

ORTHOGONALITY OF HYPOTHESE AND IDENTIFIABILITY CONSTRAINTS IN THE ANALYSIS OF VARIANCE MODEL

Washio, Yasutoshi
Faculty of Science and Technology, Keio University

<https://doi.org/10.5109/13375>

出版情報 : Bulletin of informatics and cybernetics. 22 (1/2), pp.39-54, 1986-03. Research
Association of Statistical Sciences

バージョン :

権利関係 :



ORTHOGONALITY OF HYPOTHESES AND IDENTIFIABILITY CONSTRAINTS IN THE ANALYSIS OF VARIANCE MODEL

By

Yasutoshi WASHIO*

Abstract

We consider a sequence of hypotheses $H_i: \beta_i = \mathbf{0}, i=1, 2, \dots, k$, in a linear statistical model $\mathbf{y} = X\boldsymbol{\beta} + \mathbf{e} = X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \dots + X_k\boldsymbol{\beta}_k + \mathbf{e}$. We assume that X is not of full rank—in the analysis of variance model—suitable identifiability constraints $B\boldsymbol{\beta} = \mathbf{0}$ are given in the model.

First, we give some results on the orthogonality of hypotheses by introducing the null space representation for the hypotheses. Then, we confine our model to the analysis of variance model and discuss the relationship between orthogonality of hypotheses and identifiability constraints. A method for finding the identifiability constraints which make all the considerable hypotheses orthogonal is proposed.

1. Introduction

Suppose we have a linear statistical model

$$\begin{aligned} \mathbf{y} &= X\boldsymbol{\beta} + \mathbf{e} \\ &= [X_1 \mid X_2 \mid \dots \mid X_k] \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_k \end{bmatrix} + \mathbf{e}, \end{aligned} \quad (1.1)$$

where \mathbf{y} is an $n \times 1$ vector of observation, X is a known matrix of rank r ($r \leq p$) and called a design matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, X and $\boldsymbol{\beta}$ are expressed by partitioned matrices with appropriate sizes in the last formula in (1.1) and \mathbf{e} is an $n \times 1$ vector of errors distributed as $N(\mathbf{0}, \sigma^2 I_n)$, where σ^2 is an unknown positive constant and I_n is the $n \times n$ identity matrix.

The concept of orthogonality of hypotheses is mainly associated with the analysis of variance model. In the analysis of variance model, the design matrix X is not of full rank and a set of linear constraints among parameters is usually introduced as a set of identifiability constraints to identify the parameters. Then, when X is not of full rank, we assume that suitable identifiability constraints

$$B\boldsymbol{\beta} = \mathbf{0} \quad (1.2)$$

* Faculty of Science and Technology, Keio University. 14-1, Hiyoshi, 3 Chome, Kohokuku, Yokohama 223, Japan.

are given in the model (1.1). A set of linear constraints $B\beta=0$ is called a set of identifiability constraints if the matrix B satisfies the following conditions (i) and (ii):

- (i) The composite matrix $\begin{bmatrix} X \\ B \end{bmatrix}$ has rank p .
- (ii) No linear combination of the rows of B is a linear combination of the rows of X except the zero vector.

Let's denote the expectation of y , $E(y)$, by θ . For any matrix A , the range space of A will be denoted by $\Re[A]$, i. e. $\Re[A]=\{v \mid v=Au, \text{ for some } u\}$ and the null space of A will be denoted by $\mathfrak{N}[A]$, i. e. $\mathfrak{N}[A]=\{v \mid Av=0\}$. For a subspace ω the orthogonal complement of ω will be denoted by ω^\perp .

In the linear statistical model (1.1), we are interested in testing a sequence of linear hypotheses

$$H_i: \beta_i=0, \quad i=1, 2, \dots, k. \quad (1.3)$$

The hypotheses H_i , $i=1, 2, \dots, k$, can be expressed as follows: The linear model (1.1) states

$$\theta \in \Omega \quad (1.4)$$

and the hypotheses H_i further states

$$H_i: \theta \in \omega_i \quad (\omega_i \subset \Omega), \quad (1.5)$$

$i=1, 2, \dots, k$, where $\Omega=\Re[X]$, $\omega_i=\Re[X_{(i)}]$ when X is of full rank and $\omega_i=\{v \mid v=X_{(i)}\beta_{(i)}, B_{(i)}\beta_{(i)}=0\}$ when X is not of full rank, where $X_{(i)}$ is the matrix X with the submatrix X_i deleted, $B_{(i)}$ and $\beta_{(i)}$ are similarly defined.

As to the orthogonality of hypotheses, there are two kinds of definitions: One is given by Kempthorne [3] and the other is given by Darroch-Silvey [1]. Although both of the definitions are given in a sense that all the considerable hypotheses H_1, H_2, \dots, H_k are orthogonal, we shall give their definitions in a slight different form here.

DEFINITION 1.1. (Kempthorne [3]) The two hypotheses H_i and H_j are said to be orthogonal ($H_i \perp H_j$) if $\hat{\beta}_i$ and $\hat{\beta}_j$ are uncorrelated, i. e. if $\text{cov}(\hat{\beta}_i, \hat{\beta}_j)=0$, where $\hat{\beta}_i$ and $\hat{\beta}_j$ denote the least squares estimates of β_i and β_j under the model (1.1). For hypotheses of more than two, they are said to be orthogonal if any two hypotheses of them are orthogonal.

DEFINITION 1.2 (Darroch-Silvey [1]) The two hypotheses H_i and H_j are said to be orthogonal ($H_i \perp H_j$) if the subspaces $\Omega, \omega_i, \omega_j$ satisfy the condition $\omega_i^\perp \cap \Omega \perp \omega_j^\perp \cap \Omega$, i. e. if the orthogonal complements of ω_i and ω_j with respect to Ω are perpendicular. The definition of orthogonality of more than two hypotheses is the same as that in Definition 1.1.

The equivalence between definitions 1.1 and 1.2 is proved by Seber [5].

