix} S \\ R \end{bmatrix}$ is orthogonal (for a proof of the existence of such an R, see Ito, 1969, p. 104).

LEMMA 2. Given any $N\times k$ matrix A of rank $r\leq k<N$, there exists a semiorthogonal matrix $S_1(r\times N)$ such that

$$S_1'S_1 = A(A'A)^-A'$$

PROOF. Let

$$P = A(A'A)^-A'$$

and

$$Q = (A'A)^-A'A.$$

By the rule for the rank of a product matrix, we have

$$rank\ Q = rank\ A'A = rank\ A = r.$$

Further, we observe that both the matrices P and Q are symmetric and idempotent. Therefore,

$$rank\ P = tr\ P = tr\ Q = rank\ Q = r,$$

which is indicative that exactly r characteristic roots of the matrix P are unity and the rest zero. Without loss of generality, we may assume that the first r characteristic roots of P are unity. It can then be shown (*cf.* Graybill, 1961, p. 13) that there exists an orthogonal matrix $C(N\times N)$ such that

$$P = C'\begin{bmatrix} I_r & O \\ O & O \end{bmatrix} C = C_1'C_1,$$

where C_1 is a submatrix of C defined by the partition

$$C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \begin{matrix} r\times N \\ (N-r)\times N \end{matrix}$$

Since C is orthogonal, the submatrix $C_1(r\times N)$ is necessarily semiorthogonal, *i.e.*, $C_1C_1'=I_r$, and the desired result follows if we put $S_1=C_1$. Q.E.D.

COROLLARY. If $S_1(r\times N)$ is a semiorthogonal matrix as in Lemma 2 and $S_2\{(N-r)\times N\}$ an orthogonal completion of S_1, then

$$S_2'S_2 = I_N - A(A'A^-)A'$$

and

$$S_2A = O.$$

The proof of this corollary is straightforward, thus omitted.

Now we are in a position to prove the existence of the transformation T mentioned in the beginning [see (4)]. This will be done by constructing an orthogonal matrix (for convenience, we shall make no notational distinction between an orthogonal transformation and the corresponding matrix representation) which serves our purpose. So, let us consider the matrix

$$T = \begin{bmatrix} W_1^{-1/2}B_1(A'A)^-A' \\ W_2^{-1/2}B_2(A'A)^-A' \\ S_2 \end{bmatrix} \begin{matrix} q\times N \\ (r-q)\times N \\ (N-r)\times N \end{matrix}, \tag{13}$$

where $W_i = B_i(A'A)B_i'$ $(i=1,2)$, $W_i^{-1/2}$ the ordinary inverse of the positive square root of the matrix W_i (*cf.* Ferguson, 1967, p. 106), and S_2 an orthogonal completion of S_1 as introduced in the corollary of Lemma 2. Here the existence of $W_i^{-1/2}$ $(i=1,2)$ is guaranteed by the corollary of Lemma 1 which says that W_i $(i=1,2)$ are positive-definite. By virtue of Lemma 2 and its corollary, it can be easily checked that the matrix T given by (13) is indeed orthogonal. Applying T to Y, we obtain the transformed matrix $Y^* = TY$ in the partitioned form

$$\begin{bmatrix} Y_1^* \\ Y_2^* \\ Y_3^* \end{bmatrix} = \begin{bmatrix} W_1^{-1/2} B_1 (A'A)^- A' Y \\ W_2^{-1/2} B_2 (A'A)^- A' Y \\ S_2 Y \end{bmatrix}. \tag{14}$$

Taking expectation on both sides of (14) yields

$$\begin{bmatrix} \mathcal{E}Y_1^* \\ \mathcal{E}Y_2^* \\ \mathcal{E}Y_3^* \end{bmatrix} = \begin{bmatrix} \{B_1 (A'A)^- B_1'\}^{-1/2} B_1 \Theta \\ \{B_2 (A'A)^- B_2'\}^{-1/2} B_2 \Theta \\ 0 \end{bmatrix}.$$

Since now $B_1\Theta = 0$ if and only if $\mathcal{E}Y_1^* = 0$, hence the problem of testing the hypothesis $\mathcal{H}: B_1\Theta = 0$ is equivalent to that of testing the hypothesis $\mathcal{H}_0: \mathcal{E}Y_1^* = 0$. The usual test procedures as exemplified by (1), (2) and (3) can then be applied unaltered to test the new hypothesis $\mathcal{H}_0: \mathcal{E}Y_1^* = 0$, using the hypothesis S. P. matrix

$$H = Y_1^{*\prime} Y_1^*$$

and the error S. P. matrix

$$E = Y_3^{*\prime} Y_3^* .$$

To close, let us remark that (i) the matrices H and E are invariant to the orthogonal transformation T given by (13); (ii) after the transformation T has been made on Y, the expectation $\mathcal{E}Y_2^* = \{B_2(A'A)^- B_2'\}^{-1/2} B_2\Theta$ becomes a matrix of nuisance parameters, and so Y_2^* can be ignored since the latter does not provide any useful information.

Literature Cited

BIRKHOFF, G. and S. MACLANE. 1965. A Survey of Modern Algebra, 3rd ed. Macmillan, New York.

FERGUSON, T. S. 1967. Mathematical statistics. academic Press, New York.

GRAYBILL, F. A. 1961. An Introduction to linear statistical models, vol. I, McGraw-Hill, New York.

HOTELLING, H. 1947. Multivariate quality control, illustrated by the air testing of sample bomb sights. Techniques of Statistical Analysis, Chap. II, McGraw-Hill, New York.

HOTELLING, H. 1951. A generalized T test and measure of multivariate dispersion. Proc. Second Berkeley Symp. Math. Statist. Prob., Univ. of California Press, Berkeley, California.

HSU, P. L. 1940. On generalized analysis of variance. Biometrika **31**: 221-37.

HSU, P. L. 1941. Canonical reduction of the general regression problem. Ann. Eugen. (London) **11**: 42-6.

ITO, K. 1969. Theory of multivariate analysis. Baifukan, Tokyo (in Japanese).

LAWLEY, D. N. 1938. A generalization of Fisher's Z test. Biometrika **30**: 180-7.

RAO, C. R. 1966. Generalized inverse for matrices and its applications in mathematical statistics. Research Papers in Statistics, Festschrift for J. Neyman, Ed., F. N. David. John Wiley, New York.

ROY, S. N. 1954. A report on some aspects of multivariate analysis. North Carolina Institute of Statistics Mimeograph Series No. 121.

SEARLES, S. R. 1971. Linear models. John Wiley, New York.

WILKS, S. S. 1932. Certain generalizations in the analysis of variance. Biometrika **24**: 471-94.

多變值線型假說的標準化問題

翁 登 山

美國 Mead Johnson 研究中心

本文論及多變值線型假說的標準化問題，從一般的非全階線型模式出發，試圖尋覓標準化過程中所需要的直交性轉換行列，最後證明此項行列的存在。