Although the matrix B in identifiability constraints (1.2) must satisfy some conditions, there still exist so many choices for B and we may choose any of them in the theoretical point of view (see Scheffé [4]). In testing hypotheses, however, the choice for B will alter orthogonality property of hypotheses. It is desirable to choose the matrix B so that the number of hypotheses which are orthogonal is as many as possible, because the orthogonality of hypotheses makes the computation of the sums of squares for H_i , $i=1, 2, \dots, k$, easier (See Section 3).

Seber [5] gave a necessary and sufficient condition on B for all considerable hypotheses in two-way layouts with and without replications to be orthogonal by the different methods under some further restrictions on B .

It seems to the author that there still needs a slight general and simple method to check orthogonality of hypotheses under the given identifiability constraints and to select suitable identifiability constraints for orthogonality property. This paper is motivated by this question.

In Section 2, first, the null space representation for the hypotheses are introduced and some properties of matrices being used in the representation are given as lemmas. Then, we shall give some theorems on orthogonality of hypotheses, which will be used in Section 3. Lastly, we shall give simpler proofs of two well known results on orthogonality of hypotheses.

In Section 3, we shall confine our model to the analysis of variance model, and the relationship between identifiability constraints and orthogonality of hypotheses will be considered. A theorem which proposes a method for finding the identifiability constraints which make all the hypotheses orthogonal is given. The method is illustrated by some examples.

2. Some Theorems on Orthogonality of Linear Hypotheses in the Linear Model

First, we shall give the null space representation for the hypotheses H_1, H_2, \dots, H_k . The linear model (1.1) states the expected value of y, θ , belongs to Ω .

When X is of full rank, β can be expressed as

$$\beta = (X'X)^{-1}X'\theta \quad (2.1)$$

by using θ since $\theta = X\beta$, where the prime denotes the transpose of a matrix. If we put $A \equiv (X'X)^{-1}X'$ and if we partition A into k submatrices such as $A = \begin{bmatrix} A_1 \\ \vdots \\ A_k \end{bmatrix}$ corresponding to the partition of β , (2.1) turns out to be

$$\beta = A\theta \quad \text{or} \quad \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} A_1\theta \\ \vdots \\ A_k\theta \end{bmatrix}. \quad (2.2)$$

It is clear from the definition of A, A_1, \dots, A_k that

$$\Re[A'] \subset \Omega, \Re[A_1'] \subset \Omega, \dots, \Re[A_k'] \subset \Omega. \quad (2.3)$$

By making use of matrices A_1, \dots, A_k , the hypotheses H_1, \dots, H_k can be written as

$$H_i: \theta \in \Omega \quad \text{and} \quad A_i\theta = 0, \quad (2.4)$$

$i=1, 2, \dots, k$.

When X is not of full rank, as mentioned in Section 1, suitable identifiability constraints $B\beta=0$ are given to identify the parameters, where B is a $(p-r) \times p$ matrix of rank $p-r$. From the property of the matrix B , the equation

$$\begin{bmatrix} X \\ B \end{bmatrix} \beta = \begin{bmatrix} \theta \\ 0 \end{bmatrix} \quad (2.5)$$

has a unique solution in β for every $\theta \in \Omega$. Since $\begin{bmatrix} X \\ B \end{bmatrix} \equiv G$ has rank p , from (2.5), we have

$$\beta = (G'G)^{-1}G' \begin{bmatrix} \theta \\ 0 \end{bmatrix} = (G'G)^{-1}X'\theta. \quad (2.6)$$

For this case, if we put $A \equiv (G'G)^{-1}X'$ and if A is partitioned into k submatrices as before, then β can again be expressed as (2.2). Furthermore, the property (2.3) also holds true, and the hypotheses H_1, H_2, \dots, H_k can be expressed as (2.4).

Therefore, in general, in the linear model (1.1) we can express β as

$$\beta = A\theta \quad \text{or} \quad \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} A_1\theta \\ \vdots \\ A_k\theta \end{bmatrix} \quad (2.7)$$

and the hypotheses $H_i, i=1, 2, \dots, k$, can be expressed as

$$H_i: \theta \in \Omega \quad \text{and} \quad A_i\theta = 0, \quad (2.8)$$

where A, A_1, \dots, A_k have the property such that

$$\Re[A'] \subset \Omega, \quad \Re[A'_i] \subset \Omega, \quad i=1, 2, \dots, k. \quad (2.9)$$

Let C be a matrix such that $\Omega = \Re[C]$ and the rows of C are linearly independent. The linear model (1.1) says

$$\theta \in \Omega = \Re[C] \quad (2.10)$$

and the hypothesis H_i states

$$H_i: \theta \in \omega_i = \Re \left[\begin{bmatrix} C \\ A_i \end{bmatrix} \right]. \quad (2.11)$$

Here we note that

$$\Re[A'_i] \perp \Re[C'], \quad i=1, 2, \dots, k \quad (2.12)$$

from (2.9) and the definition of C . (2.10) and (2.11) are the null space representations for the model (1.1) and the hypothesis H_i . The properties of (2.9) and (2.12) for the matrices A, A_i, C in the null space representation are important and will often be used later on.

LEMMA 2.1. *In the null space representation for the linear model and the hypothesis H_i , we have*

$$\omega_i^\perp \cap \Omega = \Re[A'_i].$$

PROOF.

$$\begin{aligned} \Omega \cap \omega_i^\perp &= \Re[C] \cap \left(\Re \left[\begin{bmatrix} C \\ A_i \end{bmatrix} \right] \right)^\perp \\ &= \Re[C] \cap \Re[C' \mid A'_i]. \end{aligned} \quad (2.13)$$

We show (2.13) is equal to $\Re[A'_i]$. It is clear from (2.9) that

$$\Re[C] \cap \Re[C' \mid A'_i] \supset \Re[A'_i].$$

If x belongs to (2.13), then $Cx=0$ and $x=C'\alpha + A'_i\gamma$ for some α and γ . Thus

$$\begin{aligned} C(C'\alpha + A'_i\gamma) &= CC'\alpha + CA'_i\gamma \\ &= CC'\alpha + \mathbf{0} = \mathbf{0}. \end{aligned}$$

Since CC' has the inverse matrix, we have $\alpha = \mathbf{0}$ from the above equation. Therefore, $\mathbf{x} = A'_i\gamma \in \mathfrak{N}[A'_i]$ and the lemma is proved. q. e. d.

LEMMA 2.2. *In the null space representation for the linear model and the hypotheses H_i , $i=1, 2, \dots, k$,*

$$H_i \perp H_j \Leftrightarrow \mathfrak{N}[A'_i] \perp \mathfrak{N}[A'_j].$$

PROOF. The lemma follows from Definition 1.2 and Lemma 2.1. q. e. d.

LEMMA 2.3. *In the null space representation for the linear model and the hypotheses H_i , $i=1, 2, \dots, k$, the equation*

$$\mathfrak{N}\left[\begin{array}{c} C \\ A_{j_1} \\ \vdots \\ A_{j_q} \end{array}\right] \cap \left(\mathfrak{N}\left[\begin{array}{c} C \\ A_{j_1} \\ \vdots \\ A_{j_q} \\ A_l \end{array}\right] \right)^\perp = \mathfrak{N}[A'_l] \quad (2.14)$$

holds true if and only if

$$H_l \perp H_{j_1}, H_l \perp H_{j_2}, \dots, H_l \perp H_{j_q}. \quad (2.15)$$

PROOF. Since the left hand side of the equation (2.14) is

$$\begin{aligned} \mathfrak{N}\left[\begin{array}{c} C \\ A_{j_1} \\ \vdots \\ A_{j_q} \end{array}\right] \cap \left(\mathfrak{N}\left[\begin{array}{c} C \\ A_{j_1} \\ \vdots \\ A_{j_q} \\ A_l \end{array}\right] \right)^\perp &= \mathfrak{N}\left[\begin{array}{c} C \\ A_{j_1} \\ \vdots \\ A_{j_q} \end{array}\right] \cap \mathfrak{N}[C' \ A'_{j_1} \dots A'_{j_q} \ A'_l] \\ &= \mathfrak{N}\left[\begin{array}{c} C \\ \bar{A}_j \end{array}\right] \cap \mathfrak{N}[C' \ \bar{A}_j' \ A'_l], \end{aligned} \quad (2.16)$$

where \bar{A}_j is the matrix consisting of linearly independent rows of $\left[\begin{array}{c} A_{j_1} \\ \vdots \\ A_{j_q} \end{array}\right]$, it is clear by Lemma 2.2 that (2.14) implies (2.15).

Conversely suppose that (2.15) is true. It is clear from (2.16) and Lemma 2.2 that

$$\mathfrak{N}\left[\begin{array}{c} C \\ A_{j_1} \\ \vdots \\ A_{j_q} \end{array}\right] \cap \left(\mathfrak{N}\left[\begin{array}{c} C \\ A_{j_1} \\ \vdots \\ A_{j_q} \\ A_l \end{array}\right] \right)^\perp \supset \mathfrak{N}[A'_l].$$

If \mathbf{x} belongs to the left-hand side of (2.14), then

$$\mathbf{x} \in \mathfrak{N}\left[\begin{array}{c} C \\ \bar{A}_j \end{array}\right] \quad \text{and} \quad \mathbf{x} = C'\alpha + \bar{A}_j'\gamma + A'_l\delta$$

for some α , γ and δ . From $C\mathbf{x} = \mathbf{0}$, we have $\alpha = \mathbf{0}$ by using the property of matrices C , \bar{A}_j' , A'_l . Similarly, from $\bar{A}_j\mathbf{x} = \mathbf{0}$, we have $\gamma = \mathbf{0}$ by using the assumption. Thus $\mathbf{x} = A'_l\delta$, which means $\mathbf{x} \in \mathfrak{N}[A'_l]$. Then (2.14) holds true. q. e. d.

In this paper, the least squares estimates of β_i under various situations will be

considered. The least squares estimate of β_i under the general model (1.1) will be denoted by $\hat{\beta}_i$. The least squares estimate of β_i under $\beta_j=0$, that is, under H_j , will be denoted by $\hat{\beta}_{i, H_j}$. Similarly, we shall use the notations $\hat{\beta}_{i, H_j \cap H_l}$, $\hat{\beta}_{i, H_j \cap H_l \cap H_m}$, etc. For example, $\hat{\beta}_{i, H_j \cap H_l}$ denotes the least squares estimate of β_i under $\beta_j=0$ and $\beta_l=0$.

THEOREM 2.1.

$$H_i \perp H_j \Leftrightarrow \hat{\beta}_i = \hat{\beta}_{i, H_j} \quad (\hat{\beta}_j = \hat{\beta}_{j, H_i}).$$

In words, H_i and H_j are orthogonal if and only if the least squares estimate of β_i (β_j) is the same whether or not we put $\beta_j=0$ ($\beta_i=0$).

PROOF. Since $\beta_i = A_i \theta$ from (2.7), we have

$$\hat{\beta}_i = A_i \hat{\theta} \quad (2.17)$$

and

$$\hat{\beta}_{i, H_j} = A_i \hat{\theta}_j, \quad (2.18)$$

where $\hat{\theta}$ and $\hat{\theta}_j$ are the least squares estimates of θ under $\theta \in \Omega$ and under $\theta \in \omega_j$ respectively. $\hat{\theta}$ and $\hat{\theta}_j$ can be expressed as

$$\hat{\theta} = P y \quad (2.19)$$

and

$$\hat{\theta}_j = P_j y, \quad (2.20)$$

where P and P_j are the orthogonal projections of R^n — n -dimensional Euclidean space—on Ω and on ω_j respectively. Then

$$\hat{\beta}_i = \hat{\beta}_{i, H_j} \quad (2.21)$$

$$\Downarrow$$

$$A_i P = A_i P_j \quad (2.22)$$

$$\Downarrow$$

$$A'_i = P_j A'_i. \quad (2.23)$$

(2.23) follows from $\Re[A'_i] \subset \Omega$. (2.23) is equivalent to

$$\Re[A'_i] \subset \omega_j. \quad (2.24)$$

Since $\omega_j = \Re\left[\begin{smallmatrix} C \\ A_j \end{smallmatrix}\right]$ and $\Re[A'_i] \perp \Re[C]$, (2.24) is equivalent to

$$\Re[A'_i] \perp \Re[A'_j]. \quad (2.25)$$

Then the theorem follows from Lemma 2.2. It is clear that the subscripts i and j are interchangeable in the proof. q. e. d.

We can see from the proof of Theorem 2.1 that $\hat{\beta}_i = \hat{\beta}_{i, H_j}$ is equivalent to $\hat{\beta}_j = \hat{\beta}_{j, H_i}$. The same property as Theorem 2.1 has been proved by John [2] for a special situation where there are only two factors (hypotheses) in connection with the orthogonality of two factors in the design of experiments.

Theorem 2.1 gives a general result.

THEOREM 2.2. *The equation*

$$\hat{\beta}_i = \hat{\beta}_{i, H_{j_1 \cap H_{j_2} \cap \dots \cap H_{j_q}}}$$

holds true if and only if

$$H_i \perp H_{j_1}, H_i \perp H_{j_2}, \dots, H_i \perp H_{j_q}.$$

PROOF. If we put

$$\begin{aligned} \omega_{j_1 \dots j_q} &\equiv \{\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \Omega, A_{j_1} \boldsymbol{\theta} = 0, \dots, A_{j_q} \boldsymbol{\theta} = 0\} \\ &= \mathfrak{N} \begin{bmatrix} C \\ A_{j_1} \\ \vdots \\ A_{j_q} \end{bmatrix}, \end{aligned} \quad (2.26)$$

then $\hat{\boldsymbol{\beta}}_{i, H_{j_1} \cap \dots \cap H_{j_q}}$ is the least squares estimate of $\boldsymbol{\beta}_i$ under $\boldsymbol{\theta} \in \omega_{j_1 \dots j_q}$. Since $\boldsymbol{\beta}_i = A_i \boldsymbol{\theta}$, we have

$$\hat{\boldsymbol{\beta}}_{i, H_{j_1} \cap \dots \cap H_{j_q}} = A_i \hat{\boldsymbol{\theta}}_{j_1 \dots j_q}, \quad (2.27)$$

where $\hat{\boldsymbol{\theta}}_{j_1 \dots j_q}$ is the least squares estimate of $\boldsymbol{\theta}$ under $\boldsymbol{\theta} \in \omega_{j_1 \dots j_q}$. Since $\hat{\boldsymbol{\theta}}_{j_1 \dots j_q}$ can be expressed by $\hat{\boldsymbol{\theta}}_{j_1 \dots j_q} = P_{j_1 \dots j_q} \mathbf{y}$, (2.27) turns out to be

$$\hat{\boldsymbol{\beta}}_{i, H_{j_1} \cap \dots \cap H_{j_q}} = A_i P_{j_1 \dots j_q} \mathbf{y}, \quad (2.28)$$

where $P_{j_1 \dots j_q}$ is the orthogonal projection of R^n on $\omega_{j_1 \dots j_q}$.

Since $\hat{\boldsymbol{\beta}}_i = A_i P \mathbf{y}$, similarly to the proof of Theorem 2.1, we have the following equivalence relations:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_i &= \hat{\boldsymbol{\beta}}_{i, H_{j_1} \cap \dots \cap H_{j_q}} \\ &\Downarrow \\ A_i P &= A_i P_{j_1 \dots j_q} \\ &\Downarrow \\ A'_i &= P_{j_1 \dots j_q} A'_i \end{aligned} \quad (2.29)$$

$$\begin{aligned} &\Downarrow \\ \mathfrak{N}[A'_i] &\subset \omega_{j_1 \dots j_q}. \end{aligned} \quad (2.30)$$

(2.30) is equivalent to

$$\mathfrak{N}[A'_i] \perp \mathfrak{N}[A'_{j_1}], \dots, \mathfrak{N}[A'_i] \perp \mathfrak{N}[A'_{j_q}], \quad (2.31)$$

because $\mathfrak{N}[A'_i] \perp \mathfrak{N}[C']$. Then, the theorem follows from Lemma 2.2.

q. e. d.

From Theorem 2.1 and Theorem 2.2, we have the following corollary.

COROLLARY. $H_{j_1}, H_{j_2}, \dots, H_{j_q}$ are orthogonal if and only if the least squares estimates of $\boldsymbol{\beta}_{j_1}$ under the model (1.1), under H_{j_2}, \dots , under H_{j_q} , under $H_{j_2} \cap H_{j_3}, \dots$, and under $H_{j_2} \cap \dots \cap H_{j_q}$ are the same and this property also holds true for other parameter vectors $\boldsymbol{\beta}_{j_2}, \dots, \boldsymbol{\beta}_{j_q}$.

THEOREM 2.3. Suppose that $H_i \perp H_{j_1}, \dots, H_i \perp H_{j_q}$. Then

$$H_i \perp H_l \Leftrightarrow \hat{\boldsymbol{\beta}}_{i, H_{j_1} \cap \dots \cap H_{j_q}} = \hat{\boldsymbol{\beta}}_{i, H_{j_1} \cap \dots \cap H_{j_q} \cap H_l}.$$

PROOF. Let $\omega_{j_1 \dots j_q}$ and $\omega_{j_1 \dots j_q l}$ be two subspaces such that

$$\omega_{j_1 \dots j_q} = \{\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \Omega, A_{j_1} \boldsymbol{\theta} = 0, \dots, A_{j_q} \boldsymbol{\theta} = 0\}$$

$$= \mathfrak{N} \begin{bmatrix} C \\ A_{j_1} \\ \vdots \\ A_{j_q} \end{bmatrix},$$

$$\begin{aligned}\omega_{j_1 \dots j_q l} &= \{\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \Omega, A_{j_1} \boldsymbol{\theta} = \mathbf{0}, \dots, A_{j_q} \boldsymbol{\theta} = \mathbf{0} \text{ and } A_l \boldsymbol{\theta} = \mathbf{0}\} \\ &= \Re \begin{bmatrix} C \\ A_{j_1} \\ \vdots \\ A_{j_q} \\ A_l \end{bmatrix}\end{aligned}$$

and let $P_{j_1 \dots j_q}$ and $P_{j_1 \dots j_q l}$ be the orthogonal projections of R^n on $\omega_{j_1 \dots j_q}$ and on $\omega_{j_1 \dots j_q l}$ respectively.

As in the proof of Theorem 2.2, we have

$$\hat{\boldsymbol{\beta}}_{i, H_{j_1} \cap \dots \cap H_{j_q}} = A_i P_{j_1 \dots j_q} \mathbf{y}$$

and

$$\hat{\boldsymbol{\beta}}_{i, H_{j_1} \cap \dots \cap H_{j_q} \cap H_l} = A_i P_{j_1 \dots j_q l} \mathbf{y}.$$

Then

$$\hat{\boldsymbol{\beta}}_{i, H_{j_1} \cap \dots \cap H_{j_q}} = \hat{\boldsymbol{\beta}}_{i, H_{j_1} \cap \dots \cap H_{j_q} \cap H_l} \quad (2.32)$$

$$\Updownarrow$$

$$(P_{j_1 \dots j_q} - P_{j_1 \dots j_q l}) A_i' = \mathbf{0}. \quad (2.33)$$

It is easily shown that $P_{j_1 \dots j_q} - P_{j_1 \dots j_q l}$ is the orthogonal projection of R^n on $\omega_{j_1 \dots j_q} \cap (\omega_{j_1 \dots j_q l})^\perp$. Therefore, (2.33) means that all the column vectors of the matrix A_i' belong to the orthogonal complement of $\omega_{j_1 \dots j_q} \cap (\omega_{j_1 \dots j_q l})^\perp$. Thus (2.33) is equivalent to

$$\Re[A_i'] \subset (\omega_{j_1 \dots j_q} \cap (\omega_{j_1 \dots j_q l})^\perp)^\perp. \quad (2.34)$$

By the assumption of the theorem and Lemma 2.3, (2.34) turns out to be

$$\Re[A_i'] \subset (\Re[A_l'])^\perp, \quad (2.35)$$

which is equivalent to

$$\Re[A_i'] \perp \Re[A_l']. \quad (2.36)$$

Thus the theorem follows by Lemma 2.2.

q. e. d.

Before concluding this section, we shall give simple proofs of two well-known and important results about the orthogonality property by using the null space representation for the hypotheses.

(1) Definitions 1.1 and 1.2 on the orthogonality of hypotheses are equivalent.

PROOF.

$$\begin{aligned}\text{cov}(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_j) &= \text{cov}(A_i \hat{\boldsymbol{\theta}}, A_j \hat{\boldsymbol{\theta}}) \\ &= \text{cov}(A_i P \mathbf{y}, A_j P \mathbf{y}) = A_i A_j' \sigma^2.\end{aligned}$$

Hence

$$\text{cov}(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_j) = 0 \Leftrightarrow \Re[A_i'] \perp \Re[A_j'].$$

Then the result follows from Lemma 2.1.

q. e. d.

(2) The sum of squares for the hypothesis H_i , Q_i , is the same as the one, $Q_{i; 1, 2, \dots, i-1}$, calculated under $H_1 \cap H_2 \cap \dots \cap H_{i-1}$ if and only if

$$H_i \perp H_1, H_i \perp H_2, \dots, H_i \perp H_{i-1}.$$

PROOF.

$$Q_i = \mathbf{y}'(P - P_i)\mathbf{y}$$

and

$$Q_{i; 1, 2, \dots, i-1} = \mathbf{y}'(P_{12\dots i-1} - P_{12\dots i})\mathbf{y},$$

where P , P_i , $P_{12\dots i-1}$, $P_{12\dots i}$ are the orthogonal projections of R^n on Ω , on ω_i , on $\omega_{12\dots i-1}$ and on $\omega_{12\dots i}$ respectively, $\omega_{12\dots i-1}$ and $\omega_{12\dots i}$ being defined as in the proof of Theorem 2.2.

Since $P - P_i$ and $P_{12\dots i-1} - P_{12\dots i}$ are the orthogonal projections of R^n on $\Omega \cap \omega_i^\perp$ and on $\omega_{12\dots i-1} \cap (\omega_{12\dots i})^\perp$, $Q_i = Q_{i; 12\dots i-1}$ is equivalent to

$$\omega_{12\dots i-1} \cap (\omega_{12\dots i})^\perp = \Omega \cap \omega_i^\perp. \quad (2.37)$$

From Lemma 2.1 and Lemma 2.3, (2.37) is equivalent to

$$H_i \perp H_1, H_i \perp H_2, \dots, H_i \perp H_{i-1},$$

and the result is proved.

q. e. d.

3. Identifiability Constraints and the Orthogonality of Hypotheses in the Analysis of Variance Model

The sum of squares for the hypothesis H_i , Q_i , is given by

$$Q_i = \mathbf{y}'(P - P_i)\mathbf{y} = (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i)'(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i). \quad (3.1)$$

Since $\hat{\boldsymbol{\theta}} = X\hat{\boldsymbol{\beta}} = \sum_j X_j\hat{\boldsymbol{\beta}}_j$, we can see from Theorem 2.1 that if H_i is orthogonal to every other hypotheses $H_1, \dots, H_{i-1}, H_{i+1}, \dots, H_k$ then Q_i can be calculated by

$$Q_i = (X_i\hat{\boldsymbol{\beta}}_i)'(X_i\hat{\boldsymbol{\beta}}_i) \quad (3.2)$$

or

$$Q_i = \hat{\boldsymbol{\beta}}_i'(X_i'\mathbf{y}), \quad (3.3)$$

using the least squares estimate $\hat{\boldsymbol{\beta}}_i$ of $\boldsymbol{\beta}_i$ under $\boldsymbol{\theta} \in \Omega$ without calculating the least squares estimates of $\boldsymbol{\beta}_j$, $j=1, 2, \dots, i-1, i+1, \dots, k$, under $\boldsymbol{\theta} \in \omega_i$.

Therefore, in order to make the calculation of the sums of squares for hypotheses simpler, it is needed that each of the hypotheses of interest is orthogonal to every other hypotheses.

Unfortunately, since orthogonality property of hypotheses depends on identifiability constraints chosen in the analysis of variance model, it is desirable to choose suitable ones for orthogonality property.

In discussing the relationship between identifiability constraints and orthogonality of hypotheses, we put the following assumptions on the design matrix and identifiability constraints according to the analysis of variance model.

ASSUMPTIONS (i) X_1 is an $n \times 1$ vector consisting solely of 1. Then β_1 is a scalar called general mean.

(ii) Identifiability constraints have the form such that

$$B_2\boldsymbol{\beta}_2 = \mathbf{0}, \dots, B_k\boldsymbol{\beta}_k = \mathbf{0}, \quad (3.4)$$

where B_i , $i=2, 3, \dots, k$, are $t_i \times p_i$ matrices, i.e. identifiability constraints are given

within each of the parameter groups β_i , $i=2, 3, \dots, k$ except β_1 .

The normal equations for obtaining the least squares estimates $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ of $\beta_1, \beta_2, \dots, \beta_k$ are given by

$$\begin{bmatrix} \mathbf{x}'_1\mathbf{x}_1 & \mathbf{x}'_1X_2 & \cdots & \mathbf{x}'_1X_k \\ X'_2\mathbf{x}_1 & X'_2X_2 & \cdots & X'_2X_k \\ \vdots & \vdots & \ddots & \vdots \\ X'_k\mathbf{x}_1 & X'_kX_2 & \cdots & X'_kX_k \\ \hline & B_2 & \cdots & B_k \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_1\mathbf{y} \\ X'_2\mathbf{y} \\ \vdots \\ X'_k\mathbf{y} \\ \hline \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \quad (3.5)$$

which are arranged in a table as (3.6).

$\hat{\beta}'_1$	$\hat{\beta}'_2$	\cdots	$\hat{\beta}'_k$	$=1$
$\mathbf{x}'_1\mathbf{x}_1$	\mathbf{x}'_1X_2	\cdots	\mathbf{x}'_1X_k	$\mathbf{x}'_1\mathbf{y}$
$X'_2\mathbf{x}_1$	X'_2X_2	\cdots	X'_2X_k	$X'_2\mathbf{y}$
\vdots	\vdots	\ddots	\vdots	\vdots
$X'_k\mathbf{x}_1$	X'_kX_2	\cdots	X'_kX_k	$X'_k\mathbf{y}$
$B_2 \cdots B_k$				$\mathbf{0}$
				\vdots
				$\mathbf{0}$

(3.6)

It is known that all the Lagrange multipliers corresponding to the identifiability constraints can be ignored in the normal equations (See, for example, Washio [6]). It is also known that the normal equations (3.6) have a unique solution for every \mathbf{y} .

THEOREM 3.1. *Suppose that the design matrix and identifiability constraints satisfy the assumptions given just before. Suppose further that the identifiability constraints satisfy the following additional assumption: In a model*

$$\mathbf{y} = [\mathbf{x}_1 \mid X_2 \mid \cdots \mid X_{i-1}] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{i-1} \end{bmatrix} + \mathbf{e} \quad (3.7)$$

which is the model that H_k, H_{k-1}, \dots, H_i are assumed to be true, the constraints

$$B_2\beta_2 = \mathbf{0}, \dots, B_{i-1}\beta_{i-1} = \mathbf{0} \quad (3.8)$$

are still identifiability constraints for the new model (3.7) and this property holds true for $i=k, k-1, \dots, 3$.

Then, all the hypotheses H_1, H_2, \dots, H_k are orthogonal if and only if all the elements above the diagonal blocks in the left-hand side in the normal equations (3.6) are crossed

off by using identifiability constraints, i.e. the normal equations (3.6) turn out to be of the form:

$\hat{\beta}'_1 \quad \hat{\beta}'_2 \quad \cdots \quad \hat{\beta}'_k$	$=1$	
<div style="display: flex; justify-content: space-between; align-items: center;"> <div style="text-align: center;"> $\boxed{x'_1 x_1}$ $X'_2 x_1$ \vdots $X'_k x_1$ </div> <div style="text-align: center;"> $\boxed{X'_2 X_2}$ \vdots $X'_k X_2$ </div> <div style="text-align: center;"> 0 \vdots $\boxed{X'_k X_k}$ </div> </div>	$x'_1 y$ $X'_2 y$ \vdots $X'_k y$	(3.9)
B_2 \vdots B_k	0 \vdots 0	

PROOF. Suppose the normal equations are of the form in (3.9). The normal equations for obtaining the least squares estimates of $\beta_1, \beta_2, \dots, \beta_{k-1}$ when H_k is true (under H_k) are given by

$\hat{\beta}'_{1, H_k} \quad \hat{\beta}'_{2, H_k} \quad \cdots \quad \hat{\beta}'_{k-1, H_k}$	$=1$	
<div style="display: flex; justify-content: space-between; align-items: center;"> <div style="text-align: center;"> $\boxed{x'_1 x_1}$ $X'_2 x_1$ \vdots $X'_{k-1} x_1$ </div> <div style="text-align: center;"> $\boxed{X'_2 X_2}$ \vdots $X'_{k-1} X_2$ </div> <div style="text-align: center;"> 0 \vdots $\boxed{X'_{k-1} X_k}$ </div> </div>	$x'_1 y$ $X'_2 y$ \vdots $X'_{k-1} y$	(3.10)
B_2 \vdots B_{k-1}	0 \vdots 0	

since the Lagrange multipliers corresponding to $B_2 \beta_{2, H_k} = 0, \dots, B_{k-1} \beta_{k-1, H_k} = 0$ can be ignored by the assumption of the theorem. By comparing (3.9) with (3.10), we get

$$\hat{\beta}_1 = \hat{\beta}_{1, H_k}, \dots, \hat{\beta}_{k-1} = \hat{\beta}_{k-1, H_k},$$

which show

$$H_k \perp H_1, H_k \perp H_2, \dots, H_k \perp H_{k-1}$$

by Theorem 2.1.

Similarly, for the least squares estimates of $\beta_1, \beta_2, \dots, \beta_{k-2}$ under H_k and under $H_k \cap H_{k-1}$, we can see

$$\hat{\beta}_{1, H_k} = \hat{\beta}_{1, H_k \cap H_{k-1}}, \dots, \hat{\beta}_{k-2, H_k} = \hat{\beta}_{k-2, H_k \cap H_{k-1}},$$

which show

$$H_{k-1} \perp H_1, H_{k-1} \perp H_2, \dots, H_{k-1} \perp H_{k-2}$$

by Theorem 2.3. Similarly, for the least squares estimates of $\beta_1, \beta_2, \dots, \beta_{k-3}$ under $H_k \cap H_{k-1}$ and under $H_k \cap H_{k-1} \cap H_{k-2}$, we can see

$$\begin{aligned}\hat{\beta}_{1, H_k \cap H_{k-1}} &= \hat{\beta}_{1, H_k \cap H_{k-1} \cap H_{k-2}} \\ &\vdots \\ \hat{\beta}_{k-3, H_k \cap H_{k-1}} &= \hat{\beta}_{k-3, H_k \cap H_{k-1} \cap H_{k-2}},\end{aligned}$$

which show

$$H_{k-2} \perp H_1, H_{k-2} \perp H_2, \dots, H_{k-2} \perp H_{k-3}$$

by Theorem 2.3.

By repeating the same procedure, we can show all the hypotheses are orthogonal.

Conversely, suppose H_1, H_2, \dots, H_k are orthogonal. From the corollary in Section 2, the least squares estimates of β_i 's under the model (1.1), under H_k , under $H_k \cap H_{k-1}, \dots$, under $H_k \cap \dots \cap H_2$ are all the same. Thus by considering normal equations for obtaining the least squares estimates of β_i 's under the various models just mentioned above, we have

$$\begin{aligned}x_1'X_2\hat{\beta}_2 &= 0 & x_1'X_3\hat{\beta}_3 &= 0 & \dots & x_1'X_k\hat{\beta}_k &= 0 \\ & & X_2'X_3\hat{\beta}_3 &= 0 & \dots & X_2'X_k\hat{\beta}_k &= 0 \\ & & & & \ddots & & \vdots \\ & & & & & X_{k-1}'X_k\hat{\beta}_k &= 0\end{aligned}\tag{3.11}$$

for every y , where $\hat{\beta}_i$'s are the least squares estimates under the model (1.1).

Since $\hat{\beta}_1$ and $\hat{\beta}_2$ are the unique solution to the equations

$$\begin{bmatrix} x_1'x_1 & x_1'X_2 \\ X_2'x_1 & X_2'X_2 \\ 0 & B_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} x_1'y \\ X_2'y \\ 0 \end{bmatrix},\tag{3.12}$$

we can show that the equations $x_1'X_2\hat{\beta}_2=0$ should be generated by the equation (3.12), and therefore, should be generated by $B_2\hat{\beta}_2=0$. Thus the cell of $x_1'X_2$ can be crossed off by the identifiability constraints.

Similarly, since $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ are also the unique solution to the equations

$$\begin{bmatrix} x_1'x_1 & x_1'X_2 & x_1'X_3 \\ X_2'x_1 & X_2'X_2 & X_2'X_3 \\ X_3'x_1 & X_3'X_2 & X_3'X_3 \\ 0 & B_2 & 0 \\ 0 & 0 & B_3 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} x_1'y \\ X_2'y \\ X_3'y \\ 0 \\ 0 \end{bmatrix},\tag{3.13}$$

then $x_1'X_3\hat{\beta}_3=0$ and $X_2'X_3\hat{\beta}_3=0$ should be generated by the equations (3.13), and then should be generated by $B_3\hat{\beta}_3=0$. Thus the cells of $x_1'X_3$ and $X_2'X_3$ can be crossed off by the identifiability constraints.

By repeating the same procedure, all the cells above the diagonal ones can be crossed off. Thus the theorem is proved. q. e. d.

Although the additional assumption on identifiability constraints in Theorem 3.1

seems tight, it will be satisfied by most of analysis of variance models by arranging the parameters in proper order. Therefore, Theorem 3.1 will provide us with an easy and simple method for finding the identifiability constraints which make all the hypotheses orthogonal.

The method proposed by Theorem 3.1 is illustrated with two models.

EXAMPLE 1. The two-way layout with one observation per cell.

Consider a two-way layout with I levels of factor A , J levels of factor B and one observation per cell. Let y_{ij} be the observation in the (A_i, B_j) cell, and assume the model

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \quad i=1, 2, \dots, I; \quad j=1, 2, \dots, J, \quad (3.14)$$

where μ represents the general mean, α_i the main effect of factor A , β_j main effect of factor B and e_{ij} the random error.

As the identifiability constraints are given within each of the parameter groups, the class of identifiability constraints is given by

$$\sum_i u_i \alpha_i = 0, \quad \sum_j v_j \beta_j = 0, \quad (3.15)$$

where $\sum_i u_i = 1$ and $\sum_j v_j = 1$. It is easy to see that the assumptions in Theorem 3.1 are all satisfied by this model having identifiability constraints (3.15).

Since the left-hand side in the normal equations for obtaining the least squares estimates of parameters is given by Table 1,

Table 1

$\hat{\mu}$	$\hat{\alpha}_1$ ----- $\hat{\alpha}_I$	$\hat{\beta}_1$ ----- $\hat{\beta}_J$
IJ	J ----- J	I ----- I
J	J \diagup 0	1 ----- 1
J	0 \diagdown J	1 ----- 1
I	1 ----- 1	I \diagup 0
I	1 ----- 1	0 \diagdown I
	u_1 ----- u_I	v_1 ----- v_J

we can conclude from Theorem 3.1 that H_1 ($\mu=0$), H_2 ($\alpha_1=\dots=\alpha_I=0$) and H_3 ($\beta_1=\dots=\beta_J=0$) are orthogonal if and only if

$$\begin{aligned} u_1 &= u_2 = \cdots = u_I = \frac{1}{I} \\ v_1 &= v_2 = \cdots = v_J = \frac{1}{J}. \end{aligned} \quad (3.16)$$

This result coincides with that of Seber [5].

EXAMPLE 2. The two-way layout with unequal numbers of observations in the cells.

Consider a two-way layout with I levels of factor A , J levels of factor B and n_{ij} observations in the (A_i, B_j) cell. If y_{ijk} denotes the k -th observation in the (A_i, B_j) cell, then y_{ijk} can be written as

$$\begin{aligned} y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \\ i &= 1, 2, \dots, I; \quad j = 1, 2, \dots, J; \quad k = 1, 2, \dots, n_{ij}, \end{aligned} \quad (3.17)$$

where μ , α_i and β_j are as in Example 1, γ_{ij} the interaction between A and B and e_{ijk} the random error.

As a class of identifiability constraints, we shall consider

$$\begin{aligned} \sum_i u_i \alpha_i &= 0, \quad \sum_j v_j \beta_j = 0, \\ \sum_i u_i \gamma_{ij} &= 0, \quad \text{for all } j, \\ \sum_j v_j \gamma_{ij} &= 0, \quad \text{for all } i, \end{aligned} \quad (3.18)$$

where $\sum_i u_i = 1$ and $\sum_j v_j = 1$.

It is easy to see that the assumptions in Theorem 3.1 are all satisfied by this model having identifiability constraints (3.18). In this example, the left-hand side in the normal equations is given by Table 2.

In order to cross off all the cells above the diagonal ones by using identifiability constraints, it is necessary that

$$\begin{aligned} \text{(i)} \quad n_{i.} &= Q u_i, \quad \text{for all } i \\ \text{(ii)} \quad n_{.j} &= R v_j, \quad \text{for all } j \\ \text{(iii)} \quad n_{ij} &= S_i v_j, \quad \text{for all } i, j \\ \text{(iv)} \quad n_{ij} &= T_j u_i, \quad \text{for all } i, j, \end{aligned} \quad (3.19)$$

where Q , R , S_i and T_j are some constants. From (i), (ii) and (iii) in (3.19), we have

$$n_{..} = Q, \quad n_{..} = R, \quad n_{i.} = S_i$$

because of $\sum_i u_i = 1$ and $\sum_j v_j = 1$. Then, again from (i), (ii) and (iii), we have

$$\begin{aligned} u_i &= \frac{n_{i.}}{n_{..}}, \quad \text{for all } i \\ v_j &= \frac{n_{.j}}{n_{..}}, \quad \text{for all } j \\ n_{ij} &= \frac{n_{i.} n_{.j}}{n_{..}} \quad \text{for all } i, j. \end{aligned} \quad (3.20)$$

Table 2

$\hat{\mu}$	$\hat{\alpha}_1 \text{ --- } \hat{\alpha}_I$	$\hat{\beta}_1 \text{ --- } \hat{\beta}_J$	$\hat{\gamma}_{11} \text{ --- } \hat{\gamma}_{IJ}$	$\hat{\gamma}_{I1} \text{ --- } \hat{\gamma}_{IJ}$
$n_{..}$	$n_{1.} \text{ --- } n_{I.}$	$n_{.1} \text{ --- } n_{.J}$	$n_{11} \text{ --- } n_{IJ}$	$n_{I1} \text{ --- } n_{IJ}$
*	$n_{1.}$ 0 $n_{I.}$	$n_{11} \text{ --- } n_{IJ}$	$n_{11} \text{ --- } n_{IJ}$	$n_{I1} \text{ --- } n_{IJ}$
*	*	$n_{.1}$ 0 $n_{.J}$	n_{11} 0 n_{IJ}	n_{I1} 0 n_{IJ}
*	*	*	n_{11} 0 n_{IJ}	n_{I1} 0 n_{IJ}
	$u_1 \text{ --- } u_I$	$v_1 \text{ --- } v_J$	$v_1 \text{ --- } v_J$	$v_1 \text{ --- } v_J$
			u_1 u_I	u_I
$n_{i.} = \sum_j n_{ij}, \quad n_{.j} = \sum_i n_{ij}, \quad n_{..} = \sum_i \sum_j n_{ij}$				

We can easily check that if u_i 's, v_j 's and n_{ij} 's are as in (3.20) then all the cells above the diagonal ones are crossed off by using identifiability constraints (3.18). Then, (3.20) is the necessary and sufficient condition for the four hypotheses H_1, H_2, H_3, H_4 ($\gamma_{11}=\gamma_{12}=\dots=\gamma_{IJ}=0$) to be orthogonal.

Seber [5] obtained the condition (3.20) under the further assumption of $u_i \geq 0$ and $v_j \geq 0$. Although even if we relax the class of identifiability constraints to

$$\begin{aligned}
 \sum_i u_i \alpha_i &= 0, \quad \sum_j v_j \beta_j = 0 \\
 \sum_i l_i \gamma_{ij} &= 0, \quad \text{for all } j \\
 \sum_j m_j \gamma_{ij} &= 0, \quad \text{for all } i,
 \end{aligned} \tag{3.21}$$

where $\sum_i u_i=1$, $\sum_j v_j=1$, $\sum_i l_i=1$ and $\sum_j m_j=1$, we can get the similar necessary and sufficient condition, we have not checked if (3.21) is identifiability constraint.

References

- [1] DARROCH, J.N. and SILVEY, S.D.: *On testing more than one hypothesis*, Ann. Math. Stat., **34** (1963), 555-567.
- [2] JOHN, P.W.M.: *Statistical design and analysis of experiments*, The Macmillan. (1971).
- [3] KEMPTHORNE, O.: *The design and analysis of experiments*, Wiley. (1952).
- [4] SCHEFFÉ, H.: *The analysis of variance*, Wiley. 1959.
- [5] SEBER, G.A.F.: *The linear hypothesis*, Griffin. (1966).
- [6] WASHIO, Y.: *On Lagrange multipliers in normal equations*, Proc. Fujihara Men. Fac. Engin. Keio Univ., **23** (1970), 87-99.

Communicated by T. Kilagawa

Received October 27, 1984